Economic Issues, Problems and Perspectives Series

Economic Forecasting

Alan T. Molnar
Editor

NOVA

ECONOMIC ISSUES, PROBLEMS AND PERSPECTIVES SERIES

ECONOMIC FORECASTING

ECONOMIC ISSUES, PROBLEMS AND PERSPECTIVES SERIES

Trust, Globalisation and Market Expansion
Jacques-Marie Aurifeille, Christopher Medlin, and Clem Tisdell
2009. ISBN: 978-1-60741-812-2

TARP in the Crosshairs: Accountability in the Troubled Asset Relief Program
Paul W. O'Byrne (Editor)
2009. ISBN: 978-1-60741-807-8

TARP in the Crosshairs: Accountability in the Troubled Asset Relief Program
Paul W. O'Byrne (Editor)
2009. ISBN: 978-1-60876-705-2 *(Online book)*

Government Interventions in Economic Emergencies
Pablo Sastre (Editor)
2010. ISBN: 978-1-60741-356-1

NAFTA Stock Markets: Dynamic Return and Volatility Linkages
Giorgio Canarella, Stephen M. Miller and Stephen K. Pollard
2010. ISBN: 978-1-60876-498-3

Economic Forecasting
Alan T. Molnar (Editor)
2010. ISBN: 978-1-60741-068-3

ECONOMIC ISSUES, PROBLEMS AND PERSPECTIVES SERIES

ECONOMIC FORECASTING

ALAN T. MOLNAR
EDITOR

Nova Science Publishers, Inc.
New York

For permission to use material from this book please contact us:
Telephone 631-231-7269; Fax 631-231-8175
Web Site: http://www.novapublishers.com

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA

Economic forecasting / editor, Alan T. Molnar.
p. cm.
Includes index.
ISBN 978-1-61122-478-8 (eBook)
1. Economic forecasting--Mathematical models. I. Molnar, Alan T.
HB3730.E243 2009
330.01'5195--dc22
2009032429

Published by Nova Science Publishers, Inc. ✢ New York

CONTENTS

PREFACE

Economic forecasting is the process of making predictions about the economy as a whole or in part. Such forecasts may be made in great detail or may be very general. In any case, they describe the expected future behaviour of all or part of the economy and help form the basis of planning. Economic forecasting is of immense importance as any economic system is a stochastic entity of great complexity and vital to the national development in the information age. Forecasts are required for two basic reasons: the future is uncertain, and the full impact of many decisions taken now might deviate later. Consequently, accurate predictions of the future would improve the efficiency of the decision-making process. In particular, the knowledge of future demand for products and services is imperative to all industries since it is a prerequisite for any viable corporate strategy. This new and important book gathers the latest research from around the globe in this field and related topics such as: the econometric modeling and forecasting of private housing demand, the nonparametric time-detrended Fisher effect, and others.

One of the biggest concerns of an economic analyst is to understand the condition that an economy is experiencing at any given time, monitoring it properly in order to anticipate possible changes. However, despite social and economic life having quickened and become more turbulent, many relevant economic variables are not available at the desired frequency. Therefore, great quantities of methods, procedures and algorithms have been specifically proposed in the literature to solve the issue of transforming a low-frequency series into a high-frequency one. Moreover, a non-negligible number of proposals have been also conveniently adapted to deal with the problem of interpolation, distribution and extrapolation of time series. Thus, in order to put some order in the subject and to comprehend the current state of the art on the topic, Chapter 1 offers a revision of the historical evolution of the temporal disaggregation problem, analysing the proposals and classifying them. This permits one to decide which method to use under what circumstances, to conclude by identifying the topics in need of further development and to make some comments on possible future research directions in this field.

Governments, corporations and institutions all need to prepare various types of forecasts before any policies or decisions are made. Particularly, serving as a significant sector of an economy, the importance of predicting the movement of the private residential market is undeniable. However, it is well recognised that the housing demand is volatile and it may fluctuate dramatically according to general economic conditions. As globalisation continues to dissolve boundaries across the world, more economies are increasingly subjected to external shocks. Frequently the fluctuations in the level of housing demand can cause significant rippling effects in the economy as the housing sector is associated with many other economic sectors. The development of econometric models is thus postulated to assist policy-makers and relevant stakeholders to assess the future housing demand in order to formulate suitable policies.

With the rapid development of econometric approaches, their robustness and appropriateness as a modelling technique in the context of examining the dynamic relationship between the housing market and its determinants are evident. Chapter 2 applies the cointegration analysis as well as Johansen and Juselius's vector error correction model (VEC) model framework to housing demand forecasting in Hong Kong. Volitality of the demand to the dynamic changes in relevant macro-economic and socio-economic variables are considered. In addition, an impulse response function and a variance decomposition analysis are employed to trace the sensitivity of the housing demand over time to the shocks in the macro-economic and socio-economic variables. This econometric time-series modelling approach surpasses other methodologies by its dynamic nature and sensitivity to a variety of factors affecting the output of the economic sector for forecasting purposes, taking into account indirect and local inter-sectoral effects.

Empirical results indicated that that the housing demand and the associated economic factors: housing prices, mortgage rate, and GDP per capita are cointegrated in the long-run. Other key macro-economic and socio-economic indicators, including income, inflation, stock prices, employment, population, etc., are also examined but found to be insignificant in influencing the housing demand. A dynamic and robust housing demand forecasting model is developed using VEC model. The housing prices and mortgage rate are found to be the most important and significant factors determining the quantity demand of housing. Findings from the impulse response analyses and variance decomposition under the VEC model further confirm that the housing price terms has relatively large and sensitive impact on the housing demand, although at different time intervals, on the volume of housing transactions in Hong Kong. Addressing these two attributes is critical to the formulation of both short- and long-term housing policies that could satisfy the expected demand effectively.

The research contributes knowledge to the academic field as currently the area of housing demand forecast using advanced econometric modelling techniques is under-explored. This study has developed a theoretical model that traces the cause-and-effect chain between the housing demand and its determinants, which is relevant to the current needs of the real estate market and is significant to the economy's development. It is envisaged that the results of this study could enhance the understanding of using advanced econometric modelling methodologies, factors affecting housing demand and various housing economic issues.

The decomposition of a time series into components representing trend, cycle, seasonal, etc., has a long history. Such decompositions can provide a formal framework in which to model an observed time series and hence enable forecasts of both the series and its components to be computed along with estimates of precision and uncertainty. Chapter 3 provides a short historical background to time series decomposition before setting out a general framework. It then discusses signal extraction from ARIMA and unobserved component models. The former includes the Beveridge-Nelson filter and smoother and canonical decompositions. The latter includes general structural models and their associated state space formulations and the Kalman filter, the classical trend filters of Henderson and Macaulay that form the basis of the X-11 seasonal adjustment procedure, and band-pass and low-pass filters such as the Hodrick-Prescott, Baxter-King and Butterworth filters. An important problem for forecasting is to be able to deal with finite samples and to be able to adjust filters as the end of the sample (i.e., the current observation) is reached. Trend extraction and forecasting under these circumstances for a variety of approaches will be discussed and algorithms presented. The variety of techniques will be illustrated by a sequence of examples that use typical economic time series.

In Chapter 4 we study the gains that securitizing companies enjoy. We expressed the gains as a spread between two costs of capital, the weighted cost of capital of the asset selling firm and the all-in, weighted average cost of the securitization. We calculate the spread for 1,713 securitizations and regress those gains on asset seller characteristics. We show that they are increasing in size in the amount of asset backed securities originated but are decreasing in the percent of the balance sheet that the originator has outstanding. Companies that off-lay the risk of the sold assets (i.e., those retaining no subordinate interest in the SPV) pick their best assets to securitize. Companies that do not off-lay risk gain more from securitization the less liquid they are. We find that securitization is a substitute for leverage and that those companies that use more conventional leverage benefit less from securitization.

Chapter 5 uses frontier nonparametric VARs techniques to investigate whether the Fisher Effect holds in the U.S. The Fisher Effect is examined taking into account structural breaks and nonlinearities between nominal interest rates and inflation, which are trend-stationary in the two samples examined. The nonparametric time-detrended test for the Fisher Effect is formed from the cumulative orthogonal dynamic multiplier ratios of inflation to nominal interest rates. If the Fisher Effect holds, this ratio statistically approaches one as the horizon goes to infinity. The nonparametric techniques developed in this paper conclude that the Fisher Effect holds for both samples examined.

Chapter 6 investigates the effect of forecasting ability on forecasting bias among Japanese GDP forecasters. Trueman (1994, Review of Financial Studies, 7(1), 97-124) argues that an incompetent forecaster tends to discard his private information and release a forecast that is close to the prior expectation and the market average forecast. Clarke and Subramanian (2006, Journal of Financial Economics, 80, 81-113) find that a financial analyst issues bold earning forecasts if and only if his past performance is significantly different from his peers. This paper examines a twenty-eight-year panel of annual GDP forecasts, and obtains supportive evidence of Clarke and Subramanian (2006). Our result indicates that conventional tests of rationality are biased toward rejecting the rational expectations hypothesis.

As explained in Chapter 7, through Monte Carlo simulations it is possible to isolate the measurement error introduced by incorrect assumptions when quantifying survey results. By means of a simulation experiment we test whether a variation of the balance statistic outperforms the balance statistic in order to track the evolution of agents' expectations and produces more accurate forecasts of the quantitative variable generated used as a benchmark.

Chapter 8 investigates the relative performance of local, foreign, and expatriate financial analysts on Latin American emerging markets. We measure analysts' relative performance with three dimensions: (1) forecast timeliness, (2) forecast accuracy and (3) impact of forecast revisions on security prices. Our main findings can be summarized as follows. Firstly, there is a strong evidence that foreign analysts supply timelier forecasts than their peers. Secondly, analyst working for foreign brokerage houses (i.e. expatriate and foreign ones) produce less biased forecasts than local analysts. Finally, after controlling for analysts' timeliness, we find that foreign financial analysts' upward revisions have a greater impact on stock returns than both followers and local lead analysts forecast revisions. Overall, our results suggest that investors should better rely on the research produced by analysts working for foreign brokerage houses when they invest in Latin American emerging markets.

Income tax revenue crucially depends on the wage distribution across and within the industries. However, many transition economies present a challenge for a sound econometric analysis due to data unavailability. Chapter 9 presents an approach to modeling and forecasting income tax revenues in an economy under missing data on individual wages within the industries. We consider the situations where only the aggregate industry-level data and sample observations for a few industries are available. Using the example of the Uzbek economy in 1995-2005, we show how the econometric analysis of wage distributions and the implied tax revenues can be conducted in such settings. One of the main conclusions of the paper is that the distributions of wages and the implied tax revenues in the economy are well approximated by Gamma distributions with semi-heavy tails that decay slower than those of Gaussian variables.

Chapter 10 analyzes the out-of-sample ability of different parametric and semiparametric GARCH-type models to forecast the conditional variance and the conditional and unconditional kurtosis of three types of financial assets (stock index, exchange rate and Treasury Note). For this purpose, we consider the Gaussian and Student-t GARCH models by Bollerslev (1986, 1987), and two different time-varying conditional kurtosis GARCH models based on the Student-t and a transformed Gram-Charlier density.

Chapter 11 argues that the transportation model of linear programming can be used to administer the Public Personnel Language Exam of Turkey in many different locations instead of just one, as is the current practice. It shows the resulting system to be much less costly. Furthermore, once the decision about number of locations is made, the resulting system can be managed either in a centralized or decentralized manner. A mixed mode of management is outlined, some historical perspectives on the genesis of the transportation model are offered and some ideas regarding the reasons for the current wasteful practices are presented. The possibility of applying the same policy reform in other MENA (Middle East and North Africa) countries is discussed in brief.

In: Economic Forecasting
Editor: Alan T. Molnar, pp. 1-27
ISBN: 978-1-60741-068-3

Chapter 1

TEMPORAL DISAGGREGATION OF TIME SERIES—A REVIEW

Jose M. Pavía*

Department of Applied Economics
University of Valencia, Spain.

Abstract

One of the biggest concerns of an economic analyst is to understand the condition that an economy is experiencing at any given time, monitoring it properly in order to anticipate possible changes. However, despite social and economic life having quickened and become more turbulent, many relevant economic variables are not available at the desired frequency. Therefore, great quantities of methods, procedures and algorithms have been specifically proposed in the literature to solve the issue of transforming a low-frequency series into a high-frequency one. Moreover, a non-negligible number of proposals have been also conveniently adapted to deal with the problem of interpolation, distribution and extrapolation of time series. Thus, in order to put some order in the subject and to comprehend the current state of the art on the topic, this chapter offers a revision of the historical evolution of the temporal disaggregation problem, analysing the proposals and classifying them. This permits one to decide which method to use under what circumstances, to conclude by identifying the topics in need of further development and to make some comments on possible future research directions in this field.

Keywords: Interpolation, Temporal Distribution, Extrapolation, Benchmarking, Forecasts.

1. Introduction

The problem of increasing the frequency of a time series has concerned economic analysts for a long time. Nevertheless, the subject did not start to receive the required attention among economists before the 1970s, despite more frequent information being of great importance for both modelling and forecasting. Indeed, according to Zellner and

---

* E-mail address: pavia@uv.es

Montmarquette (1971, p. 355): "*When the behavior of individuals, firms or other economic entities is analyzed with temporally aggregated data, it is quite possible that a distorted view of parameters' value, lag structures and other aspects of economic behavior can be obtained. Since policy decisions usually depend critically on views regarding parameter values, lag decisions, etc., decisions based on results marred by temporal aggregation effects can produce poor results*". Unfortunately, it is not unusual that some relevant variables are not available with the desired timeliness and frequency. Delays in the process of managing and gathering more frequent data, the extra costs that entails to collect variables more frequently, and practical limitations to obtain some statistics with a higher regularity deprive analysts of the valuable help that more frequent records would provide to perform a closer and more accurate short-term analysis. Certainly, having available statistical series with a higher frequency would facilitate a smaller delay and a more precise analysis of the economy (and/or of a company situation) making it easier to anticipate changes and to react to them. It is not surprising, therefore, that a number of methods, procedures and algorithms have been proposed from different perspectives in order to increase the frequency of some critical variables.

Obviously, business and economics are not the sole areas where it would be useful. Fields as diverse as engineering, oceanography, astronomy and geology also use these techniques and take advantage of these strategies in order to improve the quality of their analysis. Nevertheless, this chapter will concentrate on the contributions made and used within the economic field. There are many examples, in both macro- and microeconomics, where having available more frequent data could be useful. As examples I cite the following: (i) agents who deal within a certain region have annual aggregated information about the economic evolution of the region (e.g., annual regional accounts), although they would prefer the same information quarterly rather than annually to perform better short-term analysis; (ii) in some areas, demographic figures are known every ten years, although it would be great to have them annually making available a better match between population needs and provision of public services; (iii) a big company counts on quarterly records about its raw material necessities, although it would be much more useful to have that information monthly, or even weekly, to better manage its costs; or, (iv) in econometric modelling, where some of the relevant series are only available at lower frequencies, it could be convenient to previously disaggregate these series to estimate the model, instead of estimating the complete model at lower frequency level with the resulting loss of information (Lütkepohl, 1984; Nijman and Palm, 1985, 1990) and efficiency in the estimation of the parameters of the model (Palm and Nijman, 1984; Weiss, 1984; or, Nijman and Palm, 1988a, 1988b).

In general, inside this framework and depending on the kind of variable handled (either, flow or stock) two different main problems can be posed: the distribution problem and the interpolation problem. The distribution problem appears when the observed values of a flow low-frequency series of length T must be distributed among kT values, such that the temporal sum of the estimated high-frequency series fits the values of the low-frequency series. The interpolation problem consists in generating a high-frequency series with the values of the new series being the same as the ones of the low-frequency series for those temporal moments where the latter is observed. In both cases, when estimates are extended out of the period covered by the low-frequency series, the problem is called extrapolation. Extrapolation is used, therefore, to forecast values of the high-frequency series when no temporal constraints from short series are available; although, nevertheless, in some cases (especially in multivariate contexts) other different forms of constraints can exist. Furthermore, and related to distribution problems, they can be found benchmarking and balancing—which are mainly used in management and by official statistical agencies to adjust the values of a high-frequency series of ballpark figures (usually obtained employing sample techniques) to a more accurate low-frequency series—and other temporal disaggregation problems where the temporal aggregation function is different from the sum function. Anyway, despite the great quantity of procedures for temporal disaggregation of time series being proposed in the literature, the fulfilment of the constraints derived from the observed low-frequency series is the norm in the subject.

Although temporal disaggregation methods are currently used in a great variety of business and economic problems, the enlargement and improvement of most of the procedures have been usually developed connected with short-term analysis. They have been in fact linked to the need of solving problems related to the production of coherent quarterly national accounts (see, e.g., OECD, 1996; Eurostat, 1999; or, Bloem *et al.*, 2001) and quarterly regional accounts (see, e.g., Pavía-Miralles and Cabrer-Borrás, 2007; and, Zaier and Trabelsi, 2007). Actually, it is quite probable that some of the future fruitful developments expected in this topic get to solve the new challenges posed in this area. Nevertheless, additionally to the methods specifically proposed to estimate quarterly and monthly accounts, another significant number of methods suggested to estimate missing observations have been also adapted to this issue. Thus, classifying the large variety of methods proposed in the literature emerges as a necessary requirement in order to perform a systematic and ordered study of the numerous alternatives suggested. In fact, a classification would be in itself a proper tool for a suitable selection of the technique in each particular situation due to as DiFonzo and Filosa (1987, p. 10) pointed out: "*It also seemed opportune to stress the crucial importance of the fact that differing algorithms though derived from the same research field and using the same basic information, can give rise to series with different cyclical, seasonal and stochastic properties*".

A first division could arise attending to the plane from which the problem is faced, either the frequency domain or the temporal plane. This division, however, is not well-balanced, since the temporal perspective has been by and large more popular. On the one hand, the procedures that deal with the problem from the temporal plane will be analyzed in Sections 2, 3, and 4. On the other hand, the methods that try to solve the problem from the spectral point of view will be introduced in Section 5.

Another possible criterion of division attends to the use or not of related series, usually called *indicators*. Economic events tend to be made visible in different ways and to affect many dimensions. The economic series are therefore correlated variables that do not evolve in an isolated way. Consequently, it is not unusual that some variables available in high-frequency could display similar fluctuations than those (expected) for the target series. Some methods try to take advantage of this fact to temporally distribute the target series. Thus, the use or not of indicators has been considered as another criterion to classify.

The procedures which deal with the series in an isolated way and compute the missing data of the disaggregated high-frequency series taking into account only the information given by the objective series have been grouped under the name of *methods that do not use indicators*. Different approaches and strategies have been employed to solve the problem within this group of proposals. The first algorithms proposed were quite mechanical and distributed the series imposing some properties considered "interesting". Step by step, nevertheless, new methods (theoretically founded on the ARIMA representation of the series to be disaggregated) were progressively appearing, introducing more flexibility in the process. Section 2 is devoted to the methods classified in this group.

Complementary to the group of techniques that do not use indicators appears the procedures based on indicators, which exploit the economic relationships between indicators and objective series to temporally distribute the target series. This group is composed for an extent and varied collection of methods that have had enormous success and that have been widely used. In fact, as Chow and Lin (1976, p. 720) remarked: "*...there are likely to be some related series, including dummy variables, which can usefully serve as regressors. One should at least use a single dummy variable identically equal to one; its coefficient gives the mean of the time series.*" and moreover as Guerrero and Martínez (1995, p. 360) said: "*It is our belief that, in practice, one can always find some auxiliary data. These data might simply be an expected trend and seasonal behaviour of the series to be disaggregated*". Hence, it is not surprising the great success of these procedures and that the utilization of procedures based on indicators is a rule among the agencies and governmental statistic institutes that estimate quarterly and monthly national accounts using indirect methods. These procedures are presented in Section 3.

Finally, and independent of their use or not of indicators, it has been grouped in another category the methods that use the Kalman filter for the estimation of the non available values. The great flexibility that offers the representation of temporal processes in the state space and the enormous possibilities that these representations present to properly deal with log-transformations and dynamic approximations to the issue justify broadly its own section. The procedures based on this algorithm can be found in Section 4.

It is clear that alternative classifications could be reached if different criteria had been followed and that any classification runs the risk of being inadequate and a bit artificial. Moreover, the categorization chosen does not avoid the problem of deciding where place some procedures or methods, which could belong to different groups and whose location turns out as extremely complicated. Nevertheless, the classification of the text has been chosen because it is the belief of the author that it clarifies and makes easier the exposition. Furthermore, it is necessary to remark that no mathematical expressions have been included in the text in order to make the exposition quicker and easier to follow. The chapter makes a verbal review of the several alternatives. The interested reader can consult specific mathematical details in the numerous references cited throughout the chapter, or consult Pavía-Miralles (2000a), who offers a revision of many of the procedures suggested before 1997 unifying the mathematical terms.

In short, the structure of the chapter is as follows. Section 2 introduces the methods that do not use related information. Section 3 describes the procedures based on indicators. This section has been divided into subsections in order to handle the great quantity of methods proposed using the related variable approach. Section 4 deals with the methods that use the Kalman filter. Section 5 shows the procedures based on spectral developments. And finally, Section 6 offers some concluding remarks and comments on possible future research directions in the subject.

2. Methods that Do not Use Indicators

Despite the use of indicators being the most popular approach in the problem of temporal disaggregation of time series, a remarkable quantity of methods have been also proposed in econometric and statistical literature trying to face the problem using exclusively the observed low-frequency values of the own time series. This approach comprises purely mathematical methods and more theoretically founded model-based methods relying on the Autoregressive Integrated Moving Average (ARIMA) representation of the series to be disaggregated.

The first methods proposed in this group were developed as mere instruments, without any theoretical justification, for the elaboration of quarterly (or monthly) national accounts. These first procedures were purely ad-hoc mathematical algorithms to derive a smooth path for the unobserved series. They constructed the high-frequency series (from now on and without losing generality, they are supposed quarterly in order to lighten the language) from the low-frequency series (without losing generality, they are assumed annual) according to the properties that the series to build were supposed to follow, imposing the annual constrains.

The design of these primary methods, nevertheless, was already in those first days influenced for the need of solving one issue that appears recurrently in the subject and whose solution must tackle every method suggested to disaggregate time series: the problem of spurious steps. To prevent series with undesired discontinuities from one year to the next, the pioneers proposed to make dependent on several annual data the quarterly estimates values belonging to a particular year. The disaggregation methods proposed by Lisman and Sandée (1964), Zani (1970), and Greco (1979) were devised to estimate the quarterly series corresponding to the year t as a weighted average of the annual values of periods $t\text{-}1$, t and $t+1$. They estimate the quarterly series through a fix weight structure. The difference among these methods lies on their election of the weight matrix. Lisman and Sandée (1964) calculated the weight matrix by requiring that the estimated series verify some *a priori* "interesting" properties. Zani (1970) assumed that the curve of the quarterly estimates is located upon a second degree polynomial that passes by the origin. And Greco (1979) extended Zani's proposal to polynomials with other degrees. Furthermore, Glejser (1966) expanded Lisman and Sandée (1964) to the case of distributing quarterly or annual series into monthly ones and later Almon (1988) provided, in an econometric computer package G, a method to convert annual figures into quarterly series, assuming that a cubic polynomial is fitted to each successive set of two points of the low-frequency series. All these methods, however, are univariate and it was necessary to wait two more decades to reach from this approach a solution for the multivariate problem. Just recently, Zaier and Trabelsi (2007) has extended, for both stock and flow variables, Almon's univariate polynomial method to the multivariate case.

Using also an ad-hoc mathematical approach, although with a different line of thinking, Boot *et al.* (1967) proposed building the quarterly series by solving an optimization problem. Particularly, their method proposes to construct the quarterly series as solution of the minimization of the sum of squares of either the first or the second differences of the (unknown) consecutive quarterly values, under the condition that the annual aggregation of the estimated series adds up the available annual figures. Although Boot's *et al.* algorithms mostly reduced the subjective charge of the preceding methods, their way of solving the problem of spurious steps was still a bit subjective and therefore it was not free of criticism. On the particular, it could be consulted, among others, Ginsburg (1973), Nijman and Palm (1985), DiFonzo and Filosa (1987), and Pavía-Miralles (2000a).

In the same way that polynomial procedures were generalized, Boot's *et al.* approach was also extended in both flexibility and in the number of series to be handled. Cohen *et al.* (1971) extended Boot's *et al.* work introducing flexibility in the process in a double way: on one hand, they dealt with any pair of possible combination of high and low frequencies; and, on the other hand, they considered the minimization of the sum of the squared of the *i*th differences between successive subperiod values (not only first and second differences). The multivariate extension, nevertheless, was introduced in Pavía *et al.* (2000).

Additionally to the abovementioned ad-hoc mathematical algorithms, many other methods could be also classified within the group of techniques that base the estimates exclusively on the observed data of the target series. Among them, it will be stressed in this section Doran (1974) and Stram and Wei (1986). On the one hand, Doran (1974) assumed that there is a part of the sample period where the target series is observed in its higher frequency (it generally happening towards the end of the sample) and proposed to use this subsample to estimate the temporal characteristics of the series and to obtain the non-available values employing this information. This strategy, however, is not optimum, as Chow and Lin (1976) proved. Chow and Lin (1976) adapted the estimator suggested in Chow and Lin (1971) to the same situation treated by Doran and showed that Doran's method generated estimates with larger mean square errors. On the other hand, Stram and Wei (1986) proposed to obtain the target series minimizing a squared function defined by the inverse covariance matrix associated to the quarterly stationary ARMA(p,q) process obtained taking differences on the non-stationary one. In particular, they suggested adjusting an ARIMA model to the low-frequency series, selecting an ARIMA(p,d,q) model (an ARIMA process with autoregressive order p, integrated order d, and moving average order q) for the quarterly values compatible with the annual model and minimizing the dth differences of the objective series by the loss function, with the annual series as constraint.

Stram and Wei's proposal, besides, made possible to reassess Boot *et al.* (1967). As they showed Boot *et al.*'s algorithm is equivalent to use this procedure assuming that the series to estimate follows either a temporal integrated process of order one or two (i.e., I(1) or I(2)). According to Rodríguez-Feijoo *et al.* (2003), however, Stram and Wei method only performs well when the series are long enough to permit a proper estimation of the ARIMA process. In this line, in order to know the advantages and disadvantages of these methods and decide which to use under what circumstances, it could be consulted Rodríguez-Feijoo *et al.* (2003). They performed a simulation exercise in which the methods proposed by Lisman and Sandée (1964), Zani (1970), Boot *et al.* (1967), Denton (1971) —in its variant without indicators—, Stram and Wei (1986), and Wei and Stram (1990) were analysed.

Finally, it must be noted that although the approach using ARIMA processes has produced many others fruits, no more proposals has been set up in this section. On the one hand, those procedures based on ARIMA models which take advantage of the representation in the space of the states and use the Kalman filter and smoothing techniques to estimate both the coefficients of the process and the non-observed values of the high-frequency series have been placed in Section 4. On the other hand, those approaches that try to deduce the ARIMA process of the high-frequency series from the ARIMA model of the low-frequency as a strategy to estimate the missing values are presented in Section 3. Notice that this last strategy can be seen as a particular case of a dynamic regression model with missing observations in the dependent variable and without indicators, whose general situation with indicators is introduced in the next section.

3. Methods Based on Indicators

The methods based on related variables have been the most popular and the most widely used and successful. Thus, a great number of procedures can be found within this category. Comparing with the algorithms non-based on indicators, related variable procedures have been assigned two main principal advantages: (i) they present better foundations in the construction hypothesis, which can comparatively affect the results validation; and, (ii) they make use of relevant economic and statistical information, being more efficient. Although, as Nasse (1973) observed, in return they hide an implicit hypothesis according to which the annual relationship is accepted to be maintained in the quarterly basis. Nevertheless, as Friedman (1962, p. 731) pointed out: '*a particular series Y is of course chosen to use in interpolation because its intrayearly movements are believed to be highly correlated with the intrayearly movements of X*'. Anyway, additionally to Nasse's observation, it must be noted that the resulting estimates depend crucially on the indicators chosen and therefore special care should be taken in selecting them. To solve this issue, already in 1951, Chang and Liu (1951) tried to establish some criteria the indicators should fulfil; nevertheless, the debate far from closed has been opened during decades (see, e.g., Nasse 1970, 1973; Bournay and Laroque, 1979; INE, 1993; OECD, 1996; or, Pavía *et al.*, 2000). For example, although the use of indicators to predict the probable evolution of key series throughout the quarters of a year is the rule among the countries that estimate quarterly accounts by indirect methods (a summary of the indicators used by the different national statistics agencies can be found in OECD, 1996, pp. 22-37), there are no apparent universal criteria for selecting them. It, however, does not mean that any sound criteria have been proposed. In particular, and related to the elaboration of quarterly accounts, Pavía-Miralles and Cabrer-Borrás (2007, p. 161) pointed out that "*…indicators, available monthly or quarterly, that verified —at least in an approximate way— the following properties: (a) economic implication, (b) representation or greatest coverage, (c) maintenance of a 'constant' relation with the regional series being estimated, (d) quick availability and short lag, (f) acceptable series length, and (g) smooth profile or predominance of the trend-cycle signal*" must be chosen, to which it could be added statistical quality and having an intrayear evolution similar to the objective series. Despite the great debate about indicators, very few tests about their validity can be found in the literature. As exception, INE (1993, p.12) offers a statistical test about the accuracy of the selected indicators to estimate quarterly accounts.

This section has been divided in three subsections to better manage the large quantity of procedures classified in this category. The first subsection is devoted to those procedures, called adjusting methods, which given an initial approximation of the target series adjust their values using some penalty function in order to fulfil the annual constraints. Subsection 2 presents the procedures that take advantage of structural or econometric models—including some techniques using dynamic regression models in the identification of the relationship linking the series to be estimated and the (set of) related time series—to approximate the incompletely observed variables. According to Jacob (1994), the structural model may take the form of a (simple) time series model or a regression model with other variables where the estimates are obtained as by-product of parameter estimation of the model. Finally, subsection three shows those methods, named optimal methods, which jointly obtain the estimates of both parameters and quarterly series combining target annual series and quarterly indicators and incorporating the annual constraints in the process of estimation, basically Chow and Lin (1971) and its extensions.

3.1. Adjusting Methods

In general the adjusting methods are composed of two stages. In the first step an initial approximation of the objective series is obtained. In the second step the first estimates are adjusted by imposing the constraints derived from the available and more reliable annual series. The initial estimates are reached using either sample procedures or some kind of relationship among indicators and target series. When the initial estimates come from surveys, additionally to adjusting procedures, the so-called benchmarking and balancing techniques are usually employed (see, e.g., Dagum and Cholette, 2006; and, Särndal, 2007). Although, it must be noted that the frontier between benchmarking and adjustment algorithms is not clear and somewhat artificial (see, DiFonzo and Marini, 2005). Among those options that use related variables to obtain initial approximations both non-correlated and correlated strategies could be found. The non-correlated proposals, historically the first ones, do not explicitly take into account the existing correlation between target series and indicators —Friedman (1962) can be consulted for wide summary of those first algorithms. On the other hand, the correlation strategies usually assume a lineal relationship between the objective series and the indicators, from which an initial high-frequency series is obtained.

Once the initial approximation is available, it is adjusted to make it congruent with the observed annual series. The discrepancies between both annual series (the observed series and the series obtained by quarterly aggregation of the initial estimates) are then removed. A great quantity of adjustment procedures can be found in the literature. Bassie (1958, pp. 653-61) proposed to distribute annual discrepancies by a structure of fixed weights. Such a structure is calculated taking into account the discrepancies corresponding to two successive years and assuming that the weights function follows a third degree polynomial. Despite having no theoretical support and Bassie recognizing that the method spawns series with irregularities and cyclical components different to the initial approximations when the annual discrepancies are too big (OCDE, 1966, p. 21), Bassie's proposal has been historically applied to series of the Italian economy (ISCO, 1965; ISTAT, 1983) and currently Finland and Denmark use variants of this method to adjust their quarterly GDP series (OECD, 1996, p. 19).

Vangrevelinghe (1966) planned a different approach. His proposal (primary suggested to estimate the French quarterly familiar consumption series) consists of (i) applying Lisman and Sandée (1964) to both objective annual series and indicator annual series to obtain, respectively, an initial approximation and a control series, to then (ii) modifying the initial estimate by aggregating the discrepancies between the observed quarterly indicator and the control series, using as scale factor the Ordinary Least Squares (OLS) estimator of the linear observed annual model. Later, minimal variations of Vagrevelinghe's method were proposed by Ginsburg (1973) and Somermeyer *et al.* (1976). Ginsburg suggested obtaining the initial estimates using Boot *et al.* (1967), instead of Lisman-Sandée, and Somermeyer *et al.*

proposed generalizing Lisman and Sandée (1964) by allowing the weight structure to be different for each quarter and year, with the weight structure obtained, using annual constraints, from a linear model.

One of the most successful methods in the area (not only among adjusting procedures) is the approach proposed by Denton in 1971. The fact that, according to DiFonzo (2003a, p. 2), short-term analysis in general and quarterly accounts in particular need disaggregation techniques being "*…flexible enough to allow for a variety of time series to be treated easily, rapidly and without too much intervention by the producer;*" and that "*the statistical procedures involved should be run in an accessible and well known, possibly user friendly, and well sounded software program, interfacing with other relevant instruments typically used by data producers (i.e. seasonal adjustment, forecasting, identification of regression models,…)*" explains the great attractiveness of methods such as Denton (1971) and Chow and Lin (1971) among analysts and statistical agencies (see, e.g., Bloem *et al.*, 2001; and Dagum and Cholette, 1994, 2006); despite using more sophisticated procedures generally yielding better estimates (Pavía-Miralles and Cabrer-Borrás, 2007).

Denton (1971) suggested adjusting the initial estimates minimizing a loss function defined by a square form. Therefore, the choice of the symmetrical matrix determining the specific square form of the loss function is the crucial element in Denton's proposal. Denton concentrated on the solutions obtained minimizing the *h*th differences between the to-be-estimated series and the initial approximation and found Boot *et al.* (1967) as a particular case of his algorithm. Later on, Cholette (1984) proposed a slight modification to this function family to avoid dependence on the initial conditions. Although, nevertheless, the main extensions of Denton approach were reached by Hillmer and Trabelsi (1987), Trabelsi and Hillmer (1990), Cholette and Dagum (1994), DiFonzo (2003d) and DiFonzo and Marini (2005), they made more flexible the algorithm and extended it to the multivariate case.

Hillmer and Trabelsi (1987) and Trabelsi and Hillmer (1990) worked on the problem of adjusting a univariate high-frequency series using data obtained from different sampling sources, and found Denton (1971) and Cholette (1984) as particular cases of their proposal. In particular, they relaxed the requirements about the low-frequency series permitting it to be observed with error; although, as compensation, they had to suppose known the temporal structure of the errors caused by sampling the low frequency series (see also Weale, 1992). When benchmarks are observed without error, the problem transforms into minimizing the discrepancies between the initial estimates and the annual series according to a loss function of the square form type (Trabelsi and Hillmer, 1990). In these circumstances, they showed that the method of minimizing the *h*th differences proposed by Denton (1971) and Cholette (1984) implies to implicitly admit: (i) that the rate between the variances of the observation errors and the ARIMA modelization errors of the initial approximation tends to zero; and, (ii) that the observation errors follow a I(h) process with either null initial conditions, in Denton's approach, or with the initial values of the series of observation errors begin in a remote past, in Cholette's method.

In sample survey most time series data come from repeated surveys whose sample designs usually generate autocorrelated errors and heterocedasticity. Thus, Cholette and Dagum (1994) introduced a regression model to take into account it explicitly and showed that the gain in efficiency of using a more complex model varies with the ARMA model assumed for the survey errors. In this line, Chen and Wu (2006) showed, through a simulation exercise and assuming that the survey error series follows an AR(1) process, that Cholette and

Dagum (1994) and Dagum *et al.* (1998) have great advantages over Denton method and that they are robust to misspecification of the survey error model. On the other hand, the multivariate extension of Denton method was proposed in DiFonzo (2003d) and DiFonzo and Marini (2005) under a general accounting constraint system. They assumed a set of linear relationships among target variables and indicators from which initial estimates are obtained, to then, applying the movement preservation principle of Denton approach subject to the whole set of contemporaneous and temporal aggregation relationships, reach estimates of all the series verifying all the constraints.

Although Denton (1971) and also DiFonzo and Marini (2005) do not require any reliability measurement of survey error series, their need in many other proposals led Guerrero (1990, p. 30) to propose an alternative approach after writing that "*These requirements are reasonable for a statistical agency:…but they might be very restrictive for a practitioner who occasionally wants to disaggregate a time series*". In particular, to overcome some arbitrariness in the choice of the stochastic structure of the high frequency disturbances, Guerrero (1990) and Guerrero and Martínez (1995) developed a new adjustment procedure assuming that the initial approximation and the objective series share the same ARIMA model. More specifically, they combined an ARIMA-based approach with the use of high frequency related series in a regression model to obtain the Best Linear Unbiased Estimate (BLUE) of the objective series verifying annual constraints. This approach permits an automatic (which takes a recursive form in Guerrero and Martínez, 1995) 'revision' of the estimates with each new observation. This feature illustrates an important difference with the other procedures where the estimates obtained for the periods relatively far away from the sample final period are in practice 'fixed'. Likewise, the multivariate extension of this approach was also provided by Guerrero, who, together with Nieto (Guerrero and Nieto, 1999), suggested a procedure for estimating unobserved values of multiple time series whose temporal and contemporaneous aggregates are known using vector autoregressive models. Under this approach, moreover, it must be noted that even though the problem can be cast into a state-space formulation, the usual assumptions underlying Kalman filtering are not fulfilled in this case and that therefore Kalman filter approach cannot be applied directly.

A very interesting variant in this framework emerges when log-transformations are taken. Indeed, in many circumstances, it is strongly recommended to use logarithms or other transformations of original data (for example, most time series become stationary after applying first differences to their logarithms) to achieve better modelizations of time series and also, as Aadland (2000) showed through a simulation experiment, to obtain more accurate disaggregates because of "*…the failure to account for data transformations may lead to serious errors in estimation*" (Aadland, 2000, p. 141). However, due to the logarithmic transformation being not additive, the annual aggregation constraint can not be directly applied in a distribution problem.

The problem of dealing with log-transformed variables in the distribution framework was first considered by Pinheiro and Coimbra (1993), and later treated, among others, in Proietti (1998) and Aadland (2000). Proietti (1998) tackled the problem of adjusting estimated values to fulfil temporal aggregation constraints. On the one hand, Proietti (1998) proposed to obtain initial estimates applying the exponential function to the approximations reached using a linear relationship between the log-transformation of the target series and the indicators, to then in a second step adopt Denton's algorithm to get the final values. According to DiFonzo (2003a), however, this last step could be unnecessary as "*the disaggregated estimates present* *only negligible discrepancies with the observed aggregated values*." (DiFonzo 2003a, p. 17). On the other hand, when the linear relationship is expressed in terms of rate of change of the target variable (i.e., using the logarithmic difference), initial estimates for the non-transformed values of the objective variable could be now obtained using Fernandez (1981), being a further adjustment (using Denton's formula) for either flow or index variable eventually performed to exactly fulfil the temporal aggregation constraints (DiFonzo, 2003a, 2003b).

3.2. Structural and Econometrics Model Methods

The economic theory stands for functional relationships among variables. The econometric models express those relations by means of equations. Models based on annual data conceal higher frequency information and are not considered sufficiently informative to policy makers. Building quarterly and monthly macroeconometric models is therefore imperative and responds for one of its traditional motivations: the demand of high-frequency forecasts. Sometimes, the frequency of the variables taking part in the model is not homogeneous and expressing the model in the lower common frequency almost never offers an acceptable approximation. Indeed, with the aim of forecasting, Jacobs (2004) showed that is preferable to deal with the quarterly model with missing quarterly observations rather than generate quarterly predictions disaggregating the annual forecasts from the annual model: the quarterly estimator based on approximations is revealed as more efficient (even biased) than the annual estimator. Thus, putting the model in the desired frequency and use the same model, not only to estimate the unknown parameters but also to estimate the non-observed values of the target series, represents in general a good alternative to forecast. Furthermore, according with Vanhonacker (1990), it is also preferable to estimate the missing observations simultaneously with the econometric model rather than previously interpolated the unavailable values to directly handle the high-frequency equations, because of "*…its effects on subsequent econometric analysis can be serious: parameter estimates can be severely (asymptotically) biased…*" (Jacobs, 2004, p. 5).

There are a lot of econometric models susceptible of being formulated. And therefore, many strategies may be adopted to estimate the missing observations. As examples of this variety of model-based approach could be cited, among others, Drettakis (1973), Sargan and Drettakis (1974), Dagenais (1973, 1976), Dempster *et al.* (1977), Hsiao (1979, 1980), Gourieroux and Monfort (1981), Palm and Nijman (1982, 1984), Conniffe (1983), Wilcox (1983), Nijman and Palm (1986, 1990), and Dagum *et al.* (1998).

Drettakis (1973) formulated a multiequational dynamic model about the United Kingdom economy with one of the endogenous variables observed only annually for a part of the sample and obtained estimates for the parameters and the unobserved values by Maximum Likelihood (ML) with complete information. Sargan and Drettakis (1974) extended Drettakis (1973) to the case in which the number of unobserved series is higher than one and introduced an improvement to reduce the computational charges of the estimation procedure. The use of ML was also followed in Hsiao (1979, 1980) and Palm and Nijman (1982). As example, Palm and Nijman (1982) derived the ML estimator when data are subject to different temporal aggregations and compared its sample variance with those obtaining after applying the estimator proposed by Hsiao, Generalized Least Squares (GLS) and Ordinary Least Squares

(OLS). On the other hand, GLS estimators were be employed by Dagenais (1973), Gourieroux and Monfort (1981), and Conniffe (1983) for models with missing observations in the exogenous variables and, therefore, probably with a heteroscedastic and serially correlated disturbance term.

In the extension to dynamic regression models, the ML approach was again used in Palm and Nijman's works. Nijman and Palm (1986) considered a simultaneous equations model, not completely specified, about the Dutch labour market with some variables only annually observed and proposed to obtain initial estimates for those variables using the univariate quarterly ARIMA process that congruent with the multiequational model is derived from the observed annual series. These initial estimates were used to estimate the model parameters by ML. Palm and Nijman (1984) studied the problem of parameters identification and Nijman and Palm (1986, 1990) the estimation one. To estimate the parameters they proposed two alternatives based on ML. The first one consisted of building the likelihood function from the forecast errors, using the Kalman filter. The second alternative consisted of applying the EM algorithm adapted to incompletes samples. This adaptation was developed in a wide and long paper by Dempster *et al.* (1977) from Hartley (1958). Dagum *et al.* (1998), on the other hand, presented a general dynamic stochastic regression model, which permits to deal with the most common short-term data treatment (including interpolation, benchmarking, extrapolation and smoothing), and showed that the GLS estimator is the minimum variance linear unbiased estimator (see, also Dagum and Cholette, 2006). With respect to other temporal disaggregation procedures based on dynamic models (e.g., Santos Silva and Cardoso, 2001; Gregoir, 2003; or, DiFonzo, 2003a), they will be considered in the next subsection, since they could be observed as dynamic extensions of Chow and Lin (1971). Although, they could be also placed on the previous subsection due to they follow the classical two-step approach of adjusting methods.

3.3. Optimal Methods

Optimal methods get their name to the estimation strategy they adopt. Such procedures directly incorporate the restrictions derived from the observed annual series into the estimation process to jointly obtain the BLUE of both parameters and quarterly series. To do that, a linear relationship between target series and indicators is usually assumed. This group of methods is one of the most widely used and in fact its root proposal (Chow and Lin, 1971) has served as basis for many statistical agencies (see, e.g., ISTAT, 1985; INE, 1993; Eurostat, 1998; or DiFonzo, 2003a) and analysts (e.g., Abeysinghe and Lee, 1998; Abeysinghe and Rajaguru, 1999; Pavía and Cabrer, 2003; and, Norman and Walker, 2007) to quarterly distribute annual accounts and to provide flash estimates of quarterly growth, among other tasks.

Although many links between adjustment and optimal procedures exist, as DiFonzo and Filosa (1987, p. 11) indicated "*(i) … compared to optimal methods, adjustment methods make an inefficient (and sometimes, biased) use of the indicators; (ii) the various methods have a different capability of providing statistically efficient extrapolation…*", which points at optimal methods as more suitable to perform short-term analysis using forecasts. In compensation, the solution of this sort of methods crucially depends on the correlations structure assumed for the errors of the linear relationship. In fact, many proposals are only different in that point. All of them, nevertheless, pursue to avoid spurious steps in the estimated series.

Friedman (1962) was the first one in applying this approach. In particular and for the case of a stock variable, he obtained (assuming a linear relationship between target series and indicators) the BLUE of both coefficients and objective series. Nevertheless, Chow and Lin (1971) were who, extending Friedman's result, wrote the paper probably most influential and cited in this subject. They obtained the BLUE of the objective series for interpolation, distribution and extrapolation problems using a common notation. They focused on the case of converting a quarterly series into a monthly one and assumed an AR(1) hypothesis for the errors in order to avoid unjustified discontinuities in the estimated series. Under this hypothesis, the covariance matrix is governed by the autoregressive coefficient of order one of the high-frequency disturbance series, which is unknown. Hence, to apply the method it has to be previously estimated. Chow and Lin (1971) suggested exploiting the functional relationship between autoregressive coefficients of order one of the low- and the high-frequency errors to estimate it. Specifically, they proposed an iterative procedure to estimate the monthly AR(1) coefficient from the rate between elements (1,2) and (1,1) of the quarterly error covariance matrix.

The Chow-Lin strategy of relating the autoregressive coefficients of order one of the high and low error series, however, can not be completely generalized to any pair of frequencies (Acosta *et al.*, 1977) and consequently several other stratagems were followed to solve the issue. In line with Chow-Lin approach, DiFonzo and Filosa (1987) obtained for the annual-quarterly case a function between the two autoregressive coefficients. The problem of the relation reached by DiFonzo and Filosa is that it only has unique solution for non-negative annual autoregressive coefficient. Despite it, Cavero *et al.* (1994) and IGE (1997) took advantage of such a relation to suggest two iterative procedures to handle the Chow-Lin method in the quarterly-annual case with AR(1) errors. Cavero *et al.* (1994) even provided a solution to apply the method when an initial negative estimate of the autoregressive coefficient is obtained. Although, to handle the problem of the sign, Bourney and Laroque (1979) had already proposed to estimate the autoregressive coefficient through a two-step algorithm in which, in the first step, the element (1,3) of the covariance matrix of the annual errors is used to determinate the sign of the autoregressive coefficient. In addition to the above possibilities, strategies based on the maximum likelihood (with the hypothesis of normality for the errors) have been also tried. Examples of this approach can be found in Barbone *et al.* (1981), ISTAT (1985), and Quilis (2005).

Although the AR(1) temporal error structure has been the most extensively analyzed, other structures for the errors has been also proposed. Among the stationary structures Schmidt (1986) held MA(1), AR(2), AR(4), and a mix between AR(1) and AR(4) processes as reasonable possibilities to deal with the annual-quarterly case. Although, the Monte Carlo evidence in Pavía *et al.* (2003) showed that assuming an AR(1) hypothesis on the disturbance term does not significantly influence the quality of the estimates, despite disturbances following other stationary structures. In regard to the extensions towards no stationary structures, Fernández (1981) and Litterman (1983) can be cited. On the one hand, Fernández (1981, p. 475) recommended using Denton's approach proposing "*estimate regression coefficients using annual totals of the dependent variables, and then apply these coefficients to the high frequency series to obtain preliminary estimates…*" to afterwards "*they are 'adjusted' following the approach of Denton*" and showed that such an approach to the problem is equivalent to using the Chow-Lin method with a random walk hypothesis for the errors—a hypothesis that he defended:"*a random walk hypothesis for a series of residuals ... should not be considered unrealistic*" (Fernández, 1981, p. 475), supported by results in Nelson and Gould (1974) and Fernández (1976). On the other hand, Litterman (1983) studied the problem of monthly disaggregating a quarterly series and extended the Chow-Lin method for the case in which the residual series followed a Markov random walk. Litterman did not solve the problem of estimating the parameter of the Markov process for small samples though. Fortunately, Silver (1986) found a solution to this problem and extended Litterman's proposal to the case of annual series and quarterly indicators.

Despite DiFonzo and Filosa's abovementioned words about the superiority of optimal methods over adjustment procedures, all the previous methods can be obtained as solutions of a quadratic-linear optimization problem (Pinheiro and Coimbra, 1993), where the metric matrix that defines the loss function is the inverse of the high-frequency covariance error matrix. Therefore, theoretically other structures for the disturbances could be easily managed. In particular, in order to improve disaggregated estimates, the high-frequency covariance error matrix could be estimated, following Wei and Stram (1990), from the data by imposing an ARIMA structure and using its relationship with the low-frequency covariance matrix. Despite it, low AR order models are still systematically chosen in practice due to (i) the covariance matrix of the high-frequency disturbances cannot be, in general, uniquely identified from the low-frequency one and (ii) the typical sample sizes occurring in economics usually provide poor low-frequency error matrix estimates (Rossana and Seater, 1995; Proietti 1998; DiFonzo, 2003a). In fact, the Monte Carlo evidence presented in Chan (1993) showed that this approach would likely perform comparatively badly when the low-frequency sample size is lower than 40 (a really non infrequent size in economics).

The estimates obtained according to Chow and Lin's approach, however, are only completely satisfactory in the case where the temporal aggregation constraint is linear and there are no lagged dependent variables in the regression. Thus, to improve accuracy of estimates taking into account dynamics specifications usually encountered in applied econometrics works, several authors (e.g., Salazar *et al.*, 1997a, 1997b; Santos Silva and Cardoso, 2001; Gregoir, 2003) have proposed to generalize Chow-Lin approach (including Fernández and Litterman extensions) by the use of linear dynamic models. It permits to perform temporal disaggregation providing more robust results in a broad range of circumstances. In this line, Santos Silva and Cardoso (2001), following the way initiated by Salazar *et al.* (1997a, 1997b) and Gregoir (2003), proposed an extension of Chow-Lin—by means of a well-known transformation developed to deal with distributed lag model (e.g., Klein, 1958; Harvey, 1990)—which is particularly adequate when the series used are stationary or cointegrated (see also DiFonzo, 2003c). Their extension, furthermore, compared to Salazar *et al.* and Gregoir, solves the problems in the estimation of the first low-frequency period and produces disaggregated estimates and standard errors in a straightforward way (which was very difficult to implement in a computer program in the initial proposals). Two empirical applications of this procedure, additionally to a panoramic revision of this approach, could be found in DiFonzo (2003a, 2003b), while in Quilis (2003) a MATLAB library to perform it is offered. This library completed the MATLAB libraries that to run Boot *et al.* (1967), Denton (1971), Chow and Lin (1971), Fernández (1981), Litterman (1983), DiFonzo (1990) and DiFonzo (2003d) is provided by Instituto Nacional de Estadística (Quilis, 2002).

The Chow-Lin approach and its abovementioned extensions are all univariate, thus to handle problems with more than *J* (>1) series to-be-estimated, multivariate extensions are required. In this situations, apart from the low-frequency temporal constraints, some additional cross-section, transversal or contemporaneous aggregates among the high-frequency target series are usually available. To deal with this issue, different procedures (extending Chow-Lin method) have been proposed in the literature. Rossi (1982) was the first who faced this problem. Rossi assumed that the contemporaneous quarterly aggregate of the *J* series is known and proposed to apply an estimation procedure in two steps. In the first step, he suggested applying the Chow-Lin method, in an isolated way, to each one of the *J* series imposing only the corresponding annual constraint and assuming white noise residuals. In the second step, he proposed to apply again Chow-Lin procedure, imposing as constraint the observed contemporaneous aggregated series and under a white noise error vector, to simultaneously estimate the *J* series using as indicators the series estimated in the first step. This strategy, however, as DiFonzo (1990) pointed out, does not guarantee the fulfilment of the temporal restrictions.

DiFonzo (1990), attending to Rossi's limitation, generalized the Chow-Lin estimator and got the BLUE of the *J* series, fulfilling simultaneously the temporal and the transversal restrictions. Similar to Chow-Lin, DiFonzo (1990) again obtained that the estimated series crucially depend on the structure assumed for the disturbances. Nevertheless, he only offered a practical solution under the hypothesis of errors temporally uncorrelated. That hypothesis unfortunately is inadequate due to it can produce spurious steps in the estimated series. In order to solve it, Cabrer and Pavía (1999) and Pavía-Miralles (2000b) introduced a structure for the disturbances in which each one of the *J* error series follow either an AR(1) process or a random walk with shocks only contemporaneously correlated. Pavía-Miralles (2000b), additionally, extended the estimator obtained in DiFonzo (1990) to situations with more general contemporaneous aggregations and provided an algorithm to run such so complex disturbance structure in empirical works. Finally, DiFonzo (2003d) proposed to simplify Pavía-Miralles (2000b) suggesting a multivariate random walk structure for the error vector and Pavía-Miralles and Cabrer-Borrás (2007) extended Pavía-Miralles's (2000b) proposal to deal with the extrapolation issue.

4. Methods Based on the Representation in the State Space

One of the approaches in the study of time series is to consider the series as a realisation of a stochastic process with a particular model generator (e.g., an ARIMA process), which depends on some parameters. In order to predict how the series will behave in a future or to rebuild the series estimating the missing observation it is necessary to know the model parameters. The Kalman filter permits to take advantage of the temporal sequence of the series to implement through a set of mathematical equations a predictor-corrector type estimator, which is optimal in the sense that it minimizes the estimated error covariance when some presumed conditions are met. In particular, it is an efficient recursive filter that estimates the state of a dynamic system from a series of incomplete and noisy measurements. Within the temporal disaggregation problem, this approach appears very promising due its great versatility and presents the additional advantage of making possible that both unadjusted and seasonally adjusted series can be simultaneously estimated.

Among the different approaches to approximate the population parameters of the data generating process it stands out ML. The likelihood function of the stochastic process can be calculated in a relatively simple and very operative way by the Kalman filter. The density of the process, under a Gaussian distribution assumption for the series, can be easily derived from the forecast errors. Prediction errors can be computed in a straightforward way by representing the process in the state space, and the Kalman filter can then be used. In general, the pioneers methods based on the representation in the state space supposed an ARIMA process for the objective series and computed the likelihood of the process through the Kalman filter by employing the smooth point-fixed algorithm (details of this algorithm can be consulted, among others, in Anderson and Moore, 1979; and, Harvey, 1981; and in the multivariate extension in Harvey 1989) to estimate the not available values.

Despite the representation of a temporal process in the state space not being unique, the majority of the proposals to adapt Kalman filter to manage missing observations can be reduced to the one proposed by Jones (1980). Jones suggested building the likelihood function excluding the prediction errors associated to those temporal moments where no observation exist and proposed to use forecasts obtained in the previous instant to go on running the Kalman filter equations. Among others, this pattern was followed by Harvey and Pierse (1984), Ansley and Kohn (1985), Kohn and Ansley (1986), Al-Osh (1989), Harvey (1989), and Gómez and Maravall (1994). Additionally to Jones's approach, other approaches can be found. DeJong (1989) developed a new filter and some smooth algorithms which allow interpolating the non observed values with simpler computational and analytical expressions. Durbin and Quenneville (1997) used state space models to adjust a monthly series obtained from a survey to an annual benchmark. And, Gómez *et al.* (1999) followed the strategy of estimating missing observations considering them as outliers, while Gudmundsson (1999) introduced a prescribed multiplicative trend in the problem of quarterly disaggregating an annual flow series using its state space representation.

Jones (1980), pioneer in the estimation of missing observations from the state space representation, treated—from a representation proposed by Akaike (1974)—the case of a stock variable which is assumed to follow a stationary ARMA process. Later on, Harvey and Pierse (1984), also dealing with stationary series, extended Jones's proposal—using another representation due to Akaike (1978)—for the case of flow variables. Likewise, they adapted the algorithm to that case in which the target series follows a regression model with stationary residuals and dealt with the problem of working with logarithms of the variable. Furthermore, Harvey and Pierse also extended the procedure to the case of stock variables following non stationary ARIMA processes; although in this case, they compelled the target variable being available in a high-frequency for a large enough sample subperiod.

In the non stationary case, however, when Harvey and Pierse's hypothesis is not verified, building the likelihood of the process becomes difficult. Problems in converting the process into stationary and in defining the initial conditions arise. Thus, in order to solve it, on the one hand, Ansley and Kohn (1985) proposed to consider a diffuse initial distribution in the pre-sample and, on the other hand, Kohn and Ansley (1986) suggested transforming the observations in order to define the likelihood of the process. Kohn and Ansley's transformation made possible to generalize the previous results (including those reached by

Harvey and Pierse), although at the cost of destroying the sequentiality of the series, altering both smoothing and filtering algorithms. Fortunately, Gómez and Maravall (1994) went beyond this difficulty and solved it making possible to use the classical tools to deal with the problem of non stationary processes whatever the structure of the missing observations. However, although Kohn and Ansley (1986) and Gómez and Maravall (1994) proposals extended the issue to the treatment of regression models with non stationary residuals (allowing related variables to be included in this framework), they did not deal with the case of flow variables in an explicit way. Indeed, it was Al-Osh (1989) who handled such a problem and extended the solution to non stationary flow series. Al-Osh, moreover, suggested using the Kalman filter for the recursive estimate of the non-observed values as a tool to overcome the problem of the change of the estimates due to the increasing of the available sample. In this line, Cuche and Hess (2000) used information contained in related series to using a general approach based on the Kalman filter estimate the monthly Swiss GDP from the quarterly series, while Liu and Hall (2001) estimated a monthly US GDP series from quarterly values after testing several state space representations to, through a MonteCarlo experiment, identify which variant of the model gives the best estimates. They found the more simple representations did almost as well as more complex ones.

Most of above proposals, however, consider the temporal structure (the ARIMA process) of the objective series known. In practice, however, it is unknown and it is required to specify the orders of the process to deal with it. In order to solve it, some strategies have been followed. Some attempts have tried to infer the process of a high-frequency series from the observed process of the low-frequency one (e.g., Nijman and Palm, 1985; Al-Osh, 1989; Guerrero and Martínez, 1995); while many other studies have concentrated on analyzing the effect of aggregation over a high frequency process (e.g., among others, Telser, 1967; Amemiya and Wu, 1972; Tiao, 1972; Wei, 1978; Lütkepohl, 1984; Stram and Wei, 1986; and, more recently, Rossana and Seater, 1995) and on studying its effect over stock variables observed in fixed step times (among others, Quenouille, 1958; Werner, 1982; or Weiss, 1984). Fortunately, the necessary and sufficient conditions under which the aggregate and/or disaggregate series can be expressed by the same class of model was derived by Hotta and Vasconcellos (1999).

Both multivariate and dynamic extensions have been also tackled from this framework, although they are just incipient. On the one hand, the multivariate approach started by Harvey (1989) was continued in Moauro and Savio (2005), who suggested a multivariate seemingly unrelated time series equations model to using the Kalman filter estimate the high-frequency series when several constraints exits. The framework they proposed is flexible enough to allow for almost any kind of temporal disaggregation problems of both raw and seasonally adjusted time series. On the other hand, Proietti (2006) offered a dynamic extension providing, among others contributions, a systematic treatment of Litterman (1983), which permits to explain the difficulties commonly encountered in practice when estimating Literman's model.

5. Approaches from the Frequency Domain

From the previous sections it can be deduced that a great amount of energy has been devoted to deal with the issue from the temporal perspective. Similarly, great efforts have been also devoted from the frequential plane, although they have had less successful and have therefore done less fruits. In particular, the greatest efforts have been invested on estimating the spectral density function or spectrum of the series, the main tool of a temporal process in the frequency domain. The estimation of the spectrum of the series has been undertaken from both angles: the parametric and the non-parametric perspective.

Both Jones (1962) and Parzen (1961, 1963) were pioneers in the study of missing observations from the frequency domain. They analyzed the problem under a systematic scheme for the observed (and therefore also for the unobserved) values. Jones (1962), one of the pioneers in studying the problem of estimating the spectrum, treated the case of estimating the spectral function of a stock stationary series sampled systematically. This problem was also faced by Parzen (1963) who introduced the term of *amplitude modulation*, the key element in which later spectral developments were based on in their search for solutions. The amplitude modulation defines itself as a zeros and ones series in the sample period. The value of the amplitude modulation is one in those periods where the series is observed, whereas it is zero in case of no being observed.

Different schemes for the amplitude modulation have been considered in the literature. Scheinok (1965) considered the case in which the amplitude modulation followed a Bernoulli random scheme. This random scheme was extended to others by Bloomfield (1970, 1973). More recently, Tolio and Morettin (1993) obtained estimators of the spectral function for three types of modulation sequences: determinist, random and correlated random. On the other hand, Dunsmuir and Robinson (Dunsmuir, 1981; and Dunsmuir and Robinson, 1981a, 1981b), followed a different way, they assumed an ARIMA process and estimated its parameters with the help of the spectral approximation to the likelihood function.

Although the great majority of patterns for the missing observations can apparently be treated from the frequency domain, not all of them have a solution. This fact is due to the impossibility of completely estimating the autocovariances of the process in many practical situations. In this sense, Clinger and Van Ness (1976) studied the situations in which it is possible to estimate all the autocovariances. On the particular, it must be remembered Dunsmuir's (1981, p. 620) words: "… (the estimators) *are asymptotically efficient when compared to the Gaussian maximum likelihood estimate if the proportion of missing data is asymptotically negligible.*" Hence, the problem of disaggregating an annual time series in quarterly figures is one of those that do not still have a satisfactory solution from this perspective. Nevertheless, from a related approach, Gudmundsson (2001) have made some advances proposing a method to estimate (under some restrictive hypothesis and in a *continuous* way) a flow variable. Likewise, the efforts made to employ the spectral tools to estimate the missing values using the information given by a group of related variables have required so many restrictive hypotheses that its use has not been advisable until now.

6. Conclusion

As can be easily inferred from the references and all the above sections a really huge quantity of procedures, methods and algorithms have been proposed in the literature to try to solve the problem of transforming a low-frequency series into a high-frequency one. The first group of methods that built series through ad-hoc procedures was progressively overcome, and the methods based on indicators were progressively gaining the preference of researchers.

Within this group of methods, it highlights the Chow-Lin procedure and all its multiple extensions. Interesting solutions have been also proposed from the state space and its great flexibility makes it a proper tool to deal with the future challenges to appear in the subject and to handle situations of missing observations different from those analysed in the current document. In compensation, however, within the methods proposed from the frequency domain the progress made does not seem encouraging. Nevertheless, none of the proposals should be discarded rapidly, because according to Marcellino (2007) pooling estimates obtained from different procedures can improve the quality of the disaggregated series.

Broadly speaking, an analysis of the historical evolution of the topic seems to point towards the techniques using dynamic regression models and the techniques using formulations in terms of unobserved component models/structural time series and the Kalman filter as the two research lines that will hold a pre-eminent position in the future. On the one hand, the extension of the topic to deal with multivariate dynamic models is still waiting to be tackled; and, on the other hand, the state space methodology offers the generality that is required to address a variety of inferential issues that have not been dealt with previously. In this sense, both approaches could be combined in order to solve one of the main open problems in the area: in particular, to jointly estimate some high-frequency series of rates when the low-frequency series of rates, some transversal constraints and several related variables are available. For example, the issue of distributing regionally the quarterly national growth of a country when the annual regional growth series are known and several high-frequency regional indicators are available and, moreover, both the regional and the sectoral structure of weights change quarterly and/or annually.

Furthermore, a new emerging approach—which is taking into account the more recent developments of econometric literature (e.g., data mining, dynamic common component analyses, or time series models environment) and takes advantage of the continuous advances in computer hardware and software by making use of a large dataset available—will likely turn up in the future as a main line in the subject. Indeed, as Angelini *et al.* (2006, p. 2693) point out: "*Existing methods … are either univariate or based on a very limited number of series, due to data and computing constraints … until the recent past. Nowadays large datasets are readily available, and models with hundreds of parameters are easily estimated*". In this line, Proietti and Moauro (2006) dealt with a dynamic factor model using the Kalman filter to perform an index of coincident US economic indicators; while Angelini *et al.* (2006) modelled a large dataset with a factor model and developed an interpolation procedure that exploits the estimated factors as a summary of all the available information. This last research also shows this strategy clearly improving univariate approaches.

References

Aadland, D.M. (2000). Distribution and Interpolation using Transformed Data. *Journal of Applied Statistics,* **27**, 141-156.

Abeysinghe, T., & Lee, C. (1998). Best Linear Unbiased Disaggregation of Annual GDP to Quarterly Figures: The Case of Malaysia. *Journal of Forecasting,* **17**, 527-537.

Abeysinghe, T., & Rajaguru, G. (1999). Quarterly Real GDP Estimates for China and ASEAN4 with a Forecast Evaluation. *Journal of Forecasting,* **18**, 33-37.

Acosta, L.R., Cortigiani, J.L., & Diéguez, M.B. (1977). *Trimestralización de Series Económicas Anuales*. Buenos Aires: Banco Central de la República Argentina, Departamento de Análisis y Coordinación Estadística.

Akaike, H. (1974). Markovian Representation of a Stochastic Processes and its Application to the Analysis of Autoregressive Moving Average Processes. *Annals of the Institute of Statistical Mathematics*, **26**, 363-87.

Akaike, H. (1978). Covariance Matrix Computation of the State Variable of a Stationary Gaussian Process. *Annals of the Institute of Statistical Mathematics*, **30**, 499-504.

Almon, C. (1988). *The Craft of Economic Modeling*. Boston: Ginn Press.

Al-Osh, M. (1989). A Dynamic Linear Model Approach for Disaggregating Time Series Data. *Journal of Forecasting,* **8**, 85-96

Amemiya, T., & Wu, R.Y. (1972). The Effect of Aggregation on Prediction in the Autoregressive Model. *Journal of the American Statistical Association,* **67**, 628-632.

Anderson, B.D.O., & Moore, J.B.(1979). *Optimal Filtering*. Englewood Cliffs, New Jersey: Prentice-Hall

Angelini, E., Henry, J., &. Marcellino, M. (2006). Interpolation and Backdating with a Large Information Set. *Journal of Economic Dynamics and Control,* **30**, 2693-2724.

Ansley, C.F., & Kohn, R. (1985). Estimating, Filtering and Smoothing in State Space Models with Incompletely Specified Initial Conditions. *Annals of Statistics,* **13**, 1286-1316.

Bassie, V.L. (1958) *Economic Forecasting*. New York: Mc Graw-Hill. pp. 653-61.

Barbone, L., Bodo, G., & Visco, I. (1981). Costi e Profitti in Senso Stretto: un'Analisi du Serie Trimestrali, 1970-1980, *Bolletino della Banca d'Italia,* **36**, 465-510.

Bloem, A.M., Dippelsman, R.J., & Maehle, N.∅. (2001). *Quarterly National Accounts Manual. Concepts, Data Sources, and Compilation*. Washington DC: International Monetary Fund.

Bloomfield, P. (1970). Spectral Analysis with Randomly Missing Observations. *Journal of the Royal Statistical Society, Ser. B,* **32**, 369-380.

Bloomfield, P. (1973). An Exponential Model for the Spectrum of a Scalar Time Series. *Biometrika,* **60**, 217-226.

Boot, J.C.G., Feibes, W., & Lisman, J.H. (1967). Further Methods of Derivation of Quarterly Figures from Annual Data. *Journal of the Royal Statistical Society, Ser. C,* **16**, 65-75.

Bournay, J., & Laroque, G. (1979). Réflexions sur le Méthode d'Élaboration des Comptes Trimestriels. *Annales de L'Insée,* **36**, 3-29.

Cabrer, B., & Pavía, J.M. (1999). Estimating J(>1) Quarterly Time Series in Fulfilling Annual and Quarterly Constraints. *International Advances in Economic Research,* **5**, 339-350.

Cavero, J., Fernández-Abascal, H., Gómez, I., Lorenzo, C., Rodríguez, B., Rojo, J.L., & Sanz, J.A. (1994). Hacia un Modelo Trimestral de Predicción de la Economía Castellano-Leonesa. El Modelo Hispalink CyL. *Cuadernos Aragoneses de Economía,* **4**, 317-343.

Chan, W. (1993). Disaggregation of Annual Time-Series Data to Quarterly Figures: A Comparative Study. *Journal of Forecasting,* **12**, 677-688.

Chang, C.G., & Liu, T.C. (1951). Monthly Estimates of Certain National Product Components, 1946-49. *The Review of Economic and Statistics,* **33**, 219-227.

Chen, Z.G., & Wu, K.H. (2006). Comparison of Benchmarking Methods with and without a Survey Error Model. *International Statistical Review,* **74**, 285-304.

Cholette, P.A. (1984). Adjusting Sub-Anual Series to Yearly Benchmarks. *Survey Methodology,* **10**, 35-49.

Cholette, P.A., & Dagum, E.B. (1994). Benchmarking Time Series With Autocorrelated Survey Errors. *International Statistical Review,* **62**, 365-377.

Chow, G. C., & Lin, A. (1971). Best Linear Unbiaded Interpolation, Distribution, and Extrapolation of Time Series By Related Series. *The Review of Economics and Statistics,* **53**, 372-375.

Chow, G.C., & Lin, A. (1976). Best Linear Unbiased Estimation of Missing Observations in an Economic Time Series. *Journal of the American Statistical Association,* **71**, 719-721.

Clinger, W., & VanNess, J.W. (1976). On Unequally Spaced Time Points in Time Series. *Annals of Statistics,* **4**, 736-745.

Cohen, K.J., Müller, M., & Padberg, M.W. (1971). Autoregressive Approaches to Disaggregation of Time Series Data. *Journal of the Royal Statistical Society, Ser. C,* **20**, 119-129.

Conniffe, D. (1983). Small-Sample Properties of Estimators of Regression Coefficients Given a Common Pattern of Missing Data. *Review of Economic Studies,* **50**, 111-120.

Cuche, N.A., & Hess, M.K. (2000). Estimating Monthly GDP in a General Kalman Filter Framework: Evidence from Switzerland. *Economic and Financial Modelling,* **7**, 1-37.

Dagenais, M. G. (1973). The Use of Incomplete Observations in Multiple Regression Analysis: a Generalized Least Squares Approach. *Journal of Econometrics,* **1**, 317-328.

Dagenais, M. G. (1976). Incomplete Observations and Simultaneous Equations Models. *Journal of Econometrics,* **4**, 231-241.

Dagum, E.B., Cholette, P.A., & Chen, Z.G. (1998). A Unified View of Signal Extraction, Benchmarking, Interpolation and Extrapolation of Time Series. *International Statistical Review,* **66**, 245-269.

Dagum, E.B., & Cholette, P.A. (2006). *Benchmarking, Temporal Distribution and Reconciliation Methods for Time Series*. New York: Springer Verlag.

Dagum, E.B., Cholette, P.A., & Chen, Z.G. (1998). A Unified View of Signal Extraction, Benchmarking, Interpolation and Extrapolation of Time Series. *International Statistical Review,* **66**, 245-269.

DeJong, P. (1989). Smoothing and Interpolation with the State-Space Model. *Journal of the American Statistical Association,* **84**, 1085-1088.

Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977). Maximun Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Ser. B,* **39**, 1-38.

Denton, F.T. (1971). Adjustment of Monthly or Quarterly Series to Annuals Totals: An Approach Based on Quadratic Minimization. *Journal of the American Statistical Association,* **66**, 99-102.

DiFonzo, T. (1990). The Estimation of M Disaggregate Time Series when Contemporaneous and Temporal Aggregates are Known. *The Review of Economics and Statistics,* **72**, 178-182.

DiFonzo, T. (2003a). *Temporal Disaggregation of Economic Time Series: Towards a Dynamic Extension*. Luxembourg: Office for Official Publications of the European Communities.

DiFonzo, T. (2003b). Temporal Disaggregation Using Related Series: Log-Transformation and Dynamic Extensions. *Rivista Internazionale di Scienze Economiche e Commerciali,* **50**, 371-400.

DiFonzo, T. (2003c). Constrained Retropolation of High-frequency Data Using Related Series: A Simple Dynamic Model Approach. *Statistical Methods and Applications,* **12**, 109-119.

DiFonzo, T. (2003d). Temporal Disaggregation of System of Time Series When the Aggregates Are Known. In: Barcellan R., & Mazzi, G.L. (Eds.), *INSEE-Eurostat Quarterly National Accounts Workshop, Paris-Bercy, December 1994* (pp. 63-78). Luxembourg: Office for Official Publications of the European Communities. Available at http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-AN-03-014/EN/KS-AN-03-014-EN.PDF.

DiFonzo, T., & Filosa, R. (1987). *Methods of Estimation of Quarterly National Account Series: A Comparison*. Paper presented at "Journee Franco-Italianne de Comptabilite Nationale (Journee de Stadistique)", Lausanne 18-20 mai, 1987. pp. 1-69.

DiFonzo, T., & Marini, M. (2005). Benchmarking Systems of Seasonally Adjusted Time Series. *Journal of Business Cycle Measurement and Analysis,* **2**, 84-123.

Doran, H.E. (1974). Prediction of Missing Observations in the Time Series of an Economic Variable. *Journal of the American Statistical Association,* **69**, 546- 554

Drettakis, E.G. (1973). Missing Data in Econometric Estimation. *Review of Economic Studies,* **40**, 537-52.

Dunsmuir, W. (1981). Estimation for Stationary Time Series when Data Are Irregularly Spaced or Missing. In Findley D. F. (Ed.), *Applied Time Series Analysis II* (pp. 609-649). New York: Academic Press.

Dunsmuir, W., & Robinson, P.M. (1981a). Parametric Estimators for Stationary Time Series with Missing Observations. *Advances in Applied Probability,* **13**, 129-146.

Dunsmuir, W., & Robinson, P.M. (1981b). Estimation of Time Series Models in the Presence of Missing Data. *Journal of the American Statistical Association,* **76**, 456-467.

Durbin, J., & Quenneville, B. (1997) Benchmarking by State Space Models. *International Statistical Review,* **65**, 23–48.

Eurostat (1999). *Handbook of Quarterly National Accounts*. Luxembourg: European Commission.

Fernández, R. B. (1976). Expectativas Adaptativas vs. Expectativas Racionales en la Determinación de la Inflación y el Empleo. *Cuadernos de Economía,* **40**, 37-58.

Fernández, R. B. (1981). A Methodological Note on the Estimation of Time Series. *The Review of Economics and Statistics,* **53**, 471-478.

Friedman, M. (1962). The Interpolation of Time Series by Related Series. *Journal of the American Statistical Association,* **57**, 729-757.

Ginsburgh, V.A. (1973). A Further Note on the Derivation of Quarterly Figures Consistent with Annual Data. *Journal of the Royal Statistical Society, Ser. C,* **22**, 368-374.

Glejser, H. (1966). Une Méthode d'Evaluation de Donnés Mensuelles à Partir d'Indices Trimestriels ou Annuels. *Cahiers Economiques de Bruxelles,* **29**, 45-54.

Gómez, V., & Maravall, A. (1994). Estimation, Prediction and Interpolation for Nonstationary series with the Kalman Filter. *Journal of the American Statistical Association,* **89**, 611-624.

Gómez, V., Maravall, A. & Peña, D. (1999). Missing Observations in ARIMA Models. Skipping Strategy versus Additive Outlier Approach. *Journal of Econometrics,* **88**, 341-363

Gourieroux, C., & Monfort, A. (1981). On the Problem of Missing Data in Linear Models. *Review of Economic Studies,* **48**, 579-586.

Greco, C. (1979). Alcune Considerazioni Sui Criteri di Calcolo di Valori Trimestrali di Tendenza di Serie Storiche Annuali. *Annali della Facoltà di Economia e Commercio,* **4**, 135-155.

Gregoir, S. (2003). Propositions pour une Désagrégation Temporelle Basée sur des Modèles Dynamiques Simples. In Barcellan R., & Mazzi, G.L. (Eds.), *INSEE-Eurostat Quarterly National Accounts workshop, Paris-Bercy, December 1994* (pp. 141-166). Luxembourg: Office for Official Publications of the European Communities. Available at http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-AN-03-014/EN/KS-AN-03-014-EN.PDF.

Gudmundsson, G. (1999). Disaggregation of Annual Flow Data with Multiplicative Trends. *Journal of Forecasting,* **18**, 33-37.

Gudmundsson, G. (2001). Estimation of Continuous Flows from Observed Aggregates. *Journal of the Royal Statistical Society, Ser. D,* **50**, 285-293.

Guerrero, V. M. (1990). Temporal Disaggregation of Time Series: An ARIMA-Based Approach. *International Statistical Review,* **58**, 29-46.

Guerrero, V.M., & Martínez, J. (1995). A Recursive ARIMA-Based Procedure for Disaggregating a Time Series Variable Using Concurrent Data. *Test,* **4**, 359-376.

Guerrero, V.M., & Nieto, F.H. (1999). Temporal and Contemporaneous Disaggregation of Multiple Economic Time Series. *Test,* **8**, 459-489.

Hartley, H.O. (1958). Maximum Likelihood Estimation from Incomplete Data. *Biometrics,* **14**, 174-194.

Harvey, A.C. (1981). *Time Series Models*. Oxford: Philip Allan.

Harvey A.C. (1989). *Forecasting, Structural Time Series and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A.C. (1990). *The Econometric Analysis of Time Series*. Deddington: Philip Allan.

Harvey, A.C., & Pierse, R.G. (1984). Estimating Missing Observations in Economic Time Series. *Journal of the American Statistical Association,* **79**, 125-131.

Hillmer, S.C., & Trabelsi, A. (1987). Benchmarking of Economic Time Series. *Journal of the American Statistical Association,* **82**, 1064-1071.

Hotta, L.K., & Vasconcellos, K.L. (1999). Aggregation and Disaggregation of Structural Time Series Models. *Journal of Time Series Analysis,* **20**, 155-171.

Hsiao, C. (1979). Linear Regression Using Both Temporally Aggregated and Temporally Disaggregated Data. *Journal of Econometrics,* **10**, 243-252.

Hsiao, C. (1980). Missing Data and Maximum Likelihood Estimation. *Economics Letters,* **6**, 249-253.

IGE (1997). *Contabilidade Trimestral de Galicia. Metodoloxía e Series Históricas 1980-1991.* Santiago de Compostela: Instituto Galego de Estadística.

INE (1993). *Contabilidad Nacional Trimestral de España. Metodología y Serie Trimestral 1970-1992*. Madrid: Instituto Nacional de Estadística.

ISCO (1965). L'Aggiustamento delle Stime nei Conti Economici Trimestrali. *Rassegna del Valori Interni dell'Istituto,* **5**, 47-52.

ISTAT (1983). I Conti Economici Trimestrali dell'Italia 1970-1982. *Supplemento al Bollettino Mensile di Statistica,* **12**.

ISTAT (1985). I Conti Economici Trimestrali dell'Italia, anni 1970-1984. *Supplemento al Bollettino Mensile di Statistica,* **14**.

Jacobs, J. (2004). 'Dividing by 4': A Feasible Quarterly Forecasting Method?, *CCSO Series* 22, Center for Cyclical and Structural Research, Groningen. Available at http://www.eco.rug.nl/ccso/CCSOseries/ccso22.pdf.

Jones, R. H. (1962). Spectral Analysis with Regularly Missed Observations. *Annals of Mathematical Statistics,* **33**, 455-461.

Jones, R. H. (1980). Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations. *Technometrics,* **22**, 389-395.

Klein L.R. (1958). The Estimation of Distributed Lags. *Econometrica,* **26**, 553-65.

Kohn, R., & Ansley, C.F. (1986). Estimation, Prediction, and Interpolation for ARIMA Models with Missing Data. *Journal of the American Statistical Association,* **81**, 751-761.

Lisman, J.H.C. & Sandee, J. (1964). Derivation of Quarterly Figures from Annual Data. *Journal of the Royal Statistical Society, Ser. C,* **13**, 87-90.

Litterman, R.B. (1983). A Random Walk, Markov Model for Distribution of Time Series. *Journal of Business and Economic Statistics,* **1**, 169-173.

Liu, H., & Hall, S.G. (2001). Creating High-Frequency National Accounts with State-Space Modelling: A Monte Carlo Experiment. *Journal of Forecasting,* **20**, 441-449

Lütkepohl, H. (1984). Linear Transformations of Vector ARMA Processes. *Journal of Econometrics,* **4**, 283-293.

Marcellino, M. (2007). Pooling-Based Data Interpolation and Backdating. *Journal of Time Series Analysis,* **28**, 53–71.

Moauro, F., & Savio, G. (2005). Temporal Disaggregation Using Multivariate Structural Time Series Models. *The Econometrics Journal,* **8**, 214–234

Nasse, P. (1970). Peut-on Suivre L'évolution Trimestralle de la Consommation? *Economie et Statistique,* **8**, 33-52.

Nasse, P. (1973). Le Système des Comptes Nationaux Trimestrels. *Annales de L'Inssée,* **14**, 127-161.

Nelson, P., & Gould, G. (1974). The Stochastic Properties of the Income Velocity of Money. *American Economic Review,* **64**, 405-418.

Nijman, Th., & Palm, F.C. (1985). Series Temporelles Incompletes en Modelisation Macroeconomiques. *Cahiers Du Seminaire d'Econometrie,* **29**, 141-68.

Nijman, Th., & Palm F.C. (1986). The Construction and Use of Approximations for Missing Quarterly Observations: A Model Approach. *Journal of Business and Economic Statistics,* **4**, 47-58.

Nijman, Th., & Palm, F.C. (1988a). Efficiency Gains due to Missing Data Procedures in Regression Models. *Statististical Papers,* **29**, 249–256.

Nijman, Th., & Palm, F.C. (1988b). Consistent Estimation of Regression Models with Incompletely Observed Exogenous Variables. *The Annals of Economics and Statististics,* **12**, 151–175.

Nijman, Th., & Palm, F.C. (1990). Predictive Accuracy Gain From Disaggregate Sampling in ARIMA Models. *Journal of Business and Economic Statistics,* **8**, 189-196.

Norman, D., & Walker, T. (2007) Co-movement of Australian State Business Cycles. *Australian Economic Papers,* **46**, 360–374.

OECD (1966). La Comptabilité Nationale Trimestrelle. *Series Etudes Economiques,* **21**.

OECD (1996). *Sources and Methods used by the OECD Member Countries. Quarterly National Accounts*. Paris: OECD Publications.

Palm, F.C., & Nijman, Th. (1982). Linear Regression using both Temporally Aggregated and Temporally Disaggregated Data. *Journal of Econometrics,* **19**, 333-343.

Palm, F.C., & Nijman, Th. (1984). Missing Observations in the Dynamic Regression Model. *Econometrica,* **52**, 1415-1435.

Parzen, E. (1961). Mathematical Considerations in the Estimation of Spectra. *Technometrics,* **3**, 167-190.

Parzen, E. (1963). On Spectral Analysis with Missing Observations and Amplitude Modulation. *Sankhyä, A***25**, 383-392.

Pavía Miralles, J.M. (2000a): *La Problemática de Trimestralización de Series Anuales*. Valencia: Universidad de Valencia.

Pavía Miralles, J.M. (2000b) Desagregación Conjunta de Series Anuales: Perturbaciones AR(1) Multivariante. *Investigaciones Económicas,* **XXIV**, 727-737.

Pavía, J.M., Cabrer, B., & Felip, J.M. (2000): *Estimación del VAB Trimestral No Agrario de la Comunidad Valenciana*. Valencia: Generalitat Valenciana.

Pavía, J.M., & Cabrer, B. (2003). Estimación Congruente de Contabilidades Trimestrales Regionales: Una Aplicación. *Investigación Económica,* **LXII**, 119-141.

Pavía-Miralles, J.M., & Cabrer-Borrás, B. (2007). On Estimating Contemporaneous Quarterly Regional GDP, *Journal of Forecasting,* **26**, 155-177.

Pavía, J.M., Vila, L.E., & Escuder, R. (2003). On the Performance of the Chow-Lin Procedure for Quarterly Interpolation of Annual Data: Some Monte-Carlo Analysis. *Spanish Economic Review,* **5**, 291-305.

Pinheiro, M., & Coimbra, C. (1993). Distribution and Extrapolation of Time Series by Related Series Using Logarithms and Smoothing Penalties, *Economia,* **17**, 359-374.

Proietti, T. (1998). Distribution and Interpolation Revisited: A Structural Approach. *Statistica,* **LVIII**, 411-432.

Proietti, T. (2006). Temporal Disaggregation by State Space Methods: Dynamic Regression Methods Revisited. *The Econometrics Journal,* **9**, 357–372

Proietti, T., & Moauro, F. (2006). Dynamic Factor Analysis with Nonlinear Temporal Aggregation Constraints. *Journal of the Royal Statistical Society, Ser. C,* **55**, 281-300.

Quenouille, M-H. (1958). Discrete Autoregressive Schemes with Varing Time-Intervals. *Metrika*, **1**. 21-27.

Quilis, E. (2002). *A MATLAB Library of Temporal Disaggregation Methods: Summary*. Madrid: Instituto Nacional de Estadística. http://www.ine.es/.

Quilis, E. (2003). Desagregación Temporal Mediante Modelos Dinámicos: El Método de Santos Silva y Cardoso. *Boletín Trimestral de Coyuntura,* **88**, 1-11.

Quilis, E. (2005). *Benchmarking Techniques in the Spanish Quarterly National Accounts*. Luxembourg: Office for Official Publications of the European Communities.

Rodríguez-Feijoo, S., Rodríguez-Caro A., & Dávila-Quintana, D. (2003). Methods for Quarterly Disaggregation without Indicators; A Comparative Study Using Simulation. *Computational Statistics and Data Analysis,* **43**, 63–78.

Rossana, R.J., & Seater, J.J. (1995). Temporal Aggregation and Economic Time Series. *Journal of Business and Economic Statistics,* **13**, 441-451.

Rossi, N. (1982). A Note on the Estimation of Disaggregate Time Series when the Aggregate is Known. *The Review of Economics and Statistics,* **64**, 695-696.

Salazar, E.L., Smith, R.J., & Weale, R. (1997a). *A Monthly Indicator of GDP*. London: National Institute of Economic and Social Research/NIESR Discussion Papers.

Salazar, E.L., Smith, R.J., & Weale, R. (1997b). *Interpolation using a Dynamic Regression Model: Specification and Monte Carlo Properties*. London: National Institute of Economic and Social Research/NIESR Discussion Papers.

Santos Silva, J.M.C., & Cardoso, F.N. (2001). The Chow-Lin Method Using Dynamic Models. *Economic Modelling*, **18**, 269-80.

Särndal, C.E. (2007) The Calibration Approach in Survey Theory and Practice. *Survey Methodology,* **33**, 99-120.

Sargan, J.D., & Drettakis, E.G. (1974). Missing Data in an Autoregressive Model. *International Economic Review,* **15**, 39-59.

Scheinok, P.A. (1965). Spectral Analysis with Randomly Missed Observations: The Binomial Case. *Annals of Mathematical Statistics,* **36**, 971-977.

Schmidt, J.R. (1986). A General Framework for Interpolation, Distribution and Extrapolation of Time Series by Related Series. In *Regional Econometric Modelling* (pp. 181-194). Boston: Kluwer Nighoff Pub.

Silver, J.L. (1986). Two Results Useful for Implementing Litterman's Procedure for Interpolating a Time Series. *Journal of Business and Economic Statistics,* **4**, 129-130.

Somermeyer, J., Jansen, R., & Louter, J. (1976). Estimating Quarterly Values of Annually Know Variables in Quarterly Relationships. *Journal of the American Statistical Association,* **71**, 588-595.

Stram, D.O., & Wei, W.W.S. (1986). Temporal Aggregation in the ARIMA process. *Journal of Time Series Analysis,* **39**, 279-292.

Telser, L.G. (1967). Discrete Samples and Moving Sums in a Stationary Stochastic Process. *Journal of the American Statistical Association,* **62**, 489-499

Tiao, G.C. (1972). Asymptotic Behavior of Time Series Aggregates. *Biometrika,* **59**, 523-531.

Toloi, C.M.C. & Morettin, P.A. (1993). Spectral Analysis for Amplitude-Modulated Time Series. *Journal of Time Series Analysis,* **14**, 409-432.

Trabelsi, A. & Hillmer, S.C. (1990). Bench-marking Time Series with Reliable Bench-Marks. *Journal of the Royal Statistical Society, Ser. C,* **39**, 367-379.

Vangrevelinghe, G. (1966). L'Evolution à Court Terme de la Consommation des Ménages: Connaisance, Analyse et Prévision. *Etudes et Conjoncture,* **9**, 54-102.

Vanhonacker, W.R. (1990). Estimating Dynamic Response Models when the Data are Subject to Different Temporal Aggregation. *Marketing Letters,* **1**, 125-137.

Weale, M. (1992). Estimation of Data Measured with Error and Subject to Linear Restrictions. *Journal of Applied Econometrics,* **7**, 167-174.

Wei, W.W.S. (1978). Some Consequences of Temporal Aggregation in Seasonal Time Series Models. In Zellner A. (Ed) *Seasonal Analysis of Economic Time Series* (pp. 433-448) Washington, DC: Government Printing Office.

Wei, W.W.S. (1981). Effect of Systematic Sampling on ARIMA Models. *Communications in Statistics,* **A10**, 2389-2398.

Wei, W.W.S., & Stram, D.O. (1990). Disaggregation of Time Series Models. *Journal of the Royal Statististical Society, Ser. B*, **52**, 453–467.

Weiss, A.A. (1984). Systematic Sampling and Temporal Aggregation in Time Series Models. *Journal of Econometrics,* **26**, 271-281

Werner, H.J. (1982). On the Temporal Aggregation in Discrete Dynamical Systems. In Drenick, R.F.. & Kozin, F. (Eds.) *System Modeling and Optimatization* (pp. 819-825). New York: Springer-Verlag.

Zaier, L., & Trabelsi, A. (2007) Polynomial Method for Temporal Disaggregation of Multivariate Time Series, *Communications in Statistics - Simulation and Computation,* **36**, 741-759.

Zani, S. (1970). Sui Criteri di Calcolo Dei Valori Trimestrali di Tendenza Degli Aggregati della Contabilitá Nazionale. *Studi e Ricerche,* **VII**, 287-349.

Zellner, A., & Montmarquette, C. (1971). A Study of Some Aspects of Temporal Aggregation Problems in Econometric Analyses. *The Review of Economics and Statistics*, **53**, 335-342.

In: Economic Forecasting
Editor: Alan T. Molnar, pp. 29-54
ISBN: 978-1-60741-068-3

Chapter 2

ECONOMETRIC MODELLING AND FORECASTING OF PRIVATE HOUSING DEMAND

***James M.W. Wong* [1] *and S. Thomas Ng* [2]**
Department of Civil Engineering
The University of Hong Kong, Pokfulam, Hong Kong

Abstract

Governments, corporations and institutions all need to prepare various types of forecasts before any policies or decisions are made. Particularly, serving as a significant sector of an economy, the importance of predicting the movement of the private residential market is undeniable. However, it is well recognised that the housing demand is volatile and it may fluctuate dramatically according to general economic conditions. As globalisation continues to dissolve boundaries across the world, more economies are increasingly subjected to external shocks. Frequently the fluctuations in the level of housing demand can cause significant rippling effects in the economy as the housing sector is associated with many other economic sectors. The development of econometric models is thus postulated to assist policy-makers and relevant stakeholders to assess the future housing demand in order to formulate suitable policies.

With the rapid development of econometric approaches, their robustness and appropriateness as a modelling technique in the context of examining the dynamic relationship between the housing market and its determinants are evident. This study applies the cointegration analysis as well as Johansen and Juselius's vector error correction model (VEC) model framework to housing demand forecasting in Hong Kong. Volitality of the demand to the dynamic changes in relevant macro-economic and socio-economic variables are considered. In addition, an impulse response function and a variance decomposition analysis are employed to trace the sensitivity of the housing demand over time to the shocks in the macro-economic and socio-economic variables. This econometric time-series modelling approach surpasses other methodologies by its dynamic nature and sensitivity to a variety of factors affecting the output of the economic sector for forecasting purposes, taking into account indirect and local inter-sectoral effects.

---

[1] E-mail address: jmwwong@hkucc.hku.hk; Tel: Int+ (852) 2241 5348; Fax: Int+ (852) 2559 5337. Postdoctoral Fellow, Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong.

[2] E-mail address: tstng@hkucc.hku.hk. Associate Professor, Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong.

Empirical results indicated that that the housing demand and the associated economic factors: housing prices, mortgage rate, and GDP per capita are cointegrated in the long-run. Other key macro-economic and socio-economic indicators, including income, inflation, stock prices, employment, population, etc., are also examined but found to be insignificant in influencing the housing demand. A dynamic and robust housing demand forecasting model is developed using VEC model. The housing prices and mortgage rate are found to be the most important and significant factors determining the quantity demand of housing. Findings from the impulse response analyses and variance decomposition under the VEC model further confirm that the housing price terms has relatively large and sensitive impact on the housing demand, although at different time intervals, on the volume of housing transactions in Hong Kong. Addressing these two attributes is critical to the formulation of both short- and long-term housing policies that could satisfy the expected demand effectively.

The research contributes knowledge to the academic field as currently the area of housing demand forecast using advanced econometric modelling techniques is under-explored. This study has developed a theoretical model that traces the cause-and-effect chain between the housing demand and its determinants, which is relevant to the current needs of the real estate market and is significant to the economy's development. It is envisaged that the results of this study could enhance the understanding of using advanced econometric modelling methodologies, factors affecting housing demand and various housing economic issues.

Keywords: Economic forecasting, housing demand, impulse responses analysis, econometrics, vector error-correction modeling.

Introduction

Economic forecasting is of immense importance as any economic system is a deterministic-stochastic entity of great complexity and vital to the national development for the information age (Hoshmand, 2002). Holden *et al*. (1990) state that forecasts are required for two basic reasons: the future is uncertain; and the full impact of many decisions taken now might not be felt until later. Consequently, accurate predictions of the future would improve the efficiency of the decision-making process. In particular, the knowledge of future demand for products and services is imperative to all industries since it is a prerequisite for any viable corporate strategy (Akintoye and Skitmore, 1994).

Among the many aspects of economic forecasting, demand for residential properties has always been of great interest not only to policy-makers in the government, but also to business leaders and even the public, especially in a country with land scarcity like Hong Kong (HK). Private residential properties make up a major constituent of the private-sector wealth, and play a significant part in the whole economy (Case and Glaester, 2000; Heiss and Seko, 2001). Its large linkage effect on the economy and its anchoring function for most household activities also amplify the financial importance. In addition, housing demand has traditionally been a target for large-scale government interference. Hence, understanding both the short- and long-term future housing demand is a prerequisite for enlightened housing policy.

The Asian financial crisis started in July 1997 has indeed revealed that the overbuilding of housing in HK would cause serious financial distress on the overall economy. Foremost among those taking the brunt of the shock was the real estate brokerage sector. Others who might also be seriously impacted include decorators, lawyers, bankers, retailers, contractors, sellers of construction materials, and inevitably real estate developers (Tse and Webb, 2004).

Not only would the real estate sector be hampered, but it may also give rise to unemployment, deteriorating fiscal revenues (partially due to the drop in land sales) and sluggish retail sales. It is therefore wise and sensible to incorporate the real estate into a full macro-economic model of an economy.

However, models being developed for analysing the housing demand *per se* are limited in reliability because they cannot cater to the full set of interactions with the rest of the economy. A review of several academic papers (Arnott, 1987; Follain and Jimenez, 1985; Smith *et al.*, 1988) reveals the narrow focus of the neoclassical economic modelling of housing demand. These studies have concentrated on the functional forms, one-equation versus simultaneous equation systems, or measurement issues about a limited range of housing demand determinants, principally price and income factors. Some other estimations are made according to a projection of flats required for new households (e.g., population growth, new marriage, new immigrant, etc.) and existing families (e.g., those affected by redevelopment programmes). No doubt the demographic change would have certain implications on housing demand, yet one should not ignore the impacts of economic change on the desire for property transactions if housing units are significantly viewed as investment assets (Lavender, 1990; Tse, 1996).

Consequently, the most feasible research strategy to advance our understanding of housing consumption decisions lies in furthering the modelling of housing demand determinants to include a more conceptually comprehensive analysis of the impact of demographic and economic indicators on housing consumption decisions. However, Baffor-Bonnie (1998) stated that modelling the supply of, or demand for, housing within any period of time may not be an easy task because the housing market is subject to a dynamic interaction of both those economic and non-economic variables.

The choice of a suitable forecasting technique is therefore critical to the generation of accurate forecasts (Bowerman and O’Connell, 1993). Amongst the variety of methodologies, econometric modelling is one of the dominant methodologies of estimating macro-economic variables. Econometric modelling is readily comprehensible and has remained popular with economists and policy-makers because of its structured modelling basis and outstanding forecasting performance (Lütkepohl, 2004). This methodology is also preferred to others because of its dynamic nature and sensitivity to a variety of factors affecting the level and structure of employment, not to mention its ability to take into account the indirect and local inter-sectoral effects (Pindyck and Rubinfeld, 1998). With the rapid development of econometric approaches, their robustness and appropriateness as a modelling technique in the context of examining the dynamic relationship between the housing market and its determinants are evident.

The aim of this study is, through the application of the econometric modelling techniques, to capture the past behaviour and historical patterns of the private housing demand in HK by considering the volatility of the demand to the dynamic changes in macro-economic and socio-economic variables for forecasting purpose. The structure of the paper is as follows: the theoretical background regarding the relationship of the private housing sector and the relevant economic variables is hypothesised in the next section. The research method and data are then presented. The results of the empirical analyses are subsequently discussed prior to concluding remarks.

Housing Demand and Macro-economic Variables

Like any other business sector, the real estate market tends to move in a fluctuating pattern. Contrasting to a standard sine wave in physical science, the characteristics of real estate market fluctuations are typically complicated and they exhibit much more stochastic patterns as shown in Figure 1. Fluctuations in the real estate market do not occur at regular time intervals and do not last for the same periods of time, and each of their amplitudes also varies (Chin and Fan, 2005). As the econometric approach is proposed for developing a housing demand forecasting model, this section first attempts to identify the key determinants of housing demand.

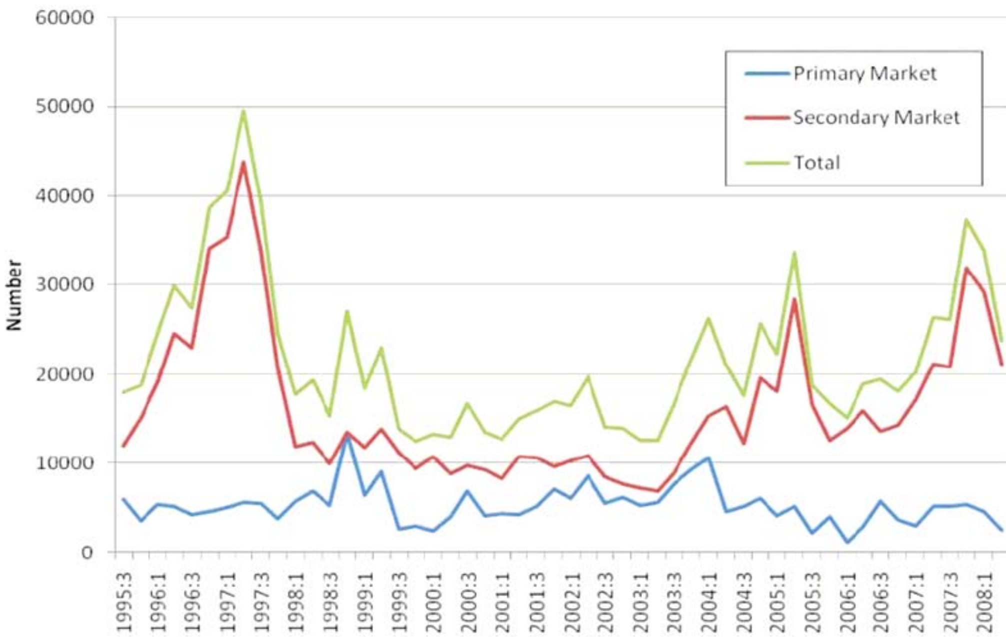


Figure 1. Number of Registrations of Sale and Purchase Agreements of Private Residential Units in HK (1995Q3-2008Q2).

The neoclassical economic theory of the consumer was previously applied to housing (Muth, 1960; Olsen, 1969) which relates to the role of consumer preferences in housing decisions to the income and price constraints faced by the household. The theory postulates that rational consumers attempt to maximise their utility with respect to different goods and services including housing in which they can purchase within the constraints imposed by market prices and their income (Megbolugbe *et al.*, 1991). The general form of the housing demand equation is:

$$Q = f(Y, P_h, P_o) \tag{1}$$

where Q is housing consumption, Y is household income, P_h is the price of housing, and P_o is a vector of prices of other goods and services.

The link between income and housing decision is indisputable for most households. Income is fundamental to explaining housing demand because it is the source of funds for homeowners' payments of mortgage principal and interest, property taxes and other relevant expenses (Megbolugbe *et al.*, 1991). Hendershott and Weicher (2002) stressed that the demand for housing is strongly related to real income. Green and Hendershott (1996) also estimated the household housing demand equations relating the demand to the income and education of the household. Hui and Wong (2007) confirmed that household income Granger causes the demand for private residential real estate, irrespective of the level of the housing stock. Kenny (1999), on the other hand, found that the estimated vector exhibits a positive sensitivity of housing demand to income based on a vector error correction model (VECM). A number of economists agree that permanent income is the conceptually correct measure of income in modelling housing decisions and housing demand. Yet, most economists often use current income in their housing demand equations because of difficulties in measuring permanent income (see Chambers and Schwartz, 1988; Muth, 1960; Gillingham and Hagemann, 1983)

Demand for housing may decline when the housing price increases (Tse *et al.*, 1999). Mankiw and Weil (1989) formulated a simple model which indicates a negative relationship between the US real house price and housing demand. However, in a broader view, trend of property price may also incorporate inexplicable waves of optimism, such as expected income and economic changes, changes in taxation policy, foreign investment flows, etc. (Tse *et al.*, 1999). For example, an expected rise in income will increase the aspiration of home owning as well as the incentive of investing in property, resulting in positive relationship between housing demand and the price.

The principal feature of housing as a commodity that distinguishes it from most other goods traded in the economy are its relatively high cost of supply, its durability, its heterogeneity, and its spatial immobility (Megbolugbe *et al.*, 1991). Initially, neoclassical economic modelling of housing market as shown in Eq. [1] ignored many of these unique characteristics of housing. Indeed, these characteristics make housing a complex market to analyse. Some research considered user costs, especially on how to model the effects of taxes, inflation, and alternative mortgage designs on housing demand decisions.

If interest rate in the economy falls while everything else being equal, the real user cost of a unit of housing services shall fall and the quantity of housing services demanded may rise. Follain (1981) demonstrated that at high interest rates, the household's liquidity constraints tend to dampen housing demand. Kenny (1999) also found that the estimated vector exhibits a negative sensitivity of housing demand to interest rates. Harris (1989) and Tse (1996), however, demonstrated that a declining real interest rate tends to stimulate house prices and thereby lead to decreases in rent-to-value ratio and housing demand.

Housing demand also depends on the inflation rate in a fundamental way (Hendershott and Hu 1981, 1983). As inflation rises, more investors are drawn into the property market, expecting continued appreciation to hedge against inflation (Kenny, 1999). Tse *et al*. (1999) stressed that the housing demand should therefore include those with adequate purchasing power to occupy available housing units as well as those desires to buy a house for renting or price appreciation. For instance, the inflation experienced by HK in the early 1990s was a period of rising speculative activities in the housing market. In addition, owner-occupied housing is largely a non-taxed asset and mortgage interest is partially deductible (Hendershott and White, 2000). As a result, when inflation and thus nominal interest rates rise, the tax subsidy reduces the real after-tax cost of housing capital and increases aggregate housing demand (Summers, 1981; Dougherty and Van Order, 1982). Harris (1989) suggested that housing consumers tend to respond to declining real costs rather than rising nominal costs. In this context, consumers' expectations about price appreciation and inflation are supposed to be an important factor in determining the demand for housing.

The findings of previous research studies (e.g. Killingsworth, 1990; Akintoye and Skitmore, 1994) realised that the building and business cycles are closely related. Bon (1989) related building cycles to business or economic cycles and postulated how economic fluctuations affect fluctuations in building activity. Swings in the general economy and stock market may thereby be treated as indicators of the prospective movement in the housing market and *vice versa* (Ng *et al*., 2008). Maclennan and Pryce (1996) also suggested that economic change shapes the housing system and that recursive links run back from housing to the economy. Housing investment is a sufficiently large fraction of total investment activity in the economy (about a third of total gross investment) to have important consequences for the economy as a whole and *vice versa* (Pozdena, 1988, p. 159).

One of the crucial foundations for residential development is employment, which serves not only as a lead indicator of future housing activity but also as an up-to-date general economic indicator (Baffor-Bonnie, 1998). The decrease in the employment that results from this process tends to reduce the demand for new housing. The macro implications for real estate activity and employment have been explored at some length in the literature, and the general consensus is that the level of employment growth tends to produce real estate cycles (Smith and Tesarek, 1991; Sternlieb and Hughes, 1977). Baffor-Bonnie (1998) applied a nonstructural vector autoregressive (VAR) model to support earlier studies that employment changes explain real estate cycles of housing demand.

In addition, a number of studies consistently have shown that housing demand is also driven mainly by demographic factors in a longer term (Rosen, 1979; Krumm, 1987; Goodman, 1988; Weicher and Thibodeau, 1988; Mankiw and Weil, 1989; Liu *et al*., 1996). Population growth captures an increase in potential housing demand, especially if the growth stems mainly from the home buying age group with significant income (Reichert, 1990). In a separate study, Muellbauer and Murphy (1996) also showed that demographic changes together with the interest rate are the two important factors causing the UK house price boom of the late 1980s. They found that demographic trends were favourable, with stronger population growth in the key house buying age group. Tse (1997), on the other hand, argued that in the steady state, rate of construction depends mainly upon the rate of household formation. Growth of population and number of households are proposed to be included as independent variables in the econometric study.

As discussed above based on a comprehensive literature of modelling specifications, the demand for housing services can be derived by assuming utility maximisation on the part of homeowners and wealth maximisation on the part of investors. The specific factors that determine the demand for housing have been previously identified and are summarised in Eq. [2].

$$Q = f(Y, P_h, MR, CPI, GDP, P_s, U, POP) \quad (2)$$

where

Q represents the quantity of housing sold;
Y is real household income;
P_h is the real price of housing;
MR measures the real mortgage interest rates;
CPI is the consumer price index to proxy inflation;
GDP is the real gross domestic product;
P_s is the stock prices proxied by the Hang Seng Index;
U is the unemployment rate; and
POP is the total resident population.

Methodology

The Econometric Model

In light of the above predictions that there will be a sluggish adjustment on the housing demand, any empirical attempt to model the housing market must clearly distinguish the long-run from the short-run information in the data. Recent advances in econometrics, in particular the development of cointegration analysis and vector error correction (VEC) model, have proven useful in help distinguishing an equilibrium as opposed to a disequilibrium relationships among economic variables (Kenny, 1999). Adopting simple statistical methods such as regression or univariate time series analysis like auto-regressive integrated moving average (ARIMA) models may be only reliable for the short-term forecast of economic time series (Tse, 1997) and may give rise to large predictive errors as they are very sensitive to 'noise' (Quevedo *et al.*, 1988; Tang *et al.*, 1991).

This study employs the Johansen cointegration technique in order to assess the extent to which the HK housing market possesses the self-equilibrating mechanisms discussed above, i.e. a well behaved long-run housing demand relationships. The HK market provides a particularly interesting case study because there have been large-scale fluctuations in the price of owner occupied dwellings over recent years. The econometric analysis takes an aggregate or macro-economic perspective and attempt to identify equilibrium relationships using key macro variables. In particular, the analysis will examine: (i) the impact of monetary policy, i.e. interest rates, on developments in the housing market; (ii) the effects of rising real incomes on house prices; (iii) the nature and speed of price adjustment in the housing market; (iv) effect of demographical change to the demand for housing; and (v) the nature and speed of stock adjustment in the housing market.

The Johansen multivariate approach to cointegration analysis and VEC modelling technique seems particularly suitable for the analysis of the above relationship as shown in Eq. [2] because it is a multivariate technique which allows for the potential endogeneity of all variables considered (Kenny, 1999). In common with other cointegration techniques, the objective of this procedure is to uncover the stationary relationships among a set of non-stationary data. Such relationships have a natural interpretation as long-run equilibrium relationships in economic sense. VEC is a restricted vector autoregressive (VAR) that has cointegration restrictions built into specification (Lütkepohl, 2004). The VEC framework developed by Johansen (1988) and extended by Johansen and Juselius (1990) provides a multivariate maximum likelihood approach that permits the determination of the number of cointegration vectors and does not depend on arbitrary normalisation rules, contrary to the earlier error correction mechanism proposed by Engle and Granger (1987).

The Johansen and Juselius's VEC modelling framework is adopted to the housing demand forecasting because of its dynamic nature and sensitivity to a variety of factors affecting the demand, and its taking into account indirect and local inter-sectoral effects. Applying conventional VAR techniques may lead to spurious results if the variables in the system are nonstationary (Crane and Nourzad, 1998). The mean and variance of a nonstationary or integrated time series, which has a stochastic trend, depend on time. Any shocks to the variable will have permanent effects on it. A common procedure to render the series stationary is to transform it into the first differences. Nevertheless, the model in its first difference level will be misspecified if the series are cointegrated and converged to stationary long-term equilibrium relationships (Engle and Granger, 1987). The VEC specification allows investigating the dynamic co-movement among variables and the simultaneous estimation of the speed with which the variables adjust in order to re-establish the cointegrated long-term equilibrium, a feature unavailable in other forecasting models (Masih, 1995). Such estimates should prove particularly useful for analysis of the effect of alternative monetary and housing market policies. Empirical studies (e.g. Anderson *et al*., 2002; Darrat *et al*., 1999; Kenny, 1999; Wong *et al*., 2007) have also shown that the VEC model achieved a high level of forecasting accuracy in the field of macro-economics.

The starting point for deriving an econometric model of housing demand is to establish the properties of the time series measuring the demand and its key determinants. Testing for cointegration among variables was preceded by tests for the integrated order of the individual series set, as only variables integrated of the same order may be cointegrated. Augmented Dickey-Fuller (ADF) unit root tests were employed which was developed by Dickey and Fuller (1979) and extended by Said and Dickey (1984) based on the following auxiliary regression:

$$\Delta y_t = \alpha + \delta t + \gamma \ y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + u_t \tag{3}$$

The variable Δy_{t-i} expresses the lagged first differences, μ_t adjusts the serial correlation errors, and α, β and γ are the parameters to be estimated. This augmented specification was used to test for $H_0 : \gamma = 0$ vs. $H_a : \gamma < 0$ in the autoregressive (AR) process.

The specification in the ADF tests was determined by a 'general to specific' procedure by initially estimating a regression with constant and trend, thus testing their significance. Additionally, a sufficient number of lagged first differences were included to remove any serial correlation in the residuals. In order to determine the number of lags in the regression, an initial lag length of eight quarters was selected, and the eighth lag was tested for significance using the standard asymptotic *t*-ratio. If the lag is insignificant, the lag length is reduced successively until a significant lag length is obtained. Critical values simulated by MacKinnon (1991) were used for the unit root tests.

Cointegration analysis and VEC model were then applied to derive housing demand specification. The econometric model attempts to link housing demand to variables in equilibrium identified with economic theory. Although many economic time series may have stochastic or deterministic trend, the groups of variables may drift together. Cointegration analysis allows the derivation of long-run equilibrium relationships among the variables. If the economic theory is relevant, it is expected that the specific set of suggested variables are interrelated in the long run. Hence, there should be no tendency for the variables to drift apart increasingly as time progresses, i.e. the variables in the model form a unique cointegrating vector.

To test for the cointegration, the maximum likelihood procedures of Johansen and Juselius were employed. Suppose that the variables in the housing demand function are in the same integrated order, these variables may cointegrate if there exists one or more linear combinations among them. A VAR specification was used to model each variable as a function of all the lagged endogenous variables in the system. Johansen (1988) suggests that the process y_t is defined by an unrestricted VAR system of order (p):

$$y_t = \delta + \Gamma_1 y_{t-1} + \Gamma_2 y_{t-2} + \ldots + \Gamma_p y_{t-p} + u_t \; t = 1, 2, 3, \ldots, T \tag{4}$$

where y_t are I(1) independent variables, Γ's are estimable parameters, and $u_t \sim \text{niid}(0, \Sigma)$ is vector of impulses which represent the unanticipated movements in y_t. However, such a model is only appropriate if each of the series in y_t is integrated to order zero, I(0), meaning that each series is stationary (Price, 1998). Using $\Delta = (I - L)$, where L is the lags operator, the above system can be reparameterised in the VEC model as:

$$\Delta y_t = \delta + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + u_t \tag{5}$$

where

$$\Pi = \sum_{i=1}^{\bar{p}} A_i - I \, , \; \Gamma_i = -\sum_{j=i+1}^{p} A_j \tag{6}$$

Δy_t is an $I(0)$ vector, δ is the intercept, the matrix Γ reflects the short-run aspects of the relationship among the elements of y_t, and the matrix Π captures the long-run information. The number of linear combinations of y_t that are stationary can be determined by the rank of Π, which is denoted as r. If there are k endogenous variables, Granger's representation theorem asserts that if the coefficient matrix Π has reduced rank $r < k$, then there exists $k \times r$ matrices, α and β, each with rank r such that $\Pi = \alpha \beta'$ and $\beta' y_t$ is stationary.

The order of r is determined by trace statistics and the maximum eigenvalue statistics. The trace statistic tests the null hypothesis of cointegrating relations r against the alternative of k cointegrating relations, where k is the number of endogenous variables, for $r = 0, 1, \ldots, k-1$. The alternative of k cointegrating relations corresponds to the case where none of the series has a unit root and a stationary VAR may be specified in terms of the levels of all of the series. The trace statistic for the null hypothesis of r cointegrating relations is computed as:

$$LR_{tr}(r|k) = -T \sum_{i=r+1}^{k} \log(1 - \lambda_i) \tag{7}$$

for $r = 0, 1, \ldots, k$-1 where T is the number of observation used for estimation, and λ_i is the i-th largest estimated eigenvalue of the Π matrix in Eq. [6] and is the test of $H_0(r)$ against $H_1(k)$.

The maximum eigenvalue statistic tests the null hypothesis of r cointegrating relations against the alternative of r+1 cointegrating relation. This test statistic is computed as:

$$LR_{tr}(r|r+1) = -T\log(1-\lambda_{r+1}) \\ = LR_{tr}(r|k) - LR_{tr}(r+1|k) \tag{8}$$

for $r = 0, 1, \ldots, k$-1.

The models will be rejected where Π has a full rank, i.e. $r = k$–1 since in such a situation y_t is stationary and has no unit root, thus no error-correction can be derived. If the rank of Π is zero, this implies that the elements of y_t are not cointegrated, and thus no stationary long-run relationship exists. As a result, the conventional VAR model in first-differenced form shown in Eq. [4] is an alternative specification.

The choice of lag lengths in cointegration analysis was decided by multivariate forms of the Akaike information criterion (AIC) and Schwartz Bayesian criterion (SBC). The AIC and SBC values[3] are model selection criteria developed for maximum likelihood techniques. In minimising the AIC and SBC, the natural logarithm of the residual sum of squares adjusted for sample size and the number of parameters included are minimised. Based on the assumption that Π does not have a full rank, the estimated long-run housing demand in HK can be computed by normalising the cointegration vector as a demand function.

While the cointegrating vectors determine the steady-state behaviour of the variables in the vector error correction model, the dynamic representation of the housing demand to the underlying permanent and transitory shocks were then completely determined by the sample data without restriction. One motivation for the VEC model(p) form is to consider the relation $\beta' y_t$ = c as defining the underlying economic relations and assume that the agents react to the disequilibrium error $\beta' y_t$ – c through the adjustment coefficient α to restore equilibrium; that is, they satisfy the economic relations. The cointegrating vector, β are the long-run parameters (Lütkepohl, 2004).

Estimation of a VEC model proceeded by first determining one or more cointegrating relations using the aforementioned Johansen procedures. The first difference of each endogenous variable was then regressed on a one period lag of the cointegrating equation(s) and lagged first differences of all of the endogenous variables in the system. The VEC model can be written as the following specification:

$$\Delta d_t = \delta + \alpha(\beta' y_{t-1} + \rho_0) + \sum_{i=1}^{p}\gamma_{1,i}\Delta y_{1,t-i} + \sum_{i=1}^{p}\gamma_{2,i}\Delta y_{2,t-i} + \ldots\ldots + \sum_{i=1}^{p}\gamma_{j,i}\Delta y_{j,t-i} + u_t \tag{9}$$

where y_t are $I(1)$ independent variables, d is the quantity of housing sold, α is the adjustment coefficient, β are the long-run parameters of the VEC function, and $\gamma_{j,i}$ reflects the short-run aspects of the relationship between the independent variables and the target variable.

[3] AIC = T ln (residual sum of squares) + 2k; SBC = T ln (residual sum of squares) + $k$ln(T) where T is sample size and k is the number of parameters included

A well known problem with VARs, and particularly important in the identification of a VEC model, is the prohibitively large number of parameters. Each equation involves estimating $k \times r$ lag coefficients plus one or more parameters for the deterministic components, and would quickly exhaust typical samples for macro-econometric research. With limited observations, the VEC approach quickly runs into the problem of severe lack of degrees of freedom. One way to address the over-parameterisation problem is to test and impose weak exogeneity assumptions. Monte Carlo results in Hall *et al*. (2002) also reveal that imposing valid weak exogeneity restrictions before testing for the cointegrating rank generally improves the power of Johansen rank tests. When the parameters of interest are the cointegrating vector β', y_t is weakly exogenous if and only if the i-th row of the α matrix is all zero with respect to the β parameters $\alpha y = 0$ (Johansen, 1991, see also Johansen (1995) for the definition and implications of weak exogeneity). Following Hall *et al*. (2002), once weak exogeneity restrictions are tested and imposed, Johansen rank tests are conducted.

Hendry and Juselius (2000), on the other hand, emphasise the importance of correct specification. If the future housing demand is not driven by the past values of the independent variables, it is more appropriate to model the demand separately from non-causal variables. The existence of a cointegrating relationship among the variables suggests that there must be unidirectional or bidirectional Granger causality. In this case, a VEC model should be estimated rather than a VAR as in a standard Granger causality test (Granger, 1988). Sources of causation can be identified by testing for significance of the coefficients on the independent variables in Eq. [7] individually.

On one hand, for instance by testing H_0: $\gamma_{2,i} = 0$ for all i, y_2 Granger weak causes housing demand can be evaluated in the short run (Asafu-Adjaye, 2000). This can be implemented by using a standard Wald test. On the other hand, long-run causality can be found by testing the significance of the estimated coefficient of α by a simple t-test. The strong Granger-causality for each independent variable can be exposed by testing the joint hypotheses H_0: $\gamma_{2,i} = 0$ and $\alpha = 0$ for all i in Eq. [6] by a joint F-test. Similar reasoning is possible for examining whether other variables Granger-cause the housing demand. In addition, in estimating the VEC model, as explained in Hendry (1995), we first use five lags of the explanatory variables, i.e. estimated unrestricted ECM. The final parsimonious causality structure of the VEC model is then established by eliminating cointegrating vectors with insignificant loading parameters.

Various diagnostic tests were applied to assess the robustness and reliability of the developed models. These included the Lagrange multiplier tests (LM) for up to respectively one and forth order serial correlation in the residuals, White's test (White, 1980) for heteroscedasticity (H) in the residuals and for model misspecification, the Jarque-Bera test for normality (NORM) of the residuals (Jarque and Bera, 1980). The forecasts were also verified by comparing the projections generated from the ARIMA model which served as a benchmark. EViews (version 6.0) was used as the statistical tool for modelling the housing demand.

Impulse Response Analysis

Traditionally, researchers have used structural models to analyse the impact of the unanticipated policy shocks, i.e. the policy and other macro-economic variables. Such models, however, impose *a priori* restrictions on the coefficients. These restrictions may inhibit researchers from revising the macro model, even when the data or historical evidence points to such a need. Hence, once the Granger causality relationship has been established, the impulse response function (IRF) is used to trace the effects of a one standard deviation shock to one of the innovations on current and future values of the endogenous variables, through the dynamic structure of the VEC specification.

The IRF is able to detect the sensitivity of the housing demand over time to the shocks in the macro-economic and socio-economic variables (QMS, 2000). It allows the data, rather than the researcher, to specify the dynamic structure of the model (Pindyck and Rubenfeld, 1998). Through normalisation, the IRF produces a fluctuating graph that has a zero line in the centre representing the equilibrium condition of the response variable. Additionally, variance decomposition (VDC) is also applied to trace the response of the endogenous variable to such shock. It provides a different method of depicting the system dynamics. While IRF traces the effects of a shock to one endogenous variable on the other variables in the VAR, VDC separates the variation in an endogenous variable into the component shocks to the VAR. In other words, the variance decomposition gives information about the relative importance of each random innovation in affecting the variables in the VAR.

Sources of Data

This section presents the time series variables used to examine the stipulated relations. Data related to the real estate market were obtained from *Hong Kong Property Review* (various issues). Other data were extracted from relevant publications issued by the Census and Statistics Department (C&SD) of the HKSAR Government and other official sources. Table 1 shows the definitions and data sources used for variables in econometric modelling. These series of data cover from the third quarter of 1995 to the second quarter of 2008, giving a total of 52 quarterly observations. Where the indicators recorded in the official data source were not exactly in quarterly form, it was necessary to estimate the quarterly figures by either aggregation or interpolation of the figures involved.

The financial variables are measured in real terms so as to measure the differences in real consumption and investment behaviour. The model is estimated in log-linear form. The major advantage of the log-linear transformation is the effect in regression computation of decreasing the relative size of observations with large variances, especially when a regression includes variables in both level and difference forms.

Table 1. Definitions and Data Sources Used for Variables in Empirical Work

Variable	Definition	Data Source
Housing Demand (Q)	Number of registrations of sale and purchase agreements of private residential units (both primary & secondary markets)	Property Review, Rating and Valuation Department
Income (Y)	Real indices of payroll per person engaged (1st Qtr 1999 = 100)	General Household Survey, Census and Statistics Department
Housing Prices (P_h)	Private domestic price index, all classes (1999=100)	Property Review, Rating and Valuation Department

Table 1. Continued

Variable	Definition	Data Source
Mortgage Rates (MR)	Best lending rate, inflation adjusted (% per annum)	Hong Kong Monetary Authority
Inflation (CPI)	Year-on-year % change of Consumer Price Index (A), based on the expenditure pattern of around 50 percent of households in HK	Monthly Digest of Statistics, Census and Statistics Department
GDP per Capita (GDP)	Real GDP per capita. Real per capita GDP is employed as the relevant economic development variable and not real GDP, since it discounts the effects of population growth.	Monthly Digest of Statistics, Census and Statistics Department
Stock Prices (P_s)	Heng Sang index	Hong Kong Monetary Authority
Unemployment (U)	Unemployment rate in HK	General Household Survey, Census and Statistics Department
Population (POP) / Number of Households (HH)	Total population in HK / Total number of domestic households	General Household Survey, Census and Statistics Department

Model Specification and Implications

Unit Root Tests

ADF tests were initially conducted to determine the integrated order of the relevant data series. Table 2 reports the results of the unit root tests. These statistics indicate that a unit root can be rejected for the first difference but not the levels for all variables at the 5% significance level, except real income and population. Thus, the housing demand, housing prices, mortgage interest rate, inflation, GDP, stock prices, and the unemployment rate are integrated of order one i.e. *I*(1) series. It is thus justified to test the long-term relationship among these *I*(1) variables using cointegration analysis for modelling and forecasting purposes.

Table 2. ADF Unit Root Tests

Variable	Test statistics	Critical values	Variable	Test statistics	Critical values
q	-2.6836 [C,1]	-2.9200	Δq	-8.4532 [1]**	-1.9475
y	-2.2695 [C,4]	-2.9266	Δy	-3.4951 [C,T,3]	-3.5107
			$\Delta^2 y$	-73.1525 [3] **	-1.9481
p_h	0.1567 [1]	-1.9475	Δp_h	-3.1082 [1] *	-1.9475
MR	-1.0190 [C,T,1]	-3.5005	ΔMR	-6.0226 [C,T,1] **	-3.5024
CPI	-0.7762 [C,T,1]	-3.5005	ΔCPI	-6.5707 [C,T,1] **	-3.5024
gdp	-2.2330 [C,T,7]	-3.5155	Δgdp	-5.0362 [C,T,6] **	-3.5155

Table 2. Continued

Variable	Test statistics	Critical values	Variable	Test statistics	Critical values
p_s	-1.1022 [1]	-1.9474	$\Delta\, p_s$	-6.9480 [1] **	-3.5155
U	-2.3309 [C,3]	-2.9238	$\Delta\, U$	-2.5217 [2] *	-1.9478
pop	-5.1357 [C,T,1] **	-3.5024			
hh	-5.7881 [C,T,2] **	-3.5043			

Note: *q*, $\log_e$ of housing demand; *y*, $\log_e$ of real income; p_h, $\log_e$ of housing prices; *MR*, mortgage interest rate; *CPI*, year-on-year % change of inflation; *gdp*, $\log_e$ of gross domestic product; p_s, $\log_e$ of stock prices; *U*, unemployment rate; *pop*, $\log_e$ of population; *hh* $\log_e$ of number of domestic households. Δ is the first difference operator. The content of the brackets [·] denotes constant, trend and the order of augmentation of the ADF test equation, respectively; * Rejection of the null at the 5% significance level; ** Rejection of the null at the 1% significance level.

Weak Exogeneity

The existence of long-run weak exogeneity among the vector of seven variables, $y_t = (q, p_h, MR, CPI, gdp, p_s, U)$ as identified previously was examined by performing the LR tests (Johansen and Juselious, 1992). We hypothesise that the three macro-economic variables, $y_t = (q, p_h, MR)$, are endogenous, and the remaining external factors, $y_t = (CPI, gdp, p_s, U)$, are exogenous. Following the strategy outlined in the methodology section, we test the null hypothesis, H_0: $\alpha y = 0$ for each candidate exogenous variable.

Table 3 shows the results of the weak exogeneity tests. We cannot reject the weak exogeneity of unemployment rate (*U*) and inflation (*CPI*) in the system: tests for both variables have p-values in excess of 10%. In contrast, weak exogeneity of stock prices (p_s) is strongly rejected at the 10% level. However, p_s was eliminated subsequently from the VEC model as the adjustment coefficient to the long-run equilibrium (α) included in the VEC model is statistically insignificant at the 10% level. These restrictions can be imposed in the restricted VEC.

Table 3. Weak ExogeneityT

Variable	χ^2	p-value
CPI	0.0725	0.7878
gdp	2.7775	0.0956
p_s	7.7712	0.0053
U	0.9208	0.3373

Note: *gdp*, $\log_e$ of gross domestic product; p_s, $\log_e$ of stock prices; *U*, unemployment rate

Cointegration Tests

Given the results of unit roots and weak exogeneity tests, Johansen's techniques were applied to test for cointegration between the y variables (q, p_h, MR, gdp) within a VEC model as specified in Eq. [5]. In implementing the Johansen procedure, it was assumed that series y have stochastic linear trends but the cointegrating equations have an intercept. This option is based on the proposition that long-run equilibrium in housing demand probably has no significant trend. The omission of the trend term is also justified by the result of testing the significance of this term in the cointegrating relation. In addition, based on the smallest AIC and SBC values, the lag-length was selected as five and the results of the cointegration tests based on the trace statistics and the max-eigen statistics are reported in Table 4. The results indicate that there is not more than one cointegrating relation, while the test rejects $r = 0$ for the alternative that $r = 1$ at the 5% significance level. It is, therefore, concluded that one cointegration relation exists among the selected variables, i.e. $r = 1$.

Table 4. Johansen Cointegration Rank Test

H_0	H_a	λ_{Trace}	Prob.**	λ_{Max}	Prob.**
$r = 0$	$r = 1$	65.5568	0.0005*	34.6904	0.0052*
$r \leq 1$	$r \geq 2$	17.8665	0.2375	15.4687	0.2574
$r \leq 2$	$r \geq 3$	15.3978	0.0617	14.2646	0.0639
$r \leq 3$	$r \geq 4$	0.7773	0.3780	0.777329	0.3780

Note: Variables q, p_h, MR, gdp, Maximum lag in $VAR = 5$

* denotes rejection of the hypothesis at the 0.05 level

**Mackinnon-Haug-Michelis (1999) p-values.

By normalising the cointegration vector of housing demand as a demand function, the estimated long-run housing demand in HK implied by the Johansen estimation is given by Eq. 10, with absolute asymptotic t-ratios in parentheses.

$$q = 16.9075 + \underset{(5.0543)}{2.3472}\, p_h - \underset{(4.9294)}{0.1998}\, MR + \underset{(2.9475)}{1.5139}\, gdp \qquad (10)$$

where q is $\log_e$ of housing demand; p_h is $\log_e$ of housing prices; MR is mortgage interest rate; gdp is $\log_e$ of gross domestic product.

The results show that the long-run equilibrium equation is valid given that the independent variables contribute significantly to the cointegrating relationship at the 5% significance level. The coefficient estimates in the equilibrium relation indicate the estimated long-run elasticity with respect to local housing demand, showing the presence of an elastic and positive link with housing prices and gross domestic product, and negative but inelastic relationship with the interest rate. Ong and Teck (1996) and Ho and Cuervo (1999) suggested that the gross domestic product (GDP) indicating the economic development, as well as income of the general public and the prime lending rate are the key determinants of private residential real estate demand. Introducing strong measures and effective strategy focusing on these aspects, land supply and monetary policies for instance, may be regarded as the mechanism for inducing changes in housing demand and thereby contributing to the development of the economy.

Interestingly, the strong relationship between the housing price and demand reflects that upsurge of the real price has a positive impact on the induction of demand in the private residential sector which contradicts the neoclassical economic theory. However, evidence suggests that private house prices drive sales of houses and not the other way around. As clearly illustrated in Salo (1994), the theoretical sign of the effect of various variables on housing demand is contingent on whether or not there are any restrictions facing borrowers in the capital market. In a market where credit is not restricted, it can be shown that housing is an 'ordinary' good in the sense that housing demand is negatively related to real house prices. However, current and future income and expected capital gains from investment tend to have an unambiguously positive effect on housing demand when credit is available. Tse *et al.* (1999) also argued that an increasing trend of property price may induce optimism for end-users and investors, i.e. a rising price is often a positive indicator of an assets' attraction in the market, anticipating continued appreciation (Harris, 1989). Under such conditions, housing is not necessarily a decreasing function of the price of housing. The results indeed support the notion that used housing consumers' expectations about appreciation are a significant factor in home-buying. House prices should, therefore, have a positive effect on transactions. Conversely, the demand for housing is a negative function of the interest rate because a higher interest rate increases the cost of consuming housing services mortgage interest costs rise.

Vector Error-Correction Model

As a cointegrating relationship has been found among the variables, the cointegration series can be represented by a VEC model according to the Granger representation theorem (Engle and Granger, 1987). In the VEC model, deviation of housing demand from its long-run equilibrium path will, in the short-term, feed on its future changes in order to force its movements towards the equilibrium state.

According to the VEC specification shown in Eq. [9], the proposed VEC model for the private residential demand in HK can be written as:

$$\Delta q_t = \delta + \alpha(\beta' y_{t-1} + \rho_0) + \sum_{i=1}^{5} \gamma_{1,i} \Delta p_{h,t-i} + \sum_{i=1}^{5} \gamma_{2,i} \Delta MR_{t-i} + \sum_{i=1}^{5} \gamma_{3,i} \Delta gdp_{t-i} + u_t \qquad (11)$$

where α is the adjustment coefficient, β represent the long-run parameters of the VEC function, and $\gamma_{j,i}$ reflects the short-run aspects of the relationship between the independent variables and the target variable.

Table 5 reports the error correction model's estimates for the housing demand. Further to the long-term relationships among the variables, the coefficients capturing the short-run dynamics are shown in the table, together with a test statistic for the significance of each estimated parameter. It is worth noticing that, the lagged variables have a significant role to play to explain the dynamic changes of housing demand, indicating a lag effect due to the changes of the independent variables in the housing market. The VEC model specification has the capability to explain 73% of the housing demand variations over the study period.

The VEC specification was used to test the Granger causality of the explanatory variables. Table 6 shows the results of the Granger-causality tests. Applying Wald tests and the joint F-tests, the null hypotheses that the independent variables do not Granger-cause the housing demand can be rejected at the 1% significance level. In addition, the significance of the coefficient of α also suggests that the independent variables Granger-cause a deviation of the housing demand from the long-run equilibrium in the previous quarter. Therefore, it is concluded that all the variables Granger-cause the housing demand, implying that the past values of these variables are useful to forecast the demand in both short-run and long-run.

Table 5. Estimation Results: Vector Error Correction (VEC) Model of the Housing Demand in Hong Kong

Variables	Δq_t				
δ	0.0420	(0.5074)			
α	0.1395	(1.0031)			
q_{t-1}	1				
$p_{h,t-1}$	-2.3472	(-5.0543)***			
MR_{t-1}	0.1998	(4.9294)***			
gdp_{t-1}	-1.5139	(-2.9475)***			
ρ_0	-16.9075				
	t-1	*t-2*	*t-3*	*t-4*	*t-5*
Δq	-0.3558 (-1.1569)	-0.6201 (-1.9323)*	-0.0415 (-0.1538)	-0.0979 (-0.3514)	-0.1076 (-0.4657)
Δp_h	3.0532 (1.9987)*	1.8561 (1.2138)	-5.3658 (-3.6620)***	2.8612 (1.7290)	0.5350 (0.4403)
ΔMR	-0.01663 (-0.4460)	0.01602 (0.5827)	-0.0879 (-2.4095)**	-0.0052 (-0.1238)	0.0758 (1.7316)*
Δgdp	-3.9437 (-1.5828)	0.4457 (0.1622)	-4.5839 (-1.7630)*	0.8120 (0.3160)	1.8212 (0.8681)

R-squared	0.7353	***	t-statistic significant at .01 level
Sum sq. resids	0.7689	**	t-statistic significant at .05 level
S.E. equation	0.1790	*	t-statistic significant at .1 level
Log likelihood	28.8316		

Note: q, is $\log_e$ of housing demand; p_h, is $\log_e$ of housing prices; MR, is mortgage interest rate; gdp, is $\log_e$ of gross domestic product; t-statistics in parentheses.

Table 6. Results of Granger-causality Tests based on the VEC Model

	Weak Granger-causality		**Strong Granger-causality**	
Null hypotheses	**Chi-square**	**Probability**	**F-statistics**	**Probability**
Housing price does not Granger-cause housing demand	196.9479	0.0000	33.9637	0.0000
Mortgage interest rate does not Granger-cause housing demand	42.6609	0.0000	10.7348	0.0002
GDP does not Granger-cause housing demand	108.7172	0.0000	20.1440	0.0000

The respective error correction term (α) reveals how the independent variables respond to the housing demand pressure in precisely the way anticipated for achieving long-term demand equilibrium. The adjustment coefficient is found to be positive but not statistically significant at the 5% level which implies that the adjustment process in the housing market is somewhat precarious and insensitive. Since the error correction term measures the deviation from the empirical long-run relationship, this result reveals that the cointegrated variables are not adjusting to the long-run equilibrium state i.e. inflows are weakly exogenous to the system. There are, however, several reasons to expect that the housing market will often be characterised by significant deviations from this long-run market demand. Principally, in light of the large transactions costs which are typically involved in buying a home, there will be significant adjustment lags on the demand side of the market (Kenny, 1999). As a result, economic agents will only adjust slowly toward their desired stock of housing following a change in exogenous demand-side variables.

Model Verification

Various diagnostic tests on the residuals of the VEC model were applied to detect any significant departure from the standard assumptions. These included the Lagrange multiplier tests (LM) for up to respectively one and forth order serial correlation in the residuals, White's test (White, 1980) for heteroscedasticity (H) in the residuals and for model misspecification, the Jarque-Bera test for normality (NORM) of the residuals (Jarque and Bera, 1980). The results of the diagnostic tests reported in Table 7 indicate that the residuals from the estimated VEC model pass the tests at 95% significance levels, and hence, there is no significant departure from the standard assumptions. The model's predictive ability was also verified using Chow's second test. Therefore, there is no evidence of problems related to serial correlation, heteroscedasticity, non-normal errors, instable parameters, or predictive failure.

Table 7. Diagnostic Tests of the Estimated VEC Model

Diagnostics	Statistics	
LM(1)	18.5478	(0.2928)
LM(4)	12.3448	(0.6991)
H	433.7705	(0.3110)
CHOW	1.4409	(0.1905)
NORM	0.66128	(0.7183)

Note: LM(p) is the Lagrange multiplier test for residual serial correlation with p lag length; H is White's test for heteroscedasticity; NORM is Jarque-Bera test for normality of the residuals; CHOW is Chow's second test for predictive failure by splitting the data at 1st quarter 2004; and figures in parentheses denote probability values.

Further to the diagnostic tests, Figure 2 shows graphically the demand (Δq_t) estimation generated from the forecasting model and the actual housing demand over the *ex post* simulation period i.e. 1995Q3–2008Q2, indicating adequate goodness of fit of the developed VEC model. Hence, the results of the diagnostic tests and the evaluation of forecasts verify that the developed VEC model is adequately efficient and robust to forecast the housing demand, both short-term and long-term, in HK.

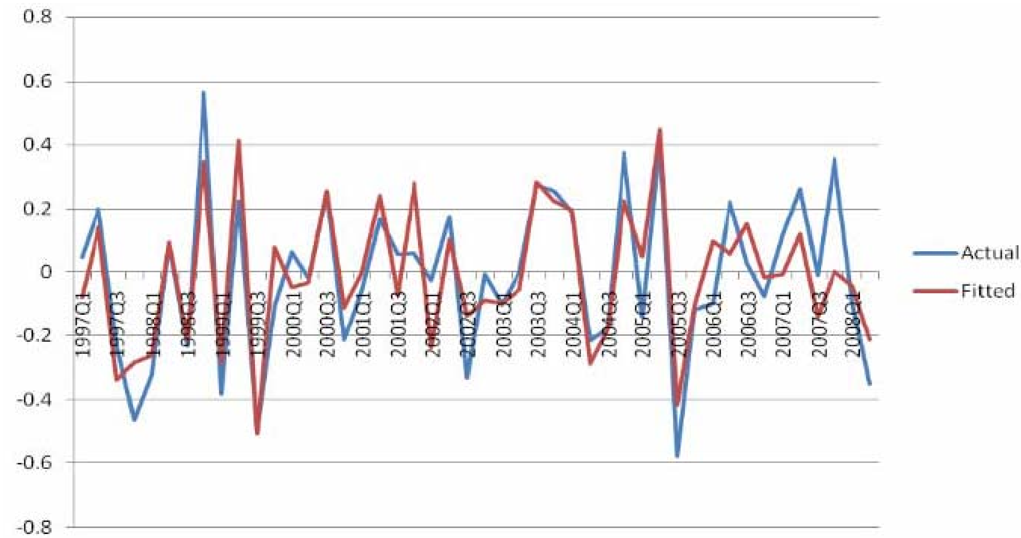


Figure 2. Goodness-of-fit of the VEC Model (variable: Δq_t).

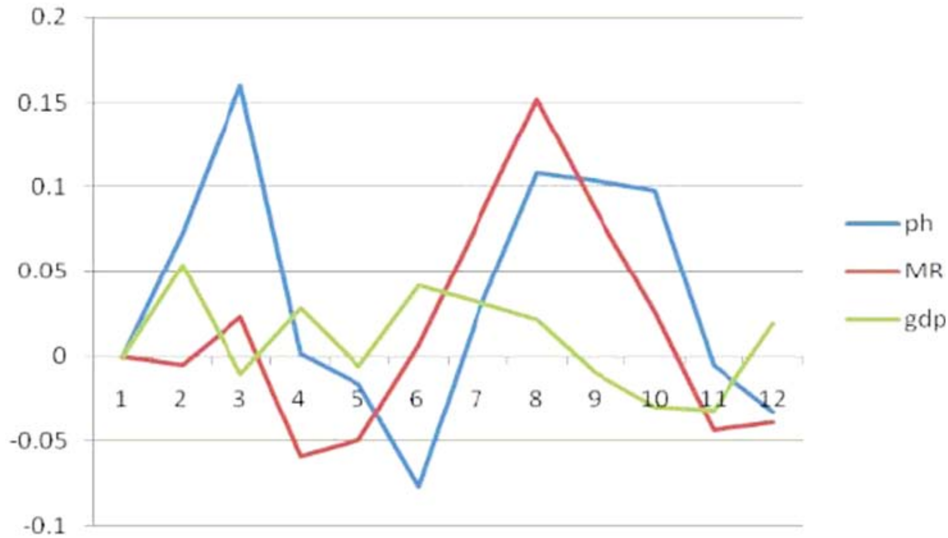
Impulse Responses and Variance Decomposition

The results in Table 5 may not give us a clear understanding of the dynamic structure of the model. The impulse response coefficients provide information to analyse the dynamic behaviour of a variable due to a random shock in other variables. The impulse response traces the effect on current and future values of the endogenous variables of one standard deviation shock to the variables[4]. Sims (1980) suggested that the graphs of the impulse response coefficients provided an appropriate devise to analyse the shocks. Figure 2 depicts the impulse response functions for the housing demand in response to changes in (i) housing prices; (ii) mortgage rate; and (iii) GDP per capita of HK. In order to capture the short-run dynamic effects, we consider the twelve-quarter responses to one standard deviation shock in each of the endogenous variables. The impulse responses are particularly important in tracing the pattern of the demand for private housing from one period to another as a result of a shock in a policy / macro-economic variable.

As revealed in Figure 3, a shock to the macro-economic variables generates different dynamic responses in the housing market. A one standard deviation shock given to housing prices produces an immediate response in the demand for housing units. This confirms the previous results that housing demand are significantly influenced by the prices. The demand increases immediately after the shock, possibly due to the speculative transactions and the

[4] The residuals were orthogonalised by the Cholesky decomposition. Cholesky uses the inverse of the Cholesky factor of the residual covariance matrix to orthogonalise the impulses. It is important to note that the Cholesky decomposition depends on the ordering of the variables, and changing the order of variable might change the impulse responses (Sims, 1980). Various orderings were therefore examined and found that changing the order did not vary the results.

optimistic expectation of investing in the residential market. Expected appreciation improves the expected rate of return on housing investment, thereby justifying a higher demand for housing assets. However, results indicated that the demand is expected to decrease to the ordinary level just after three to four quarters. The continual cyclical effect within the twelve-quarter timeframe indicates the slow adjustment process of the system.



Note: p_h is $\log_e$ of housing prices; MR is mortgage interest rate; gdp is $\log_e$ of gross domestic product.

Figure 3. Impulse Responses of Housing Demand from a One Standard Deviation Shock to the Endogenous Variables.

Similar pattern can be observed for the shock stemmed from the interest rate, except negative but insignificant effect is anticipated in the first six quarters. A plausible explanation of this relationship of this relationship may be that as the cost of financing a house (mortgage rate) is likely to dampen the demand for housing. However, the shock in mortgage rate might only discourage individuals to purchase houses in the short-run. Increase of interest rates, reflecting favourable economic condition, may therefore drive the demand for housing. The profound dynamic changes indicate that the number of houses sold is sensitive to the housing price and the mortgage market. The results also concurred with the findings of Baffoe-Bonnie (1998). In addition, inspection of Figure 2 reveals that a shock to the GDP tends to produce dynamic but relatively insignificant responses to the demand for housing. In other words, a growth of GDP has relatively less impact to the private residential market nationally.

To indicate the relative importance of the shocks requires variance decomposition of the forecast errors for each variable at horizons up to twelve quarters. The variance decompositions show the portion of variance in the prediction for each variable in the system that is attributable to its own shocks as well as shocks to other variables in the system. From the causal chain as implied by the analysis of variance decompositions (Table 8), while the forecast error variances of housing demand are mainly explained by its own shocks in the two-quarter time horizon, shocks to the housing price and mortgage rate respectively has explained 20% and 13% of the variability in housing demand beyond the ten-quarter time horizon. This provides a clear indication that the changes of housing price terms produce the largest variations in the number of houses sold.

Table 8. Variance Decompositions of Housing Demand (*q*)

Period	*S.E.*	*q*	p_h	*MR*	*gdp*
2	0.27575	89.2649	6.9018	0.0307	3.8026
4	0.39952	75.7134	19.3410	2.5687	2.3770
6	0.47383	78.0757	16.4834	2.9479	2.4920
8	0.55198	69.7585	16.1708	11.7345	2.3362
10	0.57968	63.8747	20.6811	13.0469	2.3973
12	0.59359	63.8681	20.0241	13.4297	2.6780

Note: q, is $\log_e$ of housing demand; p_h, is $\log_e$ of housing prices; *MR*, is mortgage interest rate; *gdp*, is $\log_e$ of gross domestic product.

Limitations of the Forecasting Model

The housing demand forecasting model developed in this study has several limitations:

i. The research was confined to the HK housing market. However, a similar methodology can be replicated to develop models for more complex and diverse markets. The results of this study may also serve as a basis for international studies in Asia, Europe and North America. Such a development encourages information exchange related to the mechanism of housing market and in particular markets of the Asia-Pacific regions as well as on the housing policies.
ii. Sample size might affect the efficiency and reliability of forecasting models. The developed housing demand forecasting model could be further verified and even enhanced to derive superior equations by introducing a larger sample size with a wider time horizon.
iii. In making predictions using the leading characteristics of the chosen economic indicators it is assumed that these indicators will follow a similar pattern or trend in the whole period under consideration. If there is any abrupt change in the predictors, the forecast might fail. Abrupt changes to the indicators could happen due to change in policy, housing habits and economic structure (Ng *et al*., 2008).
iv. Extensive experimentation with the macro factors is needed to assess the impact of the key external variables on the housing demand forecast. The projection of housing transaction volume is the central component of the forecasting model. The values of each of the key determinants of housing demand are indeed dependent upon the changes in exogenous factors in the macro economy. However, both the global economy and HK government policy are subject to continuous change, the link between the macro-economic forecasts and the industry-based model should, therefore, be further studied in order to generate more realistic demand forecasts.

Conclusion

The main objective of this study is to apply econometric modelling techniques to build an economic forecasting model for the estimation of future housing demand. The forecasting model also attempts to analyse the dynamic effects of macro-economic and socio-economic variables on housing demand. Applying Johansen's methodology for multivariate cointegration analysis, it is found that the housing demand and the associated economic factors, i.e., housing prices, mortgage rate, and GDP per capita are cointegrated in the long-run. Other macro-economic and socio-economic indicators, including income, inflation, stock prices, employment, population, etc., have also been examined but found to be insignificant in influencing the housing demand.

A dynamic forecasting model has been developed using the vector error correction (VEC) modelling technique to estimate the housing demand. Various diagnostic tests have been undertaken to validate the reliability and robustness of the developed forecasting models. The methods applied and the results presented in this research provide insights into the determinants of housing demand. The housing prices and mortgage rate are found to be the most important and significant factors determining the quantity demand of housing. Addressing these two attributes is critical to formulating both short and long-term housing policies tailored to deal effectively with the expected demand.

This study has also reported the effects of housing demand due to the shocks on the key factors. Empirical results from the impulse response analyses and variance decomposition under the VEC model show that housing price and mortgage rate has relatively large impact, although at different time intervals, on the volume of housing transactions in HK. This finding has important policy implication, because any fluctuations in the housing prices and mortgage rate will affect the housing demand sensitively and thereby the overall investment environment. The adjustment process to achieve the long-run equilibrium is also found to be slow, indicating the impact could be profound.

This research has made a considerable contribution to fill and update the knowledge gap in the field of housing demand forecasting using advanced econometrics techniques—an area which is currently rather underexplored. This study has developed a theoretical model that traces the cause-and-effect chain between housing demand and the its determinants, which is relevant to the current needs of the real estate market and significant to the economy's development. It is envisaged that the results of this study could enhance the understanding of using advanced econometric modelling methodologies, factors affecting housing demand, and various housing economic issues.

Acknowledgement

This study is financially supported by the Research Grants Council through the General Research Fund (grant no.: 7152/07E) and The University of Hong Kong via the CRCG Seed Funding for Basic Research (grant no.: 10207434).

References

Akintoye, A. and Skitmore, M. (1994) Models of UK private sector quarterly construction demand, *Construction Management and Economics*, **12**(1), 3-13.

Anderson, R.G., Hoffman, D.L. and Rasche, R.H. (2002) A vector error-correction forecasting model of the US economy, *Journal of Macroeconomics*, **24**, 569-598.

Arnott, R. (1987) Economic theory and housing. *In:* E.S. Mills, editor, *Handbook of Regional and Urban Economics*, Vol. 2, Ch.24. New York: Elsevier Science Publishers.

Asafu-Adjaye, J. (2000) The relationship between energy consumption, energy prices and economic growth: time series evidence from Asian developing countries, *Energy Economics,* **22**, 615-625.

Baffor-Bonnie, J. (1998) The dynamic impact of macroeconomic aggregates on housing prices and stock of houses: a national and regional analysis, *Journal of Real Estate Finance and Economics*, **17**(2), 179-197.

Bon, R. (1989) *Building as an Economic Process*, Englewood Cliffs, NJ: Prentice-Hall.

Bowerman B.L. and O'Connell, R.T. (1993) *Forecasting and Time Series: An Applied Approach*, 3ed., California: Duxbury Thomson Learning.

Case, K.E. and Glaester, E.L. (2000) Real estate and the macroeconomy, *Brookings Papers on Economic Activity*, **2**, 119-162.

Chambers, D.N. and Schwartz, M.B. (1988) *The Estimation of Housing Demand: A Review of the Evidence*, Research paper, Washington, DC: National Association of REALTORS.

Chin, L. and Fan, G.Z. (2005) Autoregressive analysis of Singapore's private residential prices, *Journal of Property Research*, **23**(4), 257-270.

Crane, S.E. and Nourzad, F. (1998) Improving local manufacturing employment forecasts using cointegration analysis, *Growth and Change*, **29**(2), 175-195.

Darrat, A.F., Shelor, R.M., and Dickens, R.N. (1999) Forecasting corporate Performance: VECM comparison with other time-series models, *Studies in Economics and Finance*, **19**(2), 49-61.

Dickey, D.A. and Fuller, W.A. (1979) Distribution of the estimators for autoregressive time series with unit roots, *Journal of American Statistical Association*, **74**, 427-431.

Dougherty, A., and R. Van Order (1982) Inflation, Housing Costs, and the Consumer Price Index, *American Economic Review*, **72**(1), 269-282.

Engle, R.F. and Granger, C.W.J. (1987) Co-integration and error correction: representation, estimation and testing, *Econometrica*, **55**(2), 251-276.

Follain, J.R. (1981) Does inflation affect real behavior? The case of housing, *Journal of Real Estate Finance and Economics*, **2**(1), 47-60.

Follain, J.R. and Jimenez, E. (1985) Estimating the demand for housing characteristics: survey and critique*, Regional Science and Urban Economic*, **15**, 77-107.

Gillingham, R. and Hagemann, R. (1983) Cross-sectional estimates of a simultaneous model of tenure choice and housing services demand, *Journal of Urban Economics*, **14**, 16-39.

Goodman, A.C. (1988) An Econometric Model of Housing Price, Permanent Income, Tenure Choice, and Housing Demand, *Journal of Urban Economic*s, **23**(3), 327-53.

Granger, C.W.J. (1988) Some recent developments in a concept of causality, *Journal of Econometrics*, **39**, 199-211.

Green, R. and Hendershott, P.H. (1996) Age, housing demand, and real house prices, *Regional Science and Urban Economics*, **26**, 465-480.

Hall, S. G., Henry, S. and Greenslade, J. (2002) On the identification of cointegrated systems in small samples: a modelling strategy with an application to UK wages and prices, *Journal of Economic Dynamics and Control*, **26**, 1517–1537.

Harris, J.C. (1989) The effect of real rates of interest on housing prices, *Journal of Real Estate Finance and Economics*, **2**(1), 61-70.

Heiss, F. and Seko, M. (2001) Housing demand in Germany and Japan, *Journal of Housing Economics*, **10**, 229-252.

Hendershott, P.H. and S.C. Hu. 1981. Inflation and Extraordinary Returns on Owner-Occupied Housing: Some Implications for Capital Allocation and Productivity Growth. *Journal of Macroeconomics* **3**: 177-203.

Hendershott, P.H. and Hu, S.C. (1983) The allocation of capital between residential and nonresidential uses: taxes, inflation, and credit market constraints, *Journal of Finance*, **38**, 795-812.

Hendershott, P.H. and Weicher, J.C. (2002) Forecasting housing markets: lessons learned, *Real Estate Economics*, **30**(1), 1-11.

Hendershott, P.H. and White, M. (2000). Taxing and subsidizing housing: the rise and fall of housing's favored investment status, *Journal of Housing Research*, **11**, 257-275.

Hendry, D. F. (1995) *Dynamic Econometrics*. Oxford: Oxford University Press, UK.

Hendry, D.F. and Juselius, K. (2000) Explaining cointegration analysis: Part 1, *The Energy Journal*, **21**(1), 1-42.

Ho, K.H.D. and Cuervo, J.C. (1999) A cointegration approach to the price dynamics of private housing - A Singapore case study, *Journal of Property Investment and Finance*, **17**(1), 36-60.

Holden, K., Peel, D.A. and Thompson, J.L. (1990) *Economic Forecasting: An Introduction*, Cambridge: Cambridge University Press.

Hoshmand, A.R. (2002) *Business and Economic Forecasting for the Information Age: A Practical Approach*. Westport, CT: Quorum Books.

Hui, E.C.M. and Wong, J.T.Y. (2007) The impact of supply of subsidized sale flats on private housing prices in Hong Kong, *Journal of Real Estate Literature*, **15**(2), 255-280.

Jarque, C.M. and Bera, A.K. (1980) Efficient tests for normality, homoscedasticity and serial independence of regression residuals, *Economics Letters*, **6,** 255-259.

Johansen, S. (1988) Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, **12**, 231-254.

Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, **59**, 1551-1580.

Johansen, S. (1995) *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford: Oxford University Press.

Johansen, S. and Juselius, K. (1990) Maximum likelihood estimation and inference on cointegration – with applications to the demand for money, *Oxford Bulletin of Economics and Statistics*, **52**(2), 169-210.

Johansen, S. and Juselious, K. (1992) Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for the UK, *Journal of Econometrics*, **53**, 211-244.

Kenny, G. (1999) Modelling the demand and supply sides of the housing market: evidence from Ireland, *Economic Modelling*, **16**(3), 389-409.

Killingsworth, M.R. (1990) *The Economics of Comparable Worth*. Kalamazoo, Michigan: W.E., Upjohn Institute for Employment Research.

Krumm, R. (1987) Intertemporal tenure choice, *Journal of Urban Economics*, **22**(3), 263-275.

Lavender, S.D. (1990) *Economics for Builders and Surveyors: Principles and Applications of Economic Decision Making in Developing the Built Environment*. New York: Longman.

Liu, E., Wu, J., Lee, V. And Lee, J. (1996) *Study of Housing Demand Model*, Research Report (RP01/96-97), Hong Kong: Research and Library Services Division, Legislative Council Secretariat.

Lütkepohl, H. (2004) Vector autoregressive and vector error correction models. *In:* H. Lütkepohl and M. Krätzig (*eds.*) *Applied Time Series Econometrics*, London: Cambridge University Press.

MacKinnon, J. G. (1991) Critical Values for Cointegration Tests. *In:* R. F. Engle and C. W. J. Granger (*eds.*), *Long-run Economic Relationships: Readings in Cointegration*, Oxford: Oxford University Press.

MacKinnon, J.G., Haug, A.A., and Michelis, L. (1999) Numerical distribution functions of likelihood ratio tests for cointegration, *Journal of Applied Econometrics*, **14**, 563-577.

Maclennan, D. and Pryce, G. (1996) Global economic change, labour market adjustment and the challenges for European housing policies, *Urban Studies*, **33**(10), 1849-1865.

Mankiw, N.G. and Weil, D. (1989) The baby boom, the baby bust, and the housing market, *Regional Science and Urban Economics*, **19**(2), 235-258.

Masih, R. (1995) Modelling the dynamic interactions among crime, deterrence and social-economic variables: evidence from a vector error-correction model, *Mathematics and Computers in Simulation*, **39**, 411-16.

Megbolugbe, I.F., Marks, A.P. and Schwartz, M.B. (1991) The economic theory of housing demand: a critical demand, *The Journal of Real Estate Research*, **6**, 381-393.

Muellbauer, J. and Murphy, A. (1996) Booms and busts in the UK housing market, in 6th Real Estate Conference, AREUEA, June, UC Berkeley.

Muth, R.F. (1960) The demand for nonfarm housing. *In:* A.C. Harberger, ed. *The Demand for Durable Goods*, 29-96. Chicago, IL: University of Chicago Press.

Ng, S.T., Skitmore, M. and Wong, K.F. (2008) Using genetic algorithms and linear regression analysis for private housing demand forecast, *Building and Environment*, **43**(6), 1171-1184.

Olsen, E.O. (1969) A competitive theory of the housing market, *American Economic Review*, **59**(4), 612-622.

Ong, S.F. and Teck, LC. (1996) Singapore residential market, *Journal of Property Valuation and Investment,* **14**(1), 50.

Pindyck R.S. and Rubinfeld D.L. (1998) *Econometric Models and Economic Forecasts*, 4^{th} ed., USA: The McGraw-Hill Companies, Inc.

Pozdena, J. R. (1988) *The Modern Economics of Housing: A Guide to Theory and Policy for Finance and Real Estate Professionals.* NY: Quarum Books.

Price, S. (1998) Cointegration and modelling the long run. *In*: Scarbrough, E. and Tanenbaum, E. (eds.) *Research Strategies in the Social Sciences: A Guide to New Approaches,* Oxford: Oxford University Press.

Quantitative Micro Software (2000) LLC, *Eviews 4.0 Users' Guide*, California, the United States of America.

Quevedo, J., Cembrano, G., Valls, A. and Serra, J. (1988) Timer series modelling of water demand – a study on short-term and long-term predictions, *Computer Applications in Water Supply*, Wiley, New York, 268-288.

Reichert, A.K. (1990) The impact of interest rates, income, and employment upon regional housing prices, *Journal of Real Estate Finance and Economics*, **3**(3), 373-391.

Rosen, H.S. (1979) Housing decisions and the U.S. income tax: an econometric analysis, *Journal of Public Economics*, **11**(1), 1-23.

Said, E.S. and Dickey, D.A. (1984) Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika*, **71**(3), 599-607.

Salo, S. (1994) Modelling the Finnish housing market, *Economic Modelling*, **11**(2), 250-265.

Sims, C. (1980) Macroeconomics and reality, *Econometrica*, **48**(1), 1-48.

Smith, B. A. and Tesarek, W. P. (1991) House prices and regional real estate cycles: market adjustments in Houston, *Journal of the American Real Estate and Urban Economics Association*, **19**, 396-416.

Smith, L.B., Rosen, K.T. and Fallis, G. (1988) Recent developments in economic models of housing markets, *Journal of Economic Literature*, **26**, 29-64.

Sternlieb, G., and Hughes, J.W. (1977) Regional market variations: the Northeast versus the South, *Journal of the American Real Estate and Urban Economics Association*, (Spring), 44-68.

Summers, L.N. (1981) Inflation, the stock market and owner occupied housing, *American Economic Review*, **71**(2), 429-434.

Tang, Z., de Almeida, C. and Fishwick, P.A. (1991) Time series forecasting using neural networks vs. Box-Jenkins methodology, *Simulation*, **57**(5), 303-310.

Tse, R.Y.C. (1996) Housing return and the determinants of capitalization rates with reference to Hong Kong, *Journal of Property Research*, **13**(2), 115-129.

Tse, R.Y.C. (1997) Construction activity and adjustment of housing stock*, Journal of Research Estate and Construction*, **7**(1), 9-18.

Tse, R.Y.C., Ho, C.W. and Ganesan, S. (1999) Matching housing supply and demand: an empirical study of Hong Kong's market, *Construction Management and Economics,* **17**(5), 625-633.

Tse, R.Y.C. and Webb, J.R. (2004) Hong Kong housing and the Asian financial crisis, *Journal of Real Estate Literature*, **12**(1), 21-32.

Weicher, J.C., and T.G. Thibodeau (1988) Filtering and housing markets: An empirical analysis, *Journal of Urban Economics*, **23**(1), 21-40.

White, H. (1980) A Heteroskedasticity – consistent covariance matrix and a direct test for heteroskedasticity, *Econometrica*, **48**, 817-838.

Wong, J.M.W. Chan, A.P.C. and Chiang, Y.H. (2007) Forecasting construction manpower demand: A vector error correction model, *Building and Environment*, **42**(8), 3030-3041.

In: Economic Forecasting
Editor: Alan T. Molnar, pp. 55-105
ISBN: 978-1-60741-068-3

Chapter 3

SIGNAL AND NOISE DECOMPOSITION OF NONSTATIONARY TIME SERIES

Terence C. Mills
Department of Economics, Loughborough University

Abstract

The decomposition of a time series into components representing trend, cycle, seasonal, etc., has a long history. Such decompositions can provide a formal framework in which to model an observed time series and hence enable forecasts of both the series and its components to be computed along with estimates of precision and uncertainty. This chapter provides a short historical background to time series decomposition before setting out a general framework. It then discusses signal extraction from ARIMA and unobserved component models. The former includes the Beveridge-Nelson filter and smoother and canonical decompositions. The latter includes general structural models and their associated state space formulations and the Kalman filter, the classical trend filters of Henderson and Macaulay that form the basis of the X-11 seasonal adjustment procedure, and band-pass and low-pass filters such as the Hodrick-Prescott, Baxter-King and Butterworth filters. An important problem for forecasting is to be able to deal with finite samples and to be able to adjust filters as the end of the sample (i.e., the current observation) is reached. Trend extraction and forecasting under these circumstances for a variety of approaches will be discussed and algorithms presented. The variety of techniques will be illustrated by a sequence of examples that use typical economic time series.

1. Introduction

The decomposition of an observed time series into unobserved components representing, for example, trend, cyclical and seasonal movements, has a long history dating from the latter part of the 19th century. Before that time, few economists recognised the existence of regular cycles in economic activity or the presence of longer term, secular movements. Rather than the former, they tended to think in terms of 'crises', a word used to mean either a financial panic or a period of deep depression. The early studies of business cycles, which began in the 1870s, are discussed in detail in Morgan (1990), who focuses on the sunspot and Venus theories of Jevons and Moore and the rather more conventional credit cycle theory of Jugler. Secular movements now tend to be referred to as trend, a term that seems to have been coined by Hooker (1901) when analysing British import and export data. The early attempts to take into account trend movements, typically by detrending using simple moving averages or graphical interpolation, are analysed by Klein (1997).

The first quarter of the 20th century saw great progress in business cycle research, most notably by Mitchell (1913, 1927) and in the periodogram studies of weather and harvest cycles by Beveridge (1920, 1921). Trends, on the other hand, were usually only isolated so that they could be eliminated. This was typically achieved by modelling the trend as a moving average spanning the period of the cycle, or by fitting a trend line or some other simple deterministic function of time. A notable example is Kitchin (1923), in which both cyclical and trend movements in data taken from the United States and Great Britain over the period from 1800 were analysed. Kitchin concluded that business cycles averaged 40 months in length (the eponymous Kitchin cycle) and that trade cycles were aggregates of typically two, and sometimes three, of these business cycles. Of equal interest is his conclusion that there had been several trend breaks - the final one, marking the commencement of a downward trend, occurring in 1920. Frickey (1934) gives a taste of the variety of methods then available for fitting trends, with twenty-three different methods used to fit a trend to pig-iron production from 1854 to 1926. Cycles were then constructed by residual, producing average cycles ranging in length from 3.3 to 45 years, thus showing how the observed properties of cyclical fluctuations could be totally dependent on the type of function used to detrend the observed data, a phenomenon that still causes great debate today.

Further difficulties with analysing economic data that appeared to exhibit cyclical behaviour were emphasised by the time series research of Yule (1926, 1927) and Slutsky (1927, 1937). Yule showed that uncritical use of correlation and harmonic analysis, both very popular at the time, was rather dangerous, as ignoring serial correlation in, and random disturbances to, time series could easily lead to erroneous claims of significance and mistaken evidence of harmonic motion. Slutsky's papers (the later one being an English translation and update of the earlier, which was written in Russian) investigated a more fundamental problem - that observed cycles in a time series could be caused entirely by the cumulation of random events. Slutsky's research was not aimed primarily at analysing business cycles, but Kuznets (1929) took up this issue and used simulation and graphical techniques to explore which shapes of distributions of random causes, which periods of moving averages, and which weighting systems produced the most cyclical effects. Indeed, Kuznets pointed out that this analysis not only removed the necessity for having a periodic cause for economic cycles, but could also make further discussion of the causes of business cycles superfluous.

These studies paved the way for the first detailed macrodynamic models of the business cycle to be developed. Frisch's (1933) 'rocking horse theory' of the business cycle, which became very influential, was built on the ideas of Yule and Slutsky (see also Frisch, 1939) and this led to the first multi-equation econometric models of the business cycle being developed (see, for example, Tinbergen, 1937). While extremely influential, such models lie outside the scope of this chapter and we therefore move on to 1946, which saw the publication of Burns and Mitchell's magnum opus for the National Bureau of Economic Research (NBER), in which they produced a new set of statistical measures of the business cycle, known as specific cycles and reference cycles, and used these to test a number of hypotheses about the long-term behaviour of economic cycles (Burns and Mitchell, 1946).

This volume created a great deal of interest and provoked the review of Koopmans (1947), which initiated the famous ‘measurement without theory’ debate in which he accused Burns and Mitchell of trying to measure economic cycles without having any economic theory about how the cycle worked. This review in turn produced Vining’s (1949) spirited defence of the Burns and Mitchell NBER position, in which he charged Koopmans with arguing from a narrow perspective, that associated with the ‘Cowles group’, which had yet to demonstrate any success in actual empirical research.

Although the ‘measurement without theory’ debate obviously focused on the measurement and theoretical modelling of business cycles, it also revealed some disquiet about the role of secular trends and the methods by which they had been removed before cyclical fluctuations could come to the forefront of the analysis. A not too distorted caricature is that data needs only to be detrended by a simple and readily available method so that attention can quickly focus on the much more interesting aspects of cyclical fluctuations. Such an approach is only justifiable if there is indeed little interaction between the trend growth of an economy and its short-run fluctuations but, even then, instability in the trend component and/or the use of an incorrect procedure for detrending will complicate the separation of trend from cycle. It took another decade for techniques to begin to be developed that would, in due course, lead to a revolution in the way trends and cycles were modelled and extracted. The groundwork was prepared by a number of papers appearing in the early 1960s, notably Box and Jenkins (1962), Cox (1961), Kalman (1960), Kalman and Bucy (1961), Leser (1961), Muth (1960) and Winters (1961), which led, after a further decade or so, to the framework for time series decomposition that we now outline.

2. A General Framework for Decomposing Time Series

The general framework that has been developed for considering issues of time series decomposition is to suppose that an observed discrete time series Z_t may be decomposed as

$$Z_t = S_t + N_t \tag{1}$$

with the objective being to use the data on Z_t to estimate the unobserved component series S_t and N_t. These component series might represent ‘signal plus noise’, or ‘seasonal plus nonseasonal’, as suggested by the choice of letters by which they are denoted, but alternative representations, such as ‘trend plus error’, are equally valid. Models admitting decompositions of this type are often referred to as *unobserved component* (UC) models. Here only two unobserved components are considered for conceptual ease but (1) will be extended to include several components in later sections.

Signal extraction results for optimal linear estimators of the components when they are assumed to be stationary and mutually uncorrelated and when an infinite realisation of Z_t (or at least a very long time series) is available may be found in Whittle (1984): optimal being used here in the minimum mean square error, MMSE, sense. This monograph also summarises, and provides references to, earlier results of Wold, Kolmogorov and Wiener.

The core result is the following. Suppose that the autocovariance generating function (ACGF) of Z_t is defined as

$$\gamma_Z(B) = \sum_{j=-\infty}^{\infty} \gamma_{Z,j} B^j \tag{2}$$

where $\gamma_{Z,j} = E(Z_t - \mu_Z)(Z_{t-j} - \mu_Z)$ is the lag *j* autocovariance of Z_t, $\mu_Z = E(Z_t)$ is the mean of Z_t, and *B* is the lag operator such that $B^j Z_t \equiv Z_{t-j}$. Note that, since $\gamma_{Z,j} = \gamma_{Z,-j}$, $\gamma_Z(B)$ is symmetric in *B*. Similarly, define the ACGF of S_t to be $\gamma_S(B)$. The MMSE estimator of S_t, based on the 'infinite' sample $\ldots, Z_{t-1}, Z_t, Z_{t+1}, \ldots$, is then

$$S_{t|\infty} = \sum_{j=-\infty}^{\infty} \varpi_j Z_{t-j} = \varpi(B) Z_t , \qquad \varpi(B) = \frac{\gamma_S(B)}{\gamma_Z(B)} \tag{3}$$

$\varpi(B)$ is often referred to as the Wiener-Kolmogorov (W-K) filter and the result (3) has been extended to incorporate non-stationary Z_t and S_t, and indeed N_t: see, for example, Bell (1984) and Burridge and Wallis (1988). If Z_t has the Wold decomposition $Z_t = \psi(B) e_t$, where e_t is serially uncorrelated (white noise) with zero mean and variance σ_e^2, then its ACGF is

$$\gamma_Z(B) = \sigma_e^2 \psi(B) \psi(B^{-1}) \equiv \sigma_e^2 |\psi(B)|^2 \tag{4}$$

and expressions of the form of (4) will be used freely throughout this chapter.

3. Signal Extraction from ARIMA Models

3.1. The Beveridge-Nelson Signal Extraction Filter

We now consider the case when Z_t is integrated of order one, so that ΔZ_t admits the Wold decomposition

$$\Delta Z_t = \mu + \psi(B) e_t = \mu + \sum_{j=0}^{\infty} \psi_j e_{t-j} \tag{5}$$

Since $\psi(1) = \sum \psi_j$ is a constant, we may write

$$\psi(B) = \psi(1) + C(B)$$

so that

$$\begin{aligned} C(B) &= \psi(B) - \psi(1) \\ &= 1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \ldots - (1 + \psi_1 + \psi_2 + \psi_3 + \ldots) \\ &= -\psi_1(1-B) - \psi_2(1-B^2) - \psi_3(1-B^3) - \ldots \\ &= (1-B)(-\psi_1 - \psi_2(1+B) - \psi_3(1+B+B^2) - \ldots) \end{aligned}$$

i.e.

$$C(B) = (1-B)\left(-\left(\sum\nolimits_{j=1}^{\infty}\psi_j\right) - \left(\sum\nolimits_{j=2}^{\infty}\psi_j\right)B - \left(\sum\nolimits_{j=3}^{\infty}\psi_j\right)B^2 - \ldots\right) = \Delta\tilde{\psi}(B)$$

Thus

$$\psi(B) = \psi(1) + \Delta\tilde{\psi}(B)$$

implying that

$$\Delta Z_t = \mu + \psi(1)e_t + \Delta\tilde{\psi}(B)e_t \tag{6}$$

This gives the decomposition due to Beveridge and Nelson (BN, 1981), with components

$$\Delta S_t^{\mathrm{BN}} = \mu + \left(\sum_{j=0}^{\infty}\psi_j\right)e_t = \mu + \psi(1)e_t \tag{7}$$

and

$$N_t^{\mathrm{BN}} = -\left(\sum_{j=1}^{\infty}\psi_j\right)e_t - \left(\sum_{j=2}^{\infty}\psi_j\right)e_{t-1} - \left(\sum_{j=3}^{\infty}\psi_j\right)e_{t-2} - \ldots = \tilde{\psi}(B)e_t \tag{8}$$

Since e_t is white noise, the BN signal is therefore a random walk with rate of drift equal to μ and an innovation equal to $\psi(1)e_t$, which is thus proportional to that of the original series. The BN noise component is clearly stationary, and is driven by the same innovation as the trend component. The BN trend will have an innovation variance of $\psi(1)^2\sigma_e^2$, which may be larger or smaller than σ_e^2 depending on the signs and patterns of the ψ_j s.

A simple way of estimating the BN components is to approximate the Wold decomposition (5) by an ARIMA$(p,1,q)$ process (see, for example, Newbold, 1990)

$$\Delta Z_t = \mu + \frac{(1 + \theta_1 B + \ldots + \theta_q B^q)}{(1 - \phi_1 B - \ldots - \phi_p B^p)}e_t = \mu + \frac{\theta(B)}{\phi(B)}e_t \tag{9}$$

so that

$$\Delta S_t^{\mathrm{BN}} = \mu + \psi(1)e_t = \mu + \frac{\theta(1)}{\phi(1)}e_t = \mu + \frac{(1+\theta_1+\ldots+\theta_q)}{(1-\phi_1-\ldots-\phi_p)}e_t \tag{10}$$

where $\psi(1)=\theta(1)/\phi(1)$, often referred to as the 'persistence' of the process. Equation (9) can also be written as

$$\frac{\phi(B)}{\theta(B)}\psi(1)\Delta Z_t = \mu + \psi(1)e_t \tag{11}$$

and comparing (10) and (11) shows that an estimate of the signal is given by

$$S_{t|t}^{\mathrm{BN}} = \frac{\phi(B)}{\theta(B)}\psi(1)Z_t = \omega^{\mathrm{BN}}(B)Z_t \tag{12}$$

The notation $S_{t|t}^{\mathrm{BN}}$ signifies that the filter uses the set of observations Z_{t-j}, $j \geq 0$. The signal is thus a weighted average of current and past values of the observed series, with the weights summing to unity since $\omega(1)=1$. The noise component is then given by

$$N_t^{\mathrm{BN}} = Z_t - \omega^{\mathrm{BN}}(B)Z_t = (1-\omega^{\mathrm{BN}}(B))Z_t = \tilde{\omega}^{\mathrm{BN}}(B)Z_t = \frac{\phi(1)\theta(B)-\theta(1)\phi(B)}{\phi(1)\theta(B)}Z_t$$

Since $\tilde{\omega}^{\mathrm{BN}}(1) = 1-\omega^{\mathrm{BN}}(1) = 0$, the weights for the noise component sum to zero. Using (9), this component can also be expressed as

$$N_t^{\mathrm{BN}} = \frac{\phi(1)\theta(B)-\theta(1)\phi(B)}{\phi(1)\phi(B)\Delta}e_t \tag{13}$$

Since N_t^{BN} is stationary, the numerator of (13) can be written as $\phi(1)\theta(B)-\theta(1)\phi(B)=\Delta\varphi(B)$, since it must contain a unit root to cancel out the one in the denominator. As the order of the numerator is $\max(p,q)$, $\varphi(B)$ must be of order $r=\max(p,q)-1$, implying that the noise has the ARMA(p,r) representation

$$\phi(B)N_t^{\mathrm{BN}} = (\varphi(B)/\phi(1))e_t$$

As an example of the Beveridge-Nelson decomposition, consider the ARIMA(0,1,1) process for Z_t

$$\Delta Z_t = \beta + e_t + \theta e_{t-1}$$

The BN decomposition of this process is

$$\Delta S_t^{\mathrm{BN}} = \mu + (1+\theta)e_t$$

$$N_t^{\mathrm{BN}} = -\theta e_t$$

The BN signal filter is

$$\begin{aligned} S_{t|t}^{\mathrm{BN}} &= (1+\theta B)^{-1}(1+\theta)Z_t = (1-\theta B + \theta^2 B^2 - \ldots)(1+\theta)Z_t = (1+\theta)\sum\nolimits_{j=0}^{\infty}(-\theta)^j Z_{t-j} \\ &= (1+\theta)Z_t - \theta S_{t-1|t-1}^{\mathrm{BN}} \end{aligned}$$

while the noise filter is

$$\begin{aligned} N_{t|t}^{\mathrm{BN}} &= \frac{(1+\theta B)-(1+\theta)}{(1+\theta B)} Z_t = \frac{-\theta(1-B)}{(1+\theta B)} Z_t = -\theta(1+\theta B)^{-1}\Delta Z_t = -\theta\sum\nolimits_{j=0}^{\infty}(-\theta)^j \Delta Z_{t-j} \\ &= -\theta \Delta Z_t - \theta N_{t-1|t-1}^{\mathrm{BN}} \end{aligned}$$

which can be seen to be exponentially weighted moving averages of past values of Z_t and ΔZ_t, respectively, with weights given by the MA parameter.

Such a model fitted to the logarithms of seasonally adjusted U.S. GNP from 1947.1 to 2007.4 (shown in figure 1) obtained

$$\Delta Z_t = \underset{(0.001)}{0.008} + e_t + \underset{(0.062)}{0.250} e_{t-1} \qquad \hat{\sigma}_e = 0.0094$$

where standard errors are shown in parentheses. The BN decomposition is thus

$$\Delta S_t^{\mathrm{BN}} = 0.008 + 1.25 e_t$$

$$N_t^{\mathrm{BN}} = -0.25 e_t$$

while the signal and noise filters are

$$S_{t|t}^{\mathrm{BN}} = 1.25\sum\nolimits_{j=0}^{\infty}(-0.25)^j Z_{t-j} = 1.25 Z_t - 0.25 S_{t-1|t-1}^{\mathrm{BN}}$$

$$N_{t|t}^{\mathrm{BN}} = -0.25\sum\nolimits_{j=0}^{\infty}(-0.25)^j \Delta Z_{t-j} = -0.25\Delta Z_t - 0.25 N_{t-1|t-1}^{\mathrm{BN}}$$

Since $\hat{\theta} = 0.25$, the changes in Z_t are positively autocorrelated and the innovations to the trend component have a larger variance, by a factor of $1.25^2 = 1.5625$, than the innovations to the series itself. The noise component, on the other hand, will have a variance just 0.0625 the size of the series innovation variance. The reasonably small value of $\hat{\theta}$ ensures that $S_{t|t}^{\mathrm{BN}}$ and Z_t are very close to each other, while the noise component $N_{t|t}^{\mathrm{BN}}$ is shown in figure 2.

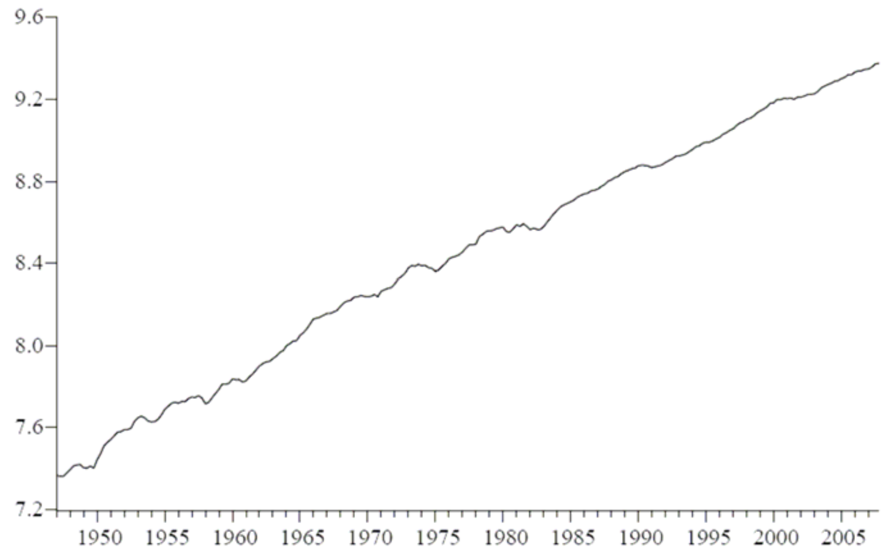


Figure 1. Logarithms of U.S. GNP: 1947.1 – 2007.4.

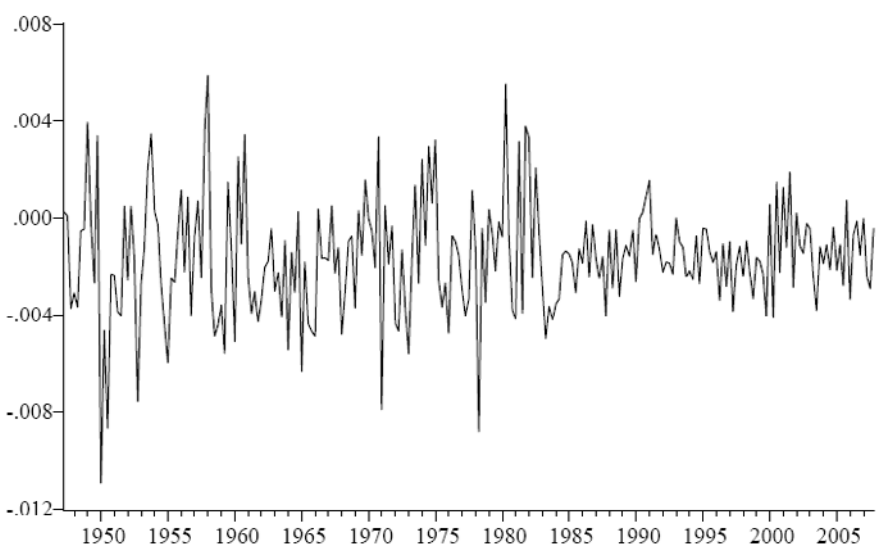


Figure 2. Beveridge-Nelson noise component obtained by fitting an ARIMA$(0,1,1)$ process to the logarithms of U.S. GNP.

Forecasts from this model are derived directly from the ARIMA$(0,1,1)$ representation and thus increase by 0.008 each period, i.e., h-step ahead forecasts made at time T are given by $\hat{Z}_T(h) = Z_T + 0.25e_T + 0.008h$, with accompanying forecast standard errors given by $0.0094(1+1.5625(h-1))^{1/2}$ (see, for example, Mills and Markellos, 2008, pages 57-64). These forecasts and 68% (one standard error) forecast intervals up to end-2012 are shown in figure 3.

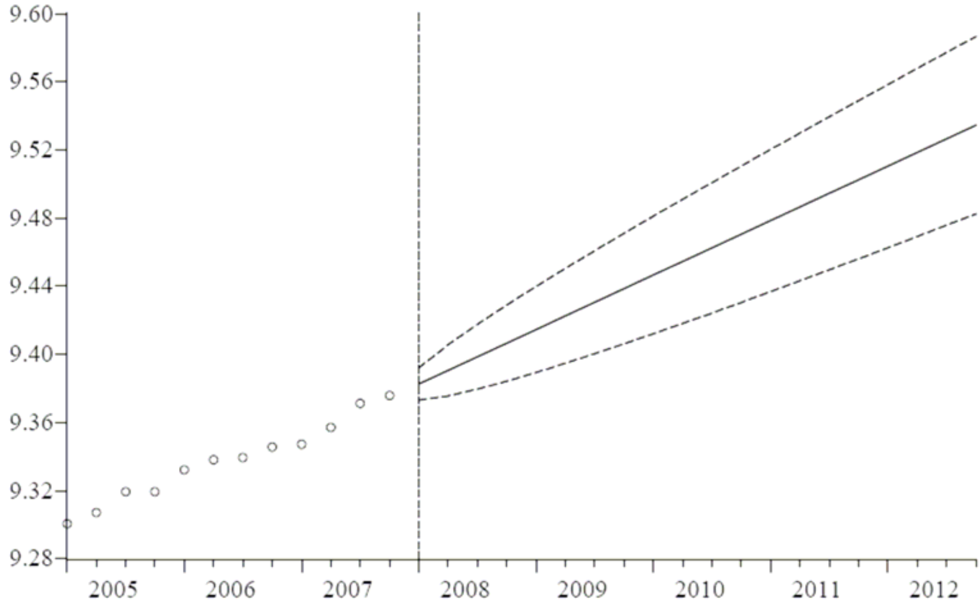


Figure 3. BN forecasts of the logarithms of U.S. GNP until 2012.4.

3.2. A Beveridge-Nelson Smoother

The signal estimator (12) is ‘one-sided’ in the sense that only current and past values of the observed series are used in its construction. Future values of Z_t may be incorporated to define a ‘two-sided’ estimator based on the filter

$$\omega^{\text{BNS}}(B) = \left|\omega^{\text{BN}}(B)\right|^2 = [\psi(1)]^2 \frac{|\phi(B)|^2}{|\theta(B)|^2} \tag{14}$$

i.e.

$$S_{t|\infty}^{\text{BNS}} = \omega^{\text{BNS}}(B) Z_t$$

What is the rationale for this estimator, termed the Beveridge-Nelson *smoother* by Proietti and Harvey (2000)? Suppose the ARIMA$(p,1,q)$ model (9) can be decomposed into a random walk signal and a stationary noise component that are mutually uncorrelated, i.e.

$$\Delta S_t = \beta + v_t \qquad\qquad N_t = \lambda(B)u_t \tag{15}$$

where

$$E(v_t) = E(u_t) = 0; \qquad E(v_t^2) = \sigma_v^2 \qquad E(u_t^2) = \sigma_u^2$$

$$E(v_t v_{t-j}) = E(u_t u_{t-j}) = 0 \text{ for all } j \neq 0; \qquad E(v_t u_{t-j}) = 0 \text{ for all } j.$$

Thus the innovations v_t and u_t are mutually uncorrelated white noise processes and the decomposition is said to be orthogonal.

The ACGF of Z_t and its components are

$$\gamma_Z = \sigma_e^2 \frac{|\theta(B)|^2}{|1-B|^2|\phi(B)|^2}$$

$$\gamma_S = \sigma_v^2 \frac{1}{|1-B|^2}$$

$$\gamma_N = \sigma_u^2 |\lambda(B)|^2$$

Thus the decomposition (1), along with the above assumptions, imply that

$$\sigma_e^2 \frac{|\theta(B)|^2}{|\phi(B)|^2} = \sigma_v^2 + \sigma_u^2 |1-B|^2 |\lambda(B)|^2 \tag{16}$$

Setting $B = 1$, which is equivalent to equating the spectra of Z_t and $N_t + S_t$, and evaluating at the zero frequency, yields

$$\sigma_v^2 = \left(\frac{\theta(1)}{\phi(1)}\right)^2 \sigma_e^2 \tag{17}$$

so that the W-K estimate of the signal is easily seen to be given by the Beveridge-Nelson smoother (14). Substituting this expression for σ_v^2 into (16) gives

$$\sigma_u^2 |1-B|^2 |\lambda(B)|^2 = \sigma_e^2 \left(\frac{|\theta(B)|^2}{|\phi(B)|^2} - \left(\frac{\theta(1)}{\phi(1)}\right)^2 \right) = \sigma_e^2 \frac{\phi(1)^2 |\theta(B)|^2 - \theta(1)^2 |\phi(B)|^2}{\phi(1)^2 |\phi(B)|^2} \tag{18}$$

Equation (17) implies that the trend can be written as

$$\Delta S_t = \beta + \frac{\theta(1)}{\phi(1)}\tilde{v}_t = \beta + \psi(1)\tilde{v}_t \tag{19}$$

where $\tilde{v}_t$ is white noise with variance σ_e^2. Although (19) is also a random walk driven by an innovation with the same variance as e_t, the innovation to the BN trend in (10), the two components differ. The latter is a function of current and past observations of Z_t, while the former is an unobserved component whose estimate will depend upon future values of Z_t as well.

Similarly, (18) shows that the noise component has the ARMA(p,r) representation

$$\phi(B)N_t = \frac{\alpha(B)}{\phi(1)}\tilde{u}_t$$

where $\alpha(B)$ is an MA polynomial of order *r* such that

$$|1-B|^2|\alpha(B)|^2 = \phi(1)^2|\theta(B)|^2 - \theta(1)^2|\phi(B)|^2 \tag{20}$$

and $\tilde{u}_t$ is a white noise component, again with variance σ_e^2. Although $\alpha(B)$ and $\varphi(B)$ are of the same order, their coefficients will differ since the requirement that the signal and noise components are uncorrelated results in parameter restrictions being embedded in (17). This may be seen by considering the orthogonal decomposition for the ARIMA$(0,1,1)$ process. Since $r = 0$, $\alpha(B) = \alpha_0$ and so

$$\Delta S_t = \beta + (1+\theta)\tilde{v}_t \qquad\qquad N_t = \alpha_0 \tilde{u}_t$$

From (20), $\alpha_0 = -\theta$. However, the decomposition implies that

$$\Delta Z_t = \beta + (1+\theta)\tilde{v}_t - \theta\tilde{u}_t + \theta\tilde{u}_{t-1}$$

from which it is straightforward to show that the first order autocorrelation of ΔZ_t is given by

$$\rho_1 = -\frac{\theta^2}{(1+\theta)^2 + 2\theta^2}$$

Clearly this is always negative, so that this model is unable to account for positive autocorrelation in ΔZ_t. Moreover, given that the lower bound of ρ_1 for a first order moving average process is -0.5, this also implies that $-1 < \theta \leq 0$. This is known as the admissability condition, which alternatively can be expressed as $0 < \psi(1) = 1 + \theta \leq 1$, i.e., that persistence must be no greater than unity. Lippi and Reichlin (1992) prove that $\psi(1) \leq 1$ is required for any orthogonal decomposition into uncorrelated random walk signal and stationary noise components.

The BN smoother for the ARIMA$(0,1,1)$ model is

$$S_{t|\infty}^{\mathrm{BNS}} = (1+\theta)^2 \frac{1}{(1+\theta B)(1+\theta B^{-1})} Z_t = \frac{1+\theta}{1-\theta} \sum_{j=-\infty}^{\infty} (-\theta)^{|j|} Z_{t-j}$$

which clearly differs from the BN filter $S_{t|t}^{\mathrm{BN}}$. Note that the BN smoother would not be admissible for the logarithms of U.S. GNP, as the ARIMA$(0,1,1)$ model fitted to this series has $\hat{\theta} = 0.25$, implying $\psi(1) = 1.25$.

3.3. A Canonical Decomposition

Consider now the more general model for Z_t

$$\zeta(B) Z_t = \theta(B) e_t \tag{21}$$

where $\zeta(B)$ is of order P and has its zeros on or outside the unit circle. It is also now assumed explicitly that $\theta(B)$ has all its zeros outside the unit circle and that $\zeta(B)$ and $\theta(B)$ have no common zero. Thus (9) has $\zeta(B) = (1-B)\phi(B)$ (the mean μ is ignored here for simplicity). If it is assumed that the noise component N_t is white, i.e., that $N_t = u_t$, then the signal S_t follows the process (see Tiao and Hillmer, 1978)

$$\zeta(B) S_t = \eta(B) v_t$$

where $\eta(B)$ is of order $R \leq \max(P, q)$ and has all its zeros on or outside the unit circle. Since $\theta(B) e_t = \eta(B) v_t + \zeta(B) u_t$,

$$|\theta(B)|^2 \sigma_e^2 = |\eta(B)|^2 \sigma_v^2 + |\zeta(B)|^2 \sigma_u^2 \tag{22}$$

Thus, MMSE estimators of the components are

$$\hat{S}_{t|\infty} = \frac{\sigma_v^2}{\sigma_e^2}\frac{|\eta(B)|^2}{|\theta(B)|^2}Z_t = \left(1 - \frac{\sigma_u^2}{\sigma_e^2}\frac{|\zeta(B)|^2}{|\theta(B)|^2}\right)Z_t = \omega(B)Z_t\,,\ \hat{N}_{t|\infty} = Z_t - \hat{S}_{t|\infty}$$

However, given the model for Z_t, the models for the components are not unique. Any choice of $\eta(B)$, $\sigma_v^2 \geq 0$ and $\sigma_u^2 \geq 0$ that satisfies (22) will be an admissable decomposition. Further, from (22) we have

$$\left|\theta\left(e^{-i\lambda}\right)\right|^2 \sigma_e^2 = \left|\eta\left(e^{-i\lambda}\right)\right|^2 \sigma_v^2 + \left|\zeta\left(e^{-i\lambda}\right)\right|^2 \sigma_u^2 \tag{23}$$

so that

$$\sigma_u^2 = \left|\frac{\theta\left(e^{-i\lambda}\right)}{\zeta\left(e^{-i\lambda}\right)}\right|^2 \sigma_e^2 - K \qquad\qquad K = \left|\frac{\eta\left(e^{-i\lambda}\right)}{\zeta\left(e^{-i\lambda}\right)}\right|^2 \sigma_v^2 \geq 0$$

and the maximum possible value of σ_u^2 is thus

$$\overline{\sigma}_u^2 = \min_{0\leq\lambda\leq\pi} \sigma_e^2 \left|\frac{\theta\left(e^{-i\lambda}\right)}{\zeta\left(e^{-i\lambda}\right)}\right|^2$$

Since the actual value of σ_u^2 is unknown, Tiao and Hillmer (1978) argue that it thus makes sense to make this variance as large as possible subject to (22) since, intuitively, this extracts the most white noise and thus extracts the strongest signal that can be recovered from the original series. Hence they refer to the decomposition corresponding to $\sigma_u^2 = \overline{\sigma}_u^2$ as the canonical decomposition. Tiao and Hillmer show that this canonical decomposition has the following properties.

i. It produces the strongest possible autocorrelation structure for the signal, in the sense that the absolute values of each autocorrelation coefficient of S_t is maximised when $\sigma_u^2 = \overline{\sigma}_u^2$.
ii. The moving average process $\eta(B)$ has at least one zero on the unit circle.
iii. The variance of the signal innovation, σ_v^2, is minimised when $\sigma_u^2 = \overline{\sigma}_u^2$, thus making S_t as nearly deterministic as possible, so that it is the strongest signal process which could lead to the known model (21) for Z_t.

The minimum variance properties of the signal carry over directly to the estimate $\hat{S}_{t|\infty} = \varpi(B)Z_t$. The filter $\varpi(B)$ has the properties that $0 \le \varpi\left(e^{-i\lambda}\right) \le 1$ and, if and only if $\sigma_u^2 = \overline{\sigma}_u^2$,

$$\min_{0\le\lambda\le\pi} \varpi\left(e^{i\lambda}\right) = 0$$

These follow from noting that

$$\varpi\left(e^{-i\lambda}\right) = 1 - \frac{\sigma_u^2}{\sigma_e^2}\frac{\left|\zeta\left(e^{-i\lambda}\right)\right|^2}{\left|\theta\left(e^{-i\lambda}\right)\right|^2} = \frac{\sigma_v^2}{\sigma_e^2}\frac{\left|\eta\left(e^{-i\lambda}\right)\right|^2}{\left|\theta\left(e^{-i\lambda}\right)\right|^2}$$

and using the properties of (23).

Consider again the ARIMA$(0,1,1)$ process

$$\Delta Z_t = e_t + \theta e_{t-1} \tag{24}$$

Since $R = P = q = 1$, the signal will also follow an ARIMA$(0,1,1)$ process:

$$\Delta S_t = v_t + \eta v_{t-1} \tag{25}$$

with $N_t = u_t$, (24) and (25) imply that

$$e_t + \theta e_{t-1} = v_t + \eta v_{t-1} + u_t - u_{t-1} \tag{26}$$

Multiplying both sides of (26) by $e_{t-k} + \theta e_{t-1-k}$, for $k = 0$ and 1, and taking expectations yields the following equations

$$\left(1+\theta^2\right)\sigma_e^2 = \left(1+\eta^2\right)\sigma_v^2 + 2\sigma_u^2$$

$$\theta\sigma_e^2 = \eta\sigma_v^2 - \sigma_u^2$$

These may be solved to yield

$$\sigma_v^2 = \frac{(1+\theta)^2}{(1+\eta)^2}\sigma_e^2$$

and

$$\sigma_u^2 = \left(\frac{\eta(1+\theta)^2 - \theta(1+\eta)^2}{(1+\eta)^2} \right) \sigma_e^2 \tag{27}$$

The restriction $\sigma_u^2 \geq 0$ requires

$$\eta(1+\theta)^2 - \theta(1+\eta)^2 = (\eta - \theta)(1 - \eta\theta) \geq 0$$

which will be satisfied if $\theta \leq \eta \leq 1$. Moreover, differentiating (27) with respect to η shows that σ_u^2 reaches a maximum at $\eta = 1$. This value will be given by

$$\overline{\sigma}_u^2 = \frac{(1-\theta)^2}{4} \sigma_e^2$$

with $\eta(B) = 1 + B$, the MMSE estimator is

$$\hat{S}_{t|\infty} = \frac{\sigma_v^2}{\sigma_e^2} \frac{|1+B|^2}{|1+\theta B|^2} Z_t = \frac{(1+\theta)^2}{4} \frac{|1+B|^2}{|1+\theta B|^2} Z_t = \frac{(1+\theta)}{4(1-\theta)} \sum_{j=-\infty}^{\infty} (-\theta)^j \left(Z_{t-j} + Z_{t-1-j} \right)$$

Setting $B = -1$ is equivalent to $\lambda = \pi$, so that $\varpi\left(e^{i\pi}\right) = 0$; otherwise $0 \leq \varpi\left(e^{-i\lambda}\right) \leq 1$, as required.

3.4. ARIMA Components

The previous section has considered the case where a known ARIMA model for the observed series Z_t can be decomposed into an ARIMA signal and a white noise. Consider now the more general case in which both the signal and noise components follow ARIMA models

$$\zeta_S(B) S_t = \eta_S(B) v_t \tag{28}$$

$$\zeta_N(B) N_t = \eta_N(B) u_t \tag{29}$$

where the pairs of polynomials $\{\zeta_S(B), \eta_S(B)\}$ and $\{\zeta_N(B), \eta_N(B)\}$ have their zeros lying on or outside the unit circle and have no common zeros, and v_t and u_t are uncorrelated. Then $Z_t = S_t + N_t$ will follow the ARIMA model given by (21), but where $\zeta(B)$ is the highest common factor of $\zeta_S(B)$ and $\zeta_N(B)$, and $\theta(B)$ and σ_e^2 can be obtained from the relationship

$$\sigma_e^2 \frac{|\theta(B)|^2}{|\zeta(B)|^2} = \sigma_v^2 \frac{|\eta_S(B)|^2}{|\zeta_S(B)|^2} + \sigma_u^2 \frac{|\eta_N(B)|^2}{|\zeta_N(B)|^2} \tag{30}$$

The MMSE estimators of the components are then

$$\hat{S}_{t|\infty} = \omega_S(B) Z_t \qquad \hat{N}_{t|\infty} = \omega_N(B) Z_t$$

where

$$\omega_S(B) = \frac{\sigma_v^2}{\sigma_e^2} \frac{|\zeta(B)|^2 |\eta_S(B)|^2}{|\theta(B)|^2 |\zeta_S(B)|^2} \qquad \omega_N(B) = 1 - \omega_S(B)$$

As an example, suppose that the signal is a random walk

$$\Delta S_t = v_t$$

and $N_t = u_t$. Thus

$$\Delta Z_t = v_t + u_t - u_{t-1} = e_t + \theta e_{t-1} \tag{31}$$

and (30) becomes

$$\sigma_e^2 |1 + \theta B|^2 = \sigma_v^2 + \sigma_u^2 |1 - B|^2$$

Thus

$$\omega_S(B) = \frac{\sigma_v^2}{\sigma_e^2} \frac{1}{|1 + \theta B|^2}$$

and, on defining the signal-to-noise ratio $\kappa = \sigma_v^2 / \sigma_u^2$, the relationship between the parameters of (31) and the component models can be shown to be

$$\theta = -\tfrac{1}{2}\left(\kappa + 2 - \sqrt{\kappa(\kappa + 4)}\right) \qquad \kappa = -(1 + \theta)^2 / \theta$$

$$\sigma_e^2 = -\sigma_u^2 / \theta$$

Thus, for $\kappa \geq 0$, $-1 \leq \theta \leq 0$ and positive first order autocorrelation in ΔZ_t (as found in the logarithms of U.S. GNP) is ruled out. When $\kappa = 0$, $\theta = -1$ and the unit roots in (31) cancel out and the overdifferenced Z_t is stationary, while $\kappa = \infty$ corresponds to $\theta = 0$, in which case Z_t is a pure random walk.

Typically, of course, S_t and N_t are unobserved, so that it is unrealistic to assume that the more general component models (28) and (29) are known. Furthermore, without further restrictions of the type considered above, equation (30) cannot be used to identify all the parameters of the component models. As a result, the filters $\omega_S(B)$ and $\omega_N(B)$ cannot be determined and the MMSE estimates cannot be calculated. Several alternative models have therefore been assumed for the components that will allow the estimates of S_t and N_t to be calculated, and these are now discussed.

4. Structural Models and State Space Forms

4.1. A General Class of Structural Models

To introduce the class of *structural* time series models, the decomposition (1) is generalised to include a wider set of unobserved components representing different features of the evolution of Z_t:

$$Z_t = \mu_t + \gamma_t + \psi_t + \nu_t + \varepsilon_t, \qquad t = 1, \ldots, T \tag{32}$$

In (32) μ_t is the trend, γ_t is the seasonal, ψ_t is the cycle, υ_t is a stationary autoregressive component, and ε_t is the irregular, taken to be normal and independently distributed with zero mean and variance σ_ε^2, which we denote as $\varepsilon_t \sim \mathrm{NID}(0, \sigma_\varepsilon^2)$.

Following Harvey and Trimbur (2003), the trend component is defined as

$$\mu_t = \mu_{t-1} + \beta_{t-1}^{(k)} + \eta_t \qquad \eta_t \sim \mathrm{NID}(0, \sigma_\eta^2) \tag{33}$$

$$\beta_t^{(j)} = \beta_{t-1}^{(j)} + \beta_t^{(j-1)} \qquad j = 1, \ldots, k$$

where $\beta_t^{(0)} = \zeta_t \sim \mathrm{NID}(0, \sigma_\zeta^2)$. The k^{th}-order stochastic trend μ_t is thus integrated of order $k+1$, i.e., it is $I(k+1)$ and becomes smoother as *k* increases. When $\sigma_\eta^2 = \sigma_\zeta^2 = 0$ μ_t reduces to a polynomial trend of order *k*: $\mu_t = c_0 + c_1 t + \ldots + c_k t^k$. For example, if $k = 1$, $\beta_t^{(1)} = \beta_{t-1}^{(1)} = \beta$ and $\mu_t = \mu_{t-1} + \beta$, so that $\mu_t = \mu_0 + \beta t = c_0 + c_1 t$. If, further, $\beta = 0$, $\mu_t = c_0$ and the trend is constant.

The seasonal component takes an evolving trigonometric form. Assuming the number of seasonal frequencies in a period (e.g., a year) is given by the even integer *s*, then

$$\gamma_t = \sum_{j=1}^{s/2} \gamma_{j,t}$$

where each $\gamma_{j,t}$, $j = 1,\ldots,s/2$, is generated by

$$\begin{bmatrix} \gamma_{j,t} \\ \gamma^*_{j,t} \end{bmatrix} = \begin{bmatrix} \cos\lambda_j & \sin\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1} \\ \gamma^*_{j,t-1} \end{bmatrix} + \begin{bmatrix} \omega_{j,t} \\ \omega^*_{j,t} \end{bmatrix} \qquad j = 1,\ldots,(s-2)/2 \tag{34}$$

$$\gamma_{s/2,t} = \cos\lambda_{s/2}\gamma_{s/2,t-1} + \omega_{s/2,t}$$

Here $\lambda_j = 2\pi j/s$ is the frequency (in radians) and $\omega_{j,t} \sim \mathsf{NID}(0,\sigma^2_\omega)$ and $\omega^*_{j,t} \sim \mathsf{NID}(0,\sigma^2_\omega)$ are mutually uncorrelated disturbances having a common variance. The component $\gamma^*_{j,t}$ appears as a matter of construction to allow the seasonal to be modelled as a stochastic combination of sine and cosine waves and its interpretation is not particularly important. The seasonal component is deterministic if $\sigma^2_\omega = 0$ and thus will evolve as a regular cycle with an annual period *s*.

The cyclical component is defined by incorporating features of both (33) and (34):

$$\begin{bmatrix} \psi_{1,t} \\ \psi^*_{1,t} \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{bmatrix} \psi_{1,t-1} \\ \psi^*_{1,t-1} \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa^*_t \end{bmatrix} \tag{35}$$

$$\begin{bmatrix} \psi_{i,t} \\ \psi^*_{i,t} \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{bmatrix} \psi_{i,t-1} \\ \psi^*_{i,t-1} \end{bmatrix} + \begin{bmatrix} \psi_{i-1,t-1} \\ \psi^*_{i-1,t-1} \end{bmatrix} \qquad i = 2,\ldots,n \tag{36}$$

This is known as an n^{th}-order stochastic cycle, with $\psi_{n,t}$ appearing as ψ_t in (32): such a cycle is thus driven by shocks that are themselves periodic. $\kappa_t \sim \mathsf{NID}(0,\sigma^2_\kappa)$ and $\kappa^*_t \sim \mathsf{NID}(0,\sigma^2_\kappa)$ are mutually uncorrelated disturbances with common variance, $0 < \rho \le 1$ is the damping factor of the cycle and $0 \le \lambda_c \le \pi$ is the frequency in radians, so that the period of the cycle is $2\pi/\lambda_c$. Equation (35) has a pair of complex conjugate roots $\rho\exp(\pm \mathrm{i}\lambda_c)$ with modulus ρ, and is therefore stationary for $|\rho| < 1$. Since the addition of each term in (36) introduces another pair of such roots into the model, the overall condition for stationarity remains the same, so that the relevant range for the damping factor is $0 < \rho < 1$.

The variance of the n^{th}-order stochastic cycle is, from Trimbur (2006, equation (16)),

$$\sigma_{\psi}^{2}(n)=\sigma_{\kappa}^{2}\frac{\sum_{i=0}^{n-1}\binom{n-1}{i}\rho^{2i}}{\left(1-\rho^{2}\right)^{2n-1}}$$

which specialises to

$$\sigma_{\psi}^{2}(1)=\sigma_{\kappa}^{2}\frac{1}{\left(1-\rho^{2}\right)},\qquad \sigma_{\psi}^{2}(2)=\sigma_{\kappa}^{2}\frac{1+\rho^{2}}{\left(1-\rho^{2}\right)^{3}},\qquad \text{etc.}$$

The τ^{th} autocorrelation of the n^{th}-order stochastic cycle is given by (Trimbur, 2006, equation (26))

$$\rho_{\tau}(n)=\rho^{\tau}\cos\lambda_{c}\tau\frac{\sum_{j=n-\tau}^{n}\left[\left(1-\rho^{2}\right)^{n-k}\binom{\tau}{n-k}\sum_{r=0}^{k-1}\left(\binom{k-1}{r}\binom{n-1}{r+n-k}\rho^{2r}\right)\right]}{\sum_{i=0}^{n-1}\binom{n-1}{i}^{2}\rho^{2i}}$$

so that

$$\rho_{\tau}(1)=\rho^{\tau}\cos\lambda_{c}\tau\qquad \rho_{\tau}(2)=\rho^{\tau}\cos\lambda_{c}\tau\left(1+\frac{1-\rho^{2}}{1+\rho^{2}}\tau\right),\qquad \text{etc.}$$

Thus, for any *n*, the autocorrelation function is a damped cosine function with the damping pattern depending on both ρ and *n*.

The autoregressive component can be specified generally as an AR(2) process

$$\nu_{t}=\left(\theta_{1}+\theta_{2}\right)\nu_{t-1}-\theta_{1}\theta_{2}\nu_{t-2}+\xi_{t}\qquad \xi_{t}\sim \mathrm{NID}\left(0,\sigma_{\xi}^{2}\right)$$

with stationarity being ensured if $-1<\theta_{1},\theta_{2}<1$. If $\theta_{2}=0$ ν_{t} will be AR(1), and will then be the limiting case of a first-order stochastic cycle when $\lambda_{c}=0$ or π. Finally, it is assumed that the disturbances driving each of the components are mutually uncorrelated.

A simple example of a structural model decomposes Z_{t} into a first order trend,

$$Z_{t}=\mu_{t}+\varepsilon_{t}$$

$$\mu_{t}=\mu_{t-1}+\beta_{t-1}+\eta_{t}$$

$$\beta_{t}=\beta_{t-1}+\zeta_{t}$$

and mutually uncorrelated irregular component ε_t. This can be written as

$$\Delta^2 Z_t = \Delta^2 \varepsilon_t + \Delta \eta_t + \zeta_{t-1}$$

thus implying an ARIMA$(0,2,2)$ model for Z_t

$$\Delta^2 Z_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} \qquad e_t \sim \mathrm{NID}\left(0, \sigma_e^2\right) \tag{37}$$

However, the structure of the component models places some quite severe restrictions on the values that can be taken by θ_1 and θ_2 and, consequently, on the two non-zero autocorrelations of $\Delta^2 Z_t$. For example, the first-order autocorrelation is restricted to the interval $\left[-\tfrac{2}{3}, 0\right]$ and the second-order autocorrelation to $\left[0, \tfrac{1}{6}\right]$, while the restrictions placed on the parameters that are needed to ensure that all innovation variances are non-negative can be shown to be $\theta_1 \le 4\theta_2 / \left(1+\theta_2\right)$, as well as the usual invertibility restrictions associated with an MA(2) process: $0 \le \theta_2 < 1$, $-2 < \theta_1 \le 0$, $1 + \theta_1 + \theta_2 > 0$ and $1 - \theta_1 + \theta_2 > 0$ (see, for example, Harvey, 1989, page 69, and Proietti, 2005).

4.2. State Space Formulation and the Kalman Filter

The complicated set of restrictions place upon 'reduced form' equations such as (37) precludes the use of the W-K filter to estimate the unobserved structural components. Instead, the structural model can be estimated by casting it into state space form (SSF), employing the Kalman (1960) filter, and using various filtering and smoothing algorithms and nonlinear optimising routines to estimate the parameters and components and to provide standard errors and confidence intervals for them (see Koopman, Shephard and Doornik, 1999). These routines are available in the commercial software package *STAMP*: see Koopman et al. (2006) for details of the estimation procedures, filtering and smoothing routines, and discussion of the supplementary diagnostic tests and graphical representations of the components and related statistics.

The SSF can be constructed by defining the *state vector* as

$$\begin{aligned}\mathbf{A}_t^{\mathsf{T}} &= \left(\mu_t, \beta_t^{(k)}, \ldots, \beta_t^{(1)}, \gamma_{1,t}, \gamma_{1,t}^{*} \ldots, \gamma_{(s-2)/2}, \gamma_{(s-2)/2}^{*}, \gamma_{s/2,t}, \psi_{n,t}, \psi_{n,t}^{*}, \ldots, \psi_{1,t}, \psi_{1,t}^{*}, \nu_t, \nu_{t-1}\right) \\ &= \left(\boldsymbol{\beta}_t^{\mathsf{T}}; \boldsymbol{\gamma}_t^{\mathsf{T}}; \boldsymbol{\psi}_t^{\mathsf{T}}; \boldsymbol{\upsilon}_t^{\mathsf{T}}\right)\end{aligned}$$

where $^{\mathsf{T}}$ denotes the transpose of a matrix or vector and

$$\begin{aligned}\boldsymbol{\beta}_t^{\mathsf{T}} &= \left(\mu_t, \beta_t^{(k)}, \ldots, \beta_t^{(1)}\right) \\ \boldsymbol{\gamma}_t^{\mathsf{T}} &= \left(\gamma_{1t}, \gamma_{1,t}^{*}, \ldots, \gamma_{(s-2)/2}, \gamma_{(s-2)/2}^{*}, \gamma_{s/t}\right)\end{aligned}$$

$$\boldsymbol{\psi}_t^{\mathsf{T}} = \left(\psi_{n,t}, \psi_{n,t}^*, \ldots \psi_{1,t} \psi_{1,t}^*\right)$$
$$\boldsymbol{\upsilon}_t^{\mathsf{T}} = \left(\nu_t, \nu_{t-1}\right)$$

Defining the vector $\mathbf{x}_t$ as

$$\mathbf{x}_t = \left(1, \underbrace{0, \ldots, 0}_{k}, \underbrace{1, 0, \ldots, 1, 0}_{2s}, 1, \underbrace{0, \ldots, 0}_{2n-1}, 1, 0\right)$$

the *measurement equation* (i.e., equation (32)) can be written as

$$Z_t = \mathbf{x}_t \mathbf{A}_t + \varepsilon_t \tag{38}$$

The *transition equation* for the trend is defined as

$$\boldsymbol{\beta}_t = \mathbf{T}_{(k+1)} \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t$$

where

$$\mathbf{T}_{(k+1)} = \begin{bmatrix} \mathbf{t} \\ \mathbf{T}_k \end{bmatrix}$$

with $\mathbf{t} = \begin{bmatrix} 1 & 1 & 0 & \ldots & 0 \end{bmatrix}$ and $\mathbf{T}_k$ a $k \times k$ upper triangular matrix of ones, and where

$$\boldsymbol{\eta}_t = \begin{bmatrix} \eta_t \\ \boldsymbol{\iota}_k \zeta_t \end{bmatrix}$$

$\boldsymbol{\iota}_k$ being a *k*-vector of ones.

The transition equation for the seasonal component is

$$\boldsymbol{\gamma}_t = \mathbf{S}_{(s)} \boldsymbol{\gamma}_{t-1} + \boldsymbol{\omega}_t$$

where

$$\mathbf{S}_{(s)} = \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} & \cdots & \mathbf{0} & 0 \\ \mathbf{0} & \mathbf{S}_2 & \mathbf{0} & \mathbf{0} & 0 \\ \vdots & \cdots & \ddots & \mathbf{0} & 0 \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{S}_{(s-2)/2} & 0 \\ 0 & \cdots & \cdots & 0 & \cos \lambda_{s/2} \\ 0 & \cdots & \cdots & 0 & 0 \end{bmatrix} \quad \mathbf{S}_j = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix}$$

$$j = 1, \ldots, (s-2)/2$$

$$\boldsymbol{\omega}_t^{\mathsf{T}} = \left(\omega_{1,t}, \omega_{1,t}^{*}, \ldots, \omega_{(s-2)/2,t}, \omega_{(s-2)/2,t}^{*}, \omega_{s/2,t}\right)$$

The transition equation for the cycle is

$$\boldsymbol{\psi}_t = \mathbf{C}_{(n)} \boldsymbol{\psi}_{t-1} + \boldsymbol{\kappa}_t$$

where, on denoting $\otimes$ as the Kroneker product,

$$\mathbf{C}_{(n)} = \mathbf{I}_n \otimes \mathbf{C} + \mathbf{C}_n \otimes \mathbf{I}_2$$

and

$$\boldsymbol{\kappa}_t = \mathbf{c}_n \otimes \begin{bmatrix} \kappa_t \\ \kappa_t^{*} \end{bmatrix}$$

$$\mathbf{C} = \rho \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \qquad \mathbf{c}_n^{\mathsf{T}} = \begin{bmatrix} \underbrace{0 \quad \cdots \quad 0}_{n-1} & 1 \end{bmatrix}$$

$$\mathbf{C}_n = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & \vdots & & \ddots & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}$$

The transition equation for the autoregressive component is

$$\boldsymbol{\upsilon}_t = \boldsymbol{\theta} \boldsymbol{\upsilon}_{t-1} + \boldsymbol{\xi}_t$$

where

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 + \theta_2 & -\theta_1\theta_2 \\ 1 & 0 \end{bmatrix} \qquad \boldsymbol{\xi}_t = \begin{bmatrix} \xi_t \\ 0 \end{bmatrix}$$

The various components may be consolidated into the system transition equation

$$\mathbf{A}_t = \begin{bmatrix} \mathbf{T}_{(k+1)} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{(s)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{(n)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\theta} \end{bmatrix} \mathbf{A}_{t-1} + \begin{bmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\omega}_t \\ \boldsymbol{\kappa}_t \\ \boldsymbol{\xi}_t \end{bmatrix}$$

i.e.,

$$\mathbf{A}_t = \mathbf{R}\mathbf{A}_{t-1} + \mathbf{\Omega}_t \tag{39}$$

where the error system covariance matrix is given by

$$V(\mathbf{\Omega}_t) = \text{diag}\left(\sigma_\eta^2, \underbrace{\sigma_\zeta^2, \ldots, \sigma_\zeta^2}_{k}, \underbrace{\sigma_\omega^2, \ldots, \sigma_\omega^2}_{s-1}, \sigma_k^2, \sigma_k^2 \right)$$

The smoothed estimates of each of the components can generically be expressed as two-sided weighted moving averages of the form, taking the trend as an example,

$$\hat{\mu}_{t|T} = \sum_{j=t-1}^{j=t-T} w_{j,t} Z_{t-j}$$

The patterns taken by the weights $w_{j,t}$ provide useful information and help the user to understand what a particular model actually does and how it compares with other models. The weights may be calculated by the technique developed in Koopman and Harvey (2003) and automatically adjust as the estimator moves through the sample: again the algorithms for computing these weights are provided in *STAMP*.

For the simple example of the first order trend plus irregular

$$\mathbf{A}_t^{\mathsf{T}} = (\mu_t, \beta_t) \qquad \mathbf{x}_t = \mathbf{x} = (1, 0) \qquad \mathbf{T}_{(2)} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \mathbf{T}$$

so that

$$Z_t = (1, 0)\begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} + \varepsilon_t$$

and

$$\begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \end{pmatrix} + \begin{pmatrix} \eta_t \\ \zeta_t \end{pmatrix}$$

with

$$V(\mathbf{\Omega}_t) = \begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\zeta^2 \end{pmatrix} = \mathbf{Q}$$

The transition equation provides insight into why this particular model is also known as the local linear trend model. Define the expectation of the state at time $t + h$, conditional on information available at time *t*, as

$$\hat{\mathbf{A}}_{t+h|t} = \mathsf{E}(\mathbf{A}_{t+h} | Z_t) \tag{40}$$

This will therefore be given by

$$\hat{\mathbf{A}}_{t+h|t} = \mathbf{T}^h \mathbf{A}_t = \begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix} \mathbf{A}_t$$

so that the estimate of the level at time $t + h$ is

$$\hat{\mu}_{t+h} = \mu_t + \beta_t h$$

i.e, a linear trend in which the intercept and slope are time varying and hence are local parameters.

Estimation of the state can be carried out via the Kalman filter. Using (40), the estimate of the state $\mathbf{A}_t$, given information up to $t-1$, is

$$\hat{\mathbf{A}}_{t|t-1} = \mathsf{E}\left(\mathbf{A}_t \middle| Z_{t-1}\right)$$

The MSE of this estimate is given by

$$\mathbf{P}_{t|t-1} = \mathsf{E}\left\{\left(\mathbf{A}_t - \hat{\mathbf{A}}_{t|t-1}\right)\left(\mathbf{A}_t - \hat{\mathbf{A}}_{t|t-1}\right)^{\mathsf{T}} \middle| Z_{t-1}\right\}$$

To compute these, we use the recursive equations of the Kalman filter, given by

$$\hat{\varepsilon}_t = Z_t - \mathbf{x}\hat{\mathbf{a}}_{t|t-1}$$

$$f_t = \mathbf{x}\mathbf{P}_{t|t-1}\mathbf{x}^{\mathsf{T}} + \sigma_\varepsilon^2$$

$$\mathbf{k}_t = \mathbf{T}\mathbf{P}_{t|t-1}\mathbf{x}^{\mathsf{T}} f_t^{-1}$$

$$\hat{\mathbf{A}}_{t+1|t} = \mathbf{A}\hat{\mathbf{a}}_{t|t-1} + \mathbf{k}_t\hat{\varepsilon}_t$$

$$\mathbf{P}_{t+1|t} = \mathbf{T}\mathbf{P}_{t|t-1}\mathbf{T}^{\mathsf{T}} - \mathbf{k}_t\mathbf{k}_t^{\mathsf{T}} f_t + \mathbf{Q}$$

and starting with $\hat{\mathbf{A}}_{1|0} = \mathbf{0}$ and $\mathbf{P}_{1|0} = \mathbf{Q}$. Here $\mathbf{k}_t$ is known as the Kalman gain, while $\hat{u}_t$ and f_t are the one-step ahead prediction error (or innovation) and its MSE respectively. For further technical details and extensions to the general class of models introduced above, see, for example, Harvey (1989), Harvey and Shephard (1992) and Koopman et al (2006).

The estimate of the current state $\hat{\mathbf{A}}_{t|t-1}$ may be thought of as a one-sided filter. Koopman and Harvey (2003) show that this estimate can be expressed as

$$\hat{\mathbf{A}}_{t|t-1} = \sum_{j=1}^{t-1} \mathbf{w}_j \left(\mathbf{A}_{t|t-1}\right) Z_{t-j}$$

where the weight vectors are computed by backward recursion as

$$\mathbf{w}_j\left(\mathbf{A}_{t|t-1}\right) = \mathbf{B}_{t,j}\mathbf{k}_t, \qquad \mathbf{B}_{t,j-1} = \mathbf{B}_{t,j} - \mathbf{w}_j\left(\mathbf{A}_{t|t-1}\right)\mathbf{x}, \qquad j = t-1, t-2, \cdots, 1$$

with $\mathbf{B}_{t|t-1} = \mathbf{I}$. A two-sided filter estimate of the state, known as the *smoothed* estimate, is obtained by running the backwards recursion

$$\mathbf{L}_t = \mathbf{T} - \mathbf{k}_t \mathbf{x}$$

$$\mathbf{r}_{t-1} = \mathbf{x}^{\mathsf{T}} f_t^{-1} \hat{\varepsilon}_t + \mathbf{L}_t^{\mathsf{T}} \mathbf{r}_t,$$

$$\mathbf{N}_{t-1} = \mathbf{x}^{\mathsf{T}} \mathbf{x} f_t^{-1} + \mathbf{L}_t^{\mathsf{T}} \mathbf{N}_t \mathbf{L}_t \qquad t = T, T-1, \cdots, 1$$

with initialisation $\mathbf{r}_T = \mathbf{N}_T = \mathbf{0}$ for a suitable choice of *T*. The smoothed state and associated MSE matrix are then calculated as

$$\hat{\mathbf{A}}_{t|T} = \hat{\mathbf{A}}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{r}_{t-1}$$

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{N}_{t-1}\mathbf{P}_{t|t-1}$$

Again, the smoothed estimate can be expressed in the form

$$\hat{\mathbf{A}}_{t|T} = \sum_{j=1}^{T-1} \mathbf{w}_j \left(\hat{\mathbf{A}}_{t|T}\right) Z_{T-j}$$

A structural model was fitted to the logarithms of U.S. GNP. As this series is already seasonally adjusted, the seasonal component γ_t is absent, and the autoregressive component ν_t was also found to be unnecessary. The best fit was obtained with a first order trend (but with $\sigma_\eta^2 = 0$: this is known as the *smooth* trend model) and a second order cycle:

$$Z_t = \mu_t + \psi_{2,t} + \varepsilon_t$$

$$\mu_t = \mu_{t-1} + \beta_{t-1}$$

$$\beta_t = \beta_{t-1} + \zeta_t$$

$$\begin{bmatrix} \psi_{1,t} \\ \psi^*_{1,t} \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{bmatrix} \psi_{1,t-1} \\ \psi^*_{1,t-1} \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa^*_t \end{bmatrix}$$

$$\begin{bmatrix} \psi_{2,t} \\ \psi^*_{2,t} \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{bmatrix} \psi_{1,t-1} \\ \psi^*_{1,t-1} \end{bmatrix} + \begin{bmatrix} \psi_{1,t-1} \\ \psi^*_{1,t-1} \end{bmatrix}$$

The estimated parameters (along with 68% (one standard error) bounds, which are asymmetric when the parameter is bounded) are

$$\hat{\sigma}_\zeta = 0.00082 \qquad 0.00037 \le \sigma_\zeta \le 0.00179$$

$$\hat{\sigma}_\kappa = 0.00521 \qquad 0.00400 \le \sigma_\kappa \le 0.00679$$

$$\hat{\sigma}_\psi(2) = 0.00758 \qquad 0.00582 \le \sigma_\psi(2) \le 0.00988$$

$$\hat{\rho} = 0.726 \qquad 0.692 \le \rho \le 0.759$$

$$\hat{\lambda}_c = 0.280 \qquad 0.175 \le \lambda_c \le 0.438$$

$$2\pi/\hat{\lambda}_c = 22.47 \qquad 14.35 \le 2\pi/\lambda_c \le 35.92$$

$$\rho_\tau(2) = 0.73^\tau \cos 0.28\tau(1 + 0.31\tau)$$

The estimated trend component is plotted superimposed on Z_t in figure 4: with $\sigma^2_\eta = 0$ the trend is indeed seen to be smooth. The cyclical and irregular components are shown in figure 5. The stochastic cycle has an average period of 22 quarters, consistent with most views of a business cycle (e.g., Baxter and King, 1999), although the amplitude of the cycles has decreased since the mid-1980s, again consistent with other models of U.S. output. Similarly, the irregular component displays decreased volatility since this time, consistent with the findings, obtained using rather different approaches, of Kim and Nelson (1999) and McConnell and Quiros (2000) concerning what has become known as the 'Great Moderation'.

Figure 6 plots the weight functions for the trend and cycle computed at the middle and at the end of the sample, and shows how the filters are modified throughout the sample. Forecasts of Z_t out to end-2012 are shown in figure 6, along with 68% forecast intervals. In comparison with the BN forecasts, the structural forecasts are slightly lower but have larger standard errors, being 30% larger by the end of the forecast period.

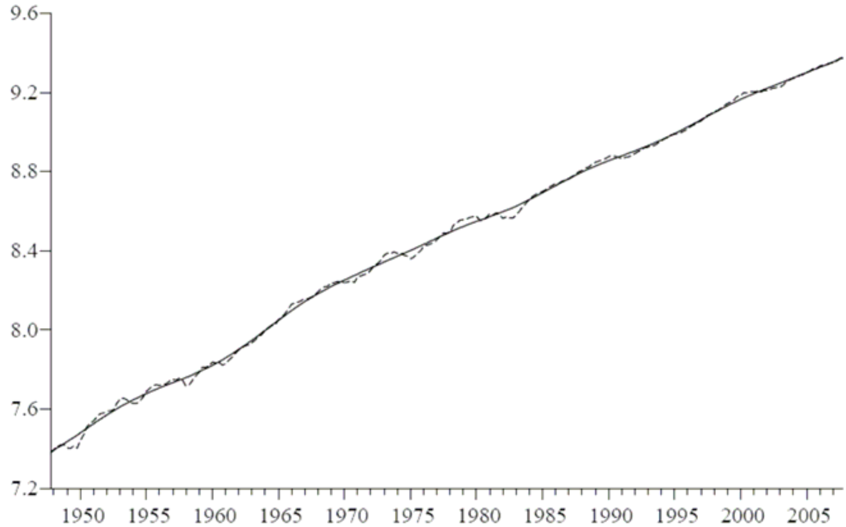


Figure 4. Structural trend component superimposed on the logarithms of U.S. GNP.

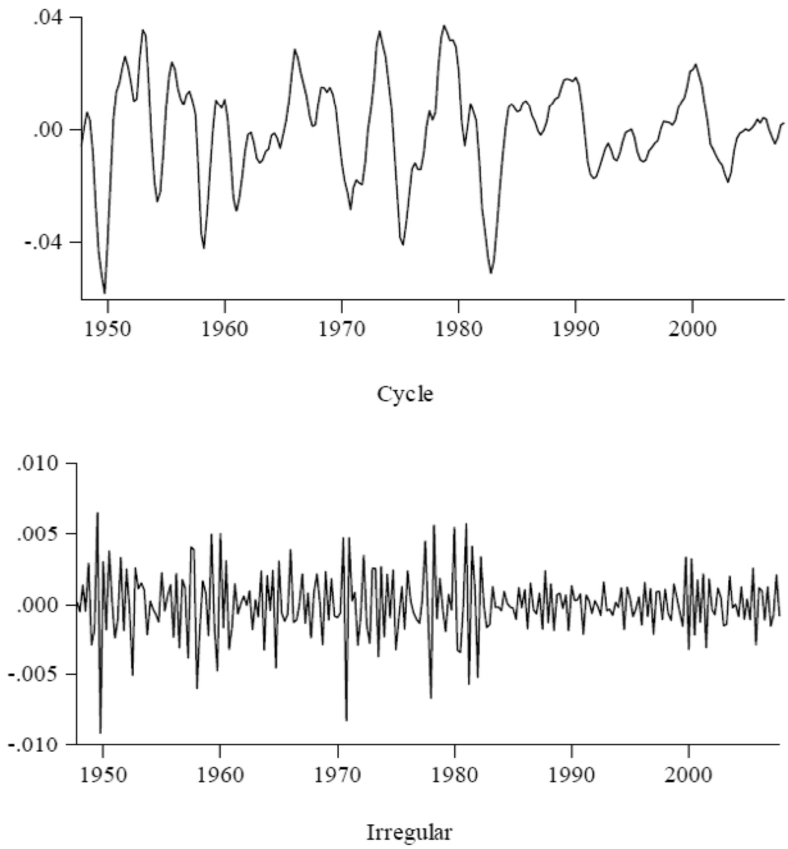


Figure 5. Structural cycle and irregular components for the logarithms of U.S. GNP.

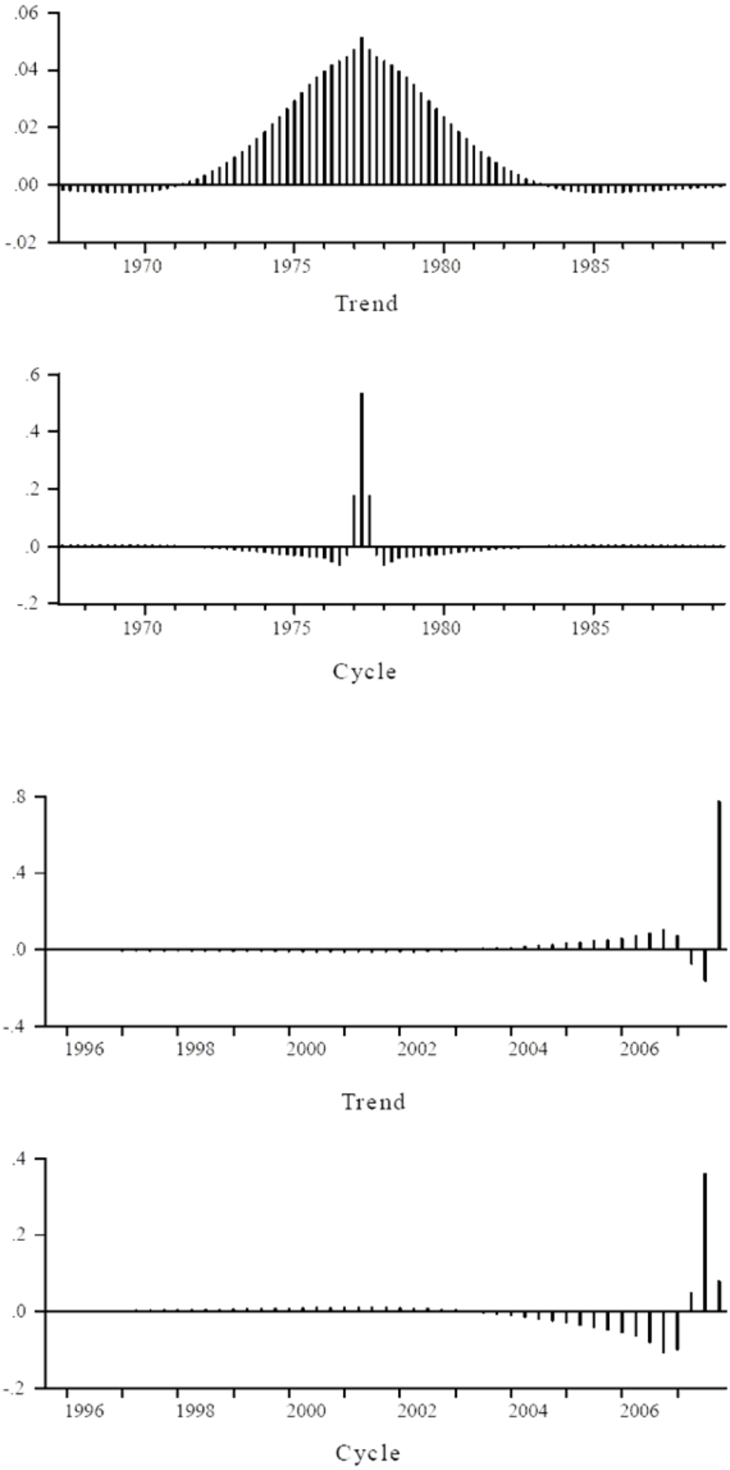


Figure 6. Trend and cycle weights. Top panel: weights at middle of sample (1977.2); bottom panel: weights at end of sample (2007.4).

5. Trend Filters

5.1. Henderson and Macaulay Trend Filters

An alternative to the structural model for representing local trend behaviour is to allow the signal to be a polynomial in time,

$$S_t = \sum_{j=0}^{p} \beta_j t^j$$

but to allow the β_j coefficients to vary across moving windows of size $2r+1$ observations. If the central time point of this window is *t*, S_t may then be estimated by the finite moving average

$$\hat{S}_t^{(r)} = \sum_{j=-r}^{r} w_j Z_{t+j}$$

The w_j weights will be particular to the choice of polynomial *p* and window $2r+1$. As a simple example (see Kendall, 1976, section 3.3, and Mills, 2007), suppose $p = r = 3$. Without loss of generality, we can take the time points in the window to be centred on $t = 0$, i.e., $t = -3, -2, -1, 0, 1, 2, 3$. As well as simplifying subsequent calculations, this normalises the central point of the window to $S_0 = \beta_0$. Fitting the β_j to the data Z_t by least squares requires minimising

$$\sum_{t=-3}^{3} N_t^2 = \sum_{t=-3}^{3} \left(Z_t - S_t\right)^2 = \sum_{t=-3}^{3} \left(Z_t - \beta_0 - \beta_1 t - \beta_2 t^2 - \beta_3 t^3\right)^2$$

Differentiating with respect to the β_j s gives the set of four equations

$$\sum Z_t t^j - \beta_0 \sum t^j - \beta_1 \sum t^{j+1} - \beta_2 \sum t^{j+2} - \beta_3 \sum t^{j+3} = 0 \qquad j = 0,1,2,3$$

Given the standardising of the window, $\sum t = \sum t^3 = 0$, $\sum t^0 = 7$, $\sum t^2 = 28$ and $\sum t^4 = 196$, so that the equations for $j = 0$ and $j = 2$ are

$$\begin{aligned} \sum Z_t &= 7\beta_0 + 28\beta_2 \\ \sum Z_t t^2 &= 28\beta_0 + 196\beta_2 \end{aligned} \tag{41}$$

Solving for β_0 yields

$$\beta_0 = \tfrac{1}{21}\left(7\sum Z_t - \sum Z_t t^2\right)$$

or

$$\beta_0 = \tfrac{1}{21}\begin{pmatrix} 7\left(Z_{-3} + Z_{-2} + Z_{-1} + Z_0 + Z_1 + Z_2 + Z_3\right) \\ -\left(9Z_{-3} + 4Z_{-2} + Z_{-1} + 0 + Z_1 + 4Z_2 + 9Z_3\right) \end{pmatrix}$$

$$= \tfrac{1}{21}\left(-2Z_{-3} + 3Z_{-2} + 6Z_{-1} + 7Z_0 + 6Z_1 + 3Z_2 - 2Z_3\right)$$

Thus, $\hat{S}_0^{(3)} = \beta_0$ and, in general,

$$\hat{S}_t^{(3)} = -\tfrac{2}{21}Z_{t-3} + \tfrac{3}{21}Z_{t-2} + \tfrac{6}{21}Z_{t-1} + \tfrac{7}{21}Z_t + \tfrac{6}{21}Z_{t+1} + \tfrac{3}{21}Z_{t+2} - \tfrac{2}{21}Z_{t+3}$$

It is seen that the weights are symmetric and sum to unity. It can also be checked that

$$\sum jw_j = \sum j^2 w_j = \sum j^3 w_j = 0$$

These results are perfectly general, so that for a *p*th order polynomial trend

$$\sum w_j = 1 \qquad \sum j^k w_j = 0 \qquad 0 < k \le p \tag{42}$$

However, the weights will depend upon both *r* and *p* and many formulae are provided in Kendall (1976, section 3.4). Note, however, that the equations in (41) show that β_0 does not depend on β_3, so that the same expression for β_0, and hence $\hat{S}_t^{(3)}$, would have resulted if a quadratic, rather than a cubic, had been fitted. This result is perfectly general: the moving average (filter) formula for an odd value of *p* also holds for the next lowest (even) *p* value.

The least squares criterion used above is equivalent to minimising

$$\mathsf{F} = \mathsf{E}\left(\left(\hat{S}_t^{(r)} - S_t\right)^2\right)$$

subject to the restrictions (42). These are sometimes referred to as Macaulay (1931) filters and F is said to measure the *fidelity* of the filter. An alternative criteria is that of *smoothness*, given by

$$\mathsf{S} = \mathsf{E}\left(\left(\Delta^{p+1}\hat{S}_t^{(r)}\right)^2\right)$$

Taking $p = 3$ as an example and recalling that σ_u^2 is the variance of the white noise component $N_t = u_t$, Kenny and Durbin (1982) show that S can be expressed as

$$\mathsf{S} = \sigma_u^2 \sum_{j=-r}^{r} \left(\widetilde{\Delta}^3 w_j\right)^2 \qquad \widetilde{\Delta} = (B^{-1} - 1) = B^{-1}\Delta$$

Minimising S subject to (42), along with the 'boundary' conditions $w_j = 0$, $j > |r|$, using Lagrange multipliers, yields

$$\widetilde{\Delta}^6 w_{j-3} = a + bj^2 \qquad j \leq |r|$$

The solution to this equation will be a polynomial of order 8 with roots $\pm(r+1)$, $\pm(r+2)$, $\pm(r+3)$, so that

$$w_j = \left((r+1)^2 - j^2\right)\left((r+2)^2 - j^2\right)\left((r+3)^2 - j^2\right)\left(a + bj^2\right)$$

The coefficients *a* and *b* are determined from the constraints (42), so that Kenny and Durbin (1982) report that

$$w_j = \left((r+1)^2 - j^2\right)\left((r+2)^2 - j^2\right)\left((r+3)^2 - j^2\right)\left(3(r+2)^2 - 16 - 11j\right)$$

The so defined moving average $\widetilde{S}_t^{(r)}$ is known as a Henderson (1924) filter and is used in the X-11 seasonal adjustment procedure, with typical choices of *r* being 9, 13 or 23. The weights for the three filters are reported in, for example, Kenny and Durbin (1982).

Gray and Thompson (1996) extend this approach in several ways. The Henderson filters used in X-11 also result if a quadratic local model is used and the filters are allowed to be asymmetric. Gray and Thompson also consider the filters that result when a linear combination of F and S is considered and when the model for S_t also includes a random walk component. They suggest that such models may be estimated by fitting local polynomial trends by generalised least squares.

5.2. Hodrick-Prescott, Butterworth and Low-Pass Filters

A filter that uses the concepts of fidelity and smoothness but in a somewhat different way is that typically associated with Hodrick and Prescott (1997), although it had been proposed earlier by Leser (1961). This filter, commonly known as the HP (trend) filter, was originally developed as the solution to the problem of minimising the variation in the noise component subject to a smoothness condition on the signal, an approach that has a long history of use in, for example, actuarial science: see Whittaker (1923). This smoothness condition penalises acceleration in the trend, so that the function being minimised is

$$\sum N_t^2 + \xi \sum \left((S_{t+1} - S_t) - (S_t - S_{t-1})\right)^2$$

with respect to S_t, and where ξ is a Lagrangean multiplier that has the interpretation of a smoothness parameter. The higher the value of ξ, the smoother the trend, so that in the limit, as $\xi \to \infty$, S_t becomes a linear trend. The first-order conditions are

$$\begin{aligned} 0 = -2(Z_t - S_t) + 2\xi((S_t - S_{t-1}) - (S_{t-1} - S_{t-2})) \\ - 4\xi((S_{t+1} - S_t) - (S_t - S_{t-1})) \\ + 2\xi((S_{t+2} - S_{t+1}) - (S_{t+1} - S_t)) \end{aligned}$$

which may be written as

$$Z_t = S_t + \xi(1-B)^2(S_t - 2S_{t+1} + S_{t=2}) = \left(1 + \xi(1-B)^2(1-B^{-1})^2\right)S_t$$

or, in terms of the signal, as

$$S_t = \frac{1}{1+\xi(1-B)^2(1-B^{-1})^2} Z_t = \frac{1}{1+\xi|1-B|^4} Z_t \tag{43}$$

When written in this form, the expression may be compared with the model of section 3.4. Recall equations (28) and (29), now with $\zeta_S(B) = 1 - B$ and $\zeta_N(B) = 1$. The ARIMA model for Z_t will then have the AR polynomial $\zeta(B) = 1 - B$ and, from (30), a moving average polynomial satisfying

$$\sigma_e^2|\theta(B)|^2 = \sigma_v^2|\eta_S(B)|^2 + \sigma_u^2|1-B|^2|\eta_N(B)|^2$$

The MMSE estimator of the signal is then

$$\hat{S}_{t|\infty}^{HP} = \frac{\sigma_v^2|\eta_S(B)|^2}{\sigma_e^2|\theta(B)|^2} Z_t = \frac{|\eta_S(B)|^2}{|\eta_S(B)|^2 + (\sigma_u^2/\sigma_v^2)|1-B|^2|\eta_N(B)|^2} Z_t = w_{HP}(B)Z_t$$

Comparing this with (43) shows that they will be identical if

$$\eta_S(B) = (1-B)^{-1} \qquad \eta_N(B) = 1 \qquad \xi = \sigma_u^2/\sigma_v^2$$

In other words, the underlying unobserved component model must have a signal given by

$$\Delta^2 S_t = v_t$$

so that it is, in effect, a 'smooth structural trend' with a white noise irregular component. In terms of the ARIMA$(0,2,2)$ model (37), $\xi = (1-\theta_2)^4 / \theta_2 (1+\theta_2)^2$. Setting $\xi = 1600$, i.e., $\sigma_u^2 = 1600\sigma_v^2$, produces the HP filter proposed by Hodrick and Prescott (1997) for quarterly observations, in which the innovations to the trend slope are extremely small compared to the noise. Figure 8 shows this HP filter superimposed on the logarithms of U.S. GNP. Compared to the structural decomposition, the HP trend is smoother and the HP cycle has correspondingly more variation.

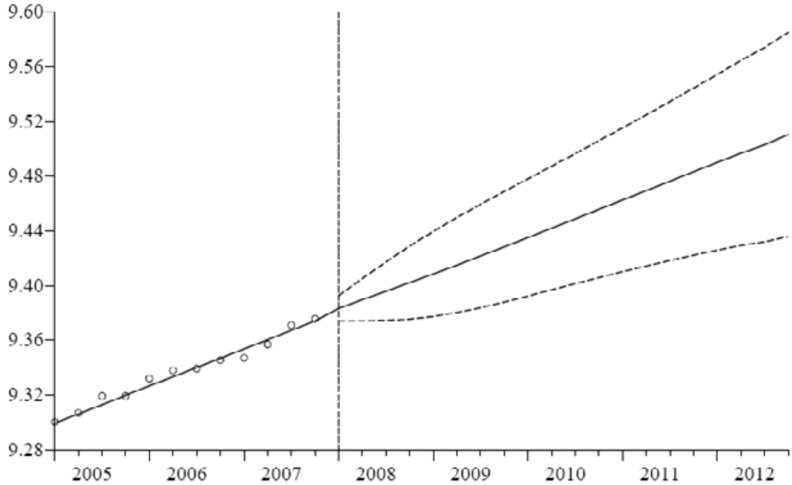


Figure 7. Forecasts of the trend component of the logarithms of U.S. GNP until 2012.4 with 68% prediction intervals.

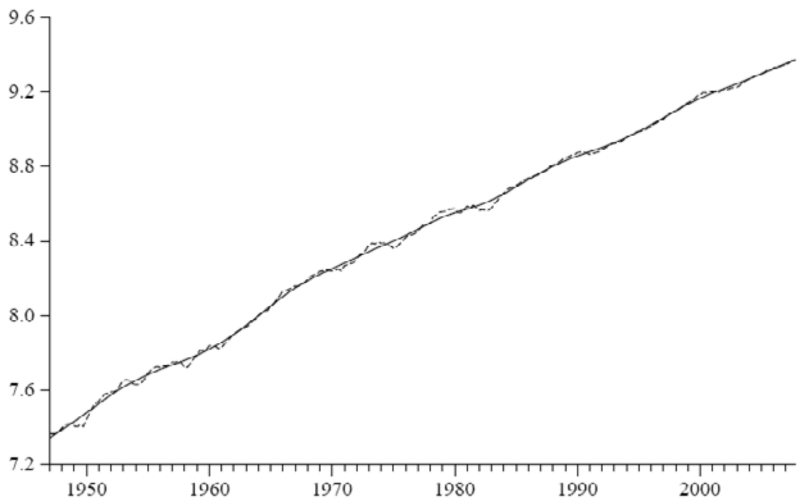


Figure 8. Continued on next page.

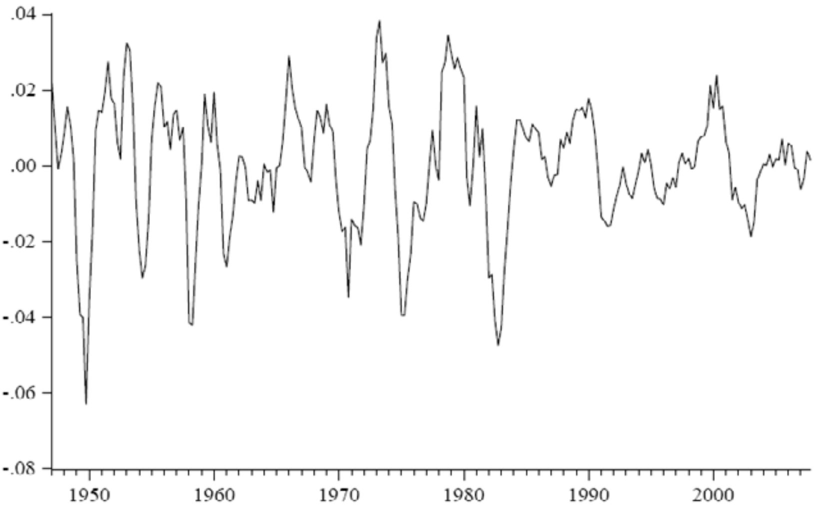


Figure 8. HP trend and cycle computed with $\xi = 1/1600$ for the logarithms of U.S. GNP.

Consider now an alternative UC model taking the form

$$\Delta^d S_t = (1+B)^n v_t$$

and

$$N_t = (1-B)^{n-d} u_t$$

The MMSE estimate of the signal is then

$$\hat{S}^B_{t|\infty} = \frac{|1+B|^{2n}}{|1+B|^{2n} + \xi^{-1}|1-B|^{2n}} Z_t = \frac{1}{1+\xi^{-1}\left(\frac{|1-B|^2}{|1+B|^2}\right)^n} Z_t = w_B(B) Z_t \tag{44}$$

The filter in (44) is known as the Butterworth (square-wave) filter and has a long history of use in electrical engineering (see Pollock, 1999, 2000, and the references therein).

The HP and Butterworth filters are closely related to the idea of a low-pass filter. Consider again the moving average filter

$$\hat{S}^{(r)}_t = \sum\nolimits_{j=-r}^{r} w_j Z_{t+j} = w(B) Z_t$$

It can be useful to consider the frequency domain properties of this filter. For a frequency λ, the *frequency response function* of the filter is defined as

$$w(\lambda) = \sum\nolimits_{j=-r}^{r} w_j e^{-i\lambda j}$$

The *gain* of the filter is defined as $|w(\lambda)|$ and, if the filter weights are symmetric, $|w(\lambda)| = w(\lambda)$ since $w(\lambda) = w(-\lambda)$.

A basic building block in the design of filters is the *low-pass* filter, $w_L(\lambda)$ say, which passes only frequencies in the interval $\lambda < \lambda_L$, where λ_L is the cut-off frequency. It therefore has the frequency response function

$$w_L(\lambda) = \begin{cases} 1 & \text{if } \lambda < |\lambda_L| \\ 0 & \text{if } \lambda > |\lambda_L| \end{cases}$$

Since the filter passes low-frequency components but retains the high frequency ones, it should have trend-extraction properties, which requires that $w(0) = 1$, which is ensured if the weights sum to unity. The complement of the low-pass filter is thus the high-pass filter having weights summing to zero.

In general, the frequency response function of the low-pass filter is given by

$$w_L(\lambda) = \sum_{j=-\infty}^{\infty} w_{L,j} e^{-i\lambda j}$$

so that the filter weights are given by the 'inverse'

$$w_{L,j} = \frac{1}{2\pi}\int_{-\pi}^{\pi} w_L(\lambda) e^{i\lambda j} d\lambda = \frac{1}{2\pi}\int_{-\lambda_L}^{\lambda_L} e^{i\lambda j} d\lambda$$

Thus, using standard trigonometric results

$$w_{L,0} = \frac{1}{2\pi}\int_{-\lambda_L}^{\lambda_L} d\lambda = \frac{\lambda_L}{\pi}$$

and

$$w_{L,j} = \frac{1}{2\pi}\left[\frac{e^{i\lambda j}}{ij}\right]_{-\lambda_L}^{\lambda_L} = \frac{\sin(\lambda_L j)}{\pi j} \qquad j \neq 0$$

The smoothing parameter ξ in the Butterworth filter is typically chosen to correspond to a frequency cut-off of λ_L. As the frequency response function of the filter is

$$w_B\left(e^{-i\lambda}\right) = \frac{1}{1 + \xi(\tan(\lambda/2))^{2n}}$$

and since the gain should equal 0.5 at λ_L, solving this equation for this value gives

$$\xi_L = \left(1/\tan(\lambda_L)\right)^{2n}$$

In this way the Butterworth filter can be made to closely approximate the ideal low-pass filter by choosing λ_L and *n* appropriately. In a similar fashion, the HP filter can also approximate the ideal low-pass filter by optimally choosing its smoothing parameter (see, for example, Pedersen, 2001, and Harvey and Trimbur, 2008). In this case the smoothing parameter and cut-ff frequency are related by $\xi_L = 1/\left(4(1-\cos\lambda_L)^2\right)$. For $\xi_L = 1600$, $\lambda_L = 0.1583$, which corresponds to a period of 39.7 quarters, so that the cyclical filter contains components with periodicity less than approximately 40 quarters.

A related idea is that of a *band pass* filter, which passes only those components having frequencies lying in the interval $\lambda_l \leq \lambda \leq \lambda_u$. A variety of band pass filters have been proposed that take their cue from the idea that the typical business cycle has a period of between 6 and 32 quarters, i.e., $0.20 \leq \lambda \leq 1.05$ (see Baxter and King, 1999, and Christiano and Fitzgerald, 2003). Thus, in comparison, the HP cyclical filter will contain more high frequency noise (components with periods less than six quarters) and some lower frequency components than would be considered part of the trend by the bandpass filters.

6. Nonlinear and Nonparametric Trends

A general specification for the signal is to allow it to be some function of time, i.e., $S_t = f(t)$. Most of the specifications so far considered can be put into this form, although with the addition of an innovation to enable them to be local and/or stochastic. We consider here specifications that are nonlinear but deterministic, in the sense that no innovations appear in the model for the signal. Although local polynomials of *t* have been considered, 'global' polynomials would seem to be unattractive. Low values of the polynomial order would force too restrictive a shape onto the signal function, while setting the order too high runs the risk of overfitting, so that short-run movements become part of the signal, which then ceases to be a smooth, only slowly changing, function of time.

One possibility is to consider the class of segmented or breaking trend models: see, for example, Bacon and Watts (1971) and the discussion and references contained in Mills (2003, chapter 2). Suppose, as is typical with this class of models, that we consider a linear function with two regimes, having a change point at time τ. This may be modelled as

$$S_t = \alpha + \beta t + \gamma d_t + \delta t d_t \tag{45}$$

where d_t is a dummy variable defined as

$$d_t = \begin{cases} 1 & t > \tau \\ 0 & t \leq \tau \end{cases}$$

The problem with this formulation is that the signal will shift abruptly at τ from $\alpha + \beta\tau$ to $\alpha + \gamma + (\beta + \delta)\tau$. Continuity can be imposed through the restriction

$$\alpha + \beta\tau = \alpha + \gamma + (\beta + \delta)\tau$$

This implies that $\gamma + \delta\tau = 0$ and imposing this restriction on (45) yields

$$S_t = \alpha + \beta t + \delta d_t (t - \tau) = \alpha + \beta t + \delta\varphi_t$$

on defining

$$\varphi_t = d_t(t - \tau) = \begin{cases} t - \tau & t > \tau \\ 0 & t \le \tau \end{cases}$$

Extensions to more than one break, to higher order polynomials and, indeed, combinations of polynomials of different orders, are straightforward. Estimation of the model is also straightforward. If the noise component is white, OLS estimation of the regression

$$Z_t = \alpha + \beta t + \delta\varphi_t + u_t \tag{46}$$

provides a MMSE estimate of S_t, while if the noise is an ARMA process, some form of feasible GLS will be required.

Although the segmented trend imposes continuity, it does not impose a continuous first derivative at the break point, so that the shift to a new regime is not smooth. The related smooth transition model allows the signal to change gradually and smoothly between the two regimes. A logistic smooth transition takes the form

$$S_t = \alpha_1 + \beta_1 t + (\alpha_2 + \beta_2 t) R_t(\theta, \tau)$$

where

$$R_t(\theta, \tau) = \left(1 + \exp(-\theta(t - \tau))\right)^{-1}$$

is the logistic smooth transition function controlling the transition between the two regimes. τ now has the interpretation of the transition mid-point since, for $\theta > 0$, $R_{-\infty}(\theta, \tau) = 0$, $R_{\infty}(\theta, \tau) = 1$ and $R_{\tau}(\theta, \tau) = 0.5$. The speed of the transition is determined by θ. If θ is small then $R_t(\theta, \tau)$ takes a long time to traverse the interval $(0,1)$ and, in the limiting case when $\theta = 0$, $R_t(\theta, \tau) = 0.5$ for all *t*. In this case

$$S_t = \alpha_1 + 0.5\alpha_2 + (\beta_1 + 0.5\beta_2)t$$

and there is just a single regime. For large values of θ, $R_t(\theta,\tau)$ traverses the interval $(0,1)$ very rapidly, and as θ approaches $+\infty$ it changes from 0 to 1 instantaneously at τ. In this case the smooth transition is tantamount to the discontinuous segmented trend model (45). If $\theta < 0$ then the initial and final regimes are reversed but the interpretation of the parameters remains the same. The smooth transition has the appealing property that the midpoint of the transition can be estimated, unlike the segmented trend model where the break times have to be determined exogenously or through a time consuming iterative process. Although only two regimes are allowed, this may not be a problem if the transition is slow.

A more general approach is to fit an unknown signal function $f(t)$ nonparametrically. This may be done by, for example, the nearest neighbours technique or by kernel local polynomial regression. Detailed discussion of nonparametric curve fitting techniques may be found in Simonoff (1996) and Wand and Jones (1995), for example, while Mills (2003, chapter 5) discusses the application of such techniques to the fitting of trends in time series.

A segmented trend with a break imposed at $\tau \equiv 1972.4$ (similar to the break imposed by Perron , 1989) was fitted to the logarithms of U.S. GNP with the noise component modelled as an AR(2) process, producing

$$Z_t = \underset{(0.020)}{7.364} + \underset{(0.00026)}{0.00944}\, t - \underset{(0.00037)}{0.00198}\, \varphi_t + N_t$$

$$N_t = \underset{(0.060)}{1.267}\, N_{t-1} - \underset{(0.060)}{0.362}\, N_{t-2} + u_t \qquad \hat{\sigma}_u = 0.00901$$

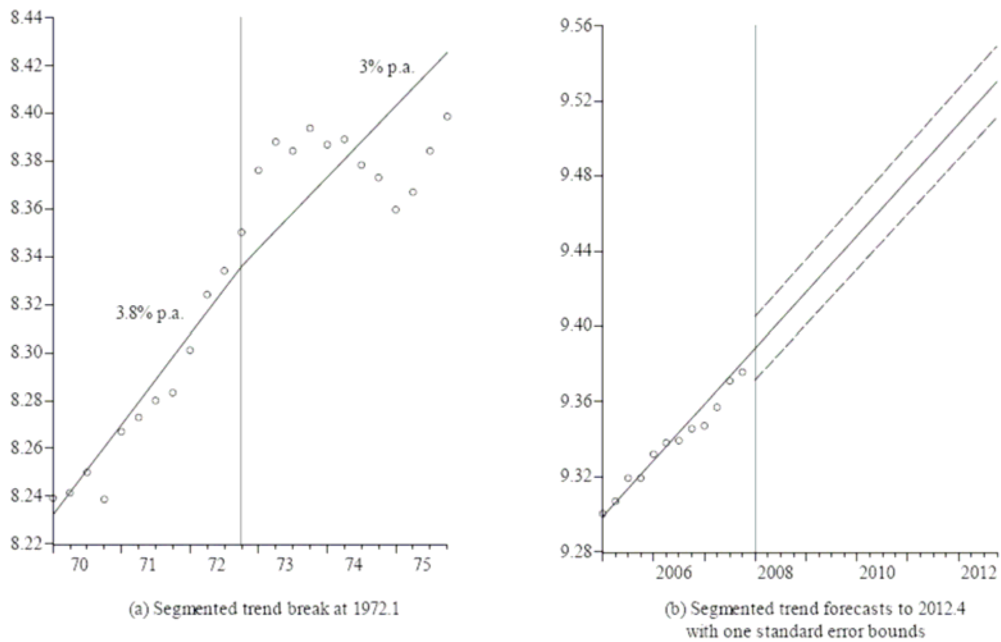


Figure 9. Segmented trend fitted to the logarithms of U.S. GNP.

Trend growth is thus 0.94% per quarter (3.8% per annum) between 1947 and 1972 and $0.944-0.198=0.75\%$ per quarter (3% per annum) from 1973 onwards. The parameters of the fitted AR(2) noise component imply a pair of real roots of magnitude 0.83 and 0.43, so that this does not exhibit (pseudo-)cyclical behaviour. The segmented trend break and forecasts out to 2012.4 are shown in figure 9.

7. Signal Extraction from Finite Samples

7.1. ARIMA Components

The use of ARIMA processes for estimating the signal and noise components in section 3 has assumed that an *infinite* sample is available: hence the summation range and notation used for the W-K MMSE estimator shown in equation (3). In this section issues arising when only a *finite* sample is available, i.e., $Z_1,\ldots,Z_{t-1},Z_t,Z_{t+1},\ldots,Z_T$, are considered. The focus here is on the 'end-point' problem: that all two-sided estimators of components will fail to provide estimates near to the start and end of the finite sample.

We begin by considering estimating S_T, so that (3) becomes the concurrent, or one-sided, filter

$$S_{T|0} = \sum_{j=0}^{\infty} \varpi_j Z_{T-j} = \varpi_S^{(0)}(B) Z_T$$

Given the ARIMA components for S_t and N_t shown in (28) and (29), then from Bell and Martin (2004), for example,

$$\varpi_S^{(0)}(B)=\frac{\sigma_v^2}{\sigma_e^2}\frac{\zeta(B)}{\theta(B)}\left[\frac{\zeta(B^{-1})}{\theta(B^{-1})}\frac{|\eta_S(B)|^2}{|\zeta_S(B)|^2}\right]_+ \tag{47}$$

The notation $[\cdot]_+$ indicates that only terms with non-negative powers of *B* are retained. If it is assumed that $\zeta_S(B)$ and $\zeta_N(B)$ have no common factors, so that $\zeta(B)=\zeta_S(B)\zeta_N(B)$, then (47) becomes

$$\varpi_S^{(0)}(B)=\frac{\sigma_v^2}{\sigma_e^2}\frac{\zeta_S(B)\zeta_N(B)}{\theta(B)}\left[\frac{\zeta_N(B^{-1})}{\theta(B^{-1})}\frac{|\eta_S(B)|^2}{\zeta_S(B)}\right]_+ \tag{48}$$

It can be shown that the expression inside $[\cdot]_+$ in (48) can be written as (Bell and Martin, 2004)

$$\frac{\zeta_N(B^{-1})\eta_S(B)\eta_S(B^{-1})}{\theta(B^{-1})\zeta_S(B)}=\frac{c(B^{-1})}{\theta(B^{-1})}+\frac{d(B)}{\zeta_S(B)} \tag{49}$$

where

$$c\left(B^{-1}\right)=c_1B^{-1}+\cdots+c_hB^{-h}$$

and

$$d(B)=d_0+d_1B+\cdots+d_kB^k$$

satisfy the relationship

$$c\left(B^{-1}\right)\zeta_S(B)+d(B)\theta\left(B^{-1}\right)=\zeta_N\left(B^{-1}\right)\eta_S\left(B^{-1}\right)\eta_S(B)$$

and

$$h=\max\left(q,p_N+q_S\right) \qquad k=\max\left(p_S-1,q_S\right)$$

Here p_N and q_S are the orders of ζ_N and η_S. Two algorithms for computing $c\left(B^{-1}\right)$ and $d(B)$ are given in the Appendix of Bell and Martin (2004) and in Bell and Martin (2002), respectively.

With the partial fraction expansion (49), the filter (48) can be rewritten as

$$\varpi_S^{(0)}(B)=\frac{\sigma_v^2}{\sigma_e^2}\frac{\zeta_S(B)\zeta_N(B)}{\theta(B)}\left[\frac{c\left(B^{-1}\right)}{\theta\left(B^{-1}\right)}+\frac{d(B)}{\zeta_S(B)}\right]_+$$

which, because the expansion of $c\left(B^{-1}\right)/\theta\left(B^{-1}\right)$ involves only negative powers of *B*, reduces to

$$\varpi_S^{(0)}(B)=\frac{\sigma_v^2}{\sigma_e^2}\frac{\zeta_N(B)d(B)}{\theta(B)} \tag{50}$$

As an example of this approach, consider again the random walk signal plus white noise model (31). In this case, the filter (48) is

$$\varpi_S^{(0)}(B)=\frac{\sigma_v^2}{\sigma_e^2}\frac{(1-B)}{(1+\theta B)}\left[\frac{1}{\left(1+\theta B^{-1}\right)(1-B)}\right]_+$$

and, since $h=1$ and $k=0$, the partial fraction expansion (49) is

$$\frac{1}{\left(1+\theta B^{-1}\right)(1-B)}=\frac{c_1B^{-1}}{\left(1+\theta B^{-1}\right)}+\frac{d_0}{(1-B)}$$

Solving

$$c_1B^{-1}(1-B)+d_0\left(1+\theta B^{-1}\right)=d_0-c_1+\left(c_1+\theta d_0\right)B^{-1}=1$$

yields

$$c_1 = -\frac{\theta}{1+\theta} \qquad d_0 = \frac{1}{1+\theta}$$

so that (50) becomes

$$\varpi_S^{(0)}(B) = \frac{\sigma_v^2}{\sigma_e^2(1+\theta)}\frac{1}{(1+\theta B)} = \frac{\sigma_v^2}{\sigma_e^2(1+\theta)}\sum_{j=0}^{\infty}(\theta B)^j = (1+\theta)\sum_{j=0}^{\infty}(\theta B)^j$$

on using $\sigma_v^2 = (1+\theta)^2\sigma_e^2$.

Consider now estimating S_{T-m}, $m > 0$. The MMSE estimator is

$$S_{T-m|m} = \sum_{j=-m}^{\infty}\varpi_j Z_{T-m-j} = \varpi_S^{(m)}(B)Z_{T-m}$$

where the asymmetric filter $\varpi_S^{(m)}(B)$ is now given by

$$\varpi_S^{(m)}(B) = B^{-m}\frac{\sigma_v^2}{\sigma_e^2}\frac{\zeta_S(B)\zeta_N(B)}{\theta(B)}\left[\frac{\zeta_N(B^{-1})}{\theta(B^{-1})}\frac{|\eta_S(B)|^2}{\zeta_S(B)}B^m\right]_+$$

It then follows that (50) extends to

$$\varpi_S^{(m)}(B) = \frac{\sigma_v^2}{\sigma_e^2}\frac{\zeta_N(B)d_S^{(m)}}{\theta(B)}B^{-m}$$

where $d_S^{(m)}$ is a polynomial in B of order $k_S^{(m)} = \max(p_S - 1, q_S + m)$. A parallel derivation establishes that the MMSE estimator of N_t is

$$\varpi_N^{(m)}(B) = \frac{\sigma_u^2}{\sigma_e^2}\frac{\zeta_S(B)d_N^{(m)}}{\theta(B)}B^{-m}$$

where $d_N^{(m)}$ is of order $k_N^{(m)} = \max(p_N - 1, q_N + m)$ and where $\varpi_N^{(m)}(B) = 1 - \varpi_S^{(m)}(B)$. For $m < 0$, (Z_t in the future), this relationship does not hold, although a generalised result does (see Bell and Martin, 2004, equation (25)).

For the random walk signal plus white noise model, it can be shown (Pierce, 1979) that, for $m \geq 0$,

$$\varpi_S^{(m)}(B) = (1+\theta)B^m\sum_{j=0}^{\infty}(\theta B)^j$$

while for $m < 0$

$$\varpi_S^{(m)}(B) = (-\theta)^{-m}(1+\theta)B^m \sum_{j=0}^{\infty}(\theta B)^j + (1+\theta B)^{-1}\sum_{j=0}^{-m-1}\theta^j B^{-j}$$

Thus, when estimating S_t for the current period ($m = 0$) or recent periods ($m > 0$), an exponentially weighted moving average is applied to Z_t, beginning with the most recent data available, but not otherwise depending on *m*. For $m < 0$, when S_t is being estimated using some, but not all, of the relevant future observations on Z_t, the filter consists of two parts: the same filter as in the $m \geq 0$ case, but applied to the furthest forward observation and with a declining weight $\left(-\theta^{-m}\right)$ placed upon it, and a second term capturing the additional influence of the observed future observations.

7.2. Moving Average Filters

As noted above, the filters constructed in section 5.1 cannot be used to estimate the trend at the beginning or end of the sample. Concentrating on the latter, which is usually of much greater interest, consider again fitting a cubic using a window of width 7 centered on $t = 0$, but where we now want to estimate S_t at $t = 1, 2, 3$ as well. Supplementing (41) with the equations for $j = 1$ and 3 gives

$$\begin{aligned}
\sum Z_t &= 7\beta_0 + 28\beta_2 \\
\sum Z_t t &= 28\beta_1 + 196\beta_3 \\
\sum Z_t t^2 &= 28\beta_0 + 196\beta_2 \\
\sum Z_t t^3 &= 196\beta_1 + 1588\beta_3
\end{aligned}$$

These equations may be solved to yield

$$\beta_0 = \tfrac{1}{21}\left(7\sum Z_t - \sum Z_t t^2\right)$$

$$\beta_1 = \tfrac{1}{1512}\left(397\sum Z_t t - 49\sum Z_t t^3\right)$$

$$\beta_2 = \tfrac{1}{84}\left(-4\sum Z_t + \sum Z_t t^2\right)$$

$$\beta_3 = \tfrac{1}{216}\left(-7\sum Z_t t + \sum Z_t t^3\right)$$

Hence, if the sample available is $Z_1,\ldots,Z_{t-1},Z_t,Z_{t+1},\ldots,Z_T$, then

$$
\begin{aligned}
\hat{S}^{(3)}_{T-3} &= \beta_0 \\
&= -\tfrac{4}{42}Z_{T-6} + \tfrac{6}{42}Z_{T-5} + \tfrac{12}{42}Z_{T-4} + \tfrac{14}{42}Z_{T-3} + \tfrac{12}{42}Z_{T-2} + \tfrac{6}{42}Z_{T-1} - \tfrac{4}{42}Z_T
\end{aligned}
$$

$$
\begin{aligned}
\hat{S}^{(3)}_{T-2} &= \beta_0 + \beta_1 + \beta_2 + \beta_3 \\
&= \tfrac{1}{42}Z_{T-6} - \tfrac{4}{42}Z_{T-5} + \tfrac{2}{42}Z_{T-4} + \tfrac{12}{42}Z_{T-3} + \tfrac{19}{42}Z_{T-2} + \tfrac{16}{42}Z_{T-1} - \tfrac{4}{42}Z_T
\end{aligned}
$$

$$
\begin{aligned}
\hat{S}^{(3)}_{T-1} &= \beta_0 + 2\beta_1 + 4\beta_2 + 8\beta_3 \\
&= \tfrac{4}{42}Z_{T-6} - \tfrac{7}{42}Z_{T-5} - \tfrac{4}{42}Z_{T-4} + \tfrac{6}{42}Z_{T-3} + \tfrac{16}{42}Z_{T-2} + \tfrac{19}{42}Z_{T-1} + \tfrac{8}{42}Z_T
\end{aligned}
$$

$$
\begin{aligned}
\hat{S}^{(3)}_{T} &= \beta_0 + 3\beta_1 + 9\beta_2 + 27\beta_3 \\
&= -\tfrac{2}{42}Z_{T-6} + \tfrac{4}{42}Z_{T-5} + \tfrac{1}{42}Z_{T-4} - \tfrac{4}{42}Z_{T-3} - \tfrac{4}{42}Z_{T-2} + \tfrac{8}{42}Z_{T-1} + \tfrac{39}{42}Z_T
\end{aligned}
$$

Note that the weights continue to sum to unity but they become more and more unequal as we move closer to the end of the sample. Obviously, a similar procedure can be used for other values of *p* and *r*, but the algebra becomes lengthy. Weights for $p \le 5$ and $r \le 10$ are given as Appendix A of Kendall (1976). Asymmetric extensions to Musgrave filters are provided in Quenneville, Ladiray, and Lefrançois, (2003); see also Gray and Thomson (2002).

To deal with the finite sample problem for the HP filter, note that (43) can be written as

$$Z_t = \xi S_{t-2} - 4\xi S_{t-1} + (1+6\xi)S_t - 4\xi S_{t+1} + \xi S_{t+2}$$

so that

$$N_t = \xi\left(S_{t-2} - 4S_{t-1} + 6S_t - 4S_{t+1} + S_{t+2}\right)$$

This expression cannot be used when $t = 1$, 2, $T-1$ or *T*, when it is modified to

$$N_1 = \xi\left(S_1 - 2S_2 + S_3\right)$$

$$N_2 = \xi\left(-2S_1 + 5S_2 - 4S_3 + S_4\right)$$

$$N_{T-1} = \xi\left(S_{T-3} - 4S_{T-2} + 5S_{T-1} - 2S_T\right)$$

$$N_T = \xi\left(S_{T-2} - 2S_{T-1} + S_T\right)$$

respectively. Define $\mathbf{Z} = (Z_1,\ldots,Z_T)^{\mathsf{T}}$, $\mathbf{S} = (S_1,\ldots,S_T)^{\mathsf{T}}$, $\mathbf{N} = \mathbf{Z} - \mathbf{S}$ and

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 0 & & & \cdots & & & 0 \\ -2 & 5 & -4 & 1 & 0 & 0 & 0 & & & \cdots & & & 0 \\ 1 & -4 & 6 & -4 & 1 & 0 & 0 & & & \cdots & & & 0 \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & & & \cdots & & & 0 \\ \vdots & & & & & & \vdots & & & & & & \vdots \\ 0 & & \cdots & & & & 0 & 1 & -4 & 6 & -4 & 1 & 0 \\ 0 & & \cdots & & & & 0 & 0 & 1 & -4 & 6 & -4 & 1 \\ 0 & & \cdots & & & & 0 & 0 & 0 & 1 & -4 & 5 & -2 \\ 0 & & & & & & 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}$$

The ‘system’ can then be written as $\mathbf{N} = \xi\mathbf{\Gamma}\mathbf{S} = \left(\mathbf{I} + \xi\mathbf{\Gamma}\right)^{-1}\xi\mathbf{\Gamma}\mathbf{Z}$, from which the signal can be recovered as $\mathbf{S} = \mathbf{Z} - \mathbf{N} = \left(\mathbf{I} - \left(\mathbf{I} + \xi\mathbf{\Gamma}\right)^{-1}\xi\mathbf{\Gamma}\right)\mathbf{Z}$. This system has been used to generate the HP estimates at the ends of the sample shown in figure 8, but it should be realised that such end of sample estimates are much less reliable than those obtained in the central block of observations.

The HP and Butterworth filters can both be written in the structural form

$$Z_t = S_t + N_t = \frac{(1+B)^m}{(1-B)^d} v_t + (1-B)^n u_t$$

The HP filter sets $d = 2$ and $m = n = 0$, while the Butterworth filter has $n = m - d$. The procedure proposed by Pollock (2000, 2001, 2006) deals with a finite sample in the following way. The operators $(1+B)^k$ and $(1-B)^k$ can be expressed as

$$(1+B)^k = 1 + \delta_1 B + \ldots + \delta_k B^k \qquad \delta_i = k!/i!(k-i)! \qquad i = 1,\ldots,k$$

and

$$(1-B)^k = 1 + \gamma_1 B + \ldots + \gamma_k B^k \qquad \gamma_i = (-1)^i \delta_i \qquad i = 1,\ldots,k$$

so that the following transformation matrices can be defined

$$\mathbf{R}_k = \begin{bmatrix} \delta_k & \ldots & \delta_1 & 1 & \ldots & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots \\ 0 & \ldots & \delta_k & \delta_{k-1} & \ldots & 1 & 0 & \ldots & 0 & 0 \\ 0 & \ldots & 0 & \delta_k & \ldots & \delta_1 & 1 & \ldots & 0 & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & \delta_k & \delta_{k-1} & \ldots & 1 & 0 \\ 0 & \ldots & 0 & 0 & \ldots & 0 & \delta_k & \ldots & \delta_1 & 1 \end{bmatrix}$$

and

$$\mathbf{Q}_k = \begin{bmatrix} \gamma_k & \dots & \gamma_1 & 1 & \dots & 0 & 0 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots \\ 0 & \dots & \gamma_k & \gamma_{k-1} & \dots & 1 & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & \gamma_k & \dots & \gamma_1 & 1 & \dots & 0 & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & \gamma_k & \gamma_{k-1} & \dots & 1 & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 & \gamma_k & \dots & \gamma_1 & 1 \end{bmatrix}$$

Defining $\mathbf{u} = (u_1, \cdots, u_T)^\top$ then, since $N_t = (1-B)^n u_t$, $\mathbf{N} = \mathbf{Q}_n \mathbf{u}$, with $\mathsf{E}(\mathbf{N}) = \mathbf{0}$ and

$$\mathsf{E}(\mathbf{N}\mathbf{N}^\top) = \mathbf{Q}_n (\mathbf{u}\mathbf{u}^\top) \mathbf{Q}_n^\top = \sigma_u^2 \mathbf{Q}_n \mathbf{Q}_n^\top = \sigma_u^2 \mathbf{\Sigma}_n$$

where $\mathbf{\Sigma}_n = \mathbf{Q}_n \mathbf{Q}_n^\top$. The decomposition $\mathbf{Z} = \mathbf{N} + \mathbf{S}$ can then be written as

$$\begin{aligned} \mathbf{Q}_d \mathbf{Z} &= \mathbf{Q}_d \mathbf{S} + \mathbf{Q}_d \mathbf{N} \\ &= \mathbf{R}_m \mathbf{v} + \mathbf{Q}_{n-d} \mathbf{u} = \mathbf{s} + \mathbf{n} = \mathbf{z} \end{aligned}$$

where $\mathbf{v} = (v_1, \cdots, v_T)^\top$ and

$$\mathbf{s} = \mathbf{R}_m \mathbf{v} = \mathbf{Q}_d \mathbf{S} = (s_{d+1}, \cdots, s_T)$$

and similarly for **n** and **z**, so that the first *d* observations are lost through differencing. Thus

$$\mathsf{E}(\mathbf{s}) = \mathbf{0} \qquad \mathsf{E}(\mathbf{s}\mathbf{s}^\top) = \mathbf{R}_m \, \mathsf{E}(\mathbf{v}\mathbf{v}^\top) \mathbf{R}_m^\top = \sigma_v^2 \mathbf{R}_m \mathbf{R}_m^\top = \sigma_v^2 \mathbf{\Omega}_\mathbf{R}$$

$$\mathsf{E}(\mathbf{n}) = \mathbf{0} \qquad \mathsf{E}(\mathbf{n}\mathbf{n}^\top) = \mathbf{Q}_d \, \mathsf{E}(\mathbf{N}\mathbf{N}^\top) \mathbf{Q}_d^\top = \sigma_u^2 \mathbf{Q}_d \Sigma_n \mathbf{Q}_d^\top = \sigma_u^2 \mathbf{\Omega}_\mathbf{Q}$$

and

$$\mathsf{E}(\mathbf{z}) = \mathbf{0} \qquad \mathsf{E}(\mathbf{z}\mathbf{z}^\top) = \mathsf{E}(\mathbf{s}+\mathbf{n})(\mathbf{s}+\mathbf{n})^\top = \mathsf{E}(\mathbf{s}\mathbf{s}^\top) + \mathsf{E}(\mathbf{n}\mathbf{n}^\top) = \sigma_v^2 \mathbf{\Omega}_\mathbf{R} + \sigma_u^2 \mathbf{\Omega}_\mathbf{Q}$$

The optimal estimate of **s** given **n** is then given by the conditional expectation

$$\hat{\mathbf{s}} = \mathsf{E}(\mathbf{s}|\mathbf{z}) = \mathsf{E}(\mathbf{s}) + \frac{\mathsf{E}(\mathbf{s}\mathbf{z}^\top)}{\mathsf{E}(\mathbf{z}\mathbf{z}^\top)} (\mathbf{z} - \mathsf{E}(\mathbf{z}))$$

Using $\mathsf{E}\left(\mathbf{s}\mathbf{z}^{\mathsf{T}}\right)=\mathsf{E}\left(\mathbf{s}(\mathbf{s}+\mathbf{n})^{\mathsf{T}}\right)=\mathsf{E}\left(\mathbf{s}\mathbf{s}^{\mathsf{T}}\right)=\sigma_v^2\mathbf{\Omega_R}$, since $\mathsf{E}\left(\mathbf{s}\mathbf{n}^{\mathsf{T}}\right)=\mathbf{0}$,

$$\hat{\mathbf{s}}=\frac{\sigma_v^2\mathbf{\Omega_R}}{\sigma_v^2\mathbf{\Omega_R}+\sigma_u^2\mathbf{\Omega_Q}}\mathbf{z}=\frac{\mathbf{\Omega_R}}{\mathbf{\Omega_R}+\xi\mathbf{\Omega_Q}}\mathbf{z}$$

Similarly,

$$\hat{\mathbf{n}}=\mathbf{z}-\hat{\mathbf{s}}=\frac{\xi\mathbf{\Omega_Q}}{\mathbf{\Omega_R}+\lambda\mathbf{\Omega_Q}}\mathbf{z}$$

Computationally, these estimates are obtained by solving the equation

$$\left(\mathbf{\Omega_R}+\xi\mathbf{\Omega_Q}\right)\mathbf{b}=\mathbf{z}$$

for the value of **b** and then computing

$$\hat{\mathbf{s}}=\mathbf{\Omega_R}\mathbf{b} \qquad \hat{\mathbf{n}}=\xi\mathbf{\Omega_Q}\mathbf{b} \tag{51}$$

However, the primary aim is to obtain the trend estimate $\hat{\mathbf{S}}$. Using $\mathbf{Q}_d^{-1}\hat{\mathbf{s}}$ will only provide estimates for the last $d+1$ observations, though. The problem of finding the d initial conditions can be circumvented by noting that the trend estimation problem is equivalent to minimising $\left(\mathbf{Z}-\hat{\mathbf{S}}\right)^{\mathsf{T}}\Sigma^{-1}\left(\mathbf{Z}-\hat{\mathbf{S}}\right)$ subject to $\mathbf{Q}_d\hat{\mathbf{S}}=\hat{\mathbf{s}}$. This requires evaluating the Lagrangean function

$$\mathsf{L}\left(\hat{\mathbf{S}},\alpha\right)=\left(\mathbf{Z}-\hat{\mathbf{S}}\right)^{\mathsf{T}}\Sigma^{-1}\left(\mathbf{Z}-\hat{\mathbf{S}}\right)+2\mathbf{a}\left(\mathbf{Q}_d\hat{\mathbf{S}}-\hat{\mathbf{s}}\right)$$

where **a** is the vector of Lagrange multipliers, for which a first order condition for a minimum is

$$\mathbf{\Sigma}^{-1}\left(\mathbf{Z}-\hat{\mathbf{S}}\right)-\mathbf{Q}_d^{\mathsf{T}}\mathbf{a}^{\mathsf{T}}=\mathbf{0}$$

Premultiplying by $\mathbf{Q}_d\mathbf{\Sigma}$ yields

$$\mathbf{Q}_d\left(\mathbf{Z}-\hat{\mathbf{S}}\right)=\mathbf{Q}_d\Sigma\mathbf{Q}_d^{\mathsf{T}}\mathbf{a}^{\mathsf{T}} \tag{52}$$

But, from (49)

$$\mathbf{Q}_d\left(\mathbf{Z}-\hat{\mathbf{S}}\right)=\mathbf{z}-\hat{\mathbf{s}}=\hat{\mathbf{n}}=\xi\mathbf{\Omega_Q}\mathbf{b}=\xi\mathbf{Q}_d\Sigma\mathbf{Q}_d^{\mathsf{T}}\mathbf{b}$$

Hence

$$\mathbf{a}^{\mathsf{T}}=\left(\mathbf{Q}_d\Sigma\mathbf{Q}_d^{\mathsf{T}}\right)^{-1}\mathbf{Q}_d\left(\mathbf{Z}-\hat{\mathbf{S}}\right)=\xi\mathbf{b}$$

Substituting into (52) thus obtains

$$\hat{\mathbf{S}} = \mathbf{Z} - \xi\boldsymbol{\Sigma}\mathbf{Q}_d^{\mathsf{T}}\mathbf{b}$$

so that

$$\hat{\mathbf{N}} = \xi\boldsymbol{\Sigma}\mathbf{Q}_d^{\mathsf{T}}\mathbf{b}$$

and hence both the signal and noise components can be estimated using matrix operations.

7.3. Structural, Nonlinear and Nonparametric Trends

These techniques do not, on the face of it, run into problems of missing values at the ends of a finite sample. Indeed, the restriction to a finite sample of size T has been explicitly made in the development of structural models, which employ the recursive estimation techniques of the Kalman filter, and for the nonlinear and nonparametric trend specifications, which are also explicitly designed for finite samples. However, modelling a smooth transition when the regime shifts near to the beginning or end of the sample may be computationally difficult. With nonparametric trends, it has been found that odd-order local polynomials have clear advantages of even-order ones for trend-fitting in these 'boundary' regions. See Wand and Jones (1995) and Simonoff (1996) for details on this. Links between non-parametric trends and UC models are developed in Harvey and Koopman (2000), who argue for a clear preference for the latter.

8. Conclusion

This chapter has considered a variety of techniques for modelling and extracting components of nonstationary time series. Rather than listing these and reiterating their properties, table 1 reports the forecasts of the logarithms of U.S. GNP out to 2012.4, along with forecast standard errors, for the three techniques that allow such quantities to be readily computed. The BN forecasts (essentially the forecasts from a drifting ARIMA$(0,1,1)$ process) produce forecasts of GNP that, by end-2012, are 2.4% larger than those from the structural model and 0.4% larger than those from the segmented trend. On the other hand, the segmented trend forecasts are accompanied by the smallest standard errors, essentially because these are forecasts from a trend stationary process for which uncertainty is bounded. Although the structural trend forecasts are accompanied by the largest standard errors, it may be argued that these best reflect the natural variability of the data and hence the uncertainty likely to be found in forecasts of GNP. In any event, what this table shows, along with the earlier figures, is that alternative models will produce different component decompositions and forecasts with different accompanying measures of imprecision. It is therefore important to understand the properties of the decomposition technique that is chosen for analysing an economic time series.

Table 1. Forecasts of the logarithms of U.S. GNP from three decomposition techniques with accompanying forecast standard errors

	BN	BN std err	Structural trend	Structural trend Std err	Segmented trend	Segmented trend std err
2008.1	9.3826	0.0094	9.3833	0.0093	9.3884	0.0168
2008.4	9.4066	0.0224	9.4022	0.0267	9.4108	0.0171
2009.4	9.4386	0.0325	9.4284	0.0405	9.4407	0.0176
2010.4	9.4706	0.0401	9.4559	0.0500	9.4705	0.0181
2011.4	9.5026	0.0465	9.4834	0.0612	9.5004	0.0186
2012.4	9.5356	0.0521	9.5107	0.0746	9.5303	0.0190

References

Bacon, D.W. and Watts, D.G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika,* **58**, 525-534.

Baxter, M. and King, R.G. (1999). 'Measuring business cycles: Approximate bandpass filters for economic time series'. *Review of Economics and Statistics*, **81**, 575-593.

Bell, W.R. (1984). Signal extraction for nonstationary time series. *Annals of Statistics,* **13**, 1286-1316.

Bell, W.R. and Martin, D.E.K. (2002). Computation of asymmetric signal extraction filters and mean squared error for ARIMA component models. *Research Report* 2002-04, Statistical Research Division, Bureau of the Census, Washington DC.

Bell, W.R. and Martin, D.E.K. (2004). Computation of asymmetric signal extraction filters and mean squared error for ARIMA component models. *Journal of Time Series Analysis,* **25**, 603-625.

Beveridge, S. and Nelson, C.R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary Economics,* **7**, 151-174.

Beveridge, W.H. (1920). British exports and the barometer'. *Economic Journal*, **30**, 13-25.

Beveridge, W.H. (1921). Weather and harvest cycles. *Economic Journal*, **31**, 429-452.

Box, G.E.P. and Jenkins, G.M. (1962). Some statistical aspects of adaptive optimization and control. *Journal of the Royal Statistical Society, Series B,* **24**, 297-331.

Burns, A.F. and Mitchell, W.C. (1946). *Measuring Business Cycles*. New York: National Bureau of Economic Research.

Burridge, P. and Wallis, K.F. (1988). Prediction theory for auto-regressive moving average processes. *Econometric Reviews,* **7**, 65-95.

Christiano, L.J. and Fitzgerald, T.J. (2003). The band pass filter. *International Economic Review,* **44**, 435-465.

Cox, D.R. (1961), Prediction by exponentially weighted moving averages and related methods. *Journal of the Royal Statistical Society, Series B*, **23**, 414-422.

Frickey, E. (1934). The problem of secular trend. *Review of Economic Statistics,* **16**, 199-206.

Frisch, R. (1933). Propagation problems and impulse problems in dynamic economics, in *Economic Essays in Honour of Gustav Cassel*, 171-205, London: George Allen & Unwin.

Frisch, R. (1939). A note on errors in time series, *Quarterly Journal of Economics,* **53**, 639-640.

Gray, A.G. and Thomson, P. (1996). Design of moving-average trend filters using fidelity, smoothness and minimum revisions criteria. *U.S. Bureau of the Census, Statistical Research Division*, Statistical Research Report No. RR96/01.

Gray, A.G. and Thomson, P. (2002). On a family of finite moving-average trend filters for the ends of series. *Journal of Forecasting,* **21**, 125-149

Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A.C. and Shephard, N. (1992). Structural time series models. In *Handbook of Statistics, Volume XI: Econometrics* (eds. G.S. Maddala, C.R. Rao and H.D. Vinod). Amsterdam: North-Holland.

Harvey, A.C. and Trimbur, T.M. (2003). General model-based filters for extracting cycles and trends in economic time series. *Review of Economics and Statistics,* **85**, 244-255.

Harvey, A.C. and Trimbur, T.M. (2008). Trend estimation and the Hodrick-Prescott filter. *Journal of the Japanese Statistical Society,* **38**, 1-9.

Henderson, R. (1924). A new method of graduation. *Transactions of the Actuarial Society of America* **25**, 29-40.

Hodrick, R.J. and Prescott, E.J. (1997). Postwar US business cycles: an empirical investigation. *Journal of Money, Credit and Banking,* **29**, 1-16.

Hooker, R.H. (1901). Correlation of the marriage rate with trade. *Journal of the Royal Statistical Society*, **64**, 485-503.

Kalman, R.E. (1960). A new approach to linear filtering and prediction theory. *Journal of Basic Engineering, Transactions of the ASME, Series D*, **82**, 35-45.

Kalman, R.E. and Bucy, R.E. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering, Transactions of the ASME, Series D*, **83**, 95-108.

Kendall, M.G. (1976). *Time Series, 2nd Edition*. London: Charles Griffin.

Kenny, P.B. and Durbin, J. (1982). Local trend estimation and seasonal adjustment of economic and social time series (with discussion). *Journal of the Royal Statistical Society, Series A,* **145**, 1-41.

Kitchen, J. (1923). Cycles and trends in economic factors. *Review of Economic Statistics,* **5**, 10-16.

Kim, C.-J. and Nelson, C.R. (1999). Has the U.S. economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle. *Review of Economics and Statistics.* **81**, 608-616.

Koopman, S.J. and Harvey, A.C. (2003). Computing observation weights for signal extraction and filtering. *Journal of Economic Dynamics and Control,* **27**, 1317-1333.

Koopman, S.J., Harvey, A.C., Doornik, J.A. and Shephard, N. (2006). *STAMP* **7***: Structural Time Series Analysis and Predictor*. London: Timberlake Consultants Ltd.

Koopman, S.J., Shephard, N. and Doornik, J.A. (1999). Statistical algorithms for models in state space form using SsfPack 2.2. *Econometrics Journal,* **2**, 113-166.

Koopmans, T.C. (1947). Measurement without theory. *Review of Economic Statistics,* **29**, 161-172.

Kuznets, S. (1929). Random events and cyclical oscillations. *Journal of the American Statistical Association,* **24**, 258-275.

Leser, C.E.V. (1961). ‘A simple method of trend construction’. *Journal of the Royal Statistical Society, Series B*, **23**, 91-107.

Lippi, M. and Reichlin, L. (1992). On persistence of shocks to economic variables. A common misconception. *Journal of Monetary Economics,* **29**, 87-93.

Macaulay, F. (1931). *The Smoothing of Time Series*. New York: National Bureau of Economic Research.

McConnell, M.M. and Quiros, G.P. (2000). Output fluctuations in the United States: What has changed since the early 1980s. *American Economic Review.* **90**, 1464-1476.

Mills, T.C. (2003). *Modelling Trends and Cycles in Economic Time Series*. Basingstoke: Palgrave Macmillan.

Mills, T.C. (2007). A note on trend decomposition: The ‘classical’ approach revisited with an application to surface temperature trends. *Journal of Applied Statistics,* **34**, 963-972.

Mills, T.C. and Markellos, R.N. (2008). *The Econometric Modelling of Financial Time Series,* **3***rd* *edition*. Cambridge: Cambridge University Press.

Mitchell, W.C. (1913). *Business Cycles and their Causes*. Berkeley: California University Memoirs, Vol. III.

Mitchell, W.C. (1927). *Business Cycles: The Problem and its Setting*. New York: National Bureau of Economic Research.

Morgan, M.S. (1990). *The History of Econometric Ideas*. Cambridge, Cambridge University Press.

Muth, J.S. (1960). ‘Optimal properties of exponentially weighted forecasts’. *Journal of the American Statistical Association*, **55**, 299-306.

Newbold, P. (1990). Precise and efficient computation of the Beveridge-Nelson decomposition of economic time series. *Journal of Monetary Economics,* **26**, 453-457.

Pedersen, T.M. (2001). The Hodrick-Prescott filter, the Slutzky effect, and the distortionary effect of filters. *Journal of Economic Dynamics and Control,* **25**, 1081-1101.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica,* **57**, 1361-1401.

Pierce, D.A. (1979). Signal extraction error in nonstationary time series. *Annals of Statistics,* **7**, 1303-1320.

Pollock, D.S.G. (1999). *A Handbook of Time Series Analysis, Signal Processing and Dynamics*. San Diego: Academic Press.

Pollock, D.S.G. (2000). Trend estimation and de-trending via rational square wave filters. *Journal of Econometrics,* **99**, 317-334.

Pollock, D.S.G. (2001). ‘Filters for short non-stationary sequences’. *Journal of Forecasting,* **20**, 341-355.

Pollock, D.S.G. (2006). ‘Econometric methods of signal extraction’. *Computational Statistics and Data Analysis*, **50**, 2268-2292.

Proietti, T. (2005). Forecasting and signal extraction with misspecified models. *Journal of Forecasting*, **24**, 539-556.

Proietti, T. and Harvey, A.C. (2000). A Beveridge-Nelson smoother. *Economics Letters,* **67**, 139-146.

Quenneville, B., Ladiray, D. and Lefrançois, B. (2003). ‘A note on Musgrave asymmetrical trend-cycle filters’, *International Journal of Forecasting,* **19**, 727-734.

Simonoff, J.S. (1996). *Smoothing Methods in Statistics*. New York: Springer-Verlag.

Slutsky, E.E. (1927). The summation of random causes as the source of cyclic processes, in *The Problems of Economic Conditions*, edited by the Conjuncture Institute, Moscow, 3:1, 34-64 (English summary, 156-161).

Slutsky, E.E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica*, **5**, 105-146.

Tiao, G.C. and Hillmer, S.C. (1978). Some consideration of decomposition of a time series. *Biometrika,* **65**, 497-502.

Tinbergen, J. (1937). *An Econometric Approach to Business Cycle Problems*. Paris: Hermann & Cie.

Trimbur, T.M. (2006). Properties of higher order stochastic cycles. *Journal of Time Series Analysis,* **27**, 1-17.

Vining, R. (1949). Methodological issues in quantitative economics. *Review of Economics and Statistics,* **31**, 77-94.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. London: Chapman and Hall

Whittaker, E.T. (1923). On a new method of graduations. *Proceedings of the Edinburgh Mathematical Society,* **41**, 63-75.

Whittle, P. (1984). *Prediction and Regulation by Linear Least Squares Methods*. 2nd Edition, Revised. Oxford: Blackwell.

Winters, P.R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science,* **6**, 324-342.

Yule, G.U. (1926). Why do we sometimes get nonsense correlations between time-series? - A study in sampling and the nature of time series. *Journal of the Royal Statistical Society*, **89**, 1-64.

Yule, G.U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London, Series A*, **226**, 267-298.

In: Economic Forecasting
Editor: Alan T. Molnar, pp. 107-133
ISBN: 978-1-60741-068-3

Chapter 4

A COST OF CAPITAL ANALYSIS OF THE GAINS FROM SECURITIZATION

***Hugh Thomas*[1] *and Zhiqiang Wang*[2]**
[1] The Chinese University of Hong Kong
Shatin, NT, Hong Kong SAR, China
[2] Dongbei University of Finance and Economics
Dalian, Liaoning, China

Abstract

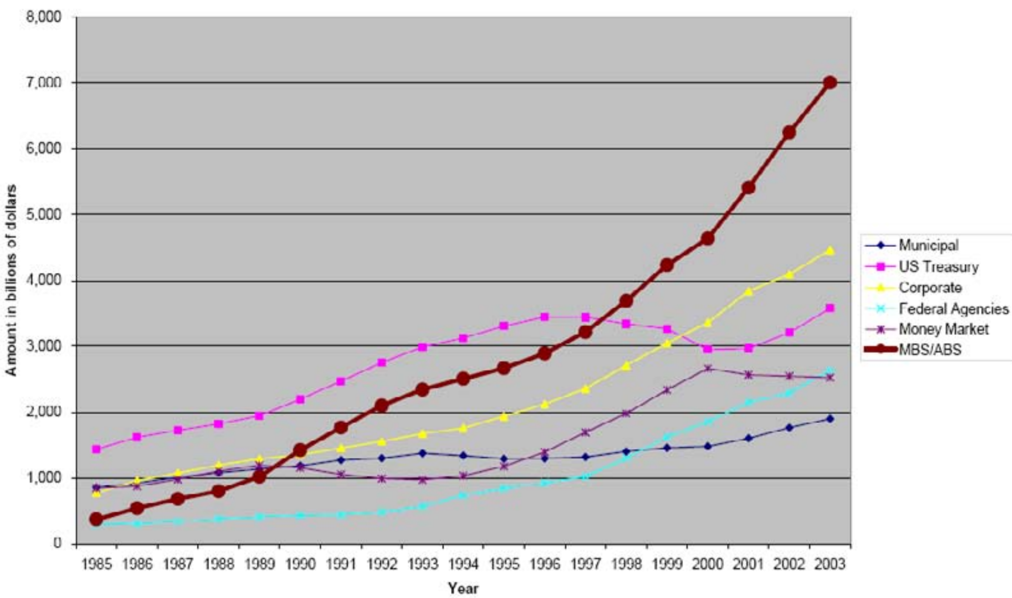
In this paper we study the gains that securitizing companies enjoy. We expressed the gains as a spread between two costs of capital, the weighted cost of capital of the asset selling firm and the all-in, weighted average cost of the securitization. We calculate the spread for 1,713 securitizations and regress those gains on asset seller characteristics. We show that they are increasing in size in the amount of asset backed securities originated but are decreasing in the percent of the balance sheet that the originator has outstanding. Companies that off-lay the risk of the sold assets (i.e., those retaining no subordinate interest in the SPV) pick their best assets to securitize. Companies that do not off-lay risk gain more from securitization the less liquid they are. We find that securitization is a substitute for leverage and that those companies that use more conventional leverage benefit less from securitization.

Introduction

Securitization, the sale of fixed income assets into bankruptcy-remote, special purpose vehicles (SPVs) funded by the issuance of new asset backed securities (ABS)[1], has achieved

[1] We use the term asset backed securities (ABS) to include all debt securities issued by bankruptcy remote special purpose entities backed primarily by fixed income assets. The term includes both pass-through securities including traditional mortgage backed securities (MBS) as well as pay-through securities. The SPV is a shell with no assets but the fixed income financial assets. It can be a limited liability company, a limited partnership or a trust account We use the term "fixed income" in a loose sense, encompassing not only traditional debt securities, but also near fixed income cash flows such as toll road revenues, lease payments. These flows must be sufficiently predictable to allow the largely debt-financed SPV's debt to attract sufficiently high debt ratings to allow the transaction to proceed economically.

tremendous popularity. From 1985 through 2003, ABS issues enjoyed an 18 percent per annum compound growth rate and rose to become the dominant fixed income security in the world. Today global outstanding ABS issues total approximately US$9 trillion. In the US, which accounts for about $7 trillion, half of all ABS are agency-based mortgaged backed securities. This article concerns the other, faster growing half, including securitizations of corporate receivables, credit card debt, auto loans and leases, home equity loans, non-agency guaranteed mortgages, etc., is approximately the same size as the treasury markets and exceeds the size of the corporate bond markets.



Source: U.S. Federal Research Board.

Figure 1. US Debt by Category. (in billions of dollars).

Securitization, is a fundamental device changing the relationships between banks (and other financial institutions), non-financial corporations and financial markets. In this paper, we investigate how securitization adds value to the asset sellers through a direct measurement of cost of capital. We do this to lend light to the banking literature that has several divergent views of securitization. In their banking text, Greenbuam and Thakor (1995) observe that securitization permits banks to unbundle the traditional lending function, allowing them to specialize in the more basic activities in which they enjoy comparative advantages. In their signaling model Greenbaum and Thakor (1987) show that grains are available to borrowers with private knowledge of their individual credit quality who can signal that quality by purchasing credit enhancements from banks in securitization. Pennacchi (1988) models a securitization like contract of a bank that sells loans but, to alleviate the moral hazard of selling fully the loans, retain the risk of the loan in a credit enhancement, or "equity position" in the sold loan. Securitization alters the capital structure of the originator. Perry (1995) states that for banks in general and for thinly capitalized banks in particular being both originator and credit enhancer increases the risk of the bank. Securitizations circumvent the bankruptcy process, possibly to the detriment of unsecured creditors [see Lupica (2000)]. Leland and

Skarabot (2003) provide a simulation model for how securitization can improve an originator's capital structure as long as the volatility of securitized assets is substantially different from the volatility of the assets that remain on the balance sheet.

Within this literature, there are several concerns voiced about securitization. If it is used as a crutch for the weak banks, or to appropriate wealth from debt-holders, or for capital arbitrage, or to deceive equity holders about the true risk of a corporation, then it should be curtailed if not abolished. Bank regulators are especially concerned with the credit, liquidity and operational risk retention and risk off-lay effects of bank securitizations on the originating bank. The disclosures prescribed by FAS 140 [Financial Accounting Standards Board (2000)], determine whether or not a securitized asset should be shown on-balance sheet by applying the test of whether effective control is retained. This attempt to show the true assets of the firm takes on even more importance in the post-Enron environment, because Enron management used securitization to hide the risks it purported to sell but in fact retained. See Burroughs et al. (2002) and Skipworth (2001).

Our tests in this paper follow in the tradition of Lockwood et al. (1996) and Thomas (2001), who measure gains to securitization to the asset originator using event studies. The former finds increases to shareholder wealth in well-capitalized banks, but finds no effect in thinly capitalized banks while the latter finds positive to securitizations in years when banks were not under pressure to increase capital levels. A recent paper by Ambrose, LaCour-Little and Sanders (2004) asks the question "does regulatory capital arbitrage, reputation or asymmetric information drive securitization?" and comes to the conclusion that capital arbitrage or reputation is driving it.

Investment bankers promoting securitization, however, answer the question of what drives securitization without hesitation. Securitization has grown rapidly because it provides companies with cheap financing.[2] In this paper, we estimate the gains in corporate wealth from this cheap financing by estimating the percentage differential costs of capital. We analyze these gains and show that they are increasing in size in the amount of ABS originated but are decreasing in the percent of the balance sheet that the originator has outstanding. Companies that off-lay the risk of the sold assets (i.e., those retaining no subordinate interest in the SPV) pick their best assets to securitize. Companies that do not off-lay risk (i.e., those who retain the concentrated first loss tranches that make up substantially all of the asset risk) gain more from securitization the less liquid they are. Securitization is a substitute for leverage; hence, those companies that use more conventional leverage benefit less from securitization.

The gains to securitization can be expressed as a spread between two costs of capital, the weighted cost of capital (WACC) of the asset selling firm and the all-in, weighted average cost of the securitization k_s.

$$Spread = WACC - k_s \tag{1}$$

---

[2] For example, Raynes and Rutledge (2003) who decry the inefficiencies of the ABS market, explained to the author the persistence of these efficiencies by their observation that corporate treasures make so much money for their firms by securitization that they are not concerned by the money left on the table by inefficient structures of the SPVs.

This paper has a very limited purpose. We calculate the spread to demonstrate its importance, run a series of linear regressions to determine the factors that influence the spread, and discuss the implications of our simple tests on further studies of securitization.

Calculating Costs of Capital

We use costs of capital to estimate gains to securitization because corporate treasurers securitize to lower their costs of capital[3]. In most securitizations with single asset sellers, the value left in the SPV after payment of all fees and principal and interest claims of the ABS purchasers reverts back to the asset seller. This entitlement to residual cash-flows is analogous to the residual rights of the equity holder if the assets remain under the ownership of the asset seller; hence, securitization can be viewed as off-balance-sheet asset financing.

We calculate the weighted average cost of capital of the asset seller as follows

$$WACC = \frac{LTD}{A} \bullet k_{LTD} + \frac{STD}{A} \bullet k_{STD} + \frac{E}{A} \bullet k_E \tag{2}$$

Where $WACC$ is the weighted average cost of capital of the asset seller; LTD is the value of the firm's long-term debts; STD is the value of the firm's short-term financial debts (i.e., excluding trade and other payables); E is the value of the firm's equity and $A = LTD + STD + E$. The terms k_{LTD}, k_{STD} and k_E denote the costs of long-term debt, short-term debt and equity respectively.

SPVs issue several classes of debt securities: A-Class (typically structured to be awarded a triple A bond rating), M-Class (high rated subordinate debt), B-Class (lower rated subordinate) and R-Class (unrated reserve trances)[4]. Each securitization is structured to maximize the amount of funding in the senior, rated class, whose yields are the lowest. The SPV is like a very simple, single purpose financial institution where management has extremely limited discretion in managing the fixed income assets. Although SPVs tend to be far more highly levered than regular financial institutions, SPVs do have an equity component. This equity component is credit support, made up of two components, a funded, un-rated reserve class, often retained by the asset seller and an "un-funded" portion. The

---

[3] Some securitizations are justified solely on the basis of increasing liquidity at the cost of increasing the cost of capital, reducing concentrations in certain risks to rebalance the credit risk of the portfolio, freeing up regulatory capital, meeting regulatory concentration limits, or improving the accounting appearance of the company, regardless of costs of capital considerations. Others are tax-motivated. But the vast majority of securitizations are justified on the basis of cost.

[4] Our terms are consistent with those of ABSNet (www.absnet.net). ABSNet is a partnership between Lewtan Technolgies and Standard and Poor's, that provides the most comprehensive database of ABS deal performance data currently available. It was set up to meet the needs of institutional investors who wish to have up-to-date information on the quality of their investments. The website gives the current performance data and structure of outstanding ABS issues.

[4] GMAC is a wholly owned subsidiary of General Motors set up to finance dealer inventories of automobiles. It has since expanded into consumer auto finance, insurance and mortgage financing to become one of the leading finance companies in the world with assets of $250 billion. Reflecting its lack bank deposit bases and relatively poor credit rating, it fund itself partly in the ABS markets with securitizations reported as $98 billion with $5 billion in continuing interests in securitizations (i.e., first loss tranches). See GMAC's 2003 10K

unfunded portion usually includes over-collateralization, injected either when the SPV was formed or built up from the excess spread between the realized yield of the assets and the funds paid out in fees and payments to the ABS holders and can also include partial guarantees.

We compute the cost of securitization as follows:

$$k_S = \sum_{i=1}^{n} \frac{D_i}{\sum_{i=1}^{n} D_i + E_R} \bullet k_i + \left[\frac{E_R}{\sum_{i=1}^{n} D_i + E_R} \bullet k_E + \frac{E_C}{\sum_{i=1}^{n} D_i + E_R} \bullet (k_E - r_f) \right] + fees \quad (2)$$

Where k_S is the SPV's all-in cost of capital, k_i is the interest rate of the *i*-th class and k_E is the cost of equity of the asset seller as calculated above; D_i is the dollar amount of the *i*-th class of the securitization, E_R and E_C denote the amount in dollars of reserve and over-collateralization, respectively; r_f is risk-free rate and *fees* denotes the out of pocket front-end and annual fees to run the securitization expressed as an annuity in terms of percentage of the securitization.

In this equation, the first term calculates the weighted cost of debt; the second term (in square brackets) calculates the weighted cost of equity of the SPV and the third term adds the floatation costs. We use the term equity somewhat loosely. E_R is legally subordinate debt, constituting the bottom of the waterfall of subordination. E_C is the portion of the credit support, made up of partial guarantees, paid in surplus and/or accumulated surpluses. Since it is not included in the ABS claims on the SPV (i.e., in the denominator $\left(\sum_{i=1}^{n} D_i + E_R \right)$), we cost it as an unfunded in cost of capital equation by subtracting off the risk-free rate. If both E_R and E_C are owned by the asset seller who also is the residual claimant, then, clearly the asset seller retains the concentrated risk of the securitization.

We assume that the cost of equity is the same for the SPV and the seller. This is likely to be an overestimation of the cost of equity of the SPV for the following reason. The costs of debt k_i of the ABS are typically less than the costs of debt of the asset seller; therefore, the equity cushion made up of E_R and E_C provides no less of a cushion to debt holders than the equity of the asset seller provides to the debt-holders of the asset seller. Therefore, the equity of the SPV should be no more expensive than the equity of the asset seller. Unfortunately, since the equity of the SPV is not traded in public markets, we do not have information on its pricing.

An investment banker giving indicative pricing to a potential asset seller would typically quote the weighted average coupon plus the cost of fees to quote an all-in-cost of securitization. In our equation, this corresponds to the first and third terms and would ignore the second term (i.e., the cost of equity[5]). Since the costs of debt of the securitization are given by the coupons, but the cost of the equity is implicit, ignoring the second term is a convenient simplification. But can be a useful simplification: if the transaction involves no appreciable risk off-lay – i.e., if the first loss classes ($E_R + E_C$) are retained by the asset seller and if the capital at risk of the securitized assets is less than or equal to the first loss classes, then one can calculate the gains to securitization simply by comparing the weighted average coupon plus floatation costs with the cost of on-balance-sheet debt.

Potential Sources of Gains from Securitization

Using a no-arbitrage Modigiani and Miller (1959) argument, one can show that there *should* be no increase in value to a firm simply by altering its capital structure—through securitization or through any other device. If one separated the assets of a firm into two pools and financed those two pools separately, the costs of capital would differ as long as the risk characteristics of those two pools differed. Unless there are tax differences, bankruptcy cost differences, or differences in the performance of the firm induced by the separation of assets into the two pools, however, the weighted average cost of capital of the combined pools should be the same as that of the company prior to separation.

The largest violation of Modiglianni and Miller invariance is the interest rate tax shield of debt. Viewing securitization as an off-balance-sheet substitute for debt, a greater use of leverage to gain the benefits of tax shields will reduce the scope for using securitization to use the same tax shields in an off-balance sheet financing.

Leyland and Skarabot (2003), following Leyland (1993) look at the case where securitization involves full sale of assets (without residual risk kept by the asset seller) where tax differences and bankruptcy costs impact on optimal capital structure achieved through securitization. Using a Merton (1974) framework and specifying that the corporation has two portfolios with different but stable asset volatilities, recovery rates in default, and correlation of returns, they show that optimal securitization conditions are those under which the merged firm has a less efficient capital structure than the separate capital structures of the two different firms. This explains the phenomenon that securitization favors sale of a firm's lowest volatility assets[6]. The sale leads to an increase in risk in the asset seller by the act of securitization that increase the seller's cost of capital and appropriates wealth from the asset seller's creditors. The differential in asset volatility within the portfolio gives rise to the gains from securitization flowing from savings of bankruptcy costs. The bankruptcy costs are calculated as a constant fraction of the bankrupt regardless of the portfolio.

Saving bankruptcy costs are probably a richer source of securitization's gains than modeled by Leyland and Skarabot. Not only are SPVs are structured to be bankruptcy remote[7], they are structured so that they will not go bankrupt themselves. In the event that the

[5] If the reserve tranche were purchased by an external party, its cost would also be included in the weighted average coupon. In our formula, we consider it in the cost of equity under the assumption that it is retained by the asset seller. The stated coupon, then, is irrelevant because residual cash-flow reverts to the seller.

[6] Greenbaum and Thakor (1987)'s signaling model came to a similar conclusion.

[7] See Cohn (1998) for a discussion of bankruptcy remoteness. The degree to which a bankruptcy court can reach into SPVs to recover value for the bankruptcy trustee of the asset seller was illustrated in Conseco in 2002-3. After purchasing the sub-prme securitiztion asset originator Green Tree, Conseco went bankrupt partly

SPV is unable to make payment on its obligations, there is no incentive for claimholders to put the SPV itself into bankruptcy (assuming it is a corporate entity). By enforcing strict priority of payout without recourse to the courts, SPV's financial claimants gain from this structure.

If securitization were a pure value transfer play from bondholders to stockholders, information of a securitization should be greeted as bad news to bondholders.

Empirical support for the proposition that securitization expropriates creditors, however is mixed. In his event study of shareholder and bondholder reactions to securitizations, Thomas (2001) finds appropriation occurs among non-investment rated asset sellers but no such effect is evident among investment grade asset sellers. Most asset sellers are investment grade.

Gains to securitization might accrue from changes in managerial incentives accompanied by creation of the SPV. SPVs are very simple business entities where management of the assets of the entity is almost entirely devoid of discretion. Unlike the managers of a corporation, who must decide marketing, personnel, logistics, manufacturing, financing and strategic policies, the administrators of an SPV need only administer the assets to realize the intrinsic value of their expected cash flows. If business management activities can be broadly divided into entrepreneurship and stewardship, then management of an SPV is largely stewardship. The separation of non-discretionary stewardship from discretionary entrepreneurship may reduce agency costs in the former achieving a more optimal capital structure in the separated (securitized) structure.

Some perceive securitization to be the realm of rascals. Lupica (2000) describes it as a scheme to rob the uninformed unsecured investor. The collapse of Enron (Timmons (2002) and Conseco has forced regulators pay increasing interest to securitization. Were this the case, the worst performing firms to gain most from securitization.

With this discussion in mind, we look at five potential sources of the gains from securitization: size, liquidity, risk, leverage and performance of the asset originator. The effect of size is obvious. Securitization is expensive in terms of front end fees. Moreover, buyers of ABS are interested in homogeneity and liquidity; hence, one would expect economies of scale to be reflected in greater gains to larger asset sellers. These gains to size, however, may be a function of the amount of securitization that the asset seller does (assuming that purchasers look to the SPV alone) or of the size of the asset seller itself.

If an asset seller is illiquid, one would expect that liquidity constraint to be evident in its higher on-balance sheet costs of capital. Hence, lower liquidity of the asset seller should be associated with greater gains to securitization. Mindful of the fact that most of our sample are financial institutions, we avoid the commonly used current ratio in place of a liquidity ratio (liquid assets to total assets) which can be used for both financial and non-financial companies.

Following the Leyland and Skarabot, the higher the risk of the asset seller, given that securitized assets are of constant risk, the higher the gains that should be available to securitization. We use the CAPM beta as our risk measure. The sign on leverage is

---

because of the decline in creditworhthiness of sub-prime lending and the fall in value of the residual first loss tranches. The bankruptcy court ruled that the service fees paid to Conseco by the SPVs were inadequate and raised the fees, imparing the creditworthiness of a large number of ABS issues, causing the some of the greatest losses in the recent history of ABS and calling into question bankruptcy-remoteness.Donovan (2003) expected to be negative if one views securitization as an off-balance sheet substitute for debt.

If securitization is a tool used by astute managers to maximize shareholder wealth, we would expect to see a positive correlation between performance and gains to securitization. If it is the preserve of rascals, however, we would expect the reverse. We use two performance measures: growth and return on equity.

Finally, since there are three distinct industries – banking, non-banking financial intermediation and non-financial industries – in which securitization operates, we control for industry effects.

Data

We downloaded securitization data from ABSNet [8] in March 2004. There were 6,470 issues outstanding, with a then-current principal amount of $5,226 billions. The largest number of issues, however, was issued by a smaller number of SPVs who were originated by a still smaller number of corporate asset sellers. Table 1 shows, that the largest single issuer, with 297 issues outstanding, comprising a total original principal of $100 billion, was Residential Funding Mortgage Securities I, Inc. a corporate SPV set up by GMAC[9] to fund residential mortgages. Within the largest 15 issuers shown in Table 1, two other GMAC-sponsored SPVs are featured – Residential Asset Securities Corporation and Capital Auto Receivables Asset Trust. Each debt series is separately followed and rated by bond ratings agencies[10]. These large SPVs are conduits that issue ABS series from time to time to meet funding needs of newly purchased assets.

From the original sample, we excluded issues that were entirely retired as of March 2004 as well as those with no listed asset seller and those with incomplete or clearly incorrect data[11] and matched them with corporate, publicly listed sellers whose data was available in the WRDS/Compustat database. This left us with 1713 issues having an aggregate principal

---

8 ABSNetTM (www.absnet.net) is a partnership between Lewtan Technolgies and Standard and Poor's, that provides the most comprehensive database of ABS deal performance data currently available. It was set up to meet the needs of institutional investors who wish to have up-to-date information on the quality of their investments. The website gives the current performance data and structure of outstanding ABS issues.

9 GMAC is a wholly owned subsidiary of General Motors set up to finance dealer inventories of automobiles. It has since expanded into consumer auto finance, insurance and mortgage financing to become one of the leading finance companies in the world with assets of $250 billion. Reflecting its lack bank deposit bases and relatively poor credit rating, it fund itself partly in the ABS markets with securitizations reported as $98 billion with $5 billion in continuing interests in securitizations (i.e., first loss tranches). See GMAC's 2003 10K.

10 Neither GNMA nor FNMA issues would feature among the largest 15 issuers because GNMA and FNMA - pools are individually much smaller than the ones listed in Table 1. Note that we did not include agency issues in this analysis and that ABSNet does not list the issue characteristics of the limited GNMA and FNMA issues given in their database. This may be because the database is constructed largely to assist in portfolio credit risk management for investors; however, the risk of agency issues is considered quasi-sovereign.

11 For example, we excluded transactions where the cost of securitization was less than 1 percent. The deal summary contains information on the closing date, the original amount, remittance frequency, placement type, lead managers and trustee. For each class in each deal, information includes the class name, currency, original balance of outstanding securities, current balance of outstanding securities and original subordination (in terms of classes below the class in question that can be viewed as a cushion), current subordination, collateral type, original weighted average maturity, asset seller and asset servicer. Data are also given on initial and current credit support, broken down into over-collateralization, excess spread, etc., and the classes to which the credit support applies. The database gives information on delinquencies, loss rate, prepayment rate and the collateral pool balances.

outstanding of $810 billion originated by 120 asset sellers composed of 61 banks, 35 non-bank financial institutions and 24 non-financial companies. From these companies, we obtained information about their securitization activity, liquidity, profitability, risk and performance. We considered the 1,713 issues in aggregate as if they were a single securitization for each of the 120 asset sellers by weighting them by outstanding principal amount as of March 2004.

Table 1. Top 15 Issuers of ABS: March 2004 Outstanding Series

Issuer	Number of debt series outstanding	Total Principal ($ millions)
Residential Funding Mortgage Securities I, Inc (GMAC)	297	100,805.5
Ford Credit Auto Owner Trust	33	90,075.6
SLM Funding Corporation (Sallie Mae)	44	77,446.1
Conseco Finance Corp.	142	71,456.1
Countrywide Home Loans, Inc.	128	63,910.8
Citibank Credit Card Master Trust I	43	57,266.5
MBNA Master Credit Card Trust II	106	56,104.3
Prudential Home Mortgage Company	153	47,435.7
Residential Asset Securities Corporation (GMAC)	55	45,072.0
Capital Auto Receivables Asset Trust (GMAC)	21	44,236.0
CS First Boston Mortgage Securities Corp.	46	43,079.1
Discover Card Master Trust I	52	42,941.6
Chase Credit Card Master Trust	44	41,518.0
Wells Fargo Mortgage Backed Securities	81	41,400.2
Washington Mutual Mortgage Securities Corp.	24	38,650.5

The above table summarizes the top issuers listed by ABSNet within the total 6,470 outstanding issues before reduction of sample size for matching with originators. The name in brackets gives main asset originator in instances where it is not obvious from the title.

Source: ABSNet March 2004.

Costs of capital are calculated for each of the asset sellers by using the bond rating for the asset seller's long term debt and short term debt respectively and the average yield of such rated securities in March 2004 as provided by Datastream. Costs of capital of the securitization are given by the coupons on each of the classes in each of the securitizations as reported by ABSNet. The cost of equity for both the asset seller and the securitization is calculated by using the capital asset pricing model and betas given by WRDS/Compustat. We use a risk premium of eight percent[12] and the risk free rate of the 10-year bond rate in March 2004.

Table 2 shows the summary statistics.

[12] Varying the equity market premium from five percent to eight percent did not substantially alter the results.

Table 2. Summary Statistics

Panel A: Total Sample of 120 Asset Sellers

Variable	Mean	Std Dev	Skewness	Kurtosis	Minimum	Maximum
Cost of Securitization	5.39	2.60	0.75	0.02	1.43	12.70
WACC	6.63	2.47	3.28	17.64	2.77	23.27
Spread	1.24	3.41	1.37	7.08	-5.64	20.33
Total Securitization	6.34	14.64	4.07	17.77	0.003	91.41
Securitize Ratio	0.46	1.35	4.43	21.23	0.00	9.25
Beta	0.82	0.52	0.86	1.31	-0.28	2.88
Size	10.20	2.10	-0.30	-0.84	5.15	13.59
Leverage	87.43	13.01	-2.47	6.16	32.98	98.17
Liquid Asset Ratio	47.87	33.42	-0.35	-1.30	0.00	98.86
Assets Growth Rate	11.33	18.39	-0.52	9.38	-89.99	86.63
ROE	11.14	14.66	-1.68	9.86	-63.75	63.77
I1	0.51	0.50	-0.03	-2.03	0.00	1.00
I2	0.29	0.46	0.93	-1.16	0.00	1.00
I3	0.20	0.40	1.52	0.31	0.00	1.00

Panel B: Sub-Sample of 63 Assets Sellers without Retention of Equity Risk

Variable	Mean	Std Dev	Skewness	Kurtosis	Minimum	Maximum
Cost of Securitization	5.64	2.62	0.57	-0.34	1.70	11.89
WACC	6.53	2.81	3.80	20.48	2.77	23.27
Spread	0.90	3.84	2.01	9.44	-5.45	20.33
Total Securitization	2.37	3.23	1.94	3.93	0.003	15.37
Securitize Ratio	0.53	1.61	4.14	17.66	0.00	9.25
Beta	0.73	0.51	1.28	3.80	-0.28	2.88
Size	9.86	2.24	-0.19	-1.09	5.15	13.49
Leverage	88.17	12.31	-2.50	6.08	37.08	97.61
Liquid Asset Ratio	44.17	35.31	-0.13	-1.54	0.00	98.86
Assets Growth Rate	10.44	18.67	-1.41	4.34	-28.06	86.63
ROE	8.77	17.03	-1.75	8.87	-63.75	63.77
I1	0.48	0.50	0.10	-2.06	0.00	1.00
I2	0.37	0.49	0.57	-1.73	0.00	1.00
I3	0.16	0.37	1.91	1.72	0.00	1.00

Table 2. Continued

Panel C: Sub-Sample of 57 Asset Sellers with Retention of Equity Risk

Variable	Mean	Std Dev	Skewness	Kurtosis	Minimum	Maximum
Cost of Securitization	5.12	2.58	0.99	0.73	1.43	12.70
WACC	6.73	2.07	1.74	4.32	3.94	14.87
Spread	1.61	2.85	-0.05	-0.24	-5.64	8.34
Total Securitization	10.74	20.16	2.70	7.02	0.044	91.41
Securitize Ratio	0.39	1.00	4.09	17.84	0.00	5.62
Beta	0.92	0.52	0.52	-0.32	0.04	2.11
Size	10.59	1.87	-0.27	-0.66	6.79	13.59
Leverage	86.62	13.80	-2.47	6.47	32.98	98.17
Liquid Asset Ratio	51.95	30.99	-0.62	-0.82	0.00	98.32
Assets Growth Rate	12.31	18.19	-2.85	17.77	-89.99	54.22
ROE	13.77	11.05	-0.37	3.69	-21.53	46.37
I1	0.54	0.50	-0.18	-2.04	0.00	1.00
I2	0.21	0.41	1.46	0.13	0.00	1.00
I3	0.25	0.43	1.21	-0.55	0.00	1.00

Panel D: Test of Difference Between Means of Sub-Samples

Variable	Without Equity Risk	Without Equity Risk	Difference	t Value	Pr >|t|
Cost of Securitization	5.64	5.12	0.52	1.09	0.28
WACC	6.53	6.73	-0.20	-0.44	0.66
Spread	0.90	1.61	-0.71	-1.17	0.25
Total Securitization	2.37	10.74	-7.37	-3.10***	0.00
Securitize Ratio	0.53	0.39	0.14	0.61	0.54
Beta	0.73	0.92	-0.19	-2.03**	0.04
Size	9.86	10.59	-0.73	-1.95*	0.05
Leverage	88.17	86.62	1.55	0.65	0.52
Liquid Asset Ratio	44.17	51.95	-7.78	-1.28	0.20
Assets Growth Rate	10.44	12.31	-1.87	-0.55	0.58
ROE	8.77	13.77	-5.00	-1.93*	0.06
I1	0.48	0.54	-0.06	-0.74	0.46
I2	0.37	0.21	0.16	1.89*	0.06
I3	0.16	0.25	-0.09	-1.18	0.24

Note: Sample statistics of 1,713 series with a total outstanding principal of $810 billions sponsored by 120 asset sellers. Panel B sub-sample shows sellers who did not retain equity risk being the first-loss reserve tranche and/or other guarantees. Panel C subsample shows sellers who did retain equity risk being the first loss reserve tranche and/or other guarantees. *Spread* is WACC – cost of securitization; *Total securitization* is the total principal of outstanding securitized assets of the asset seller expressed in billions of dollars; *Securitize Ratio* is the outstanding amount of securitized assets expressed as a percent of total seller assets; *Beta* is the CAPM Beta of the seller as provided by Compustat; *Size* is the natural logarithm of total assets of the seller in millions of dollars; *Leverage* is Total Liabilities / Total Assets of the seller; *Current Ratio* is current assets / current liabilities of the seller; *Liquid Asset Ratio* is Liquid Assets /Total Assets of the seller; *Assets Growth Rate* is the one year growth rate of the sellers assets for the previous year; *ROE* is the return on equity of the seller in the last year.

Results

Summary Statistics

Gains from securitization can be read directly from Table 2. The average cost of securitization is 5.39 percent. The average cost of capital is 6.63 percent. The 1.24 percent difference is the gain. If one multiplies the average spread by total sample securitized principal of $810 billion, the annual savings are a not inconsiderable $10 billion.

Although GMAC described above is one of the most active originators of securitizations, it is by no means an outliner among the 120. The average principal of ABS outstanding for each asset seller is $6.3 billion, being 46 percent of on-balance sheet assets of the asset sellers. The most active asset sellers (in terms of outstanding ABS as a proportion of assets) raise more than nine times their on-balance sheet funding with securitization. On average, the asset sellers exhibit lower than one betas. Their high leverage (87 percent debt) reflects the fact that most members of the sample are banks or non-bank financial institutions. Liquidity varies widely. On average the firms are profitable (11.33 percent return on equity) and enjoying a moderately rapid growth of 11.14 percent per annum.

It is instructive to compare two sub-samples based on risk retention versus risk off-lay. Just over half of the asset sellers (52.5 percent) sell off their risk exposure to their assets when they securitize them. Their summary statistics are reported in Panel B as “Asset Sellers without Retention of Equity Risk”. The other 57 asset originators, reported in Panel C, retain equity risk, i.e., the first loss classes that contribute credit support in securitizations. Although the risk retention group contains only 47.5 percent of the asset sellers by number, it accounts for 80.4 percent of the ABS by dollar amount of principal outstanding. In this important aspect, the vast majority of securitizations involve no substantial off-lay of risk of the securitized assets: for them securitization is simply an off-balance sheet, debt-financing technique. As Table 2 Panel D shows, the asset sellers without retention of equity risk are significantly smaller, securitize less assets, are less risky (in terms of beta) and are less profitable than those that retain the equity risk of the assets securitized. The asset sellers that do not retain equity risk have substantially lower gains from securitization measured by the spread between the asset seller WACC and the SPV cost of securitization, (90 basis points versus 161 basis points), but this substantial difference is not statistically significant. The asset sellers that do not retain equity risk are more likely to be non-bank financial institutions (non-bank financial institutions make up about one third of the sup-sample compared with about one fifth of the equity risk retention sub-sample). In other respects, however, the sub-samples are not much different.

Sources of Gains in Total Sample

Table 3 presents the regression where the dependent variable is the spread between the WACC of the asset seller and the cost of securitization and a set of independent variables. Since the spread is the difference between two costs of capital which themselves are combinations of similar independent variables, we present the regression equations for the WACC and the cost of securitization as well as the spread. We only discuss the regressions for the spread.

Table 3. Sources of Gains from Securitization

Panel A: Total Sample of 120 Asset Sellers

Dependent Variable	Independent Variables										
	Intercept	Log (Assets)	Log (Securitization)	Securitize Ratio	Beta	Leverage	Liquid Asset Ratio	Assets Growth Rate	ROE	Dummy Variable I1	Dummy Variable I2
Cost of Securitization	10.17***	-0.24	0.52***	0.73***	1.81***	0.04	0.01	0.00	0.01	-0.42	-1.44**
	4.57	-1.49	3.58	2.67	3.75	1.56	1.64	0.17	0.56	-0.62	-2.03
WACC	12.03***	-0.36***	0.07	-0.21	2.81***	-0.05***	-0.00	-0.00	-0.02	0.06	0.00
	6.63	-2.79	0.60	-0.95	7.14	-2.69	-0.83	-0.37	-1.42	0.10	0.01
Spread	1.86	-0.13	0.59***	-0.94**	1.00	-0.09***	-0.02*	-0.01	-0.03	0.48	1.44
	0.62	-0.58	3.00	-2.54	1.53	-2.77	-1.71	-0.35	-1.27	0.52	1.50

Panel B: Sub-Sample of 63 Assets Sellers without Retention of equity Risk

Dependent Variable	Independent Variables										
	Intercept	Log (Assets)	Log (Securitization)	Securitize Ratio	Beta	Leverage	Liquid Asset Ratio	Assets Growth Rate	ROE	Dummy Variable I1	Dummy Variable I2
Cost of Securitization	10.95***	-0.53**	-0.37*	0.56	2.32***	0.05	-0.01	0.01	0.04	-2.00*	-2.51*
	2.84	-2.38	-1.73	1.38	3.05	1.41	-0.64	0.69	1.67	-1.71	-2.01
WACC	10.98***	-0.46**	0.20	-0.81**	2.85***	-0.05	-0.01	-0.01	-0.04*	0.64	1.65
	3.09	-2.23	1.01	-2.18	4.09	-1.43	-0.97	-0.33	-1.77	0.60	1.44
Spread	0.03	0.07	0.58*	-1.38**	0.53	-0.11*	-0.00	-0.02	-0.07**	2.64	4.16**
	0.01	0.22	1.78	-2.28	0.47	-1.83	-0.16	-0.67	-2.21	1.51	2.24

Table 3. Continued

Panel C: Subsample of 57 Asset Sellers with Retention of equity Risk

Dependent Variable	Independent Variables										
	Intercept	**Log (Assets)**	**Log (Securitization)**	**Securitize Ratio**	**Beta**	**Leverage**	**Liquid Asset Ratio**	**Assets Growth Rate**	**ROE**	**Dummy Variable I1**	**Dummy Variable I2**
Cost of Securitization	10.26***	-0.29	0.49**	0.79*	2.25***	0.01	0.03***	-0.02	-0.04	1.08	-0.66
	3.89	-1.08	2.27	1.88	3.33	0.50	3.31	-0.87	-1.33	1.25	-0.72
WACC	13.15***	-0.16	0.01	0.46*	2.03***	-0.07***	-0.01	-0.00	-0.00	-0.10	-1.45***
	8.79	-1.07	0.06	1.96	5.30	-4.18	-1.26	-0.20	-0.16	-0.21	-2.79
Spread	2.88	0.13	0.49**	-0.32	-0.22	-0.08**	-0.04***	0.01	0.04	-1.18	-0.79
	1.00	0.43	2.10	-0.70	-0.29	-2.62	-3.67	0.69	1.13	-1.25	-0.78

Notes:

The bold-faced number in each line is the coefficient estimates below which T statistics are given. *, **, *** significant in a 2-tailed test at the 10 percent, 5 percent and 1 percent levels of confidence.

Spread is WACC – cost of securitization; *Total securitization* is the total principal of outstanding securitized assets by the asset seller expressed in billions of dollars; *Securitize Ratio* is the outstanding amount of securitized assets expressed as a percent of total seller assets; *Beta* is the CAPM Beta of the seller as provided by Compustat; *Size* is the natural logarithm of total assets of the seller in millions of dollars; *Leverage* is Total Liabilities / Total Assets of the seller; *Current Ratio* is current assets / current liabilities of the seller; *Liquid Asset Ratio* is Liquid Assets /Total Assets of the seller; *Assets Growth Rate* is the one year growth rate of the sellers assets for the previous year; *ROE* is the return on equity of the seller in the last year; *Bank* is a dummy variable that equals 1 when the asset seller is a bank and zero otherwise; Non-Bank FI is a dummy variable that equals 1 when the asset seller is a non-bank financial institution and zero otherwise

The R-squared for the three regressions are: for Panel A 0.29, 0.47 and 0.24; for Panel B 0.30, 0.48 and 0.28l and for Panel C 0.47,0.74 and 0.48, respectively.

The Chow Test's that F value (Pr>F) are 1.65(0.0956), 1.69(0.0858) and 1.39(0.1919).

Size Counts

The higher the dollar amount of ABS outstanding originated by the asset seller, the greater the spread enjoyed by the asset seller. An increase from $6 billion ABS to $16 billion in ABS outstanding is associated with a very significant and economically substantial rise of 59 basis points[1]. The market appears to be rewarding the liquidity of aggregate ABS issued, not the size of the asset seller (which is measured by the log of the asset seller's assets, an independent variable that remains insignificant). But a rise in the *proportion* of assets that an asset seller securitizes leads to a significant fall in the spread. Remembering that the average asset seller has ABS outstanding of about 46 percent of on-balance-sheet assets, raising that one standard deviation to 180 percent would lower the spread by over 100 basis points, wiping out the positive spread to securitization. Too much securitization – i.e., securitization out of proportion to the asset seller – is penalized by the market.

The higher the leverage of the asset seller the lower the gains from securitization, lending support to the view that gains from on-balance sheet leverage of debt and the off-balance securitization are substitutes for each other.

We hypothesized that the riskier the company, the higher the gains to securitization. The sign of the beta coefficient is correct, but insignificant.

Finally, asset sellers that are less liquidity-constrained have less to gain from securitization. An increase in the liquid asset ratio from the average 48 percent to 58 percent will decrease spreads by 20 basis points.

There is no significant effect for either performance or industry.

With or without Equity Risk Retention

We gain greater insight by dividing our sample into two sub-samples: those asset sellers who do not retain equity risk of the assets sold and those who do retain equity risk. Turn first to liquidity. Among those asset sellers who retain equity risk (Table 3 Panel C), liquidity becomes the most significant determinant of gains to securitization. This group has a high 52 percent average liquidity. Decreasing that liquidity to 42 percent leads to a 40 basis point improvement in spread gains. For this sub-sample, too, the on-balance sheet – off-balance-sheet substitution of leverage is strongly evident. A company that retains the risk of assets whether or not it securitizes them can take the interest rate tax shield of debt with either on or off the balance sheet financing.

Turn now to Table 3 Panel B, the larger sub-sample in terms of number of asset sellers, but the much smaller sub-sample in terms of total ABS dollar volume, where the asset sellers off-lay the equity risk when the sell assets. Here liquidity is an insignificant explainer of gains to securitization and leverage drops substantially in terms of significance. Return on equity, however, takes on a significant negative sign, and the securitization ratio's sign is large and negative. Both of these signs lend support to the hypothesis that securitization that truly isolates two parts of the formerly integrated portfolio of the asset seller is a method by which poorly performing companies can gain by selling their best assets into SPVs. Their gains,

---

[1] The size factors are in natural logs of the ABE outstanding expressed in thousands of dollars . The average principal of ABS outstanding for each asset seller is $6.3 billion or 15.6 expressed in natural logs. Increasing the coefficient to 16.6 to is an increase to $16.2 billion dollars.

measured in the lower cost of capital of the securitization, comes possibly at the expense of existing asset seller creditors. These companies correspond more to the theoretical models of securitization of Leyland and Skarabot which model SPVs and post securitization asset sellers as truly separated. The highly negative and significant sign on the securitization ratio (not evident in Panel C) may be interpreted as the market's increasing doubt in the quality of the assets sold as the ratio of securitized assets increases.

Industry Sub-Samples

In the above analysis, we have treated the 61 asset-selling banks identically to the 35 non-bank financial institutions and the 24 non-financial institutions. We used no more than a dummy variable to separate their average spread effects. Although we cannot reject the null hypothesis that jointly these three types of corporations gains from securitization are the same (See Table 5). The strong tradition of academic banking literature that views banks as special leads us to examine industry sub-samples. Table 4 presents the industry sub-samples' regressions. Panel A suggests that liquidity and liquidity alone explains the gains to securitizations that banks enjoy. The less liquid the bank, the more it gains from securitization.

The 35 non-bank financial institutions drive the sample results that show gains to securitization decreasing in the securitization ratio. If the reduced benefits for increased proportion of the balance sheet in securitized assets is a market penalty for suspected lower quality assets being securitized (and, from Table 3 Panel B, securitized without retention of equity risk by the asset seller), then suspicions are falling on the non-bank financial institutions. Turning to the smallest sub-sample, the 24 non-financial institutions, gains to securitization are strongly and significantly decreasing in leverage. Given the higher leverage that banks and other financial institutions display, and the greater differential between typical fixed income securitized assets and the typical on-balance-sheet assets of non-financial corporations, it is not surprising that we observe the tradeoff between on balance sheet and off balance sheet leverage to be greatest among the non-financial institutions.

These industry conclusions, however, are tentative because we have insufficiently large samples to allow division of our sub-samples by the criteria of both Table 3 and Table 4.

Finally, our sample contains both US and foreign asset sellers. We ran separate regressions of gains from securitization of the 74 US asset sellers and the 46 non-US asset sellers in Table 6. As one would expect from the reduced sample size, the non US sample shows reduced significance of all coefficients, with only the size variables retaining significance: as with the overall sample, gains to securitization are increasing in the amount of ABS issued by an issuer but decreasing in the securitization outstanding as a percentage of assets. An interesting change occurs with the (larger) US sample, however. Throughout the sample, the previous tests, the risk of the asset seller measured by beta had no significant effect on the gains to securitization. For US securitizes, when isolated in their own sample, however the riskier the asset seller, the greater the gains to securitization.

Table 4. Sources of Gains from Securitization: Industry Subsamples

Panel A: Banks

Dependent Variable	Independent Variables								
	Intercept	**Log (Assets)**	**Log (Securitization)**	**Securitize Ratio**	**Beta**	**Leverage**	**Liquid Asset Ratio**	**Assets Growth Rate**	**ROE**
Cost of Securitization	16.01	-0.39	-0.40	-1.32	3.12***	-0.04	0.02	-0.01	-0.04
	1.18	-1.58	-1.50	-0.65	3.44	-0.24	1.57	-0.34	-1.05
WACC	13.54***	0.06	-0.14	1.33*	2.05***	-0.08	-0.01**	-0.00	0.00
	3.05	0.68	-1.66	2.00	6.94	-1.66	-2.56	-0.36	0.23
Spread	-2.46	0.45*	0.25	2.64	-1.07	-0.04	-0.03**	0.01	0.04
	-0.18	1.80	0.96	1.30	-1.18	-0.30	-2.41	0.22	1.12

Panel B: Non-Bank Financial Institutions

Dependent Variable	Independent Variables								
	Intercept	**Log (Assets)**	**Log (Securitization)**	**Securitize Ratio**	**Beta**	**Leverage**	**Liquid Asset Ratio**	**Assets Growth Rate**	**ROE**
Cost of Securitization	6.82	-0.29	-0.45	0.60*	1.07	0.07*	0.00	0.00	-0.02
	1.59	-0.89	-1.46	1.74	1.44	1.75	0.11	0.23	-0.71
WACC	12.93**	-1.27**	0.22	-0.83*	3.60***	0.03	-0.02	-0.00	-0.04
	2.15	-2.73	0.51	-1.71	3.45	0.48	-0.79	-0.09	-1.06
Spread	6.10	-0.98	0.68	-1.43**	2.54	-0.05	-0.02	-0.01	-0.02
	0.71	-1.47	1.09	-2.06	1.70	-0.54	-0.61	-0.18	-0.39

Table 4. Continued

Panel C: Non-FI Company

Dependent Variable	Independent Variables								
	Intercept	**Log (Assets)**	**Log (Securitization)**	**Securitize Ratio**	**Beta**	**Leverage**	**Liquid Asset Ratio**	**Assets Growth Rate**	**ROE**
Cost of Securitization	9.85*	-0.93	-0.25	1.31*	1.64	0.09	-0.02	0.02	0.03
	1.94	-1.44	-0.49	1.82	1.00	1.55	-0.33	0.62	0.93
WACC	9.87***	-0.02	0.13	0.43	3.66***	-0.08***	-0.03	-0.00	-0.02
	4.72	-0.06	0.64	1.46	5.40	-3.31	-1.33	-0.32	-1.56
Spread	0.02	0.92	0.38	-0.87	2.02	-0.17**	-0.01	-0.02	-0.06
	0.00	1.21	0.65	-1.05	1.05	-2.50	-0.19	-0.64	-1.35

The R-squared for the three regressions are: for Panel A 0.33,0.61 and 0.35; for Panel B 0.32,0.51 and 0.32; and for Panel 0.52,0.86 and 0.33, respectively.

Table 5. Tests of Structural Change: Chow Test

Difference between	Dependent Variable	F Value	Pr> F
Banks and Non-Bank Financial Institutions	Cost of Securitization	0.85	0.5693
	WACC	2.07	0.0425
	Spread	1.44	0.1863
Non-Bank Financial Institutions and Non-FI Companies	Cost of Securitization	1.16	0.3448
	WACC	0.96	0.4894
	Spread	0.52	0.8490
Banks and Non-FI Companies	Cost of Securitization	0.48	0.8819
	WACC	1.31	0.2469
	Spread	0.95	0.4888

Note: These tests measure the joint hypothesis that all coefficients *except the intercept* are the same.

Table 6. Sources of Gains from Securitization

Panel A: Sub-Sample of 74 Assets Sellers from the US

Dependent Variable	Independent Variables										
	Intercept	Log (Assets)	Log (Securitization)	Securitize Ratio	Beta	Leverage	Liquid Asset Ratio	Assets Growth Rate	ROE	Dummy Variable I1	Dummy Variable I2
Cost of Securitization	12.74***	-0.12	-0.54***	0.48**	1.02*	0.01	0.02	0.01	0.00	-1.55	-2.43***
	5.00	-0.59	-3.00	2.04	1.73	0.40	1.31	0.87	0.01	-1.66	-2.68
WACC	13.82***	-0.54**	0.13	-0.31	3.03***	-0.08***	0.01	-0.01	-0.01	0.50	0.35
	5.53	-2.62	0.72	-1.35	5.26	-2.81	0.70	-0.38	-0.26	0.55	0.39
Spread	1.08	-0.41	0.66**	-0.78**	2.01**	-0.09**	-0.01	-0.02	-0.01	2.06	2.78**
	0.28	-1.32	2.47	-2.23	2.28	-2.11	-0.42	-0.83	-0.17	1.46	2.04

Note: The R-squared for the three regressions are 0.32,0.53 and 0.29, respectively.

Panel B: Subsample of 46 Asset Sellers from non-US

Dependent Variable	Independent Variables										
	Intercept	Log (Assets)	Log (Securitization)	Securitize Ratio	Beta	Leverage	Liquid Asset Ratio	Assets Growth Rate	ROE	Dummy Variable I1	Dummy Variable I2
Cost of Securitization	7.81	-0.08	-0.56	1.05	3.09***	0.03	-0.00	-0.02	0.01	0.09	-0.51
	1.27	-0.22	-1.53	0.41	3.03	0.70	-0.06	-0.76	0.31	0.08	-0.34
WACC	6.42***	-0.18	0.14	-0.74	2.20***	-0.01	-0.00	-0.01	-0.01	-0.46	1.09*
	2.92	-1.40	1.04	-0.81	6.04	-0.63	-0.50	-0.68	-0.89	-1.11	2.01
Spread	-1.39	-0.10	0.69*	-1.79	-0.89	-0.05	-0.00	0.01	-0.02	-0.55	1.60
	-0.21	-0.26	1.76	-0.65	-0.80	-0.86	-0.11	0.48	-0.57	-0.43	0.98

Note: The R-squared for the three regressions are 0.27,0.66 and 0.21, respectively.
The Chow Test's F value (Pr>F) are 1.06(0.4010), 1.33(0.2208) and 1.02(0.4322).

Conclusion

Our paper has contributed to the literature in several respects. We have shown that gains from securitization are considerable. This is not surprising, given securitization's rapid rise to become the largest fixed income asset class in the world. Gains to the asset seller are increasing in the size (and hence liquidity) of the amount of an asset seller's ABS outstanding. Further analysis of the gains, however, take into consideration differences between securitization where the asset seller retains (versus off-lays) equity risk.

The vast majority of securitizations by dollar volume (80 percent) are in the former category where the asset originator enjoys no off-lay of the equity risk of the sold assets. Because the sellers retain the equity of the SPV, these securitizations are really off-balance sheet debt financing. These asset sellers primarily source their gains from liquidity. These asset sellers can use substitution of on-balance leverage with off-balance sheet leverage. This phenomenon is especially evident among non-financial institutions.

Some 20 percent of securitizations by dollar volume, but slightly over half by asset seller number, involve asset sellers off-laying equity risk of the sold assets. These securitizations resemble more closely non-recourse securitization modeled in the optimal capital structure literature. Here the declining returns to performance and the securitization ratio suggest that investor concerns of asset quality and the potential for wealth transfers from creditors to equity holders are substantial.

We undertook this research to publish and analyze the magnitude of the gains from securitization. Our metric for calculating those gains, suggested by the investment banking method of reporting securitizations costs as a cost of capital, has, we believe, provided some insights. But our conclusions are limited by our lack of theoretical development. Theories of optimal capital structure can and should be expanded to include structured finance. The special purpose entities of structured finance—securitizations, real estate investment trusts, project finance, mutual funds, hedge funds etc.—give academics rich empirical testing grounds on which we can increase our understanding of the determinants of optimal capital structure.

Acknowledgements

We gratefully acknowledge the support of an Earmark Grant from the Hong Kong Research Grants Council that made this research possible. Please address all correspondence regarding to this paper to Hugh Thomas. Hugh-Thomas@cuhk.edu.hk.

Appendix: 120 Securitizing Companies with Average Sizes of Securitization Tranches

Company Name	Industry	Country	Total Securitiztion	Senior Class (%)	Subordinate Class (%)	Other Class (%)	Credit Support (%)
AAMES FINANCIAL CORP.	2	United States	410,871	0.6733	0.3016	0.0252	0.0000
ABBEY NATIONAL PLC -ADR	1	England	884,820	0.7909	0.1045	0.1045	0.0300
ABN AMRO HLDG N V -SPON ADR	1	Netherlands	2,265,361	0.8217	0.0656	0.1128	0.0272
ACE LIMITED	2	Cayman Islands	1,073,690	0.7895	0.2105	0.0000	0.0000
ADVANTA CORP -CL B	2	United States	229,275	1.0000	0.0000	0.0000	0.0000
AEGON NV	2	Netherlands	2,848,772	0.9341	0.0319	0.0340	0.0000
AMERICAN EXPRESS	2	United States	15,372,188	0.8576	0.0753	0.0671	0.0000
AMSOUTH BANCORPORATION	1	United States	94,362	0.5467	0.3021	0.1511	0.0823
ANGLO IRISH BANK	1	Ireland	575,367	0.6737	0.1328	0.1935	0.0301
AUTONATION INC	3	United States	177,071	1.0000	0.0000	0.0000	0.0000
BANCA ANTONVENETA POPOLARE	1	Italy	717,890	0.8976	0.0334	0.0689	0.0000
BANCA INTESA SPA	1	Italy	2,239,635	0.8728	0.0421	0.0851	0.0120
BANCA POPOLARE DI BERGAMO	1	Italy	738,686	0.9222	0.0778	0.0000	0.0306
BANCA POPOLARE DI MILANO	1	Italy	851,958	0.9105	0.0596	0.0299	0.0157
BANCA POPOLARE DI VERONA	1	Italy	373,227	0.9237	0.0697	0.0067	0.0204
BANCO COMERCIAL PORTGE -ADR	1	Portugal	2,436,089	0.9151	0.0448	0.0400	0.0189
BANCO GUIPUZCOANO	1	Spain	889,898	0.9490	0.0510	0.0000	0.0000
BANCO PASTOR	1	Spain	803,642	0.9468	0.0453	0.0079	0.0000
BANCO POPULAR ESPANOL	1	Spain	432,566	0.9458	0.0542	0.0000	0.0000
BANCO SANTANDER CENT -ADR	1	Spain	1,080,000	0.9000	0.0350	0.0650	0.0170
BANK OF AMERICA CORP	1	United States	10,679,721	0.3690	0.0974	0.5335	0.0000
BANK OF NOVA SCOTIA	3	Canada	294,750	0.9500	0.0195	0.0305	0.0000
BANK ONE CORP	1	United States	752,667	0.9756	0.0244	0.0000	0.0030
BANKINTER	1	Spain	5,099,273	0.9099	0.0303	0.0599	0.0136

Appendix. Continued

Company Name	Industry	Country	Total Securitiztion	Senior Class (%)	Subordinate Class (%)	Other Class (%)	Credit Support (%)
BARCLAYS PLC/ENGLAND -ADR	1	England	2,631,993	0.9692	0.0308	0.0000	0.0147
BEAR STEARNS COMPANIES INC	2	United States	6,094,611	0.5511	0.0292	0.4197	0.0000
BMW-BAYER MOTOREN WERKE AG	3	Germany	2,344,520	0.9522	0.0478	0.0000	0.0184
BNL-BANCA NAZIONALE LAVORO	1	Italy	3,040,507	0.8068	0.0425	0.1507	0.0000
BOMBARDIER INC	3	Canada	1,571,868	0.7968	0.2032	0.0000	0.0013
CANADIAN IMPERIAL BANK	1	Canada	3,075,000	1.0000	0.0000	0.0000	0.0000
CANADIAN TIRE CORP	3	Canada	2,362,500	0.0000	0.0000	1.0000	0.0000
CAPITAL ONE FINL CORP	2	United States	91,414,812	0.7996	0.1019	0.0985	0.0025
CAPITALIA SPA	1	Italy	3,495,758	0.2619	0.1445	0.5937	0.0000
CAPSTEAD MORTGAGE CORP	2	United States	10,436	0.0000	0.7127	0.2873	0.0000
CARMAX INC	3	United States	5,157,735	0.9115	0.0273	0.0611	0.0174
CATERPILLAR INC	3	United States	2,356,311	0.8976	0.0615	0.0409	0.0474
CENDANT CORP	2	United States	1,691,280	0.8016	0.1958	0.0026	0.0002
CENTEX CORP	3	United States	10,553,295	0.6452	0.1409	0.2139	0.0000
CENTRO PROPERTIES GROUP	2	Australia	1,031,300	0.9280	0.0720	0.0000	0.0126
CHARMING SHOPPES	3	United States	500,000	0.6470	0.1740	0.1790	0.0000
CITICORP	1	United States	14,253,706	0.8983	0.0974	0.0044	0.0504
CITY NATIONAL CORP	1	United States	448,107	0.9280	0.0532	0.0188	0.0248
CNH GLOBAL NV	3	Netherlands	3,092,736	0.8898	0.0330	0.0772	0.0000
COLONIAL BANCGROUP	1	United States	472,815	0.9603	0.0397	0.0000	0.0000
COLONIAL PROPERTIES TRUST	2	United States	251,000	0.7100	0.2050	0.0850	0.0464
COUNTRYWIDE FINANCIAL CORP	2	United States	31,518,423	0.3072	0.0713	0.6215	0.0000
CREDIT AGRICOLE INDOSUEZ	1	France	43,705	0.8645	0.1355	0.0000	0.0000
CREDIT LYONNAIS SA	1	France	400,442	0.7967	0.2033	0.0000	0.0000

Appendix. Continued

Company Name	Industry	Country	Total Securitiztion	Senior Class (%)	Subordinate Class (%)	Other Class (%)	Credit Support (%)
CREDIT SUISSE FIRST BOS USA	2	United States	9,653,397	0.9327	0.0466	0.0207	0.0000
DAIMLERCHRYSLER AG	3	Germany	27,393,340	0.6426	0.0000	0.3574	0.0012
DEUTSCHE BANK AG	1	Germany	6,469,731	0.8785	0.1019	0.0196	0.0189
DVI INC	2	United States	1,853,869	0.8709	0.0162	0.1129	0.0053
EGG PLC	1	United Kingdom	500,000	0.8700	0.0500	0.0800	0.0000
EMC INSURANCE GROUP INC	2	United States	6,242,178	0.4901	0.1895	0.3204	0.0000
ENI S P A -SPON ADR	3	Italy	6,637,000	0.7900	0.1293	0.0808	0.0000
EUROHYPO AG	1	Germany	674,826	0.1843	0.2464	0.5693	0.0000
FEDERAL HOME LOAN MORTG CORP	2	United States	40,685	1.0000	0.0000	0.0000	0.0000
FIAT SPA -ADR	3	Italy	850,005	0.9000	0.1000	0.0000	0.0000
FIRST ACTIVE PLC	1	Ireland	6,843,393	0.9276	0.0724	0.0000	0.0220
FIRST KEYSTONE FINL INC	1	United States	195,312	0.2567	0.7433	0.0000	0.0000
FIRST UNION RE EQ & MTG INV	2	United States	1,011,799	0.6336	0.0765	0.2899	0.0000
FLEETBOSTON FINANCIAL CORP	1	United States	7,680,100	0.8394	0.0606	0.1000	0.0035
FLEETWOOD ENTERPRISES	3	United States	34,822	0.9650	0.0350	0.0000	0.0000
FORD MOTOR CO	3	United States	22,084,746	0.8404	0.0675	0.0921	0.0109
FREMONT GENERAL CORP	2	United States	712,356	0.1404	0.1650	0.6946	0.0000
GENERAL ELECTRIC CAPITAL SVC	2	United States	242,550	0.5517	0.2988	0.1494	0.0000
GENERAL MOTORS CORP	3	United States	65,643,254	0.8552	0.0324	0.1124	0.0127
GREENPOINT FINANCIAL CORP	1	United States	1,846,205	0.3742	0.1438	0.4820	0.0015
HARLEY-DAVIDSON INC	3	United States	2,384,351	0.9387	0.0483	0.0129	0.0098
HONDA MOTOR LTD -AM SHARES	3	Japan	9,036,282	0.9673	0.0000	0.0327	0.0114
HOUSEHOLD FINANCE CORP	2	United States	13,799,854	0.8921	0.0670	0.0409	0.0154
HUNTINGTON BANCSHARES	1	United States	3,185	0.0000	0.0603	0.9397	0.0000

Appendix. Continued

Company Name	Industry	Country	Total Securitiztion	Senior Class (%)	Subordinate Class (%)	Other Class (%)	Credit Support (%)
HYUNDAI MOTOR CO LTD	3	Korea	640,737	0.8568	0.0609	0.0822	0.0082
IKON OFFICE SOLUTIONS	3	United States	1,273,363	1.0000	0.0000	0.0000	0.0245
IMPAC MORTGAGE HLDGS INC	2	United States	4,503,112	0.4164	0.1131	0.4704	0.0000
INDEPENDENT BANK CORP/MA	1	United States	149,178	0.0740	0.2627	0.6633	0.0000
INDYMAC BANCORP INC	2	United States	1,244,053	0.4024	0.2309	0.3666	0.0000
IRWIN FINL CORP	1	United States	468,645	0.4515	0.2347	0.3138	0.0000
J P MORGAN CHASE & CO	1	United States	80,341,023	0.8015	0.0671	0.1314	0.0044
KEYCORP	1	United States	774,177	0.6922	0.0388	0.2690	0.0127
KREDITANSTALT FUER WIEDERAUF	1	Germany	164,725	0.4688	0.2793	0.2519	0.0000
LEHMAN BROTHERS HOLDINGS INC	2	United States	17,380,347	0.7006	0.2113	0.0881	0.0009
MACQUARIE BANK LTD	1	Australia	700,000	0.9750	0.0250	0.0000	0.0000
MARSHALL & ILSLEY CORP	1	United States	298,670	0.9256	0.0744	0.0000	0.0175
MBIA INC	2	United States	60,040	0.0000	0.0000	1.0000	0.0201
MBNA CORP	1	United States	63,877,006	0.8365	0.0768	0.0867	0.0014
MELLON FINANCIAL CORP	1	United States	1,270,113	0.7365	0.0345	0.2290	0.0000
MERRILL LYNCH & CO	2	United States	7,144,618	0.4563	0.0922	0.4515	0.0000
METRIS COMPANIES INC	2	United States	5,984,789	0.7844	0.1038	0.1118	0.0000
MID-STATE BANCSHARES	1	United States	605,991	0.6544	0.3456	0.0000	0.0033
MORGAN STANLEY	2	United States	11,537,741	0.6629	0.2231	0.1139	0.0002
NATIONAL BANK CANADA	3	Canada	5,115,192	0.5961	0.0273	0.3766	0.0025
NEW CENTURY FINANCIAL CORP	2	United States	10,638,675	0.8102	0.1882	0.0016	0.0008
NORTHERN ROCK PLC	1	United Kingdom	24,788,765	0.8986	0.0522	0.0492	0.0045
NOVASTAR FINANCIAL INC	2	United States	4,237,120	0.8901	0.0765	0.0334	0.0000
ONYX ACCEPTANCE CORP	2	United States	2,362,054	1.0000	0.0000	0.0000	0.0000

Appendix. Continued

Company Name	Industry	Country	Total Securitiztion	Senior Class (%)	Subordinate Class (%)	Other Class (%)	Credit Support (%)
POPULAR INC	1	United States	4,699,374	0.4044	0.1032	0.4925	0.0000
PROVIDENT FINL HLDGS INC	1	United States	380,420	1.0000	0.0000	0.0000	0.0000
PRUDENTIAL PLC -ADR	2	England	163,118	0.7736	0.2264	0.0000	0.0000
PSB BANCORP INC	1	United States	36,832	0.1915	0.8085	0.0000	0.0000
REGIONS FINL CORP	1	United States	1,333,060	0.9283	0.0404	0.0313	0.0194
ROYAL BANK OF CANADA	3	Canada	2,675,000	0.0000	0.0000	1.0000	0.0000
SAXON CAPITAL INC	2	United States	989,288	0.6984	0.3011	0.0005	0.0000
SKY FINANCIAL GROUP INC	1	United States	344,574	1.0000	0.0000	0.0000	0.0443
SLM CORP	2	United States	42,443,938	0.9376	0.0315	0.0308	0.0022
ST GEORGE BANK LTD	1	Australia	4,803,282	0.8946	0.0401	0.0653	0.0000
STERLING BANCORP/NY	1	United States	44,355	0.9422	0.0578	0.0000	0.0062
SUPERIOR FINANCIAL CORP DE	1	United States	100,714	0.5736	0.4264	0.0000	0.0000
SYNOVUS FINANCIAL CP	1	United States	65,519	0.0750	0.1919	0.7331	0.0000
TEXTRON FINANCIAL CORP	2	United States	212,875	0.8866	0.1134	0.0000	0.0000
UNIPOL	2	Italy	149,552	0.7939	0.0843	0.1219	0.0000
UNITED COMMUNITY FINL CORP	1	United States	185,284	0.6292	0.3708	0.0000	0.0233
WACHOVIA CORP	1	United States	4,447,901	0.9137	0.0434	0.0429	0.0000
WASHINGTON MUTUAL INC	1	United States	7,539,417	0.4295	0.0432	0.5273	0.0000
WELLS FARGO & CO	1	United States	12,639,919	0.4494	0.0450	0.5056	0.0005
WESTCORP	1	United States	300,781	1.0000	0.0000	0.0000	0.0879
WILSHIRE FINL SVCS GROUP INC	1	United States	302,805	0.9270	0.0565	0.0164	0.0000
XEROX CORP	3	United States	57,010	0.0000	0.0000	1.0000	0.1312
YAMAHA MOTOR CO LTD	3	Japan	200,000	0.8550	0.0600	0.0850	0.0306
ZIONS BANCORPORATION	1	United States	579,381	0.9025	0.0975	0.0000	0.0000

Note: For Industry, 1,2,3 denote bank, non-bank financial institution, and other industry (non-financial institution), respectively.

References

Ambrose, Brent W., Michael LaCour-Little and Anthony Saunders (2004) "Does Regulatory Capital Arbitrage, Reputation or Asymmetric Information Drive Securitization?" *working paper.* August 24

Burroughs, Katherine, Robert Grady and Kathleen Bailey (2002). "Select Recent Bankruptcy Cases Affecting Commercial Real Estate Finance and Securitization." *Finance & Real Estate* (July), Dechert, www.dechert.com.

Cohn, Michael J. (1998) "Asset Securitization: How Remote Is Bankruptcy Remote?" *Hofstra Law Review.* **26**: 4 Summer.

Donovan, Kevin.(2003) "Conseco Bankruptcy Heats Up: ABS players fight fee-related sleight of hand" *The Investment Dealers' Digest*: IDD. New York (Feb 24) 1

Financial Accounting Standards Board (2000). "Statement No. 140:Accounting for Transfers and Servicing of Financial Assets and Extinguishments of Liabilities." Norwalk, CT (September 2000).

Greenbaum, Stuart and Anjan Thakor (1987). "bank lending models: securitization versus deposits". *Journal of Banking and Finance.* **11** (3) 379-401.

-- -- (1995). *Contemporary Financial Intermediation.* Dryden. Fort Worth.

Leyland, Hayne and Jure Skarabot (2003) "On Purely Financial Synergies and the Optimal Scope of the Firm: Implications for Mergers, Spin-offs and Structured Finance" *working paper*.

Leyland, Hayne (1994) "Corporate Debt Value, Bond Covenants and Optimal Capital Structure" *The Journal of Finance.* **49**:4, 1213-1252.

Leland, Hayne and Jure Skarabot (2003). "On purely financial synergies and the optimal scope of the firm: implications for mergers, spinoffs and off-balance sheet finance." *working paper*.

Lockwood, Larry, Ronald Rutherford and Martin Herrera (1996). "Wealth Effects of Asset Securitization." *Journal of Banking and Finance.* **20**, 151-164.

Lupica, Lois (2000). "Circumvention of the bankruptcy process: the statutory institutionalization of securitization." *Connecticut Law Review.* **33**: 199.

LaCour-Little and Sanders (2004) asks the question "does regulatory capital arbitrage, reputation or asymmetric information drive securitization?" and comes to the conclusion that capital arbitrage or reputation is driving it.

Merton, RC. (1974) "On the pricing of corporate debt: The risk structure of interest rates" *The Journal of Finance.* **29**(2): 449-70.

Modigliani, F., and M. H. Miller(1959). "Corporation Finance, and the Theory of Investment: Reply." *American Economic Review.* vol. 49, no. 4 (September):655–669.

Pennacchi, George (1988). "Loan Sales and the Cost of Bank Capital." *The Journal of Finance.* **43**:2, 375-96.

Perry, Debra (1995). "Can asset securitization squeeze any juice out of your capital structure?" Speech for the Investors Forum on Asset Securitization given on October 23, 1995 in Bermuda reprinted in Moody's Investor Service Special Comment.

Raynes, Sylvain and Ann Rutledge (2003). *The Analysis of Structured Securities: precise risk measurement and capital allocation.* Oxford University Press, New York.

Skipworth, Douglas (2001). "Statement 140: a study of securitization disclosure." *FASB Study*. December 26.

Thomas, Hugh (2001). "Effects of Asset Securitization on Seller Claimants." *Journal of Financial Intermediation.* **10**, 306-330.

Timmons, Heather (2002) Everybody out of the risk pool? Enron and other problems are taking a toll on the structured-finance business" *Business. Week* New York (September 2, 2002) 3797, 86.

In: Economic Forecasting
Editor: Alan T. Molnar, pp. 135-162
ISBN: 978-1-60741-068-3

Chapter 5

THE NONPARAMETRIC TIME-DETRENDED FISHER EFFECT

Heather L.R. Tierney
Department of Economics and Finance, College of Charleston, Charleston, SC

Abstract

This paper uses frontier nonparametric VARs techniques to investigate whether the Fisher Effect holds in the U.S. The Fisher Effect is examined taking into account structural breaks and nonlinearities between nominal interest rates and inflation, which are trend-stationary in the two samples examined. The nonparametric time-detrended test for the Fisher Effect is formed from the cumulative orthogonal dynamic multiplier ratios of inflation to nominal interest rates. If the Fisher Effect holds, this ratio statistically approaches one as the horizon goes to infinity. The nonparametric techniques developed in this paper conclude that the Fisher Effect holds for both samples examined.

Keywords: Fisher Effect, nonparametrics, dynamic multipliers, monetary policy, trend-stationarity.

JEL Classification Code: E40, E52, E58.

1. Introduction

The dynamics of nominal interest rates and inflation are fundamental forces at the core of economic and financial decisions. The Fisher Effect, which relates these two variables, has several consequences on market rationality and efficiency, option pricing, portfolio allocation, monetary policy, and international trade just to name a few illustrations.[1]

[1] The Fisher Effect is a theoretical proposition based on the Fisher Equation, which defines nominal interest rates as being equal to the ex ante real interest rate plus expected inflation. According to the Fisher Effect in the long run, the real interest rate is constant, which means that there should be a long run one-to-one relationship between inflation and nominal interest rates.

Some economic and financial models implicitly assume constancy of the real interest rate (e.g. the capital asset pricing model, CAPM). Thus, the legitimacy of the decisions based on these models is tied to the assumption that the Fisher Effect holds. Monetary policy is also closely related to the Fisher Effect. For instance, if the Fisher Effect holds, short-term changes to nominal interest rates can be made to adjust for changes in inflation and unemployment without impacting the long-term real interest rate.

This paper investigates whether the Fisher Effect holds in the U.S. for two sample periods: for the first quarter of 1960 to the third quarter of 1995, and for the first quarter of 1960 to the second quarter of 2004. These long samples are necessary to investigate the long run relationship between nominal interest rates and inflation. The latter sample encompasses a period in which there is a substantial decrease in the volatility of nominal interest rates. This has allowed markets to have a more perceptible signal of likely future monetary policy action by the Federal Reserve, especially since 1995.

Frontier nonparametric techniques are used to investigate whether the Fisher Effect holds. In particular, the local linear least squares nonparametric method (LLLS) is applied to investigate the effect. This method is better equipped to deal with outliers, since each nonparametric estimator is locally estimated with more weight given to observations closer to the data point. This methodology is applied to study the dynamics of nominal interest rates and inflation in bivariate VARs, and the results are compared with the ones obtained from parametric techniques.

The advantage of using nonparametric over parametric VARs is that the nonparametric version yields cumulative dynamic multipliers obtained from nominal interest rates and inflation that capture the nonlinearities present in these series. The consideration of nonlinearities in the system generally generates on average larger estimated coefficients than the parametric counterpart, which are stationary and close enough to unity to produce the Fisher Effect. In addition, it also yields on average larger estimated variance-covariance matrices, which are necessary for the orthogonalization of the dynamic multipliers.

The related literature examining the time series properties of nominal interest rates and inflation do not present a consensus on whether or not the Fisher Effect holds. In addition, the literature is also divided into two broad groups: those who find a cointegration relationship between nominal interest and inflation, and those who find that the series are not integrated.

There are several possible explanations for the divergence of these conclusions. The possible econometric reasons could be related to not taking into consideration the potential non-stationarity and/or nonlinearities in the individual series or breaks in the relationship between nominal interest rates and inflation. These statistical features may arise from changes in the structure of the economy, such as different monetary policy regimes over time, or breaks in the level of inflation and in inflation expectations, among other things.

Along the lines of the cointegration examination of the Fisher Effect, Mishkin (1992) finds a cointegration relationship through the use of the Engle and Granger (1987) model, which indicates that the long run Fisher Effect exists. Wallace and Warner (1993) have some reservations that inflation is integrated of order one, I(1), but nonetheless proceed with the cointegration model as proposed by Johansen (1988) and Johansen and Juselius (1990). Upon establishing cointegration between the nominal interest rate and inflation, they conclude that the long run Fisher Effect holds. Crowder and Hoffman (1996) and Daniels, Nourzad, and Toutkoushian (1996) report similar findings for nominal interest rates and inflation. In particular, Crowder and Hoffman find that inflation is an I(1) moving average process, and

Daniels et al. find that inflation is I(1), based on the Dickey and Pantula (1987) test for unit roots. These two papers report that inflation Granger-causes nominal interest rates, and since cointegration exists between the series, the conclusion is that the long run Fisher Effect holds. On the other hand, King and Watson (1997) use a bivariate model projecting the nominal interest rate onto inflation and vice-versa. The model produces an estimated coefficient less than one, which indicates that the long run Fisher Effect does not hold.

In regards to the non-cointegration literature, the seminal work of Fama (1975) examines the Fisher Effect with the purpose of determining market efficiency. Fama finds that the Fisher Effect does hold. Garcia and Perron (1996) study the constancy of the ex ante real interest rate under the assumption of rational expectations in order to investigate the Fisher Effect. Non-stationarity is accounted for by allowing changes in the mean and variance through the use of Hamilton's (1989) Markov Switching model. The conclusion is that the Fisher Effect sporadically holds if infrequent breaks in the mean are permitted. Another non-cointegration technique to study the Fisher Effect is proposed by Malliaropulos (2000), which uses dynamic multipliers from a bivariate VAR with the assumption that inflation and nominal interest rates are trend-stationary once structural breaks and deterministic trends have been taken into account.

The main difficulty in determining the Fisher Effect in a cointegration framework is the low power of unit root and cointegration tests, which tend to erroneously fail to reject the null of non-stationarity too frequently. The power of these tests is further weaken if structural breaks are not taken into account or if the model is also misspecified due to erroneous number of lags lengths, etc.

In this paper, the Fisher Effect is further studied, taking into account possible misspecifications, the presence of potential structural breaks, and nonlinearities. The results of parametric and nonparametric techniques are compared, and the nonparametric time-detrended relationship between nominal interest rates and inflation is presented as an alternative tool to examine the 'traditional' Fisher Effect.

The key to investigating the Fisher Effect lies in the time series properties of nominal interest rates and inflation. This paper attempts to resolve the economic and econometric issues of testing for the Fisher Effect by first undertaking a careful investigation of the univariate properties of inflation and nominal interest rates. The findings are that these series present structural changes in the early 1980s. When these breaks are taken into account, nonstationarity tests indicate that they are trend-stationary.

In the second stage, the residuals from the detrending regressions of nominal interest rates and inflation are used to form a stationary, nonparametric, time-detrended VAR with median or averaged coefficients. This framework allows for the investigation of nonlinearities in the relationship between nominal interest rates and inflation through the study of the estimated coefficients and the dynamic multipliers of the impulse response functions. The nonparametric time-detrended test for the Fisher Effect is formed from the cumulative dynamic multiplier ratios of inflation to nominal interest rates inflation, under a shock to nominal interest rates. If this ratio statistically approaches one as the horizon goes to infinity, this signifies that, even with the removal of the time component, the changes in nominal interest rate are being matched by the changes in inflation in the long run, and hence, the Fisher Effect holds.

Two variations of the relationship between nominal interest rates and inflation are investigated in this paper: a nonparametric time-detrended Fisher effect using standardized data for the first sample period, and a nonparametric time-detrended Fisher effect using level data for the second sample period. Both techniques conclude that the Fisher effect holds. In particular, using a median, nonparametric, time-detrended VAR(3), the nonparametric time-detrended Fisher Effect statistically holds at approximately the twelfth quarter for the first sample period (1960:Q1-1995:Q3) using standardized detrended nominal interest rates and inflation. Using level detrended inflation and nominal interest rates in a nonparametric VAR(4) with averaged coefficients for the second period (1960:Q1-2004:Q2), the nonparametric time-detrended Fisher Effect is statistically achieved at approximately the fifteenth quarter.

The recent widespread use of inflation targeting rules by different countries has spurred some debate as to whether the Federal Reserve should follow a monetary policy rule or discretionary monetary policy. The question of the validity of the Fisher Effect can be applied to assess inflation targeting rules. For example, a proposed rule would be to implement tight or loose monetary policy depending on whether inflation is above or below the long run equilibrium given by the Fisher Effect. That is, the monetary policy rule could be implemented based on investigating whether or not the movements in inflation exceed the movements in nominal interest rates in the long run – changes in policy would be warranted in order to maintain the long run equilibrium between these series, as reflected in the Fisher Effect. The findings that the Fisher Effect holds, especially for the more recent period, suggest that current monetary policy can be implemented based on minor discrete changes vis-à-vis discrete changes in nominal interest rates in order to maintain the long run equilibrium between nominal interest rates and inflation.

The structure of this paper is as follows: Section 2 investigates the univariate dynamic properties of inflation and nominal interest rates. Section 3 presents the parametric and nonparametric techniques and the VAR models used to test for the Fisher Effect. The empirical results are presented in Section 4, and Section 5 concludes.

2. Univariate Analysis – Modeling Inflation and Nominal Interest Rates

A first important step before composing the VAR(*p*) models is to study the individual dynamics of each time series to be used in the system. This section investigates univariate model specifications of inflation and nominal interest rates, including nonstationarity tests in the presence of structural breaks.

The data are analyzed in annualized quarterly frequency. The three-month Treasury bill rate is used as the nominal interest rate, whereas the log of the first difference of the seasonally adjusted Consumer Price Index (CPI) is used as a measure of inflation.[2] The analysis is carried out for two sample periods: from the second quarter of 1960 to the third quarter of 1995 (142 observations), and from the second quarter of 1960 to the second quarter of 2004 (177 observations).[3]

[2] The data are obtained from the St. Louis F.R.E.D. For CPI (CPIAUCSL) and the 3-month T-Bill rates (TB3MS), monthly data are converted to quarterly data.

[3] The first quarter of 1960 is used to construct inflation.

Each series, inflation and nominal interest rates, is first tested in order to identify potential time intervals containing structural breaks. The recursive residuals test, CUSUM test, CUSUMQ test, and the recursive coefficients test are applied to an AR(1) model of the series.[4] Once an interval containing a potential structural break is identified by at least one of the previously mentioned tests, the log likelihood ratio of the Chow breakpoint test and the Chow forecast test is then applied to specific times within the indicated periods. Based upon agreement of the two Chow tests for both sample periods, a structural break is found in 1981:Q3 for inflation and in 1980:Q2 for nominal interest rates.

The optimal univariate model for each time series is chosen taking into consideration the parsimony of lag length, using the Akaike Information Criteria (AIC) and the Schwartz's Bayesian Criteria (SBC), the statistical significance of each regressor as well as the whiteness of the residuals.

The Augmented Dickey-Fuller (ADF) test is used to test for stationarity of inflation and nominal interest rates. The test indicates that inflation and nominal interest rates do not have unit roots for the first sample period (1960:Q2-1995:Q3); it fails to reject the hypothesis of nonstationarity in these series for the second sample period (1960:Q2 and 2004:Q2). A summary of the results is reported in tables 1A and 2B.

The conflicting results in the unit root tests across samples point to a possible model misspecification, which could lead the ADF test to erroneously fail to reject the null of nonstationarity. Hence, the dynamics of the series are further investigated by including an appropriate deterministic trend. Based on the findings of breakpoints in these series, Perron's test (1989) for nonstationarity against the alternative of a deterministic trend in the presence of sudden structural changes is used in the next step.

The inclusion of a time trend is just as important in Perron's test, since otherwise it could also mistakenly lead to failing to reject the null of nonstationarity. The test is estimated under the alternate hypothesis of trend stationarity in the residuals of the detrended series. For this reason, three specific types of deterministic time trends are considered as alternate hypotheses: Model A – taking into account a break in the mean (intercept); Model B – taking into account a break in the drift (slope); and Model C – taking into account a break in the mean and in the drift. Model C, which encompasses Models A and B, is:

$$x_t = a_0 + a_2 t + \mu_2 D_L + \mu_3 D_T + x_t^a \qquad (1)$$

where x_t is either nominal interest rates or inflation, t refers to the time trend, and x_t^a are the residuals of the detrended series. D_L is a dummy variable that takes a value of *0* for $t < T_B$ and a value of *1* for $t \geq T_B$, where T_B is the time of the structural break. D_T is a dummy variable multiplied by the time trend, which takes a value of *0* for $t < T_B$, and a value of the time trend, t, for any $t \geq T_B$.[5]

---

[4] All hypotheses are tested at the 5% significance level.
[5] A more complete report of the results for Perron's test can be found in Tables 2A and 2B.

Table 1A. Inflation (π_t)—Results of the ADF Test

Sample Period	***Model***	**Estimated Coefficient of** ϕ	***Estimated* t-statistic of** $\hat{\phi}$	**ADF Critical Value**	**Result Unit Root Process (Yes or No)**
1960:Q2 - 1995:Q3	Intercept Only (3 lags $\Delta\pi_t$) Removing $\Delta\pi_{t-1}$	-0.137	-3.2733	-2.882	No
1960:Q2 - 2004:Q2	Intercept Only (3 lags $\Delta\pi_t$)	-0.111	-2.758	-2.878	Yes
1960:Q2 - 1979:Q1	Trend & Intercept; Eliminating $\Delta\pi_{t-1}$ and $\Delta\pi_{t-2}$, (4 lags $\Delta\pi_t$)	-0.338	*-4.117*	-3.473	No
1984:Q4 - 1995:Q3	Intercept Only (0 lags $\Delta\pi_t$)	-0.596	*-4.167*	-2.929	No
1984:Q4 - 2004:Q2	Intercept Only (0 lags $\Delta\pi_t$)	-0.538	*-5.277*	-2.898	No

Table 1B. Nominal Interest Rates (i_t)—Results of the ADF Test

Sample Period	***Model***	**Estimated Coefficient of** ϕ	***Estimated* t-statistic of** $\hat{\phi}$	**ADF Critical Value**	**Result Unit Root Process (Yes or No)**
1960:Q2 - 1995:Q3	Intercept Only (5 lags Δi_t) Removing Δi_{t-4}	-0.083	-3.241	-2.882	No
1960:Q2 - 2004:Q2	Intercept Only (5 lags Δi_t) Removing Δi_{t-4}	-0.111	-2.815	-2.878	Yes

Table 1B. Continued

Sample Period	***Model***	**Estimated Coefficient of ϕ**	***Estimated* t-statistic of $\hat{\phi}$**	**ADF Critical Value**	**Result Unit Root Process (Yes or No)**
1960:Q2 - 1979:Q1	Trend & Intercept (3 lags Δi_t) Removing Δi_{t-2}	-0.237	-3.952	-3.471	No
1984:Q4 - 1995:Q3	Trend & Intercept (6 lags Δi_t) Removing Δi_{t-2}, Δi_{t-3}, Δi_{t-4}, Δi_{t-5}	-0.227	-4.746	-3.514	No
1984:Q4 - 2004:Q2	Trend & Intercept (6 lags Δi_t) Removing Δi_{t-2}, Δi_{t-3}, Δi_{t-4}, Δi_{t-5}	-0.159	-4.691	-3.467	No

Table 2A. Detrended Inflation (π_t^a)

Results of the Perron Test for Structural Change

Sample Period	**Time Series**	**Model**	**Break Fraction[1] λ**	**Estimated Coefficient of a_1**	**Estimated t-statistic of $\hat{a}_1$**	**Perron Critical Value**	**Result: Unit Root (Yes or No)**
1960:Q2 to 1995:Q3	π_t^a	Model A; Eliminating $\Delta\pi_{t-1}^a$ (3 lags $\Delta\pi_t^a$)	0.60	0.715	-4.775	-3.76	No

[1] The break fraction, λ, is rounded off to the tenths since only these critical values are provided in Perron (1989). For the sample period of 1960:Q2 to 1995:Q3 for inflation, λ = 85/142 =0.598, and for nominal interest rates, λ = 80/142 =0.563. For the sample period of 1960:Q2 to 1995:Q3 for inflation, λ = 85/177 =0.480, and for nominal interest rates, λ = 80/177 =0.452.

Table 2A. Continued

Sample Period	Time Series	Model	Break Fraction[2] λ	Estimated Coefficient of a_1	Estimated t-statistic of $\hat{a}_1$	Perron Critical Value	Result: Unit Root (Yes or No)
1960:Q2 to 1995:Q3	π_t^a	Model B; Eliminating $\Delta\pi_{t-1}^a$, & $\Delta\pi_{t-2}^a$ (3 lags $\Delta\pi_t^a$)	0.60	0.572	-6.381	-3.95	No
1960:Q2 to 1995:Q3	π_t^a	Model C; Eliminating $\Delta\pi_{t-1}^a$, & $\Delta\pi_{t-2}^a$ (3 lags $\Delta\pi_t^a$)	0.60	0.550	-6.622	-4.24	No
1960:Q2 to 2004:Q2	π_t^a	Model A (3 lags $\Delta\pi_t^a$)	0.50	0.853	-3.015	-3.76	Yes
1960:Q2 to 2004:Q2	π_t^a	Model B; Eliminating $\Delta\pi_{t-1}^a$, & $\Delta\pi_{t-2}^a$ (3 lags $\Delta\pi_t^a$)	0.50	0.609	-6.755	-3.96	No
1960:Q2 to 2004:Q2	π_t^a	Model C Eliminating $\Delta\pi_{t-1}^a, \Delta\pi_{t-2}^a$ (3 lags $\Delta\pi_t^a$)	0.50	0.529	-7.525	-4.24	No

[2] The break fraction, λ, is rounded off to the tenths since only these critical values are provided in Perron (1989). For the sample period of 1960:Q2 to 1995:Q3 for inflation, λ = 85/142 =0.598, and for nominal interest rates, λ = 80/142 =0.563. For the sample period of 1960:Q2 to 1995:Q3 for inflation, λ = 85/177 =0.480, and for nominal interest rates, λ = 80/177 =0.452.

Table 2B. Detrended Nominal Interest Rates (i_t^a)

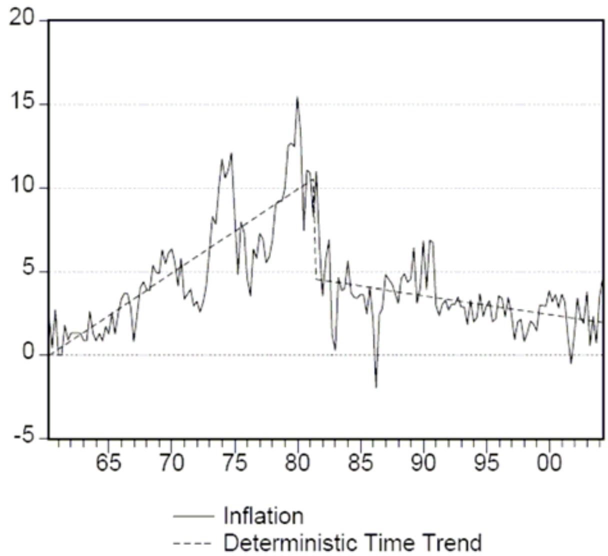
Results of the Perron Test for Structural Change

Sample Period	**Time Series**	**Model**	**Break Fraction λ**	**Estimated Coefficient of a_1**	**Estimated t-statistic of $\hat{a}_1$**	**Perron Critical Value**	**Result: Unit Root (Yes or No)**
1960:Q2 to 1995:Q3	i_t^a	Model A; Eliminating Δi_{t-4}^a (5 lags Δi_t^a)	0.50	0.908	-3.090	-3.76	Yes
1960:Q2 to 1995:Q3	i_t^a	Model B; Eliminating Δi_{t-4}^a (5 lags Δi_t^a)	0.50	0.906	-3.487	-3.95	Yes
1960:Q2 to 1995:Q3	i_t^a	Model C; Eliminating Δi_{t-4}^a (5 lags Δi_t^a)	0.50	0.691	-5.005	-4.24	No

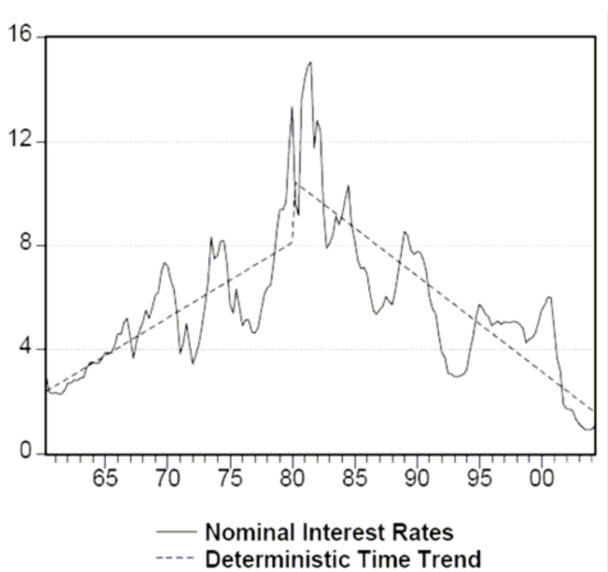
Table 2B. Continued

Sample Period	Time Series	Model	Break Fraction λ	Estimated Coefficient of a_1	Estimated t-statistic of $\hat{a}_1$	Perron Critical Value	Result: Unit Root (Yes or No)
1960:Q2 to 2004:Q2	i_t^a	Model A; Eliminating Δi_{t-4}^a 5 lags Δi_t^a)	0.50	0.917	-2.987	-3.76	Yes
1960:Q2 to 2004:Q2	i_t^a	Model B; Eliminating Δi_{t-4}^a , Δi_{t-5}^a , & Δi_{t-6}^a (7 lags Δi_t^a)	0.50	0.953	-2.139	-3.96	Yes
1960:Q2 to 2004:Q2	i_t^a	Model C Eliminating Δi_{t-4}^a (5 lags Δi_t^a)	0.50	0.775	-5.237	-4.24	No

The results of Perron's test are reported in tables 2A and 2B. The test selects Model C as the best specification for both series and for both sample periods. This result is illustrated in Graphs 1A and 1B. Since the residuals of the detrended regressions of inflation and nominal interest rates, x_t^a, are found to be trend stationary, these detrended residuals – which take into account the structural breaks found in the individual series – will be used in the VAR in the next section.



Graph 1A. Inflation. 1960:Q2 to 2004:Q2.



Graph 1B. Nominal Interest Rates. 1960:Q2 to 2004:Q2.

3. Parametric and Nonparametric VAR Models

The methodology followed in this paper to test for the Fisher Effect consists of first determining the best model specification for the univariate series. Since trend stationarity has been determined for both series, the next step is to set up a bivariate VAR of inflation and nominal interest rates. The residuals from the detrending regression, x_t^a, which include a break in the mean and in the drift as specified in Equation (1), are used in the VAR, and the analysis of the Fisher Effect is carried out using parametric and nonparametric techniques.

The cumulative sums of the orthogonalized dynamic multipliers are used to study the long run behavior of the VAR. These sums capture the long run effect of a shock in the variable of the system. The time-detrended Fisher Effect can be investigated using the ratio of the sum of the orthogonalized dynamic multipliers of the responses of inflation to the sum of the responses of nominal interest rates due to a shock in nominal interest rates. If this ratio converges to unity, the Fisher Effect holds, which indicates that the changes in detrended nominal interest rates and detrended inflation are in synchronization in the long run. This section presents the theoretical VAR models used to examine the Fisher Effect.

3.1. The Parametric Model

The Fisher Effect essentially relates the long run movement of nominal interest rates and inflation. A linear, stationary, bivariate VAR(p) framework can be used to study the interaction between the variables in the system, without the complications that arise from estimating and testing a non-stationary system.

The impulse response functions use the history of the system in order to capture the average behavior of a variable to an isolated shock in the system. Orthogonalizing the impulse response functions results in dynamic multipliers that are not history or shock-dependent when the size of the shock is standardized to one standard deviation.

Hence, a one-standard-deviation shock has no direct impact on the dynamic multipliers. Furthermore, the orthogonalization of the impulse response functions removes the composition effect by not including the impact of the shock on the other variables in the VAR, when the covariances are different from zero.

In forming the VAR(p) to be used to test the Fisher Effect, the lag p and the order of causality are determined based on statistical tests and economic theory. Inflation is found to follow nominal interest rates, based on the persistence of inflation and on the Granger causality test. This test finds that nominal interest rates uni-directionally causes inflation for both sample periods. The optimal lag length of the VAR(p) is determined using the AIC, SBC, and the stationarity of the VAR(p). The structural breaks of inflation and nominal interest rates are taken into account indirectly in the VAR(p) through the residuals of the detrending regressions obtained from Equation (1), which are stationary.

Concerning the variables in the system, the regressand matrix is denoted as Y, which is an $\left(\left(n \times T\right) \times 1\right)$ matrix, and with each iteration of the VAR being denoted as $y_t = \left(y_{1t} \quad \cdots \quad y_{rt} \quad \cdots \quad y_{n-1t} \quad y_{nt}\right)'$ where T equals the total number of observations, n is the total number of equations in each iteration of the VAR with r representing the r^{th}

regressand of the VAR with $r = 1,\ldots, n$. For this paper, y_{rt} represents the regressand, which is either detrended nominal interest rates $\left(i_t^a\right)$ or detrended inflation $\left(\pi_t^a\right)$. The set of regressors is denoted as $X = \begin{pmatrix} Z & X_1 & \cdots & X_m & \cdots & X_{k-1} & X_k \end{pmatrix}$ which is an $\left(\left(n \times T\right) \times q\right)$ matrix with $q = (k + 1)$.[1] Z is a scalar column matrix with the dimensions of $\left(\left(n \times T\right) \times 1\right)$, and X_m is a column matrix with the dimensions of $\left(\left(n \times T\right) \times 1\right)$ with m = 1,…, k.[2] In regards to the stability of the system, this can be verified by converting the VAR(p) to a VAR(1), which is weakly-stationary, if the modulus of each eigenvalue of the system is less than one, i.e., $\left|\lambda_j\right| < 1$.

The parametric VAR(p) model is represented by:

$$\Phi(L)\, y_t = u_t \tag{2}$$

where y_t is a vector $y_t = \begin{pmatrix} i_t^a & \pi_t^a \end{pmatrix}'$ containing the detrended series of nominal interest rates and inflation, $\Phi(L)$ is a $(n \times p)$ matrix of the p^{th} order polynomials in the lag operator, and $u_t = \begin{pmatrix} u_{it} & u_{\pi t} \end{pmatrix}'$ is the residual vector which are $u_t \sim N(0, \Sigma)$.

In order to test for the existence of the time-detrended Fisher Effect, the coefficients of the VAR(p) are used to form its MA(∞) representation, which are orthogonalized through the application of the Choleski's decomposition. Specifically,

$$y_t = M(L)w_t \tag{3}$$

where $M(L) \equiv \Psi(L)P \equiv \Phi(L)^{-1}P$, P is the lower triangular Choleski matrix that satisfies $\Sigma = PP'$ and $w_t = (w_{\pi t}\ w_{it})' = P^{-1}u_t$, with $E(w_t) = 0$, $Var(w_t) = I$. The orthogonalized MA(∞) coefficients are used to form the dynamic multiplier ratios, which will be discussed in more detail in Section 3.3. Concerning the implementation of the theoretical model, if a parametric VAR(p) has the same regressors in each equation and the error terms, u_t, are uncorrelated with the set of regressors, then the VAR can be consistently estimated by n ordinary least squares (OLS) equations.

3.2. The Nonparametric Model

The nonparametric model is an empirical kernel density, which is a weighted smoothing function of the data. The nonparametric model presented in this paper is estimated using the local linear least squares nonparametric method (LLLS). LLLS is a kernel weighted least squares regression model, which locally fits a straight line for each data point, with more weight given to observations close to the data point and less weight to observations farther away.

---

[1] The set of regressors can either be lag operators of y_t or exogenous variables.
[2] When using standardized data, the scalar matrix is not included.

As opposed to other nonparametric methodologies, such as using the Nadaraya-Watson estimator, the LLLS is better able to utilize the information in the tail regions. Hence, we have a more efficient use of information, particularly when compared to a linear parametric model. In this instance, LLLS is able to generally produce on average larger coefficients than the parametric VAR(*p*). Since the nonparametric model does not specify a functional form of the model, the LLLS is capable of exploiting the non-linearity present in the model to produce large stationary coefficients, which are important in testing for the Fisher Effect. Another benefit of using LLLS is that the coefficients can be analyzed using standard regression methods such as ordinary least squares by using the average or median coefficients.

As in the parametric VAR(*p*), the nonparametric VAR can be estimated equation-by-equation provided that each equation has the same set of regressors, and the error terms of each equation are not correlated with the set of regressors.

The implementation of the nonparametric VAR(*p*) requires an equation-by-equation methodology. The Gaussian kernel measures the distance between each observation in each regressor to the j^{th} element of each regressor with a weight assigned to each observation with the intent that more weight is given to observations closer to the j^{th} element and progressively less weight to observations farther away. The form of the Gaussian kernel is as follows:

$$K = \sum_{j=1}^{T} K\left(\psi_j\right), \tag{4}$$

where

$$K\left(\psi_j\right) = \frac{1}{\left(2\pi\right)^{\frac{k}{2}}} exp\left(-\frac{1}{2}\left(\left(\frac{x_1 - x_{1j}}{h_1}\right)^2 + \cdots \left(\frac{x_m - x_{mj}}{h_m}\right)^2 + \cdots + \left(\frac{x_k - x_{kj}}{h_k}\right)^2\right)\right)$$

with $$\psi_j = \left(\frac{x_1 - x_{1j}}{h_1} \quad \cdots \quad \frac{x_m - x_{mj}}{h_m} \quad \cdots \quad \frac{x_k - x_{kj}}{h_p}\right).$$

In order to prevent over-smoothing, which results in a loss of information, or under-smoothing, which causes too much ‘noise’ in the empirical density, the standard window width, h_m, for $m = 1 \ldots, k$, is used:

$$h_m = C\sigma_m\left(T\right)^{-\left[\left(k+1\right)+4\right]} \tag{5}$$

where k equals the number of regressors with the exclusion of the constant term in the model, C is a constant term that depends of the type of data, and σ_m is the standard deviation of X_m. It should be noted that the optimal window width used in Equation (5) is obtained from the mean integrated squared approach (MISE). Pertaining to this paper, the method of cross validation is used to obtain the optimal window width, which is further discussed in the empirical portion of this paper, which is Section 4 (Pagan and Ullah, 1999).

In matrix notation, the (*k x 1*) vector of nonparametric coefficients for the r^{th} regressand and t^{th} iteration of the VAR(p) is denoted as

$$\beta_{rt} = \left(X'KX\right)^{-1} X'Ky_{rt} \tag{6}$$

Using Equation (6) to calculate all T iterations for each r^{th} regressand, the compilation of the VAR(p) results in a $\left(T \times k\right)$ matrix of nonparametric coefficients.

Once the nonparametric coefficients for the r^{th} regressand is obtained, the VAR(p) can be re-written as a linear combination of the coefficients and regressors since the LLLS nonparametric method fits a line within the window width:

$$y_{rt} = X\beta_{rt} + u_{rt}, \tag{7}$$

where the regressand y_{rt} and the residual u_{rt} are scalars, with $u_{rt} \sim \left(0, \sigma_{rt}^2\right)$.

The median or average of each column is used to form the aggregated nonparametric version of vector $\Phi(L)$, an $(n \times p)$ matrix of the p^{th} order polynomials in the lag operator. The aggregate nonparametric coefficients could consist of either the mean or median measures. The nonparametric orthogonalized MA(∞) representation of the VAR(p) is best described by Equation (3) is used to form the dynamic multipliers used to test for the Fisher Effect. This is presented in the next section.

3.3. Testing for the Fisher Effect

The time-detrended Fisher Effect holds statistically if the cumulative orthogonalized dynamic multiplier ratios converge to one as g – the lag length of the impulse response function – goes to infinity. The orthogonalized MA(∞) coefficients are used to form the nonparametric conditional orthogonalized impulse response functions. For instance, the form of the orthogonalized dynamic multiplier of the $\left(t+s\right)$ response of the $(r')^{th}$ regressand caused by a shock to the r^{th} regressand at time t, are of the form:

$$\frac{d\left(y_{r't+s}\right)}{dw\left(x_t\right)_r} = M_{r'r}. \tag{8}$$

The g^{th} cumulative dynamic multiplier ratio is referred to as Γ_g where Γ_g is the sum of the responses of detrended inflation caused by a shock to detrended nominal interest rates $\left(M_{\pi i}\right)$ to the sum of the responses of detrended nominal interest rates caused by a shock to detrended nominal interest rates $\left(M_{ii}\right)$. Specifically for this paper, for each sample period, the g^{th} ratio of nonparametric orthogonalized cumulative dynamic multipliers of the $(r')^{th}$ regressand, which is denoted as Γ_g, is of the general form of

$$\Gamma_g = \frac{\sum_{s=0}^{g} M_{\pi i,s}}{\sum_{s=0}^{g} M_{ii,s}} , \tag{9}$$

where $g = 1, 2, \ldots, \infty$.

For both the parametric and nonparametric models, up to one hundred lag lengths of the impulse response functions are calculated in order to determine whether the dynamic multiplier ratios converge to unity. This is needed since potential non-stationarity can mistakenly indicate an early appearance of the Fisher Effect in the medium run, which may break down in the very long run. The convergence to unity signifies that the changes in detrended nominal interest rate are being matched by the changes in detrended inflation, which indicates that the Fisher Effect holds.

Three different methodologies to compute the orthogonal dynamic multipliers are presented in order to test for the Fisher Effect. One is based on the parametric VAR and the orthogonal dynamic multiplier ratios described in Equation (9). The other two methods are based on the nonparametric VAR. Methods 2 and 3 can be computed using the median nonparametric and the average nonparametric coefficients. The three techniques are as follows:

METHOD 1: The parametric VAR of detrended nominal interest rates and inflation is estimated equation-by-equation, in order to obtain the sum of the orthogonalized dynamic multipliers of the responses of inflation divided by the responses of nominal interest rates to a shock in nominal interest rates.

METHOD 2: The orthogonal dynamic multiplier ratios of the Fisher Effect are obtained from the nonparametric estimation of the VAR using the median of the T' estimated coefficients for each regressor as a measure of central tendency where T' denotes the total number of observations once the lags of the VAR are taken into account. The median coefficients are used to form the MA(∞) coefficients, which are then used to form the dynamic multipliers. The error terms from the T' nonparametric equation estimation of the VAR are used to obtain the (2×2) unconditional variance-covariance matrix of the error terms, which is calculated by using the residual sum of squares. The Choleski decomposition can then be calculated in order to form the orthogonalized impulse responses – the orthogonal dynamic multipliers.

METHOD 3: The orthogonal dynamic multiplier ratios of the Fisher Effect are obtained from the nonparametric estimation of the VAR. The median of T' estimated coefficients for each regressor is obtained. The median nonparametric coefficients are used to form the MA(∞) coefficients that comprises the dynamic multiplier ratios of the Fisher Effect. The error terms of each equation in the VAR are then obtained from the regression:

$$\varepsilon = Y - X\beta_{np\_med} , \tag{10}$$

which is then used to obtain the (2×2) variance-covariance matrix of the error terms. Once the variance-covariance matrix of the error terms is obtained the Choleski decomposition can then be calculated to form orthogonal dynamic multipliers.

The average, orthogonal, nonparametric dynamic multiplier ratios of the time-detrended Fisher Effect are obtained by replacing the median nonparametric coefficients in the formation of the MA(∞) version of the VAR(p) with the average nonparametric coefficients. If the orthogonal dynamic multiplier ratios of the Fisher Effect converge to unity and are within the 95% bootstrapped confidence band, the Fisher Effect statistically holds. For each test of the Fisher Effect, the bootstrapped confidence band is constructed from the empirical density based on five thousand iterations of re-sampling with replacement. For each run, the VAR is estimated and the test for the Fisher Effect is formed using the bootstrapped data with the average used to construct the confidence band.

By definition, the Fisher Effect means that the real interest rate is constant in the long run and is not impacted by either the short-term movements of nominal interest rates or inflation. Hence, if the Fisher Effect holds, the long run dynamics as represented by the impulse response functions, the convergence movement of nominal interest rates should match the convergent movement of inflation. Intuitively, that is:

$$r_{t+s} = i_{t+s} - \pi_{t+s} = \delta \tag{11}$$

$$\frac{d\sum_{s=0}^{g} \pi_{t+s}^{a}}{dw_{it}} = \frac{d\sum_{s=0}^{g} i_{t+s}^{a}}{dw_{it}} \tag{12}$$

$$\left(\frac{d\sum_{s=0}^{g} \pi_{t+s}^{a}}{dw_{it}} \right) \Bigg/ \left(\frac{d\sum_{s=0}^{g} i_{t+s}^{a}}{dw_{it}} \right) = 1 \tag{13}$$

where δ is some constant. For monetary policy purposes, examining Equation (13) can be an informative tool for inflation-targeting regimes, since an inequality indicates that inflation and nominal interest rates are in disequilibria. If the orthogonal dynamic multiplier ratios are greater than unity, the cumulative responses of inflation are greater than the cumulative responses of nominal interest rates, which would act as a signal to the monetary authorities that anti-inflationary measures in the form of tight monetary policies might be needed. This is particularly the case if the ratio exceeds unity by a preset amount. On the other hand, if the orthogonal dynamic multiplier ratios are less than unity, then the monetary authorities might consider implementing loose monetary policies that could bring inflation and nominal interest rates back into synchronization, given that the Fisher Effect holds in the long run.

4. Empirical Results

4.1. First Sample Period

The first sample period is from the second quarter of 1960 to the third quarter of 1995. In order to show the advantages of the nonparametric model, the parametric model is used as a benchmark. The data used in the estimated parametric and nonparametric VAR(p) models are the standardized detrended residuals of nominal interest rates and inflation generated from the estimation of Equation (1).

In choosing the lag length of the VAR(p), the estimated parametric VAR(3) has a higher AIC but a lower SBC, when compared to the parametric VAR(4). Due to the conflicting results of the AIC and SBC, various lag lengths of the VAR(p) were tested in an attempt to obtain a more parsimonious model. The optimal estimated model is selected based upon the stationarity of the VAR, as indicated by the moduli of the eigenvalues (whether they are less than unity). Even though the detrended series of inflation and nominal interest rates are stationary, the VAR(p) is not necessarily stationary as indicated by the moduli of the eigenvalues of the VAR. This is due to the dynamic interaction of the two series in the VAR(p) (tables 6A and 6B).

Table 3A. Sample Period—1960:Q2 to 1995:Q3

Parametric and Median Nonparametric Estimated Coefficients of a VAR(3)[3]

Estimated Coefficients	Parametric Method 1		Nonparametric Method 2		Nonparametric Method 3	
	EQ 1 of VAR	EQ 2 of VAR	EQ 1 of VAR	EQ 2 of VAR	EQ 1 of VAR	EQ 2 of VAR
i_{t-1}^{a}	1.1377 (0.0847)	0.6940 (0.1123)	1.3139 (0.0029)	0.7644 (0.0034)	1.3139 (0.0067)	0.7644 (0.0101)
i_{t-2}^{a}	-0.6132 (0.1196)	-0.4627 (0.1587)	-0.3561 (0.0057)	-0.3146 (0.0069)	-0.3561 (0.0133)	-0.3146 (0.0202)
i_{t-3}^{a}	0.2834 (0.0951)	0.0785 (0.1262)	-0.1661 (0.0036)	-0.0625 (0.0043)	-0.1661 (0.0084)	-0.0625 (0.0127)
π_{t-1}^{a}	-0.0956 (0.0651)	0.2972 (0.0864)	0.0796 (0.0017)	0.3209 (0.0020)	0.0796 (0.0039)	0.3209 (0.0060)
π_{t-2}^{a}	0.2609 (0.0675)	0.0085 (0.0896)	0.07523 (0.0018)	0.0558 (0.0022)	0.07523 (0.0042)	0.0558 (0.0064)
π_{t-3}^{a}	-0.1617 (0.0657)	0.1494 (0.0871)	-0.0505 (0.0017)	0.1779 (0.0021)	-0.0505 (0.0040)	0.1779 (0.0061)

[3] For both sample periods, the nonparametric estimated coefficients of the VAR(p) using Method 2 and Method 3 are the same, but the methods of obtaining the estimated standard deviations are different and are discussed in Section 3.3.

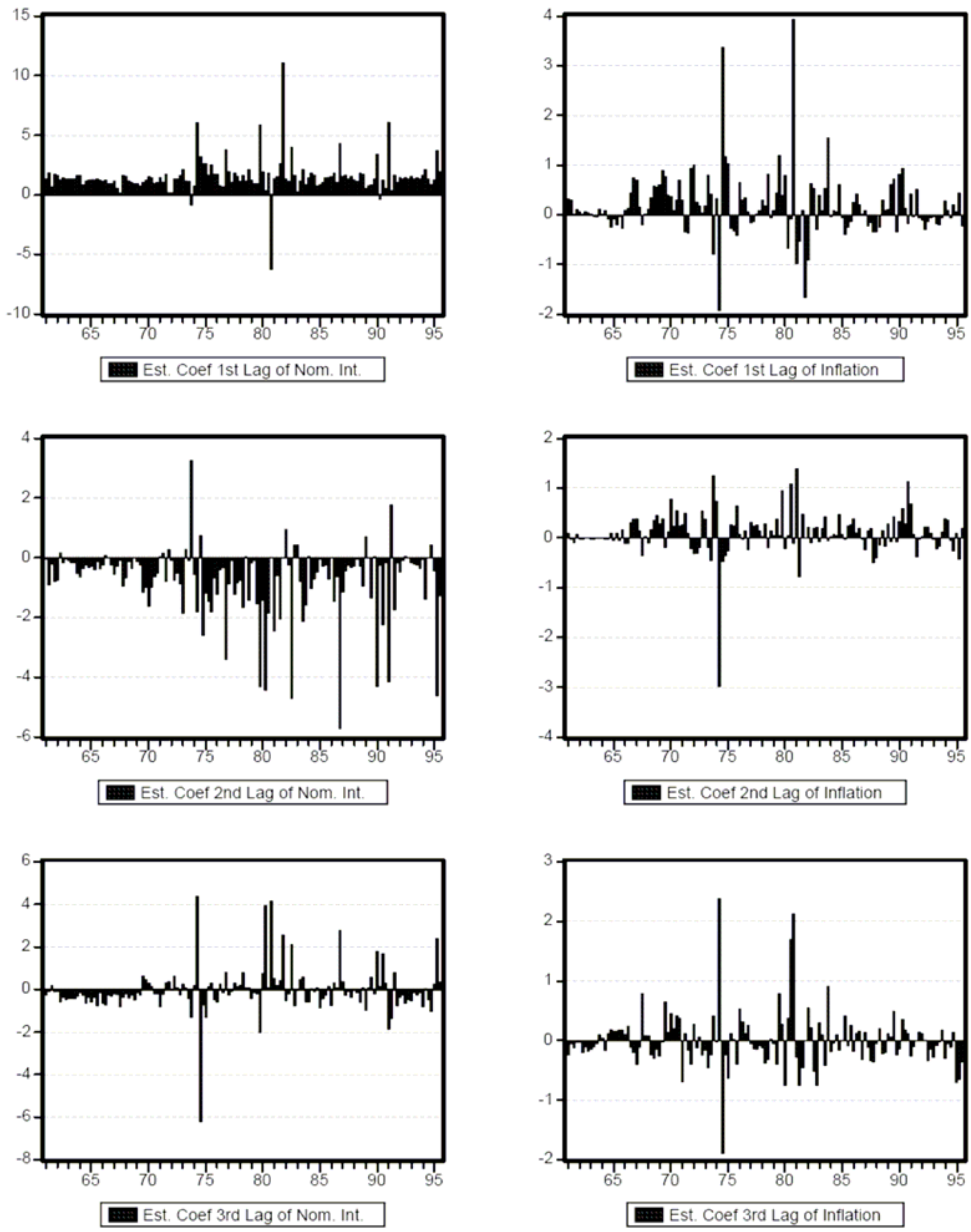
Table 3B. Sample Period—1960:Q2 to 2004:Q3

Parametric and Average Nonparametric, Estimated Coefficients of a VAR(4)

Estimated Coefficients	**Parametric Method 1**		**Nonparametric Method 2**		**Nonparametric Method 3**	
	EQ 1 of VAR	EQ 2 of VAR	EQ 1 of VAR	EQ 2 of VAR	EQ 1 of VAR	EQ 2 of VAR
i_{t-1}^{a}	1.2377 (0.0771)	0.8517 (0.1357)	1.3051 (0.0108)	0.8192 (0.0014)	1.3051 (0.0043)	0.8192 (0.0108)
i_{t-2}^{a}	-0.6849 (0.1179)	-0.7110 (0.2073)	-0.4290 (0.0252)	-0.5161 (0.0033)	-0.4290 (0.0101)	-0.5161 (0.0251)
i_{t-3}^{a}	0.5050 (0.1201)	0.4318 (0.2113)	0.2711 (0.0262)	0.1981 (0.0035)	0.2711 (0.0105)	0.1981 (0.0261)
i_{t-4}^{a}	-0.1990 (0.0838)	-0.4247 (0.1474)	-0.1605 (0.0127)	-0.0895 (0.0017)	-0.1605 (0.0051)	-0.0895 (0.0127)
π_{t-1}^{a}	-0.0701 (0.0429)	0.3814 (0.0754)	0.1179 (0.0033)	0.4008 (0.0004)	0.1179 (0.0013)	0.4008 (0.0033)
π_{t-2}^{a}	0.1461 (0.0446)	-0.0554 (0.0784)	0.0199 (0.0036)	-0.0661 (0.0005)	0.0199 (0.0015)	-0.0661 (0.0036)
π_{t-3}^{a}	-0.0994 (0.0455)	0.2864 (0.0801)	-0.0446 (0.0038)	0.3062 (0.0005)	-0.0446 (0.0015)	0.3062 (0.0037)
π_{t-4}^{a}	-0.0169 (0.0437)	-0.1041 (0.0770)	-0.1047 (0.0035)	-0.0902 (0.0005)	-0.1047 (0.0014)	-0.0902 (0.0035)

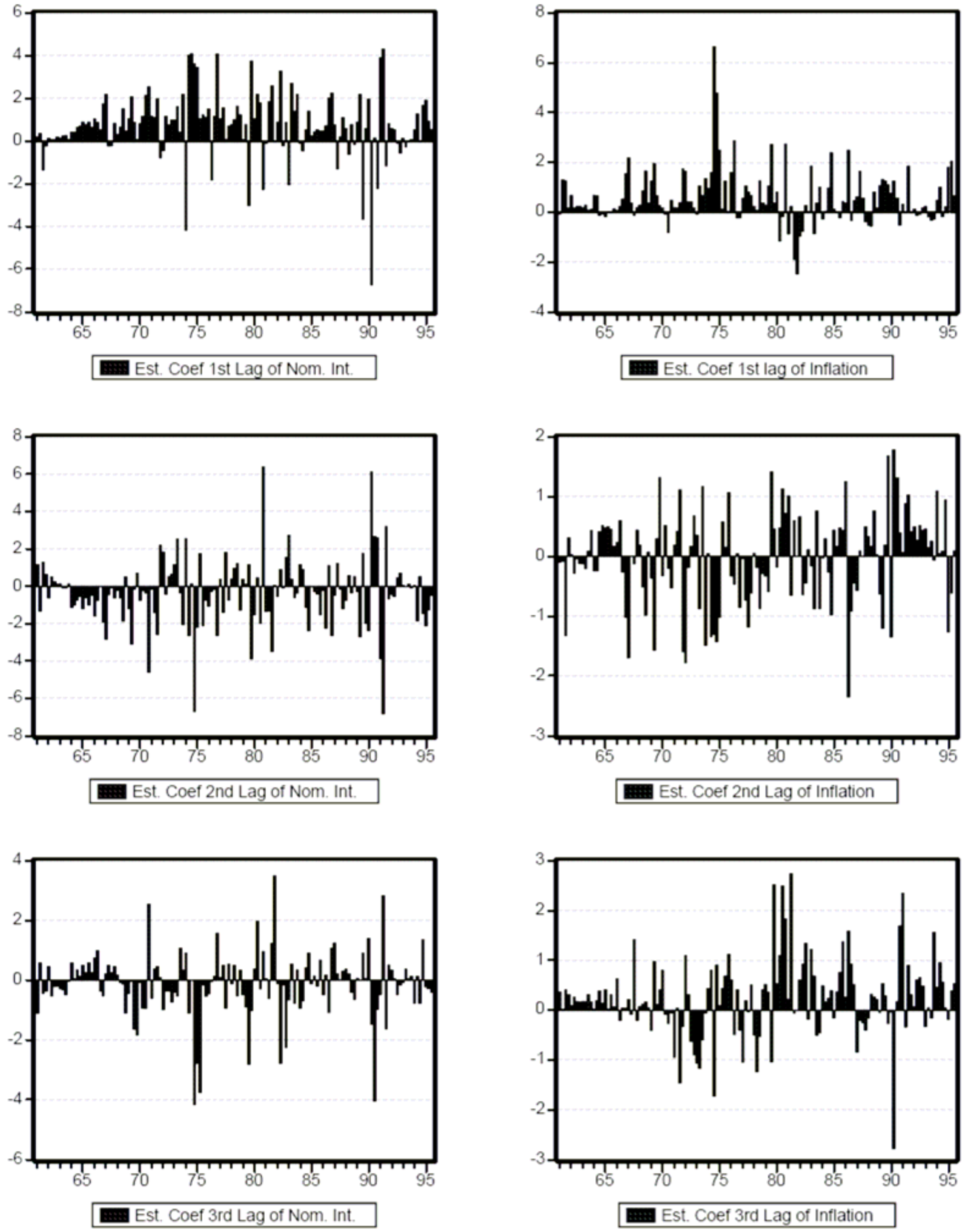
Concerning the estimation of the nonparametric models, the choice of window width is critical since too large of a window width over-smoothes the data, and too small of a window width causes the data to become very erratic. For this paper, the choice of window width is based upon cross validation of the error terms. More specifically, the window width, which produces the smallest sum of squared errors, as calculated by Methods 2 and 3, is chosen as the optimal window width as opposed to using the standard window width as provided by Equation (5). The optimal window with for the VAR(3) with median nonparametric coefficients is 0.40 while the optimal window with for the VAR(4) with median nonparametric coefficients is 0.48.

The nonparametric VAR(3) and VAR(4) models, which are able to produce the dynamic multiplier ratios of the Fisher Effect, have the characteristics that the sum of the AR coefficients and the modulus of the eigenvalue of each variable of the system are not very close to unity (i.e., not equal to 0.99 or exceed unity). As shown in table 3A, the majority of the estimated nonparametric coefficients is larger than their parametric counterparts and the non-linearity can be seen in Graph 2A.[4] The moduli of the eigenvalues are, on average, larger than their analogous parametric versions (table 4A).



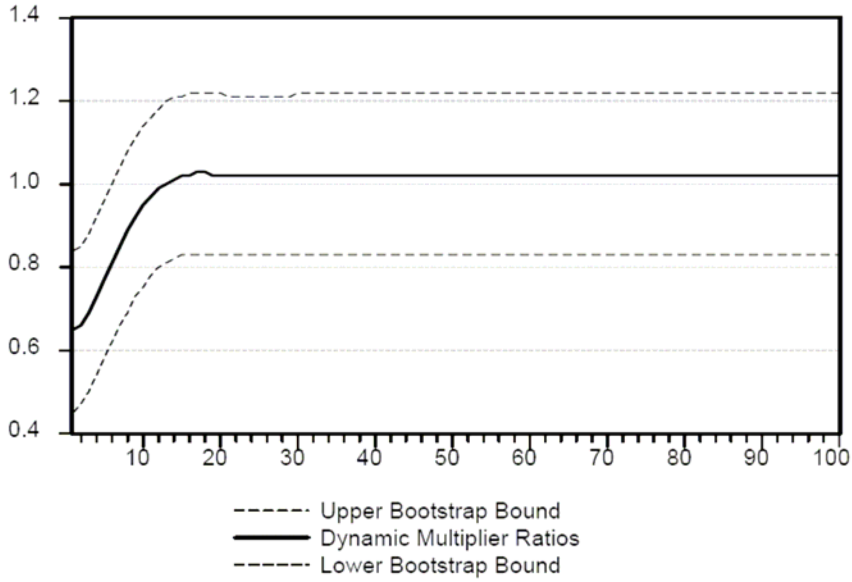
Graph 2A. Graphs of Estimated Nonparametric Coefficients with Detrended Infaltion as the Regressand Sample Period: 1960:Q2 to 1995:Q3.

[4] The intent of Graphs 2A and 2B are to demonstrate the non-linearity present in both sample periods. Hence for demonstration purposes, four nonparametric coefficients greater than |9| were removed in order to make the scaling of the graphs more homogeneous in Graphs 2A and 2B.

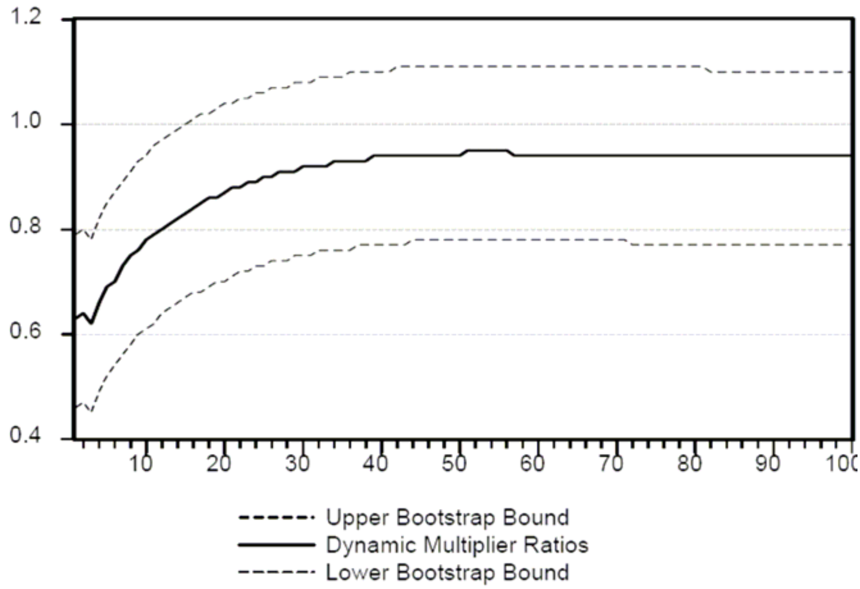


Graph 2B. Graphs of Estimated Nonparametric Coefficients with Detrended Nominal Interest Rates as the Regressand Sample Period: 1960:Q2 to 1995:Q3.

As illustrated in Graph 3A, the Fisher Effect holds in the case of the nonparametric VAR models, since they are able to exploit the non-linearity in the system, which produces larger estimated coefficients while maintaining stationarity (i.e., they do not exceed unity).



Graph 3A. Graph of the Fisher Effect for the First Sample Period.



Graph 3B. Graph of the Fisher Effect for the Second Sample Period.

In addition, the estimated nonparametric VAR(3) models generally produce larger unconditional variance-covariance matrices than the parametric versions, which are important for the estimation of the Choleski decompositions. The Choleski decomposition and its corresponding variance-covariance matrix of the optimal nonparametric VAR(3) model used to orthogonalize the dynamic multipliers are larger than the parametric counterpart using either Method 2 or Method 3 (table 5A).

Table 4A. Sample Period—1960:Q2 to 1995:Q3

Eigenvalues of Parametric VAR(3)	Modulus	Eigenvalues of Median Nonparametric VAR(3)	Modulus
-0.1974+0.6373i	0.6671	-0.2331+0.4848i	0.5379
-0.1974-0.6373i	0.6671	-0.2331-0.4848i	0.5379
0.1727+0.4422i	0.4747	0.7583+0.2575i	0.6237
0.1727-0.4422i	0.4747	0.7583-0.2575i	0.6237
0.7925	0.7925	0.8038	0.8038
0.6918	0.6918	-0.2193	0.2193

Table 4B. Sample Period—1960:Q2 to 2004:Q2

Eigenvalues of Parametric VAR(4)	Modulus	Eigenvalues of Median Nonparametric VAR(4)	Modulus
-0.3428+0.6761i	0.7580	-0.2820+0.6519i	0.7103
-0.3428-0.6761i	0.7580	-0.2820-0.6519i	0.7103
-0.0330+0.6432i	0.6440	-0.1363+0.4234i	0.4448
-0.0330-0.6432i	0.6440	-0.1363-0.4234i	0.4448
0.7117+0.2555i	0.7562	0.9393+0.0727i	0.9421
0.7117-0.2555i	0.7562	0.9393-0.0727i	0.9421
0.8275	0.8275	0.5609	0.5609
0.1199	0.1199	0.1029	0.1029

Table 5A. Sample Period—1960:Q2 to 1995:Q3

	Parametric Method 1		Nonparametric Method 2		Nonparametric Method 3	
Variance-Covariance Matrix	0.2992	0.1007	0.5554	-0.0701	0.4785	0.1507
	0.1007	0.5263	-0.0701	0.8448	0.1507	0.5731
Choleski Decomposition Matrix	0.5468	0.0000	0.7452	0.0000	0.6917	0.0000
	0.1841	0.7017	-0.0941	0.9143	0.2179	0.7250

Table 5B. Sample Period—1960:Q2 to 2004:Q2

	Parametric Method 1		Nonparametric Method 2		Nonparametric Method 3	
Variance-Covariance Matrix	0.5482	0.1613	1.9859	0.0054	0.8007	0.2416
	0.1613	1.6956	0.0054	0.2621	0.2416	1.9826
Choleski Decomposition Matrix	0.7405	0.0000	1.4092	0.0000	0.8948	0.0000
	0.2177	1.2838	0.0038	0.5119	0.2700	1.3819

In the case on nonparametrics, using both Methods 2 and 3, the VAR(3) model statistically produce the Fisher Effect when using standardized data with median nonparametric coefficients while the counterpart parametric model is unable to produce the Fisher Effect.[5] As shown in table 6A, the dynamic multiplier ratios of the nonparametric VAR(3) using standardized data and median nonparametric coefficients statistically converge to 0.97 at approximately the ninth quarter using Method 2. Using Method 3, the dynamic multiplier ratios statistically converge to 1.02 using Method 3 at approximately the sixth quarter, and the dynamic multiplier ratios converge to 1.0 at approximately the twelfth quarter as demonstrated in Graph 3A and table 6A.

Table 6A. Sample Period—1960:Q2 to 1995:Q3

Parametric/ Nonparametric Coefficients	No. of Lags	Window Width	Stationary (Yes/No)	Convergence of Γ_k	Existence of Fisher Effect (Yes/No)
Parametric --Level Data	3	N/A	Yes	0.83	No
Parametric --Standardized Data	3	N/A	Yes	0.77	No
Parametric --Level Data	4	N/A	Yes	1.57	No
Parametric --Standardized Data	4	N/A	Yes	0.57	No
Nonparametric --Level Data (Median)	3	0.78	Yes	Method 2 → -3.64 Method 3 → -0.36	No
Nonparametric --Level Data (Mean)	3	0.73	Yes	Method 2 → 0.33 Method 3 → 0.48	No
Nonparametric -- Standardized Data (Median)	3	0.40	Yes	Method 2 → 0.97 Method 3 → 1.02	Yes
Nonparametric -- Standardized Data (Mean)	3	0.43	Yes	Method 2 → 0.71 Method 3 → 0.74	No
Nonparametric --Level Data (Median)	4	0.90	Yes	Method 2 → 0.69 Method 3 → 0.69	No
Nonparametric --Level Data (Mean)	4	0.92	Yes	Method 2 → 0.40 Method 3 →0.40	No

[5] For the first sample period, using Methods 2 and 3 and a nonparametric VAR(4) using standardized data with either mean or median nonparametric coefficients, the test for the Fisher Effect converges to 1.05. Regarding the parametric counterpart model, the test for the Fisher Effect converges 0.57. Hence the Fisher Effect is statistically achieved in the nonparametric VAR(4) models and not in the parametric VAR(4) model.

Table 6A. Continued

Parametric/ Nonparametric Coefficients	No. of Lags	Window Width	Stationary (Yes/No)	Convergence of Γ_k	Existence of Fisher Effect (Yes/No)
Nonparametric -- Standardized Data (Median)	4	0.48	Yes	Method 2 → 1.01 Method 3 →0.93	Yes
Nonparametric -- Standardized Data (Mean)	4	0.52	Yes	Method 2 → 1.05 Method 3 → 1.05	Yes

Table 6B. Sample Period—1960:Q2 to 2004:Q2

Parametric/ Nonparametric Coefficients	No. of Lags	Window Width	Stationary (Yes/No)	Convergence of Γ_k	Existence of Fisher Effect (Yes/No)
Parametric --Level Data	3	N/A	Yes	0.58	No
Parametric --Standardized Data	3	N/A	Yes	0.51	No
Parametric --Level Data	4	N/A	Yes	1.35	No
Parametric --Standardized Data	4	N/A	Yes	0.48	No
Nonparametric --Level Data (Median)	3	0.95	Yes	Method 2 →0.54 Method 3 → 0.56	No
Nonparametric --Level Data (Mean)	3	0.90	Yes	Method 2 → 0.17 Method 3 → 0.21	No
Nonparametric -- Standardized Data (Median)	3	0.46	Yes	Method 2 → 0.07 Method 3 → 0.10	No
Nonparametric -- Standardized Data (Mean)	3	0.53	Yes	Method 2 →0.34 Method 3 → 0.33	No
Nonparametric --Level Data (Median)	4	0.87	Yes	Method 2 → 0.94 Method 3 → 0.95	Yes
Nonparametric --Level Data (Mean)	4	0.99	Yes	Method 2 → 0.83 Method 3 → 0.67	No
Nonparametric -- Standardized Data (Median)	4	0.45	Yes	Method 2 → 1.09 Method 3 → 1.08	Yes
Nonparametric -- Standardized Data (Mean)	4	0.54	Yes	Method 2 → 0.52 Method 3 → 0.51	No

While the Fisher Effect statistically holds for Methods 2 and 3, Method 3 is better representative of the Fisher Effect due to the smaller variances. The smaller variances produce a smaller bootstrap confidence band, which leads to more reliable results of the test for the Fisher Effect. Thus, the nonparametric VAR(3), using standardized detrended data and median coefficients, from this point forward will be referred to as the optimal nonparametric VAR(3).

The optimal nonparametric VAR(3) produces orthogonalized dynamic multiplier ratios that statistically converge to 1.02 at approximately the twelfth quarter, while the parametric VAR(3) using standardized data produces orthogonalized dynamic multiplier ratios that converge to 0.77. Hence, the time-detrended Fisher Effect holds in the very long run, which can be explained by the sluggish nature of inflation as well as by the long lag monetary policy takes to affect inflation.

4.2. Second Sample Period

As in the case of the first sample period, the results of the nonparametric model are compared to the corresponding parametric model. The data used in the parametric and nonparametric VAR(*p*) models for the second time period (from the second quarter of 1960 to the second quarter of 2004) are the level detrended residuals of nominal interest rates and inflation as generated by the estimated detrending regression of Equation (1).

Various lag lengths are tested which produced conflicting AIC and SBC results in the second sample period as is the case in the first sample period. Even with the inclusion of trend-stationary data in the VAR(*p*) and changing the lag length, the dynamics of the VAR(*p*) cause the system to become non-stationary or very close to non-stationary in the second sample. Thus, the VAR(*p*) produces non-convergent equilibria for detrended nominal interest rates and inflation.

The level nonparametric VAR(4) model, in which the Fisher Effect statistically holds, is able to exploit the non-linearity in the system to generate larger estimated coefficients whose sum does not exceed unity (Graph 2B and table 3B). Furthermore, the VAR(4) produces moduli of the eigenvalues that are large but do not exceed unity (table 4B). Concerning, the window width, which produces the smallest sum of squared errors, as calculated by Methods 2 and 3 for the VAR(4) with median nonparametric coefficients is 0.45 while the optimal window with for the VAR(4) with mean nonparametric coefficients is 0.54.

In regards to the variance-covariance matrices, the estimated nonparametric VAR(4) produces larger variance-covariance matrices than the estimated parametric counterpart, which also enables the system to achieve the Fisher Effect through the use of Methods 2 or 3. By using Method 2, the dynamic multiplier ratios statistically converges to 0.94 at approximately the fifteenth quarter. Method 2 also produces a smaller estimated variance-covariance matrix, which produces a smaller bootstrap confidence band thereby making the results of the test for the Fisher Effect more reliable when compared to the results obtained by using Method (Graph 3B and table 5B). Alternatively, Method 3 produces dynamic multiplier ratios that statistically converge to 0.95 at approximately the twenty-fifth quarter (table 6B).

Hence the nonparametric VAR(4) using level data and median nonparametric coefficients is the optimal model that produces the Fisher Effect.[6]

The difference in results between the first and second samples indicates that some changes have taken place in the dynamics of nominal interest rates and inflation. In particular, the use of level detrended data is necessary to create a stable VAR(p) with large enough coefficients and variance-covariance matrix in the second sample, which permits investigation of the Fisher Effect in a stationary setting. Furthermore, the Fisher Effect statistically takes a longer time to go into effect in the second sample period when compared to the first sample period. These changes are partially related to the decreased volatility in nominal interest rates. In regards to the inflation-targeting rule, the results of the second sample period indicate that there is a one-to-one movement between nominal interest rates and inflation. Consequently, the Fisher Effect still holds, which indicates that monetary policy should only follow discrete movements in inflation.

In summary, the results are that the Fisher Effect statistically holds, regardless of the use of standardized detrended data or level detrended data, even with the noticeable decrease in the volatility of nominal interest rates during the second period.

5. Conclusion

This paper investigates the Fisher Effect for two sample periods: for the first quarter of 1960 to the third quarter of 1995, and for the first quarter of 1960 to the second quarter of 2004. Nonparametric techniques are used to investigate whether the Fisher Effect holds. The methodology is applied to study the dynamics of nominal interest rates and inflation in bivariate VARs, and the results are compared with the ones obtained from parametric techniques. The advantages of using nonparametric over parametric techniques are that the nonparametric version is better equipped to deal with outliers, and is able to capture nonlinearities in the underlying system. The empirical investigation of the Fisher Effect pursued in this paper also takes into account possible misspecifications and the presence of potential structural breaks.

The first findings are that nominal interest rates and inflation present structural changes in the early 1980s. When these breaks are taken into account, nonstationarity tests indicate that they are trend-stationary. Thus, a time-detrended Fisher Effect is warranted to measure the long run relationship of nominal interest rates to inflation. Notice that if these breaks are not taken into account, this could lead to the erroneous conclusion that the respective time series contain a stochastic trend due to the low power of unit root and cointegration tests.

In the second stage, the residuals from the detrending regressions of inflation and nominal interest rates are used to form a stationary, nonparametric, time-detrended VAR with median or mean coefficients. This framework permits the investigation of nonlinearities in the relationship between nominal interest rates and inflation through the study of the estimated coefficients and the dynamic multipliers of the orthogonalized impulse response functions. The nonparametric time-detrended test for the Fisher Effect is formed from the cumulative

[6] For the second sample period, a VAR(4) using standardized data and median nonparametric estimated coefficients is able to produce the Fisher Effect while the parametric VAR(4) with standardized data is unable to produce the Fisher Effect. For the nonparametric model converges to 1.09 using Method 2 and 1.08 using Method 3. The parametric model converges to 0.48.

orthogonalized dynamic multiplier ratios of inflation to nominal interest rates pursuant to a shock to nominal interest rates. If the Fisher Effect holds, this ratio statistically approaches one as the horizon goes to infinity, which means that the changes in detrended nominal interest rates are matched by the changes in detrended inflation in the long run.

Two variations of the relationship between inflation and nominal interest rates are investigated in this paper: a nonparametric time-detrended Fisher Effect using standardized data for the first sample period, and a nonparametric time-detrended Fisher Effect using level data for the second sample period. Thus, both nonparametric techniques conclude that the Fisher Effect holds.

A monetary policy rule could be implemented based on investigating whether or not movements in inflation exceed movements in nominal interest rates in the long run – changes in policy would be warranted in order to maintain the long run equilibrium between these series, as reflected in the Fisher Effect. The findings that the Fisher Effect holds, especially for the more recent period, suggest that current monetary policy can be implemented based on minor discrete changes vis-à-vis discrete changes in nominal interest rates, in order to maintain the long run equilibrium between nominal interest rates and inflation.

References

Crowder, W. J. and D. L. Hoffman, 1996, The long run relationship between nominal interest rates and inflation: the Fisher equation revisited, *Journal of Money, Credit, and Banking,* **28**, 103-118.

Daniels, J. P., Nourzad, F. and R. K. Toutkoushian, 1996, Testing the Fisher Effect as a long-run equilibrium relation, *Applied Financial Economics,* **6**, 115-120.

Engle, R. F., and C. W. Granger, 1987, Co-integration and error correction: representation, estimation and testing, *Econometrica,* **55**, 251-276.

Fama, E. F., 1975, Short-term interest rates as predictors of inflation, *American Economic Review,* **65**, 269-282.

Garcia, R. and P. Perron, 1996, An analysis of the real interest rate under regime Shifts, *Review of Economics and Statistics,* **78**, 111-125.

Hamilton, J. D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica,* **57**, 357-384.

King, R. G. and M. W. Watson, 1997, Testing long run neutrality, *Economic Quarterly*, **83**, 69-101.

Malliaropulos, D., 2000, A note on non-stationarity, structural breaks, and the Fisher Effect, *Journal of Banking and Finance,* **24**, 695-707.

Mishkin, F., 1992, Is the Fisher Effect for real a re-examination of the relationship between inflation and interest rates, *Journal of Monetary Economics,* **30**, 95-215.

Perron, P., 1989, The great crash, the oil price shock, and the unit root hypothesis, *Econometrica*, **6,** 1361-1401.

Wallace, M.S. and J.T. Warner, 1993, The Fisher Effect and the term structure of interest rates: tests of cointegration. *Review of Economics and Statistics,* **75**, 320-324.

In: Economic Forecasting
Editor: Alan T. Molnar, pp. 163-179
ISBN: 978-1-60741-068-3

Chapter 6

FORECASTING ABILITY AND STRATEGIC BEHAVIOR OF JAPANESE INSTITUTIONAL FORECASTERS

Masahiro Ashiya*
Faculty of Economics, Kobe University, 2-1 Rokko-dai,
Nada, Kobe, 657-8501, Japan.

Abstract

This paper investigates the effect of forecasting ability on forecasting bias among Japanese GDP forecasters. Trueman (1994, Review of Financial Studies, 7(1), 97-124) argues that an incompetent forecaster tends to discard his private information and release a forecast that is close to the prior expectation and the market average forecast. Clarke and Subramanian (2006, Journal of Financial Economics, 80, 81-113) find that a financial analyst issues bold earning forecasts if and only if his past performance is significantly different from his peers. This paper examines a twenty-eight-year panel of annual GDP forecasts, and obtains supportive evidence of Clarke and Subramanian (2006). Our result indicates that conventional tests of rationality are biased toward rejecting the rational expectations hypothesis.

Keywords: Forecast evaluation; Rational expectations hypothesis; Reputation; Herd behavior; Economic forecasts.

JEL Classification Codes: E37; C53; D84.

1. Introduction

A central premise of economic analysis is that people seek to maximize their utility. Various factors influence their utility and hence their behavior. The recent development of information economics has shown that those with private information signal what they know through their actions. For example, a forecaster with private information based on his own ability will intentionally release biased forecasts. Numerous empirical studies have

* E-mail address: ashiya@econ.kobe-u.ac.jp

investigated this kind of forecasting bias, but the vast majority of them are on financial analysts[1] and few are on macroeconomic forecasters.[2] The purpose of this paper is to fill this gap: we examine the relation between forecasting ability and forecasting bias among Japanese GDP forecasters.

More specifically, we consider the signaling hypotheses of Trueman (1994) and Clarke and Subramanian (2006). Trueman (1994) argues that a forecaster of poor ability has little confidence in his private information, and that he prefers to release a forecast that is close to the prior expectation and the market average forecast. Clarke and Subramanian (2006) suggest that the forecaster's payoff is convex in her reputation and that, with the presence of employment risk, there is a U-shaped relation between the forecast boldness and prior performance. They confirm their prediction using earnings forecast data from I/B/E/S over the period 1988 to 2000.

We test these hypotheses using a twenty-eight-year panel of annual GDP forecasts. Although the result is not consistent with the hypothesis of Trueman (1994), it is consistent with the hypothesis of Clarke and Subramanian (2006). To the best of our knowledge, this is the first study to find a supportive evidence of Clarke and Subramanian (2006) among macroeconomic forecasters. Our result indicates that rational forecasters sometimes issue biased forecasts and hence biased forecasts do not necessarily contradict the rational expectations hypothesis.

There are several existing studies on the strategic behavior of Japanese GDP forecasters. Ashiya and Doi (2001) investigate the relation between forecasters' age and degree of herding, and find that Japanese forecasters herd to the same degree regardless of their forecasting experience. Ashiya (2002, 2003) demonstrates that reputation models cannot explain the biases he found in the forecast revisions of Japanese GDP forecasters. Ashiya (2009) found that forecasters in industries that emphasize publicity most tend to issue the most extreme and least accurate predictions. Our results complement these previous works.

The plan of the paper is as follows. Section 2 explains the data. Section 3 introduces the variables. Section 4 tests the hypothesis of Trueman (1994), and Section 5 tests the hypothesis of Clarke and Subramanian (2006). Section 6 concludes.

2. Data

Toyo Keizai Inc. has published the forecasts of about 70 Japanese institutions (banks, securities firms, trading companies, insurance companies, and research institutions) in the February or March issue of "Monthly Statistics (Tokei Geppo)" since the 1970s. In every December, institution *i* releases forecasts of the Japanese real GDP growth rate for the ongoing fiscal year and for the next fiscal year. We call the former $f_{t,t}^i$ and the latter $f_{t,t+1}^i$. For example, the February 2008 issue contains forecasts for fiscal year 2007 (from April 2007 to March 2008) and for fiscal year 2008 (from April 2008 to March 2009). We treat the former as $f_{2007,2007}^i$ and the latter as $f_{2007,2008}^i$.

[1] Ramnath et al. (2008) survey the literature exhaustively.

[2] Fildes and Stekler (2002, p.461) review the literature. See also Pons-Novell (2003) and Elliott et al. (2008).

Since the participation rate was very low throughout the 1970s (on average 13.8 institutions per year), we use the forecasts published from February 1981 on. That is, we use $f_{t,t}^{i}$ for the fiscal years 1980 through 2007 and $f_{t,t+1}^{i}$ for the fiscal years 1981 through 2008. We exclude institutions that participated in less than eleven surveys, leaving 53 institutions. The average number of observations per institution is 20.28 for the current-year forecasts ($f_{t,t}^{i}$) and 20.55 for the year-ahead forecasts ($f_{t,t+1}^{i}$).

As for the actual growth rate g_t, Keane and Runkle (1990) argue that the revised data introduces a systematic bias because the extent of revision is unpredictable for the forecasters [see also Stark and Croushore (2002)]. For this reason, we use the initial announcement of the Japanese government, which was released in June through 2004 and was released in May from 2005 on. We obtain the same results by using the revised data of g_t released in June of year $t+2$.

Figure 1 shows the forecast distributions and the actual growth rates. The vertical line shows the support of the forecast distribution of the 53 institutions. The horizontal line shows the mean forecast. The closed diamond shows the actual growth rate. Table 1 shows the descriptive statistics of the absolute forecast errors (AFE) for each year. For example, 19 institutions released the current-year forecasts ($f_{t,t}^{i}$) in 1980, and the average of the AFE is 0.974. The difference between the worst forecast and the realization is 1.3 percentage points, and the difference between the best forecast and the realization is 0.7 percentage points.[3] The standard deviation of the AFE in 1980 is 0.152. Table 2 shows the descriptive statistics of the mean absolute forecast error (MAFE) of each institution. As for the current-year forecast, the MAFE of the best institution is 0.373%, and that of the worst institution is 0.692%. The average MAFE is 0.503% and the standard deviation is 0.066. As for the year-ahead forecast, the best is 1.100% and the worst is 1.616%.

Table 1. Absolute forecast errors

(a) Current-year forecasts

	Avg.	Max.	Min.	S.D.	Obs.
1980	0.974	1.3	0.7	0.152	19
1981	0.815	1.4	0.5	0.222	26
1982	0.497	0.9	0.0	0.219	34
1983	0.185	0.4	0.0	0.116	34
1984	0.213	0.5	0.0	0.136	38
1985	0.241	0.7	0.0	0.163	39
1986	0.418	0.8	0.1	0.176	44
1987	1.221	1.8	0.6	0.251	33
1988	0.133	0.6	0.0	0.123	48
1989	0.181	0.7	0.0	0.173	48
1990	0.518	1.1	0.1	0.227	50
1991	0.117	0.4	0.0	0.099	53

[3] Although the value of the realization is between the highest forecast and the lowest one in 1980 [as Figure 1 (a) shows], there is no forecast that coincides with the realization.

Table 1. Continued

	Avg.	Max.	Min.	S.D.	Obs.
1992	0.948	2.0	0.4	0.294	50
1993	0.294	0.9	0.0	0.240	51
1994	0.438	0.8	0.0	0.208	50
1995	1.355	1.9	0.6	0.202	51
1996	0.562	1.3	0.0	0.275	50
1997	0.763	1.7	0.2	0.268	49
1998	0.391	1.8	0.0	0.296	47
1999	0.226	0.8	0.0	0.184	43
2000	1.025	1.6	0.2	0.261	40
2001	0.246	0.7	0.0	0.172	35
2002	0.668	1.4	0.2	0.285	19
2003	0.844	1.4	0.3	0.328	25
2004	0.193	0.7	0.0	0.166	27
2005	0.346	0.8	0.1	0.173	26
2006	0.170	0.5	0.0	0.174	23
2007	0.174	0.5	0.0	0.136	23

(b) Year-ahead forecasts

	Avg.	Max.	Min.	S.D.	Obs.
1981	1.668	3.0	0.4	0.551	19
1982	0.496	1.2	0.0	0.324	26
1983	0.965	1.9	0.2	0.409	34
1984	1.432	2.1	0.8	0.361	34
1985	0.437	1.8	0.0	0.374	38
1986	0.446	1.3	0.0	0.319	39
1987	2.423	3.3	1.4	0.420	44
1988	1.472	2.5	0.7	0.324	47
1989	0.669	1.6	0.0	0.377	48
1990	1.260	2.1	0.5	0.309	48
1991	0.242	0.7	0.0	0.184	50
1992	2.172	3.2	1.5	0.310	53
1993	2.608	3.5	1.7	0.331	50
1994	0.376	1.0	0.0	0.245	51
1995	0.422	1.6	0.0	0.338	50
1996	1.092	2.1	0.4	0.345	51
1997	2.082	2.9	1.2	0.395	50
1998	2.859	4.0	1.5	0.493	49
1999	1.028	2.7	0.0	0.571	47
2000	0.309	1.2	0.0	0.252	43
2001	3.083	4.0	2.1	0.353	40
2002	2.263	3.6	1.2	0.473	35
2003	3.068	4.1	2.2	0.452	19
2004	0.356	0.9	0.0	0.235	25
2005	1.919	2.8	1.3	0.378	27
2006	0.281	1.0	0.0	0.283	26
2007	0.365	0.8	0.0	0.233	23

Table 2. Mean absolute forecast error of 53 institutions

	Current-year	Year-ahead
Average	0.503	1.339
Maximum	0.692	1.616
Minimum	0.373	1.100
S.D.	0.066	0.118

3. Definition of Variables

As table 1 has shown, some years are more difficult to forecast than others. The variance of the absolute forecast errors (AFE) tends to be larger in these difficult-to-forecast years. It follows that the level of the MAFE is mainly determined by the performance in the difficult-to-forecast years. Therefore the MAFE is not an appropriate measure of forecast accuracy. We deal with this issue by considering the accuracy *ranking* of the institutions. Following Kolb and Stekler (1996) and Skillings and Mack (1981), we employ the adjusted ranking that is robust to changes in the variance of the forecast variables.

Suppose the panel data consists of *N* institutions and *M* periods. Let N_t ($\leq N$) be the number of institutions that release forecasts in year *t*. $AFE^i_{t,t} \equiv \left| f^i_{t,t} - g_t \right|$ denotes the absolute forecast error of institution *i* in year *t*. Let $r^i_t \in \{1, \cdots, N_t\}$ be the relative rank of $AFE^i_{t,t}$. If ties occur, we use average ranks. If institution *i* does not participate in year *t*, we assume $r^i_t = 0.5(1 + N_t)$. We define the adjusted rank of institution *i* in year *t*, $R^i_t\left(AFE^i_{t,t}\right)$, as

$$R^i_t\left(AFE^i_{t,t}\right) \equiv \left(\frac{12}{1+N_t}\right)^{0.5} \left(r^i_t - \frac{1+N_t}{2}\right). \tag{1}$$

The first term of $R^i_t\left(AFE^i_{t,t}\right)$ compensates for the difference in observations. The second term measures relative performance. A negative (positive) $R^i_t\left(AFE^i_{t,t}\right)$ indicates that the rank of institution *i* in year *t* is above (below) the median and $f^i_{t,t}$ is relatively accurate (inaccurate). $R^i\left(AFE^i_{t,t}\right) \equiv \sum_{t=1}^{M} R^i_t\left(AFE^i_{t,t}\right)$ denotes the sum of the adjusted ranks. If $R^i\left(AFE^i_{t,t}\right)$ is close to zero, forecast accuracy of institution *i* is on average similar to other institutions. If $R^i\left(AFE^i_{t,t}\right)$ is significantly smaller (larger) than zero, the forecast accuracy of institution *i* is on average better (worse) than that of other institutions.

We also define the relative boldness of forecasts. Let $\bar{f}^{-i}_{t,t}$ be the forecast average excluding institution *i* in year *t*. Then $Bold^i_{t,t} \equiv \left| f^i_{t,t} - \bar{f}^{-i}_{t,t} \right|$, the forecast boldness, indicates the extremeness of *i*'s forecast in year *t*.[4] Let $R_t^i\left(Bold_{t,t}^i\right)$ be the adjusted rank of $Bold_{t,t}^i$ defined in equation (1). Institution *i* issues a forecast similar to others in year *t* if $R_t^i\left(Bold_{t,t}^i\right)$ is significantly smaller than zero. Institution *i* issues a forecast different from others in year *t* if $R_t^i\left(Bold_{t,t}^i\right)$ is significantly larger than zero. $R^i\left(Bold_{t,t}^i\right) \equiv \sum_{t=1}^{M} R_t^i\left(Bold_{t,t}^i\right)$ denotes the sum of the adjusted ranks of boldness.

4. Trueman's Hypothesis

4.1. Overview of Trueman's Theory

Trueman (1994) examines the optimal earning forecasts for security analysts given their private information. This section reviews his argument in the context of macroeconomic forecasts.

Consider a group of economists who are paid by clients to prepare GDP forecasts. The growth rate takes one of four possible values compared with the prior expectations: "very high", "high", "low", and "very low". The prior distribution is symmetric, and the prior probability that the growth rate is "high", *t*, satisfies $0.25 < t < 0.5$. Each economist obtains private information and releases his forecast at the beginning of the year. There are two types of economists: weak and strong. The economists know their own types with certainty, but the clients do not know the economists' abilities.

The private information is correct with probability $p_S > 0.5$ when the economist is strong, and is correct with probability $p_W \in (0.5, p_S)$ when the economist is weak. If the economist's private information is either "very high" or "high" ("low" or "very low"), then the probability that the actual growth rate is either "low" or "very low" ("very high" or "high") is zero. For example, if a strong economist obtains the signal of "high", then "high" growth rate is realized with probability p_S, "very high" growth rate with $1 - p_S$, "low" with probability zero, and "very low" with probability zero. Similarly, if a weak economist obtains the signal of "very low", then "very low" growth rate is realized with probability p_W, and "low" with $1 - p_W$.

While an economist's compensation is usually based on several factors, it is natural to assume that his compensation is increasing in the clients' perception of his forecasting ability, which is inferred from his past forecast record. If this is the case, then the weak economists have a strong incentive to discard his (relatively inaccurate) private information. More specifically, Trueman (1994, Proposition 2) demonstrates that, given $p_W < 2t$, a weak economist releases moderate forecasts with positive probability even if his information suggests extreme growth rates. The inequality $p_W < 2t$ indicates that the accuracy rate of the signal the weak economist obtains (p_W) is lower than the prior

[4] Clarke and Subramanian [2006, p.93, equation (12)] employ the same definition of forecast boldness.

probability of moderate growth rate ($2t$). When this inequality is satisfied, the weak economist avoids extreme forecasts because his signal is uninformative and moreover the prior probability of extreme outcomes ($1-2t$) is very small. Hence there is a downward (upward) bias in the forecasts of weak economists that are larger (smaller) than the prior expectations.

Trueman also analyzes the case in which economists can observe others' forecasts. The strong economist always releases a forecast that is consistent with his private information because his private signal is accurate. The weak economist, on the other hand, mimics others because he is not confident in his private information (Proposition 4). Therefore, on average, the forecasts of the weak economists are more similar to the mean forecast of the market than those of the strong economists.

To summarize, Trueman argues that

(H1) a weak economist avoids forecasts different from the prior expectations, and that
(H2) a weak economist releases forecasts similar to the market average.

Using the variables defined in Section 3, we can restate these hypotheses as follows:

(H1) a forecaster whose $R_t^i\left(AFE_{t,t}^i\right)$ (the adjusted rank of the absolute forecast error) is large will avoid forecasts different from the prior expectations.

(H2) a forecaster with large $R_t^i\left(AFE_{t,t}^i\right)$ tends to choose smaller $R_t^i\left(Bold_{t,t}^i\right)$ (i.e., a forecast similar to others).

We will test these hypotheses in the next section.

4.2. Empirical Results

To test the hypothesis (H1) described in the last section, we consider the following regression:

$$\left|f_{t,t}^i - g_{t-1}\right| = \alpha + \beta \cdot R_{t-1}^i\left(AFE_{t,t}^i\right) + \gamma \cdot \left|g_t - g_{t-1}\right| + u_{t,t}^i . \tag{2}$$

If the growth rate follows a random walk process, the optimal forecast for year *t* is the latest realization. Hence we use g_{t-1} as the prior expectation for the current-year forecast. We use g_{t-2} as the prior expectation for the year-ahead forecast because forecasters do not know the value of g_{t-1} when they release $f_{t-1,t}^i$. The dependent variable, $\left|f_{t,t}^i - g_{t-1}\right|$, shows the degree of deviation from the prior expectation. To investigate the effect of the forecasting ability, we use $R_{t-1}^i\left(AFE_{t,t}^i\right)$ (the adjusted rank of the absolute forecast error for the latest realization) as an independent variable. We use $R_{t-2}^i\left(AFE_{t-1,t}^i\right)$ for the year-ahead forecast because forecasters do not know the value of $R_{t-1}^{i}\left(AFE_{t-1,t}^{i}\right)$ when they release $f_{t-1,t}^{i}$. If a weak economist tends to release a forecast closer to the prior, then β should be significantly negative. We add $\left|g_t - g_{t-1}\right|$ as an independent variable to control on specific factors in each year (we add $\left|g_t - g_{t-2}\right|$ for the year-ahead forecast). If $R_{t-1}^{i}\left(AFE_{t,t}^{i}\right)$ is irrelevant and $f_{t,t}^{i}$ is accurate, $\alpha = \beta = 0$ and $\gamma = 1$ will hold.

Table 3 shows the result of regression (2). Standard errors of estimated coefficients are in parentheses. As for the current-year forecasts, β is negative but insignificant. As for the year-ahead forecasts, β is significant but *positive*. These results are not consistent with the hypothesis (H1). We obtain the same results when we use $R^{i}\left(AFE_{t,t}^{i}\right) \equiv \sum_{t=1}^{M} R_{t}^{i}\left(AFE_{t,t}^{i}\right)$ instead of $R_{t-1}^{i}\left(AFE_{t,t}^{i}\right)$ as an independent variable (See table 4).

Table 3. Forecast deviation from the prior expectation

(a) Current-year forecasts

Dependent variable: $\left|f_{t,t}^{i} - g_{t-1}\right|$

Constant	0.199 (0.031)***
$R_{t-1}^{i}\left(AFE_{t,t}^{i}\right)$	-0.0016 (0.0028)
$\left\lvert g_t - g_{t-1}\right\rvert$	0.728 (0.018) ***[a]
observations	979
$\overline{R}^2$	0.613

(b) Year-ahead forecasts

Dependent variable: $\left|f_{t-1,t}^{i} - g_{t-2}\right|$

Constant	0.706 (0.039)***
$R_{t-2}^{i}\left(AFE_{t-1,t}^{i}\right)$	0.0077 (0.0036)**
$\left\lvert g_t - g_{t-2}\right\rvert$	0.312 (0.015)***[a]
observations	910
$\overline{R}^2$	0.310

Notes: Standard errors of estimated coefficients are in parentheses.

a: The null hypothesis is $\gamma = 1$.

***: Significant at the 0.01 level.

**: Significant at the 0.05 level.

*: Significant at the 0.10 level.

Table 4. Forecast deviation from the prior expectation (robustness check)

(a) Current-year forecasts

Dependent variable: $\left\|f^i_{t,t} - g_{t-1}\right\|$	
Constant	0.180 (0.030)***
$R^i\left(AFE^i_{t,t}\right)$	-0.0005 (0.0006)
$\left\|g_t - g_{t-1}\right\|$	0.726 (0.018) ***[a]
observations	1075
$\overline{R}^2$	0.610

(b) Year-ahead forecasts

Dependent variable: $\left\|f^i_{t-1,t} - g_{t-2}\right\|$	
Constant	0.672 (0.036)***
$R^i\left(AFE^i_{t-1,t}\right)$	-0.0001 (0.0009)
$\left\|g_t - g_{t-2}\right\|$	0.321 (0.014)***[a]
observations	1066
$\overline{R}^2$	0.315

Notes: Standard errors of estimated coefficients are in parentheses.
a: The null hypothesis is $\gamma = 1$.
***: Significant at the 0.01 level.
**: Significant at the 0.05 level.
*: Significant at the 0.10 level.

Next we test the hypothesis (H2) by the following regression:

$$R^i_t\left(Bold^i_{t,t}\right) = \alpha + \beta \cdot R^i_{t-1}\left(AFE^i_{t,t}\right) + u^i_{t,t} \,. \tag{3}$$

$R^i_t\left(Bold^i_{t,t}\right)$ is the adjusted rank of the forecast deviation from the market mean. Negative $R^i_t\left(Bold^i_{t,t}\right)$ indicates that forecaster *i* makes a forecast similar to the market consensus in year *t*. Positive $R^i_{t-1}\left(AFE^i_{t,t}\right)$ indicates that $f^i_{t-1,t-1}$ was inaccurate. Therefore β should be significantly *negative* for (H2) to be true. Table 5 shows the result of (3). The coefficients of $R^i_{t-1}\left(AFE^i_{t,t}\right)$ are insignificant for both the current-year and year-ahead forecasts.

Table 5. Forecast deviation from the market consensus

(a) Current-year forecasts

Dependent variable:	$R_t^i\left(Bold_{t,t}^i\right)$
Constant	0.024 (0.202)
$R_{t-1}^i\left(AFE_{t,t}^i\right)$	0.032 (0.032)
observations	979
$\overline{R}^2$	0.000

(b) Year-ahead forecasts

Dependent variable:	$R_t^i\left(Bold_{t-1,t}^i\right)$
Constant	0.003 (0.213)
$R_{t-2}^i\left(AFE_{t-1,t}^i\right)$	-0.007 (0.033)
observations	932
$\overline{R}^2$	-0.001

Notes: Standard errors of estimated coefficients are in parentheses.
***: Significant at the 0.01 level. **: Significant at the 0.05 level.*: Significant at the 0.10 level.

Table 6. Forecast deviation from the market consensus [robustness check (a)]

(a) Current-year forecasts

Dependent variable:	$R_t^i\left(Bold_{t,t}^i\right)$
Constant	0.021 (0.189)
$R^i\left(AFE_{t,t}^i\right)$	0.030 (0.007)***
observations	1075
$\overline{R}^2$	0.019

(b) Year-ahead forecasts

Dependent variable:	$R_t^i\left(Bold_{t-1,t}^i\right)$
Constant	0.001 (0.193)
$R^i\left(AFE_{t-1,t}^i\right)$	0.003 (0.008)
observations	1089
$\overline{R}^2$	-0.001

Notes: Standard errors of estimated coefficients are in parentheses.
***: Significant at the 0.01 level. **: Significant at the 0.05 level.*: Significant at the 0.10 level.

We check the robustness of the above results by substituting R^i for R^i_{t-1}:

$$R^i_t\left(Bold^i_{t,t}\right) = \alpha + \beta \cdot R^i\left(AFE^i_{t,t}\right) + u^i_{t,t}\text{, and} \tag{3'}$$

$$R^i\left(Bold^i_{t,t}\right) = \alpha + \beta \cdot R^i\left(AFE^i_{t,t}\right) + u^i\text{.} \tag{3''}$$

Tables 6 and 7 show the results: no coefficient is significantly negative.

Since (H1) and (H2) both are not supported, our empirical results are inconsistent with Trueman's hypothesis.

Table 7. Forecast deviation from the market consensus [robustness check (b)]

(a) Current-year forecasts

Dependent variable: $R^i\left(Bold^i_{t,t}\right)$

Constant	-0.033 (4.867)
$R^i\left(AFE^i_{t,t}\right)$	0.601 (0.166)***
observations	53
$\overline{R}^2$	0.188

(b) Year-ahead forecasts

Dependent variable: $R^i\left(Bold^i_{t-1,t}\right)$

Constant	0.040 (5.702)
$R^i\left(AFE^i_{t-1,t}\right)$	0.060 (0.237)
observations	53
$\overline{R}^2$	-0.018

Notes: Standard errors of estimated coefficients are in parentheses.
***: Significant at the 0.01 level.
**: Significant at the 0.05 level.
*: Significant at the 0.10 level.

5. Clarke and Subramanian's Hypothesis

5.1. Overview of Clarke and Subramanian's Theory

Clarke and Subramanian (2006) (CS hereafter) analyze a multi-period Bayesian learning model in which a forecaster issues her forecast based on her noisy private information and the consensus forecast. Her ability when she first enters the market is common knowledge, but it can change over time as she gains experience. All agents dynamically update their assessments of the forecaster's ability based on her observed performance.

Then, with some additional conditions, CS show that the forecaster's compensation is convex in her average perceived ability and that the forecaster could be fired only if her average perceived ability falls below a certain level. The presence of employment risk induces underperformers (i.e. those whose perceived abilities are below the termination threshold) to issue extreme forecasts because they have little to lose. On the other hand, the employment risk keeps intermediate performers from gambling behavior. Superior forecasters, however, do not face the employment risk because their perceived abilities are far above the termination threshold. Consequently the payoff functions of superior forecasters are convex in the ability, which lead them to risk-taking behavior. To sum up, the forecaster in CS model issues bold (i.e. extreme) forecasts if and only if she outperforms or underperforms her peers significantly.

Let us restate this hypothesis using the variables defined in Section 3:

(H3) a forecaster with significantly positive or negative $R_{t-1}^{i}\left(AFE_{t,t}^{i}\right)$ chooses positive $R_{t}^{i}\left(Bold_{t,t}^{i}\right)$, and a forecaster with near-zero $R_{t-1}^{i}\left(AFE_{t,t}^{i}\right)$ chooses negative $R_{t}^{i}\left(Bold_{t,t}^{i}\right)$.

The next section tests this hypothesis.

5.2. Empirical Results

Clarke and Subramanian (2006) argue that forecasters with superior or inferior track records tend to release extreme forecasts. Namely, those with significantly positive or negative $R_{t-1}^{i}\left(AFE_{t,t}^{i}\right)$ choose positive $R_{t}^{i}\left(Bold_{t,t}^{i}\right)$, and those with near-zero $R_{t-1}^{i}\left(AFE_{t,t}^{i}\right)$ choose negative $R_{t}^{i}\left(Bold_{t,t}^{i}\right)$. We test this hypothesis, (H3), by the following regression:

$$R_{t}^{i}\left(Bold_{t,t}^{i}\right)=\alpha+\beta\cdot R_{t-1}^{i}\left(AFE_{t,t}^{i}\right)+\gamma\cdot\left[R_{t-1}^{i}\left(AFE_{t,t}^{i}\right)\right]^{2}+u_{t,t}^{i}\,. \tag{4}$$

(H3) predicts a U-shaped relation between $R_{t-1}^{i}\left(AFE_{t,t}^{i}\right)$ (prior performances) and $R_{t}^{i}\left(Bold_{t,t}^{i}\right)$ (the forecast boldness). Since $R_{t-1}^{i}\left(AFE_{t,t}^{i}\right)=0$ corresponds to the medium performance, the bottom of the quadratic curve should be on the vertical axis. Therefore we expect $\alpha<0$, $\beta=0$, and $\gamma>0$.

Table 8 shows the result of regression (4). We have $\alpha<0$, $\beta=0$, and $\gamma>0$ for both the current-year and the year-ahead forecasts. Although α for the current-year forecasts is insignificant, α for the year-ahead forecasts and γ for both the current-year and the year-ahead forecasts are statistically significant. Furthermore, we have positive $R_{t}^{i}\left(Bold_{t,t}^{i}\right)$ for both sufficiently positive $R^i_{t-1}\left(AFE^i_{t,t}\right)$ and sufficiently negative $R^i_{t-1}\left(AFE^i_{t,t}\right)$.[5] All in all, the results are consistent with (H3).

We check the robustness of the above results by substituting $Bold^i_{t,t}$ for $R^i_t\left(Bold^i_{t,t}\right)$:

$$Bold^i_{t,t} = \alpha + \beta \cdot R^i_{t-1}\left(AFE^i_{t,t}\right) + \gamma \cdot \left[R^i_{t-1}\left(AFE^i_{t,t}\right)\right]^2 + u^i_{t,t}\,. \qquad (4')$$

Since $Bold^i_{t,t}$ is always positive by definition, we expect $\alpha > 0$, $\beta = 0$, and $\gamma > 0$. The results of regression (4') are summarized in table 9: we have $\beta = 0$ and significantly positive α and γ for both the current-year and the year-ahead forecasts. Namely, (H3) is supported again with this specification.

Table 8. Forecast boldness and prior performance

(a) Current-year forecasts

Dependent variable: $R^i_t\left(Bold^i_{t,t}\right)$

Constant	-0.365 (0.291)
$R^i_{t-1}\left(AFE^i_{t,t}\right)$	0.031 (0.032)
$\left[R^i_{t-1}\left(AFE^i_{t,t}\right)\right]^2$	0.010 (0.005)*
observations	979
$\overline{R}^2$	0.003

(b) Year-ahead forecasts

Dependent variable: $R^i_t\left(Bold^i_{t-1,t}\right)$

Constant	-0.989 (0.305)***
$R^i_{t-2}\left(AFE^i_{t-1,t}\right)$	-0.009 (0.033)
$\left[R^i_{t-2}\left(AFE^i_{t-1,t}\right)\right]^2$	0.024 (0.005)***
observations	932
$\overline{R}^2$	0.019

Notes: Standard errors of estimated coefficients are in parentheses.
***: Significant at the 0.01 level.
**: Significant at the 0.05 level.
*: Significant at the 0.10 level.

[5] The estimation result of Table 8(a) indicates that $R^i_t\left(Bold^i_{t,t}\right) = 0.601$ when $R^i_{t-1}\left(AFE^i_{t,t}\right) = -11.5$ [which corresponds to $\left(N_{t-1}, r^i_{t-1}\right) = (47,1)$]. The estimation result of Table 8(b) indicates that $R^i_t\left(Bold^i_{t-1,t}\right) = 2.0815$ when $R^i_{t-2}\left(AFE^i_{t-1,t}\right) = 11.5$ [which corresponds to $\left(N_{t-2}, r^i_{t-2}\right) = (47,47)$].

Table 9. Forecast boldness and prior performance (robustness check)

(a) Current-year forecasts

Dependent variable: $Bold^i_{t,t}$

Constant	0.171 (0.008)***
$R^i_{t-1}\left(AFE^i_{t,t}\right)$	0.001 (0.001)
$10^2 \times \left[R^i_{t-1}\left(AFE^i_{t,t}\right)\right]^2$	0.030 (0.014)**
observations	979
$\overline{R}^2$	0.004

(b) Year-ahead forecasts

Dependent variable: $Bold^i_{t-1,t}$

Constant	0.258 (0.012)***
$R^i_{t-2}\left(AFE^i_{t-1,t}\right)$	-0.000 (0.001)
$10^2 \times \left[R^i_{t-2}\left(AFE^i_{t-1,t}\right)\right]^2$	0.109 (0.021)***
observations	932
$\overline{R}^2$	0.025

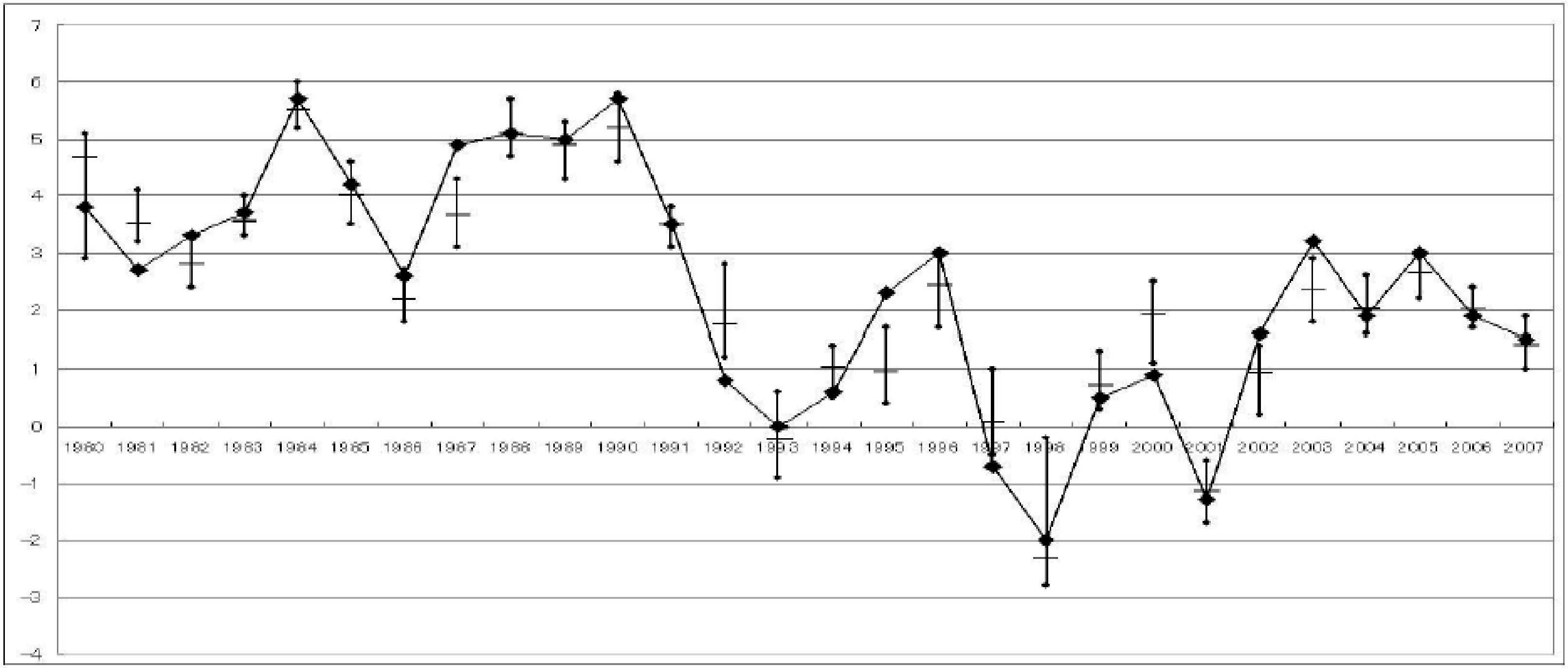
Notes: Standard errors of estimated coefficients are in parentheses.
***: Significant at the 0.01 level.
**: Significant at the 0.05 level.
*: Significant at the 0.10 level.

6. Conclusions

This paper has analyzed whether forecasters' abilities affect the extent of their forecasting bias using a twenty-eight-year panel of Japanese GDP forecasts. The empirical evidence is inconsistent with the hypothesis of Trueman (1994), but is consistent with the hypothesis of Clarke and Subramanian (2006). Namely, a forecaster issues bold forecasts if and only if his past performance is significantly better or worse than his peers. This result indicates that rational forecasters sometimes issue biased forecasts. Since conventional tests of rationality do not consider this kind of strategic behavior, they are biased toward rejecting the rational expectations hypothesis.[6]

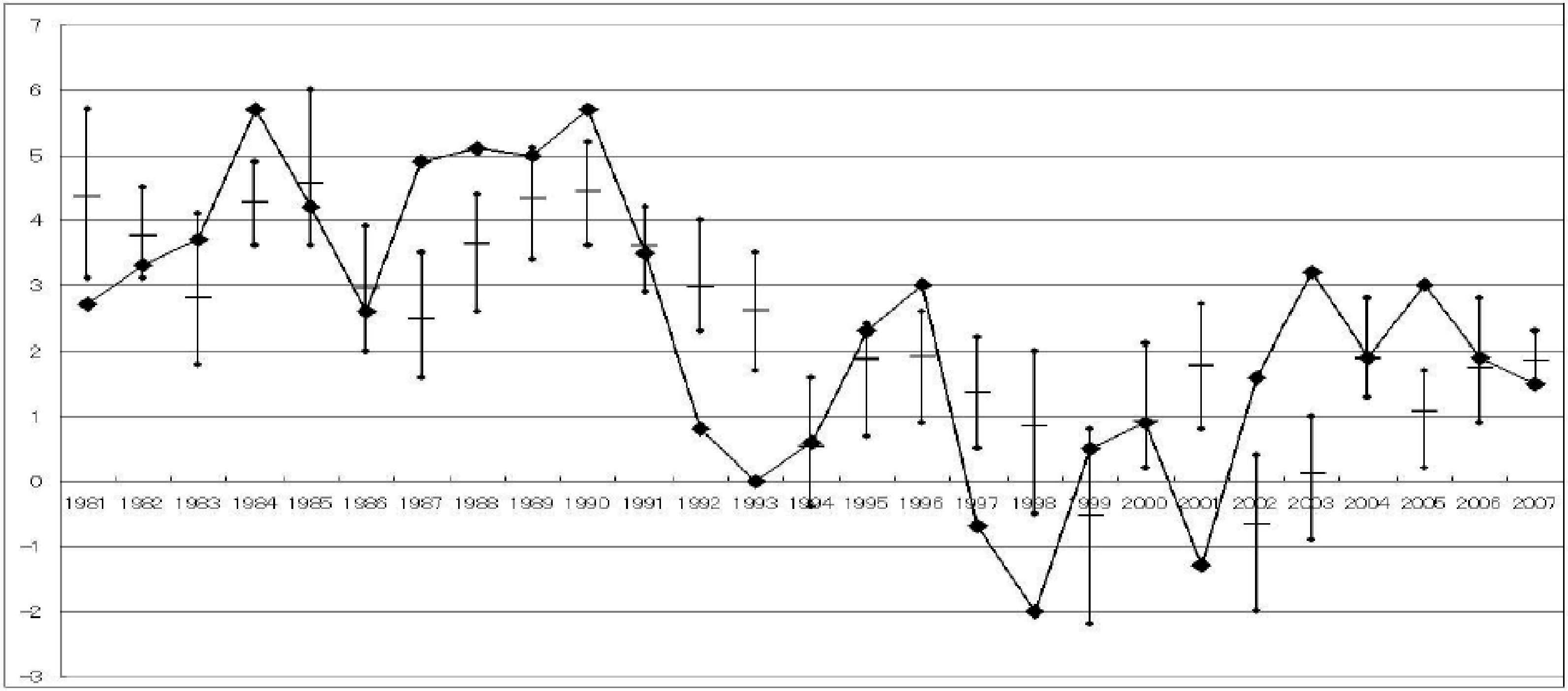
---

[6] Ashiya (2005, pp.80-81) finds that about 80% of the current-year forecasts and the year-ahead forecasts made by 38 Japanese GDP forecasters pass various conventional tests for rationality.



(a) Current-year forecasts

Figure 1. Continued on next page.



(b) Year-ahead forecasts

Notes:
Black diamond: the realization. Horizontal line: the mean forecast.
The vertical line shows the support of the forecast distribution of 53 forecasters.

Figure 1. Forecast distributions and the actual growth rates. (a) Current-year forecasts.

References

Ashiya, Masahiro, and Doi, Takero (2001) “Herd Behavior of Japanese Economists.” *Journal of Economic Behavior and Organization*, **46**(3), 343-346.

Ashiya, Masahiro (2002) “Accuracy and Rationality of Japanese Institutional Forecasters.” *Japan and the World Economy*, **14**(2), 203-213.

Ashiya, Masahiro (2003) “Testing the Rationality of Japanese GDP Forecasts: The Sign of Forecast Revision Matters.” *Journal of Economic Behavior and Organization*, **50**(2), 263-269.

Ashiya, Masahiro (2005) “Twenty-two Years of Japanese Institutional Forecasts.” *Applied Financial Economics Letters*, **1**(2), 79-84.

Ashiya, Masahiro (2009) “Strategic Bias and Professional Affiliations of Macroeconomic Forecasters.” *Journal of Forecasting*, **28**(2), 120-130.

Clarke, Jonathan, and Subramanian, Ajay (2006) “Dynamic Forecasting Behavior by Analysts: Theory and Evidence.” *Journal of Financial Economics*, **80**, 81-113.

Elliott, Graham, Timmermann, Allan, and Komunjer, Ivana (2008) “Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?” *Journal of the European Economic Association*, **6**(1), 122-157.

Fildes, Robert, and Stekler, Herman (2002) “The State of Macroeconomic Forecasting.” *Journal of Macroeconomics*, **24**, 435-468.

Keane, Michael P., and David E. Runkle (1990) “Testing the Rationality of Price Forecasts: New Evidence from Panel Data.” *American Economic Review*, September, **80**(4), pp.714-735.

Kolb, R.A., and Stekler, H.O. (1996) “How Well Do Analysts Forecast Interest Rates?” *Journal of Forecasting*, **15**, 385-394.

Pons-Novell, Jordi (2003) “Strategic Bias, Herding Behaviour and Economic Forecasts.” *Journal of Forecasting*, **22**, 67-77.

Ramnath, Sundaresh, Rock, Steve, and Shane, Philip (2008) “The Financial Analyst Forecasting Literature: A Taxonomy with Suggestions for Further Research.” *International Journal of Forecasting*, **24**, 34-75.

Skillings, John H., and Mack, Gregory A. (1981) “On the Use of a Friedman-Type Statistic in Balanced and Unbalanced Block Designs,” *Technometrics*, **23**(2), 171-177.

Stark, Tom, Croushore, Dean (2002) “Forecasting with a Real-time Data Set for Macroeconomists.” *Journal of Macroeconomics*, **24**, 507-531.

Toyo Keizai Inc. *Monthly Statistics* (Tokei Geppo).

Trueman, Brett (1994) “Analyst Forecasts and Herding Behavior,” *Review of Financial Studies*, *Spring*, **7**(1), pp. 97-124.

In: Economic Forecasting
Editor: Alan T. Molnar, pp. 181-189
ISBN: 978-1-60741-068-3

Chapter 7

QUALITATIVE SURVEY DATA ON EXPECTATIONS. IS THERE AN ALTERNATIVE TO THE BALANCE STATISTIC?

Oscar Claveria *
Research Institute of Applied Economics (IREA),
University of Barcelona, Barcelona (Spain)

Abstract

Through Monte Carlo simulations it is possible to isolate the measurement error introduced by incorrect assumptions when quantifying survey results. By means of a simulation experiment we test whether a variation of the balance statistic outperforms the balance statistic in order to track the evolution of agents' expectations and produces more accurate forecasts of the quantitative variable generated used as a benchmark.

Keywords: Quantification; Expectations; Forecasting.

JEL: classification: C42, C51.

1. Introduction

Survey results are presented as weighted percentages of respondents expecting a particular variable to rise, fall or remain unchanged. Survey results have often been quantified making use of official data. The differences between the actual values of a variable and quantified expectations may arise from three different sources (Lee, 1994): measurement or conversion error due to the use of quantification methods, expectational error due to the agents' limited ability to predict the movements of the actual variable, and sampling errors. Since survey data are approximations of unobservable expectations, they inevitably entail a measurement error.

Monte Carlo simulations can distinguish between these three sources of error, but there have been few attempts in the literature to compare quantification methods in a simulation context. Common (1985) and Nardo (2004) analyse different quantification methods focusing on rational expectation testing rather than on their forecasting ability. Löffler (1999) estimates the measurement error introduced by the probabilistic method, and proposes a linear correction.

In this paper we design a simulation experiment in order to test whether a variation of the balance statistic outperforms the balance statistic in order to track the evolution of agents' expectations and produces more accurate forecasts of the quantitative variable generated used as a benchmark.

The paper is organised as follows. The second section describes the purposed variation of the balance statistic. Section three presents the simulation experiment. Section four analyses the relative forecasting performance of the estimated series of expectations generated by the two different quantification methods applied. Section five concludes.

2. The Weighted Balance Statistic

Unlike other statistical series, survey results are weighted percentages of respondents expecting an economic variable to increase, decrease or remain constant. As a result, tendency surveys contain two pieces of independent information at time t, R_t and F_t, denoting the percentage of respondents at time $t-1$ expecting an economic variable to rise or fall at time t. The information therefore refers to the direction of change but not to its magnitude.

A variety of quantification methods have been proposed in the literature in order to convert qualitative data on the direction of change into a quantitative measure of agents' expectations. The output of these quantification procedures (estimated expectations) can be regarded as one period ahead forecasts of the quantitative variable under consideration. In this paper we apply the following quantification methods using agents' expectations about the future (prospective information).

The first attempt to quantify survey results is due to Anderson (1951). Assuming that the expected percentage change in a variable remains constant over time for agents reporting an increase and for those reporting a decrease, Anderson (1951) defined the balance statistic ($R_t - F_t$) as a measure of the average changes expected in the variable.

The balance statistic (B_t) does not take into account the percentage of respondents expecting a variable to remain constant (C_t). As C_t usually shows the highest proportions and high levels of dispersion, we purpose a non-linear variation of the balance statistic (WB_t, weighted balance) that accounts for this percentage of respondents:

$$WB_t = \frac{R_t - F_t}{R_t + F_t} = \frac{B_t}{1 - C_t} \tag{1}$$

* E-mail address: oclaveria@ub.edu Tel.: +34-934-021-825; fax.: +34-934-021-821 (Corresponding author)

Weighting the balance statistic by the proportion of respondents expecting a variable to rise or fall allows discriminating between two equal values of the balance statistic depending on the percentage of respondents expecting a variable to remain constant.

3. The Simulation Experiment

By Monte Carlo simulations we compare the forecasting performance of the two quantification methods: the balance and the weighted balance. The experiment is designed in five consecutive steps:

(i)The simulation begins by generating a series of actual changes of a variable. We consider 1000 agents and 300 time periods. Let y_{it} indicate the percentage change of variable Y_{it} for agent i from time $t-1$ to time t. Additionally we suppose that the true process behind the movement of y_{it} is given by:

$$y_{it} = d_{it} + \varepsilon_{it} \tag{2}$$

$i = 1,\ldots,1000$, $t = 1,\ldots,300$ and $d_{it} = \mu + \varphi y_{i,t-1}$, where d_{it} is the deterministic component. The initial value, $y_{i0} = 0.9$, is assumed to be equal for all agents[1]. ε_{it} is an identical and independent normally distributed random variable with mean zero and variance σ_{ε}^2 =1, 10, 20. The average rate of change, y_t, is given by $y_i = \frac{1}{1000}\sum_i y_{it}$. The same weight is given to all agents. We assume different values of μ and φ.

(ii)Secondly, we generate a series of agents' expectations about y_t under the assumption that individuals are rational in Muth's sense[2]:

$$y_{it}^e = d_{it} + \zeta_{it} \qquad\qquad \zeta_{it} \sim N\left(0, \sigma_{\zeta}^2\right) \tag{3}$$

where y_{it}^e has the same deterministic part as y_{it} but a different stochastic term ζ_{it}. We derive $y_i^e = 1/1000\sum_i y_{it}^e$. Additionally, we assume that $\sigma_{i\zeta}^2 = \sigma_{\zeta}^2 = 1$. All the values given to σ_{ε}^2 and σ_{ζ}^2, and to the indifference interval are set to simulate actual business survey series.

---

[1] To check the robustness of the results, we chose different values for the autoregressive parameter, ranging from 0 to 1 with an increase of 0.1 each time. As the final results did not vary significantly from one specification to the other we presented the results for 0, 0.1, 0.5 and 0.9.

[2] Muth (1961) assumed that rationality implied that expectations had to be generated by the same stochastic process that generates the variable to be predicted.

(iii)The third step consists of constructing the answers to the business surveys. The answers are given in terms of the direction of change, i.e., if the variable is expected to increase, decrease or remain equal. We assume that agents' answers deal with the next period and that all agents have the same constant indifference interval $(-a, b)$ with $a = b = 5$. If $y_{it}^{e} \geq 5$, agent i answers that Y_{it} will increase; if $y_{it}^{e} \leq -5$, i expects Y_{it} to decrease; while the agent will report no change if $-5 < y_{it}^{e} < 5$. With these answers, qualitative variables R_{it} and F_{it} can be constructed. R_{it} (F_{it}) takes the value 1 (0) whenever the agent expects an increase (decrease) in Y_{it}. R_t and F_t are then constructed by aggregation.

(iv)The fourth step of the simulation experiment consists of using the two different quantification methods to trace back the series of actual changes of the generated quantitative variable, y_t, from the qualitative variables. We will refer to these expectations as estimated expectations in order to distinguish them from the unobservable ones. With the aim of analysing the performance of the different proxy series, we use the last 200 generated observations. Keeping the series of actual changes fixed, the experiment of generating the rational expectations series as well as the proxy series is replicated 1000 times[3].

(v)To test the robustness of the results, we repeat the simulation experiment for different values of μ, therefore assuming y_{it} is generated by a random walk and by an autoregressive process with different drifts.

4. Evaluation of the Estimated Expectations

In order to evaluate the relative performance and the forecasting accuracy of the different quantification procedures, we keep the series of actual changes fixed and we replicate the experiment of generating the rational expectations series as well as the qualitative variables R_t and F_t 1000 times. The specification of the quantification procedures is based on information up to the first 100 periods; models are then re-estimated each period and forecasts are computed in a recursive way. In each simulation, forecast errors for all methods are obtained for the last 200 periods.

To summarize this information, we calculate the Root Mean Squared Error (RMSE), the Mean Error (ME), the Theil Coefficient (TC) and the three components of the Mean Square Error (MSE): the bias proportion of the MSE (U1), the variance proportion (U2) and the covariance proportion (U3). With the aim of testing whether the reduction in RMSE when comparing both methods is statistically significant, we calculate the measure of predictive accuracy proposed by Diebold-Mariano (1995). Given these two competing forecasts and the series of actual changes of the generated quantitative variable, we have calculated the DM measure which compares the mean difference between a loss criteria (in this case, the root of the mean squared error) for the two predictions using a long-run estimate of the variance of

[3] All simulations are performed with Gauss for Windows 6.0.

the difference series. Table 1 shows the results of an off-sample evaluation for the last 200 periods when $\mu = 0$.

Table 1. Forecasts evaluation ($\mu = 0$)

$\mu = 0$	$\sigma_\varepsilon^2 = 1$		$\sigma_\varepsilon^2 = 10$		$\sigma_\varepsilon^2 = 20$	
$\varphi = 0$	***B***	***WB***	***B***	***WB***	***B***	***WB***
RMSE [a]	2.48	1.53	2.47	1.74	2.42	1.92
ME [b]	-0.01	-0.01	-0.11	-0.08	0.01	0.01
% U1 [c]	0.5	0.5	0.5	0.5	0.2	0.2
% U2 [c]	98.9	98.1	96.3	93.4	95.2	92.6
% U3 [c]	0.5	1.4	3.1	6.1	4.6	7.2
TC [d]	6.23	2.41	6.26	3.20	6.00	3.83
DM [e]	0.81		1.10		0.03	
$\varphi = 0.1$	***B***	***WB***	***B***	***WB***	***B***	***WB***
RMSE [a]	2.49	1.54	2.47	1.75	2.51	1.98
ME [b]	0.01	0.01	-0.06	-0.04	0.00	0.01
% U1 [c]	0.6	0.5	0.4	0.4	0.2	0.3
% U2 [c]	98.9	98.0	96.6	93.7	96.0	93.8
% U3 [c]	0.5	1.4	3.0	5.9	3.8	6.0
TC [d]	6.26	2.42	6.28	3.21	6.41	4.05
DM [e]	0.72		0.53		0.28	
$\varphi = 0.5$	***B***	***WB***	***B***	***WB***	***B***	***WB***
RMSE [a]	2.49	1.54	2.50	1.81	2.41	1.97
ME [b]	0.00	0.00	-0.03	-0.02	0.00	-0.01
% U1 [c]	0.5	0.5	0.3	0.3	0.2	0.2
% U2 [c]	99.0	98.1	96.6	93.8	95.3	93.1
% U3 [c]	0.5	1.4	3.1	5.9	4.5	6.7
TC [d]	6.26	2.42	6.42	3.42	5.96	3.99
DM [e]	0.34		0.31		0.02	
$\varphi = 0.9$	***B***	***WB***	***B***	***WB***	***B***	***WB***
RMSE [a]	2.51	1.56	2.28	1.84	2.37	2.08
ME [b]	0.11	0.06	0.15	0.12	0.04	0.01
% U1 [c]	0.6	0.6	0.7	0.6	0.2	0.1
% U2 [c]	98.0	96.0	93.7	90.8	91.5	88.8
% U3 [c]	1.4	3.4	5.7	8.6	8.3	11.0
TC [d]	6.42	2.54	5.30	3.51	5.29	3.99
DM [e]	1.44		0.73		0.71	

Notes: [a] RMSE = root mean square error; [b] ME = mean error; [c] Decomposition of the mean square error: (i) %U1 = percentage of mean error (bias proportion of the MSE), (ii) %U2 = percentage of regression error (variance proportion of the MSE), (iii) %U3 = percentage of disturbance error (covariance proportion of the MSE); [d] TC = Theil coefficient; [e] DM = results of the Diebold-Mariano test (the statistic uses a NW estimator. Null hypothesis: the difference between the two competing series is non-significant. A positive sign of the statistic implies that the Balance has bigger errors, and is worse. When the t-stat is significant, the second model is statistically better.

* Significant at the 5% level.

Table 2 and Table 3 show the results of an off-sample evaluation for the last 200 periods when $\mu = 1$ and $\mu = -1$ respectively.

Table 2. Forecasts evaluation ($\mu = 1$)

$\mu = 1$	$\sigma_\varepsilon^2 = 1$		$\sigma_\varepsilon^2 = 10$		$\sigma_\varepsilon^2 = 20$	
$\varphi = 0$	***B***	***WB***	***B***	***WB***	***B***	***WB***
RMSE [a]	6.51	3.70	5.06	3.40	3.49	2.70
ME [b]	-6.02	-3.36	-4.36	-2.88	-2.54	-1.91
% U1 [c]	85.4	82.2	73.9	70.9	52.5	49.6
% U2 [c]	14.5	17.6	25.3	27.4	45.3	46.8
% U3 [c]	0.1	0.2	0.8	1.7	2.2	3.6
TC [d]	42.46	13.72	25.74	11.73	12.30	7.40
DM [e]	397.93*		42.01*		20.09*	
$\varphi = 0.1$	***B***	***WB***	***B***	***WB***	***B***	***WB***
RMSE [a]	7.12	4.03	5.46	3.67	3.54	2.73
ME [b]	-6.67	-3.72	-4.82	-3.19	-2.66	-2.00
% U1 [c]	87.8	85.0	77.4	74.6	56.0	52.9
$\mu = 1$	$\sigma_\varepsilon^2 = 1$		$\sigma_\varepsilon^2 = 10$		$\sigma_\varepsilon^2 = 20$	
% U2 [c]	12.1	14.8	21.9	23.9	41.8	43.5
% U3 [c]	0.1	0.2	0.7	1.4	2.2	3.6
TC [d]	50.75	16.33	30.03	13.64	12.68	7.59
DM [e]	467.73*		41.53*		22.03*	
$\varphi = 0.5$	***B***	***WB***	***B***	***WB***	***B***	***WB***
RMSE [a]	12.18	6.91	8.30	5.69	4.97	3.88
ME [b]	-11.93	-6.73	-7.92	-5.39	-4.36	-3.35
% U1 [c]	95.9	94.7	90.8	89.4	76.4	74.3
% U2 [c]	4.1	5.3	8.9	9.9	22.5	23.9
% U3 [c]	0.0	0.1	0.3	0.6	1.1	1.8
TC [d]	148.48	47.80	69.09	32.55	24.86	15.15
DM [e]	587.85*		57.10*		25.50*	
$\varphi = 0.9$	***B***	***WB***	***B***	***WB***	***B***	***WB***
RMSE [a]	51.59	36.80	20.79	16.31	6.93	5.57
ME [b]	-51.56	-36.75	-20.62	-16.15	-6.61	-5.26
% U1 [c]	99.9	99.7	98.3	98.1	91.6	89.8
% U2 [c]	0.1	0.3	1.6	1.8	7.3	8.5
% U3 [c]	0.0	0.0	0.1	0.1	1.1	1.7
TC [d]	2661.95	1354.63	432.36	266.04	47.77	30.86
DM [e]	6287.05*		79.01*		43.20*	

Notes: [a] RMSE = root mean square error; [b] ME = mean error; [c] Decomposition of the mean square error: (i) %U1 = percentage of mean error (bias proportion of the MSE), (ii) %U2 = percentage of regression error (variance proportion of the MSE), (iii) %U3 = percentage of disturbance error (covariance proportion of the MSE); [d] TC = Theil coefficient; [e] DM = results of the Diebold-Mariano test.

[*] Significant at the 5% level.

Table 3. Forecasts evaluation ($\mu = -1$)

$\mu = -1$	$\sigma_{\varepsilon}^{2} = 1$		$\sigma_{\varepsilon}^{2} = 10$		$\sigma_{\varepsilon}^{2} = 20$	
$\varphi = 0$	***B***	***WB***	***B***	***WB***	***B***	***WB***
RMSE [a]	6.49	3.68	4.90	3.30	3.33	2.58
ME [b]	6.00	3.34	4.24	2.80	2.25	1.69
% U1 [c]	85.3	82.1	74.1	70.9	45.1	42.2
% U2 [c]	14.6	17.6	25.1	27.3	52.7	54.1
% U3 [c]	0.1	0.2	0.8	1.8	2.2	3.7
$\mu = -1$	$\sigma_{\varepsilon}^{2} = 1$		$\sigma_{\varepsilon}^{2} = 10$		$\sigma_{\varepsilon}^{2} = 20$	
TC [d]	42.17	13.63	24.24	11.05	11.20	6.77
DM [e]	451.90*		50.23*		17.85*	
$\varphi = 0.1$	***B***	***WB***	***B***	***WB***	***B***	***WB***
RMSE [a]	7.11	4.03	5.48	3.68	3.48	2.69
ME [b]	6.66	3.72	4.86	3.21	2.56	1.92
% U1 [c]	87.8	85.0	78.3	75.5	53.5	50.4
% U2 [c]	12.2	14.8	21.1	23.1	44.2	45.9
% U3 [c]	0.1	0.2	0.6	1.4	2.2	3.7
TC [d]	50.61	16.29	30.16	13.69	12.28	7.35
DM [e]	416.73*		48.41*		22.91*	
$\varphi = 0.5$	***B***	***WB***	***B***	***WB***	***B***	***WB***
RMSE [a]	12.14	6.89	8.43	5.79	4.78	3.73
ME [b]	11.89	6.70	8.05	5.49	4.23	3.25
% U1 [c]	95.9	94.7	90.9	89.5	77.9	75.7
% U2 [c]	4.1	5.3	8.9	9.9	20.7	22.2
% U3 [c]	0.0	0.1	0.3	0.6	1.3	2.2
TC [d]	147.42	47.46	71.28	33.68	22.98	14.00
DM [e]	622.52*		53.87*		29.03*	
$\varphi = 0.9$	***B***	***WB***	***B***	***WB***	***B***	***WB***
RMSE [a]	51.50	36.70	20.82	16.34	6.86	5.49
ME [b]	51.46	36.64	20.69	16.22	6.55	5.18
% U1 [c]	99.9	99.7	98.8	98.5	91.7	90.0
% U2 [c]	0.1	0.3	1.2	1.4	7.5	8.6
% U3 [c]	0.0	0.0	0.1	0.1	0.9	1.4
TC [d]	2651.99	1347.12	433.54	267.06	46.81	29.80
DM [e]	6782.75*		114.47*		40.03*	

Notes: [a] RMSE = root mean square error; [b] ME = mean error; [c] Decomposition of the mean square error: (i) %U1 = percentage of mean error (bias proportion of the MSE), (ii) %U2 = percentage of regression error (variance proportion of the MSE), (iii) %U3 = percentage of disturbance error (covariance proportion of the MSE); [d] TC = Theil coefficient; [e] DM = results of the Diebold-Mariano test.

[*] Significant at the 5% level.

Table 1 shows that the weighted balance (WB) obtains lower RMSE, ME and TC in all cases. Although the proportion of systematic error (U1) is not very different, the balance shows higher proportions of regression error (U2). As σ_{ε}^{2} increases, forecasting results tend to worsen for both methods. Nevertheless if we look at the results of the DM test, we can see that there is no significant difference between both methods.

However, in Table 2 and Table 3, when $\mu \neq 0$, the difference is significant and is always in favor of the weighted balance. Comparison of the results in Table 1 with those in Table 2 and Table 3 highlights several differences, in particular regarding the forecasting results as the value of φ increases from 0 to 0.9: while they worsen when $\mu \neq 0$, this effect is not clear when $\mu = 0$. Another difference is that when $\mu \neq 0$, as σ_{ε}^{2} increases the forecasting results improve for both methods.

These results suggest that taking into account the percentage of respondents expecting a variable to remain constant may improve the use of the balance statistic with forecasting purposes. Though it is impossible to completely eliminate the measurement error introduced when converting qualitative data on the direction of change into quantitative estimations of agents' expectations, the weighted balance shows lower measurement errors and better forecasts.

5. Conclusion

We purpose a variation of the balance statistic (weighted balance) in order to take into account the percentage of respondents expecting no change in the evolution of an economic variable. By means of a simulation experiment we test whether this variation of the balance statistic outperforms the balance statistic in order to track the evolution of agents' expectations and produces more accurate forecasts of the quantitative variable generated used as a benchmark. In all cases, the weighted balance outperforms the balance statistic and provides more accurate forecasts of the quantitative variable generated as a benchmark.

References

Anderson, O. (1951). Konjunkturtest und Statistik. *Allgemeines Statistical Archives*, **35**, 209-220.

Anderson, O. (1952). *The business test of the IFO-Institute for economic research, Munich, and its theoretical model. Revue de l'Institut International de Statistique,* **20**, 1-17.

Claveria, O., Pons, E. and Ramos, R. (2007). Business and consumer expectaions and macroeconomic forecasts. *International Journal of Forecasting,* **23**, 47-69.

Claveria, O., Pons, E. and Suriñach, J. (2006). Quantification of expectations. Are they useful for forecasting inflation? *Economic Issues,* **11**, 19-38.

Common, M. (1985). Testing for rational expectations with qualitative survey data. *Manchester School of Economic and Social Statistics,* **53**, 138-148.

Diebold, F.X. and Mariano, R. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, **13**, 253-63.

Lee, K.C. (1994). Formation of price and cost inflation expectations in British manufacturing industries: a multi-sectoral analysis. *Economic Journal,* **104**, 372-385.

Löffler, G. (1999). Refining the Carlson-Parkin method. Economics Letters **64**, 167-171.

Muth, J. (1961). Rational expectations and the theory of price movements. *Econometrica,* **29**, 315-335.

Nardo, M. (2004). The quantification of qualitative survey data: a critical assessment. *Journal of Economic Surveys,* **17**, 645-668.

Smith, J. and McAleer, M. (1995). Alternative procedures for converting qualitative response data to quantitative expectations: an application to Australian manufacturing. *Journal of Applied Econometrics,* **10**, 165-185.

In: Economic Forecasting
Editor: Alan T. Molnar, pp. 191-211
ISBN: 978-1-60741-068-3

Chapter 8

ANALYST ORIGIN AND THEIR FORECASTING QUALITY ON THE LATIN AMERICAN STOCK MARKETS*

Jean-François Bacmann[1,†] and Guido Bolliger[2,‡]
[1]RMF Investment Management, Quantitative Analysis, Huobstrasse 16, CH-8808 Pfäffikon SZ.
[2]Olympia Capital Management, 21-25 Rue Balzac, F-75008 Paris, France

Abstract

This paper investigates the relative performance of local, foreign, and expatriate financial analysts on Latin American emerging markets. We measure analysts' relative performance with three dimensions: (1) forecast timeliness, (2) forecast accuracy and (3) impact of forecast revisions on security prices. Our main findings can be summarized as follows. Firstly, there is a strong evidence that foreign analysts supply timelier forecasts than their peers. Secondly, analyst working for foreign brokerage houses (i.e. expatriate and foreign ones) produce less biased forecasts than local analysts. Finally, after controlling for analysts' timeliness, we find that foreign financial analysts' upward revisions have a greater impact on stock returns than both followers and local lead analysts forecast revisions. Overall, our results suggest that investors should better rely on the research produced by analysts working for foreign brokerage houses when they invest in Latin American emerging markets.

Keywords: analysts' forecasts, home bias, international diversification, emerging markets, herding behaviour.

JEL Classification: G14, G15, G24

---

* The first version of this article was entitled "Who are the best? Local versus foreign analysts on Latin American stock markets". It was written at a time when Bacmann was affiliated to the University of Neuchâtel and Bolliger to the University of Neuchâtel as well as the International Center FAME. The views expressed in this article are individual views of the authors and do not necessarily reflect their employers' opinions.

† E-mail address: Jean-Francois.Bacmann@rmf.Chapter. Phone: ++4155 417 77 10 . Fax: ++4155 417 77 11.
‡ E-mail address: guidobolliger@olympiagroup.com. Phone: ++331 4953 7426. Fax: ++331 4256 7009.

1. Introduction

Past research suggests that geographic proximity is related to information flow. However, the empirical evidence on the impact of geographic proximity on the quality of investors' information is mixed. Brennan and Cao (1997) report that US investors are less informed about foreign markets conditions than are local investors. Kang and Stulz (1997) find no evidence that foreign investors outperform in Japan. Using US mutual fund holdings, Coval and Moskowitz (2001) show that investors located near potential investments have significant informational advantages relative to the rest of the market. According to Choe et al. (2000), foreign investors on the Korean market are disadvantaged relative to domestic individual investors. Inversely, Seasholes (2000) reports that foreigners act like informed traders in emerging markets. He finds that foreign investors profits come from trading stocks of large firms with low leverage and liquid shares. Similarly, Grinblatt and Keloharju (2000) exihibit evidence that foreign investors on the Finnish stock market generate superior performance than local investors. It is likely that the previous mixed findings are driven by the information available to the investors. This is why our research does not focus on the relative performance of investors but on the relative performance of analysts located at the upstream side of them.

Research devoted to financial analyst forecast accuracy documents that some groups of analysts display a better forecasting ability than others. Stickel (1992) finds that Institutional Investor All-American analysts provide more accurate earnings forecasts and tend to revise their forecasts more frequently than other analysts. Clement (1999) investigates the origin of financial analysts differential accuracy. He documents a negative relationship between financial analysts relative accuracy and the complexity of their stock portfolio. On the other hand, he shows that analysts' performance improves with their age and that analysts working for big research houses with more resources available, outperform their peers. Agency problems such as corporate financing business conflicts, have also an impact on financial analysts' performance. Lin and McNichols (1999) and Michaely and Womack (1999) show that analysts whose employer is affiliated with a company through an underwriting relationship issue more optimistic forecasts than unaffiliated analysts.

The present paper is directly related to these two streams of research. The objective is to investigate the relative performance of local, expatriate, and foreign analysts on Latin American emerging markets. Local analysts are those who work for local research firms. Expatriate analysts work for foreign brokerage houses but are located in the country. Finally, foreign analysts work for foreign research firms with no local presence. Ex-ante, three main reasons may be at the origin of differential performance across the three groups of analysts: geographical distance, agency problems, and available resources.

Residence may give local and expatriate analysts several advantages compared to foreign ones. First, they may have a better knowledge of the local economy. Local economy has been shown to have a significant impact on emerging stock markets ;see Harvey (1995). Second, they may be more familiar with the institutional context in which the companies evolve. Institutional factors have a significant influence on the properties of financial analyst forecasts ; see Hope (2003). Third, they may have a better knowledge of the local culture. Finally, they may have a better human network in the country. This network may give them access to relevant private information. On the other hand, being closer from the analyzed firms, they may be more subject to agency problems such as conflict of interests. Foreign and expatriate analysts usually work for important international research firms. These big research firms have more resources available, they have the financial capacity to attract the best analysts, and their international expertise may help them to better anticipate international macro-economic fluctuations. Overall, if geographic proximity improves the quality of the information available to analysts, local and expatriate analysts should outperform their foreign counterparts. On the other hand, if the quantity of resources available to the analysts, their reputations as well as their expertise are the key determinant of their performance, foreign and expatriate analysts should outperform local ones. Finally, if conflict of interests, caused by tighter investment banking relationships between firms and banks having a local representation, have an important influence on the quality of financial analysts' output in these markets, foreign analysts should outperform both local and expatriate ones.

We conduct our investigation on Latin American markets for two reasons. First, due to geographical considerations, Latin American markets have always presented a great interest for US institutional investors. As a consequence, they create an important demand for financial analysts services on these markets. Second, as underlined by Choe et al. (2002), private information is likely to be more important on emerging stock markets than on developed ones.

We measure analysts' relative performance with three dimensions: (1) forecast timeliness, (2) forecast accuracy and (3) impact of forecast revisions on security prices. Our main findings can be summarized as follows. Firstly, there is a strong evidence that foreign analysts supply timelier forecasts than their peers. In particular, we detect a greater number of leaders among foreign analysts than among analysts with local residence. This finding suggests that both local and expatriate analysts have a tendency to revise their earnings forecasts in order to accommodate the opinions of foreign analysts. Secondly, analyst working for foreign brokerage houses (i.e. expatriate and foreign ones) produce less biased forecasts than local analysts. Lead foreign and expatriate analysts produce much more accurate forecasts than other analysts suggesting that leaders have an important informational advantage over other analysts. Finally, after controlling for analysts' timeliness, we find that foreign financial analysts' upward revisions have a greater impact on stock returns than both followers and local lead analysts forecast revisions. This suggests that the market considers forecast revisions provided by foreign leader analysts as being more informative than the revisions provided by their local counterparts.

Our research has important practical implication: investors should better rely on the research produced by analysts working for foreign brokerage houses when they invest in Latin American emerging markets. Moreover, our paper complements previous research in three ways. Firstly, we contribute to the literature on the importance of geography in economics by showing that location has an impact on the quality of the information provided by analysts. If foreign (local) investors rely mostly on foreign and expatriate (local) analysts' research in order to take their investment decisions, our results may explain the superior performance of foreign investors on some markets; see Seasholes (2000) and Grinblatt and Keloharju (2000). Secondly, by showing that analysts' location/affiliation has a significant impact on their forecast accuracy, we contribute to the large amount of literature which investigates the origins of financial analysts forecasts' bias. Thirdly, we complement, and somehow contradict, the recent research which also investigate the impact of analysts' location on forecast accuracy. Malloy (2003), Chang (2003), and Orpurt (2002) document that analysts located closer to the companies they follow make more accurate forecasts than their more distant counterparts. As underlined by Kini et al. (2003), the almost opposite conclusion drawn from our investigation may be due to differences in the industrial structure of the countries examined in these different papers. If, in Latin America, a good understanding of the sectors is a major determinant of forecast accuracy, a foreign (and to some extent an expatriate) analyst who focuses on a sector in multiple countries may have an advantage over a local analyst who focuses on multiple local firms across multiple sectors. Of course, the reverse may be true for other markets. This shows that the conclusions drawn from these studies may not be generalized to all countries.

The paper proceeds as follows. Section 2 presents the data used in this study. Section 3 investigates the relative timeliness of financial analysts. Section 4 tests for differences in forecast accuracy. Section 5 examines the impact of forecast revisions on security prices; and Section 6 concludes.

2. Data and Overview Statistics

The analysts' forecasts[1] are provided by Institutional Broker Estimate System (I/B/E/S) for 7 Latin American emerging markets: Argentina, Brazil, Chile, Colombia, Mexico, Peru and Venezuela. One year earning per share (EPS) forecasts are used from 1993 to 1999. We use the Nelson Directory of Investment Research to classify financial analysts. The Nelson Directory of Investment Research provides the name and the coordinates of each analyst that follows a particular company. Financial analysts who work for local brokerage houses are classified as local, those who work for foreign brokerage houses with residence in the country as classified as expatriate, and those who work for foreign brokerage houses without residence in the country are classified as foreign. Stock prices are extracted from Datastream. To be included in the sample, a forecast should meet the following conditions:

1. Realized EPS has to figure in the I/B/E/S Actual File.
2. The forecast must be issued between the end of previous fiscal year and current year earning reporting date.
3. The forecast must be issued by an analyst listed in the Nelson Directory of Investment Research.
4. The company for which the forecast is issued must be followed by at least 3 analysts of each group during a given year.

The last condition restricts the sample to big and medium-sized companies. The final sample includes 61'209 EPS forecasts. Table 1 shows that local analysts have produced 59% more forecasts than their foreign counterparts and more than twice much forecasts than expatriate analysts. The number of analysts and brokerage houses active on Latin American markets has sensibly increased between 1993 and 1999. This is due to the increasing coverage of the I/B/E/S database but also to the increasing attractiveness of these markets for foreign investors.

[1] Note that we make no distinction between individual analysts and team of analysts.

Table 1. Summary statistics by year

Year	No. Forecasts			No. Analysts			No. Brokers			No. Stocks
	Local	Foreign	Expatriate	Local	Foreign	Expatriate	Local	Foreign	Expatriate	
1993	1670	783	432	74	56	41	35	18	10	84
1994	4937	2345	2263	114	84	87	49	36	16	208
1995	4999	2526	1989	236	123	122	51	32	20	200
1996	4764	2864	1899	257	163	147	57	37	17	180
1997	5229	3888	2056	245	238	170	56	33	16	212
1998	4508	3694	2141	244	232	175	50	24	15	205
1999	3674	2624	1924	182	176	148	41	19	11	170
Total	29781	18724	12704	719	584	365	93	61	27	351

This table reports yearly statistics for the data. No. Forecasts represents the number of annual earnings forecasts made each year. No. Analyst represents the number of analysts who produced a forecast during the fiscal year t. The total number of analysts who produced an earning forecast during the entire period is indicated in the last row. No. Brokers represents the number of banks (or brokerage companies) for which analysts work each year. The total number of brokers identified during the entire period is indicated in the last row. No. Stocks is the number of firms in the sample. The total number of firms for which forecasts were produced during the period is indicated in the last row.

Table 2. Summary statistics by country

Country	No. Forecasts			No. Analysts			No. Brokers			No. Stocks
	Local	Foreign	Expatriate	Local	Foreign	Expatriate	Local	Foreign	Expatriate	
Argentina	5114	2685	1835	135	215	86	22	36	9	45
Brazil	11897	7238	6349	293	244	191	30	31	19	160
Chile	2224	1530	697	67	150	39	11	25	4	29
Colombia	160	364	174	6	43	15	2	17	2	11
Mexico	12905	7700	3753	242	286	128	21	35	12	82
Peru	651	927	226	27	111	27	7	32	3	17
Venezuela	110	279	97	1	66	15	1	18	2	7

This table reports statistics by country and by industry. *No. Forecasts* represents the number of annual earnings forecasts made each year. *No. Analyst* represents the number of analysts who produced a forecast during the fiscal year *t*. *No. Brokers* represents the number of banks (or brokerage companies) for which analysts work in each country. *No. Stocks* is the number of firms in the sample.

Table 2 shows that most of the forecasts (81%) are concentrated on Brazil and Mexico. In addition, in each country excepting Brazil, foreign analysts tend to be more numerous than local and expatriate ones. However, from Table 1, we see that this finding is reversed at the aggregated level: local analysts are more numerous than foreign ones and the difference between foreign and expatriate is smaller. Thus, foreign analysts tend to follow several different markets while local and expatriate analysts are more focused on specific markets.

Non-tabulated results indicate that the average number of analysts employed by foreign brokerage houses amounts to 7.9 while it amounts to 5.5 for local ones suggesting that, on average, foreign brokerage houses are bigger than local ones. Our sample contains 91 companies out of 351 that have quoted American Depositary Receipts (ADR). Lang et al.

(2002) show that non-U.S. companies listed on U.S. exchanges have richer informational environment than other non-U.S. firms. Therefore, we will control for ADR listing in the subsequent analysis.

Table 3 shows that expatriate analysts are the less active ones. On average, they produce a forecast every 77 day while their foreign and local peers do it every 73, respectively 71 day. Similarly, expatriate analysts revise less frequently than their counterparts: on average 1.33 times per firm each year against 1.92 times for foreigners and 2.45 for locals. Although the frequency of forecast revisions gives an insight on the activity of financial analysts, this does not indicate that more active analysts have advantages in collecting and processing information. They may simply change their mind several times to accommodate the opinions of others. Therefore, in the subsequent section, we propose to measure analysts' relative activity with their timeliness.

Table 3. Frequency of forecast issuance and revision

Panel A: number of calendar days elapsed between forecasts				
	Mean	**Min**	**Median**	**Max**
Local analysts	70.63	1.00	59.00	344.00
Foreign analysts	73.00	1.00	59.75	372.00
Expatriate analysts	77.29	1.00	65.75	362.00
Panel B: number of revisions per analyst				
	Mean	**Min**	**Median**	**Max**
Local analysts	2.45	0.00	1.00	50.00
Foreign analysts	1.92	0.00	1.00	19.00
Expatriate analysts	1.33	0.00	1.00	11.00

This table reports summary statistics on financial analysts' activity. Panel A presents statistics about the number of calendar days that separate two consecutive forecasts by analyst for a particular company in a given year. Panel B reports statistics on the number of revisions by analyst for a particular company in a given year.

3. Analysts' Timeliness

3.1. Empirical Design

Cooper, Day and Lewis (2001, thereafter CDL) show that timely analysts' (leaders) forecast revisions provide greater value to investors than other analysts' (followers) forecasts. They argue that timeliness is an important and necessary indicator of financial analysts' relative performance. Using forecast accuracy alone to assess the relative performance of financial analysts can lead to misclassification errors because less informed analysts can improve the accuracy of their forecasts by simply mimicking timely skilled analysts.

The leader to follower ratio (*LFR*) developed by CDL is used to distinguish leaders from followers.[2] This ratio is computed for each analyst/firm/year unit. It is distributed as

[2] A precise description of the *LFR* computation methodology is given in the Appendix.

$F_{(2KH,2KH)}$,[3] where H is the number of other analysts following a particular firm in a given year and K is the total number of forecasts provided by the analyst during the year for that firm. Similar to CDL, analysts having LFR significantly greater than 1 at the 10% level are considered as leaders. Moreover, each analyst is required to produce at least 3 forecasts per year for the firm under consideration. As mentioned CDL, this restriction minimizes the possibility for an analyst to be classified as leader thanks to a single lucky forecast.

In order to test whether a group (local or foreign) tends to lead the other one, we compare the number of local leaders to the foreign ones. However, since the total number of analysts is different between the 2 groups, such a comparison is not directly possible. Thus, the proportion of leaders in a given group g, L_g, is compared to the proportion of analysts in group g in the sample, P_g. In order to determine whether a group of analysts has significantly more (less) leaders than its proportion in the population suggests, we test the following hypothesis:

$$H_0 : L_g = P_g \ vs \ H_1 : L_g \neq P_g .$$

Consequently, the following normally distributed statistic is computed:

$$Time_g = \frac{\left(L_g - P_g\right)}{\sqrt{P_g \cdot \left(1 - P_g\right)}} \cdot \sqrt{N} ,$$

where:

$$L_g = \frac{Number\ of\ leaders\ in\ group\ g}{Total\ number\ of\ leaders},$$

$$P_g = \frac{Number\ of\ analysts\ from\ group\ g}{N},$$

$$N = Total\ number\ of\ analysts .$$

Table 4. Financial analysts' timeliness

Panel A: LFR for Latin America						
	Analysts' origin	**No. observations**	**No. leaders**	**% leaders**	**% observations**	**Difference**
		N		L_g	P_g	
Latin America	Local	5599	621	47.7	49.6	-1.9***
	Foreign	3457	444	34.1	30.6	3.5***
	Expatriate	2226	236	18.1	19.7	-1.6***

[3] CDL derive the distribution of the *LFR* by assuming that the time elapsed between the arrival of two subsequent revisions follows an exponential distribution.

Table 4. Continued

Panel B: LFR by country

Country	Analysts' origin	No. observations	No. leaders	% leaders	% observations	Difference
		N		L_g	P_g	
Argentina	Local	938	90	45.9	53.6	-7.7***
	Foreign	476	62	31.6	27.2	4.4***
	Expatriate	337	44	22.4	19.2	3.2***
Brazil	Local	1948	231	44.6	46.5	-1.9***
	Foreign	1247	176	34.0	29.7	4.2***
	Expatriate	998	111	21.4	23.8	-2.4***
Chile	Local	315	26	43.3	48.1	-4.8***
	Foreign	246	29	48.3	37.6	10.8***
	Expatriate	94	5	8.3	14.4	-6.0***
Colombia	Local	4	0	0.0	20.0	-20.0***
	Foreign	12	3	100.0	60.0	40.0***
	Expatriate	4	0	0.0	20.0	-20.0***
Mexico	Local	2323	264	52.6	52.0	0.6
	Foreign	1388	163	32.5	31.0	1.4***
	Expatriate	760	75	14.9	17.0	-2.1***
Peru	Local	65	10	58.8	44.8	14.0***
	Foreign	58	6	35.3	40.0	-4.7
	Expatriate	22	1	5.9	15.2	-9.3***
Venezuela	Local	6	0	0.0	12.8	-12.8***
	Foreign	30	5	100.0	63.8	36.2***
	Expatriate	11	0	0.0	23.4	-23.4***

This table reports the number of analysts identified as leaders as well as the test of the null hypothesis, which is stating that the proportion of leaders in a given group equals the proportion of analysts from the given group in the total sample. The last column represents the difference between the percentage of leaders in a given group, L_g, and the percentage of analysts from the given group, P_g. The significance of this difference is determined by the following normally distributed statistic: $Time_g = \frac{(L_g - P_g)}{\sqrt{P_g \cdot (1-P_g)}} \cdot \sqrt{N}$.

Panel A reports results for all Latin American markets. Panel B reports results by country
***, **, * denote significance at the 1%, 5%, and 10% levels, respectively.

3.2. Results for Analysts' Timeliness

According to the LFR statistic, 1301 leaders out of 11282 observations are detected. Table 4 shows the breakdown of the leaders according to their origin. The proportions of local and expatriate analysts within the leaders are significantly smaller than their proportions within the full sample.[4] On the other hand, there are more leaders among foreign analysts than their proportion in the sample would suggest. These results indicate that, on average, foreign analysts lead while local and expatriate analysts herd. Analysts with local residence have a tendency to issue their forecasts shortly after foreign analysts and their revisions do not induce other analysts to revise their own forecasts.

[4] The inverse is automatically true for foreign leaders.

Panel B of Table 4 shows the breakdown of the leaders across the different countries. The individual country results are consistent with those obtained for Latin America. The exceptions are Brazil and Peru. In Brazil, the proportion of expatriate analysts identified as leaders is significantly more important than their proportion in the population. The same is true for local analysts in Peru.

In summary, the above results indicate that foreign analysts have a greater tendency to lead than analysts with local residence. This holds at the aggregated level as well as for most of the individual stock markets. The implications of these findings in terms of forecast accuracy and earnings forecasts' informativeness are investigated in the following two sections.

4. Forecast Accuracy

4.1. Empirical Design

Forecast accuracy is the most widely used measure of the quality of an analyst's research. Indeed, the more accurate earnings forecast is, the more accurate the price extracted from any valuation model will be. Forecast accuracy is measured using the average percentage forecast error adjusted for the horizon bias.[5] Analyst i 's percentage forecast error at date t is,

$$FE_{ijt} = \frac{FEPS_{it} - EPS}{|EPS|},$$

where:

$FEPS_{it} =$ analyst i 's EPS forecast for company *j* at date t,

$EPS =$ reported earning per share at the end of the forecast horizon.

In order to correct for the horizon bias, CDL forecast accuracy regression is used. Compared to the matching forecasts methodology used by Stickel (1992), this operation is much less data-consuming and better suited for our study. Each FE_{ijt} is regressed on the length of time from forecast release to earning announcement date. The residuals from this regression are used to measure forecast accuracy. Formally,

$$FE_{ijt} = \alpha + \beta \cdot T + \varepsilon_{ijt}, \tag{1}$$

where:

$T =$ number of days until the earnings announcement date,

$\varepsilon_{ijt} =$ residual forecast error for analyst i on firm j at date t.

---

[5] Prior studies such as Kang, O'Brien and Sivaramkarishnan (1994) show that forecast bias increases with forecast horizon.

The relative accuracy of each group of analysts is computed in three successive steps. First, for a given firm, the average residual forecast error is computed for each analyst,

$$MFE_{ij} = \sum_{t=1}^{K} \left| \varepsilon_{ijt} \right| \Big/ K \,,$$

where:

$MFE_{ij} =$ mean forecast error by analyst i for firm j,

$K =$ number of forecasts issued by analyst i for firm j during a given year.

Second, for each firm/year, individual analysts' mean forecast errors are averaged over all analysts of a given group g,

$$MGFE_{gj} = \sum_{i \in g} MFE_{ij} \Big/ N_j^g \,,$$

where:

$MGFE_{gj} =$ mean group forecast error for firm j,

$N =$ number of analysts from group g following firm j during a given year.

Finally, the mean difference forecast error between 2 groups is computed as

$$MDFE = \sum_{j=1}^{J} \left[MGFE_{Aj} - MGFE_{Bj} \right] \Big/ J$$

where J is the number of company/year units. In order to assess whether one group of analysts produces more (less) accurate forecasts than the other, the following hypothesis is tested:

$$H_0 : MDFE = 0 \; vs \; H_1 : MDFE \neq 0 \,.$$

A parametric mean test, a Wilcoxon sign rank test of equality of medians as well as a non-parametric binomial sign test are performed to test the hypothesis.

4.2. Results for Forecast Accuracy

The slope coefficient of equation (1) equals 0.01 and is significantly different from zero.[6] Emerging market analysts' bias decreases significantly with the distance between forecast release date and earnings announcement date. The intercept is not statistically different from zero.

[6] Results are not shown. They are available on request by the authors.

Hypothesis tests and descriptive statistics for the mean difference forecast errors ($MDFE$) are reported in Table 5. Panels A through C report the difference across each category of analysts for all Latin American countries, for individual countries as well as for different security categories. The distribution of the MDFEs appears to be highly skewed by the presence of some extreme observations. As this may bias the results of the parametric tests, we will only consider the non-parametric results of column 7 and 8 for our analysis and conclusions.

Panel A shows that the median MDFE is positive for the whole sample and each individual countries indicating that local analysts' average forecast error is greater than foreign analysts' one. The Wilcoxon sign rank test and the binomial test reject the null hypothesis of equal forecasting skills at the aggregate level as well as for five of the individual countries. The superior ability of foreign analysts to predict firms earnings does not depend on size. Surprisingly, this superior ability disappears for American Depositary Receipts, which have a richer information environment and are the least distant firms for foreign analysts. Conflicts of interest due to increased investment and commercial banking relationship with foreign banks following U.S. exchange listing may explain this finding.

Table 5. Financial analysts relative forecast accuracy

Distribution of the Mean Difference Forecast Errors (*MDFE*)									**Sign of *MDFE***	
Panel A: Difference in forecast accuracy between local and foreign analysts										
Sample	**N**	**Mean**		**Stdev**	**Min**	**Median**		**Max**	**% Local > Foreign**	
Latin America	1263	-0.16		7.98	-238.88	0.02	***	62.10	54.95	***
Argentina	191	-0.01		0.88	-9.22	0.02	**	4.59	58.64	***
Brazil	557	-0.55		11.68	-238.88	0.02		47.11	53.68	**
Mexico	332	0.22		3.52	-4.97	0.02		62.10	53.31	
Chili	112	0.11	**	0.50	-1.96	0.03	**	3.31	55.36	
Peru	38	0.23		0.98	-0.42	0.05		5.46	60.53	*
Colombia	21	0.32	*	0.83	-0.38	0.10	**	3.54	71.43	**
Venezuela	12	-0.06		0.43	-1.13	0.02		0.36	50.00	
High Market Value	493	-0.01		6.54	-121.27	0.02	**	62.10	54.56	**
Small Market Value	323	0.16		2.25	-9.22	0.02	*	35.44	53.25	
ADR	277	0.04		0.88	-5.35	0.02		10.51	53.07	
Panel B: Difference in forecast accuracy between local and expatriate analysts										
Sample	**N**	**Mean**		**Stdev**	**Min**	**Median**		**Max**	**% Local > Expatriate**	
Latin America	1263	0.61	*	12.89	-20.69	0.01	**	402.14	52.26	*
Argentina	191	0.05		0.89	-4.24	0.02		10.27	53.40	
Brazil	557	1.04		18.48	-20.69	0.01		402.14	51.35	
Mexico	332	0.45		7.65	-13.31	0.01		136.34	52.41	
Chili	112	0.08	*	0.46	-0.76	0.00		3.03	50.89	
Peru	38	0.29	*	1.03	-0.65	0.08	*	5.59	60.53	*
Colombia	21	0.24		0.80	-0.36	0.02		3.20	61.90	*
Venezuela	12	-0.09		0.33	-0.75	-0.07		0.55	41.67	
High Market Value	493	0.63		9.60	-3.35	0.00		163.62	49.49	
Small Market Value	323	0.00		2.65	-20.69	0.02		35.23	52.94	
ADR	277	0.05		0.67	-3.61	0.01		6.81	52.71	

Table 5. Continued

Panel C: Difference in forecast accuracy between expatriate and foreign analysts

Sample	N	Mean	Stdev	Min	Median	Max	% Expatriate > Foreign
Latin America	1263	-0.77	18.81	-641.02	0.00	19.46	50.75
Argentina	191	-0.06	1.35	-12.48	0.01	4.41	52.88
Brazil	557	-1.59	28.10	-641.02	-0.01	19.46	48.29
Mexico	332	-0.23	4.45	-74.23	0.01	18.11	52.41
Chili	112	0.03	0.38	-2.58	0.05 **	0.91	58.04 **
Peru	38	-0.06	0.22	-0.45	-0.08 **	0.53	34.21 **
Colombia	21	0.08	0.22	-0.25	0.04	0.50	52.38
Venezuela	12	0.03	0.47	-1.28	0.10	0.52	66.67 *
High Market Value	493	-0.65 *	8.29	-120.94	0.01	3.71	52.54
Small Market Value	323	0.16	1.97	-11.16	0.01	19.46	51.39
ADR	277	-0.02	0.67	-4.70	0.01	3.71	52.71

Panel D: Difference in forecast accuracy between leaders and followers

Sample	N	Mean	Stdev	Min	Median	Max	% Leaders > Followers
Local Leaders	426	-0.04	1.31	-11.04	-0.05 ***	17.08	39.20 ***
Foreign Leaders	350	-3.74	70.33	-1315.50	-0.58 ***	12.09	40.57 ***
Expatriate Leaders	198	0.07	6.13	-40.74	-0.02	73.38	44.95 *

This table presents descriptive statistics as well as hypothesis tests for the Mean Difference in Forecast Errors (*MDFE*). In Panel A, the third column reports the average difference between local analysts' forecast errors and foreign analysts' forecast errors. Column 6 reports the median difference between local analysts' forecast errors and foreign analysts' forecast errors. Column 8 reports the percentage of firm/year units for which the average forecast error of local analysts was greater than the average forecast error of foreign ones. In Panel B, the third column reports the average difference between local analysts' forecast errors and expatriate analysts' forecast errors. Column 6 reports the median difference between local analysts' forecast errors and expatriate analysts' forecast errors. Column 8 reports the percentage of firm/year units for which the average forecast error of local analysts was greater than the average forecast error of expatriate ones. In Panel C, the third column reports the average difference between expatriate analysts' forecast errors and foreign analysts' forecast errors. Column 6 reports the median difference between expatriate analysts' forecast errors and foreign analysts' forecast errors. Column 8 reports the percentage of firm/year units for which the average forecast error of expatriate analysts was greater than the average forecast error of foreign ones. In Panel D, the third column reports the average difference between lead analysts' forecast errors and follower analysts' forecast errors. Column 6 reports the median difference between lead analysts' forecast errors and follower analysts' forecast errors. Column 8 reports the percentage of firm/year units for which the average forecast error of lead analysts was greater than the average forecast error of follower ones. A parametric mean test is performed on column 3 numbers, a Wilcoxon signed rank test of equality of medians is performed on column 6 numbers, and a non-parametric sign test is performed on column 8 numbers. Note that in Panel D, the total number of firm/year units for each group of leader is lower than the number of leaders that has been identified. This is explained by the fact that there can be several leaders for a particular company in a given year.

***, **, * denote significance at the 1%, 5%, and 10% levels, respectively.

The results in panel B indicate that, excepting for Venezuela, the average error is greater for local analysts forecasts than for expatriate ones. The difference between both groups of analysts is statistically significant at the Latin American level but only weakly or not significant at the country and security category levels.

As indicated by the results in panel C, no difference between the forecasting skills of expatriate and foreign analysts can be found. As reported, in panel D, there is a strong evidence that leaders produce more accurate forecasts than follower analysts. Their mean forecast error appears to be much smaller than that of follower analysts. This is particularly true for local and foreign leaders for which the null hypothesis is rejected at the 1% level. The leader-follower criterion appears more important than the geographical one. However, no comparison is performed across leaders from each analyst group as the number of firm/year units for which leaders of both types are simultaneously identified is very low. Two important conclusions can be drawn about the behavior of financial analysts on Latin American markets. First, contrary to what has been documented by CDL, leader analysts do not "trade accuracy for timeliness". Indeed, foreign analysts are able to release timelier and more accurate forecasts. Second, follower analysts do not exactly reproduce the earnings per share forecasts issued by leader analysts. Even if their forecast releases closely follow leader analysts' ones, they avoid to reproduce exactly the information released by leader analysts.

Overall, this section shows that foreign analysts foreign analysts have a better ability to analyze Latin American firms' earnings potential than their local peers. There is no significant difference between the performance of foreign and expatriate analysts and only a weak difference between expatriate and local analysts in the favor of expatriate ones. These finding indicate that analysts who work for foreign institutions may have greater resources, expertise and/or talent than their local peers. Finally, timely analysts are the most accurate ones. Consequently, lead analysts do not give up forecast accuracy when releasing more timely forecasts.

5. Impact of Forecast Revisions on Security Prices

5.1. Empirical Design

This section investigates whether one group of analysts' revisions provides more information to investors. The objective is to determine whether the stock price reaction following forecast revisions differs between the different groups of analysts. The reaction around forecast revisions for a given firm is proxied by the cumulative excess return during the forecast release period (days 0 and +1). This cumulative excess return is computed as the difference between the buy-and-hold returns for the firm's common stock and the value-weighted Datastream country index.

The incremental information content of each revision is measured by the scaled distance relative to the consensus forecast.[7] More precisely:

$$FSUR_{ijt} = \frac{FEPS_{ijt} - CF_{jt-1}}{\sigma(CF_{jt-1})}$$

where:

$FSUR_{ijt}$ = forecast surprise following analyst i 's revision for firm j at date t,

CF_{jt-1} = consensus EPS forecast for firm j at date $t-1$,

$\sigma(CF_{jt-1})$ = standard deviation of the consensus forecast[8] at date $t-1$.

---

[7] Our results are not sensitive to the choice of the scaling factor.

The consensus forecast is based on the average of the forecasts issued by analysts (excluding analyst i) during the 2 months preceding date t. Each analyst is required to provide at least 3 forecasts per year for the firm and each consensus forecast is required to contain at least 2 individual forecasts.

The impact of forecast revisions on security prices is measured by the following cross-sectional regression equations:

$$CAR_{jt} = \beta_0 + \beta_1 FSUR_{ijt} + \beta_2 LOC_i + \beta_3 FOR_i + \beta_4 LNSIZE_{jt} + \varepsilon_{jt} \quad (2)$$

$$\begin{aligned} CAR_{jt} = {} & \beta_0 + \beta_1 LOC_i \times FSUR_{ijt} + \beta_2 FOR_i \times FSUR_{ijt} + \beta_3 EXPAT_i \times FSUR_{ijt} \\ & + \beta_4 LNSIZE_{jt} + \varepsilon_{jt} \end{aligned} \quad (3)$$

$$\begin{aligned} CAR_{jt} = {} & \beta_0 + \beta_1 FSUR_{ijt} + \beta_2 LOCLEAD_{ij} + \beta_3 FORLEAD_{ij} + \beta_4 EXPATLEAD_{ij} \\ & + \beta_5 LNSIZE_{jt} + \varepsilon_{jt} \end{aligned} \quad (4)$$

$$\begin{aligned} CAR_{jt} = {} & \beta_0 + \beta_1 LOCLEAD_{ij} \times FSUR_{ijt} + \beta_2 FORLEAD_{ij} \times FSUR_{ijt} \\ & + \beta_3 EXPATLEAD_{ij} \times FSUR_{ijt} + \beta_4 FOL_i \times FSUR_{ijt} + \beta_5 LNSIZE_{jt} + \varepsilon_{jt} \end{aligned} \quad (5)$$

where:

$CAR_{jt} =$ cumulative excess return for firm j during the forecast release period (days 0 and +1),

$LOC_i =$ dummy variable set to 1 if analyst i is a local one and 0 otherwise,

$FOR_i =$ dummy variable set to 1 if analyst i is foreign and 0 otherwise,

$LNSIZE_j =$ logarithm of the market value (in USD) of common stock at fiscal year end,

$EXPAT_i =$ dummy variable set to 1 if analyst i is an expatriate and 0 otherwise,

$LOCLEAD_{ij} =$ dummy variable set to 1 if analyst i is a local analyst that has been identified as leader for company j and 0 otherwise,

$FORLEAD_{ij} =$ dummy variable set to 1 if analyst i is a foreign analyst that has been identified as leader for company j and 0 otherwise,

$EXPATLEAD_{ij} =$ dummy variable set to 1 if analyst i is an expatriate analyst that has been identified as leader for company j and 0 otherwise,

$FOL_{ij} =$ dummy variable set to 1 if analyst i has been identified as follower for company j and 0 otherwise,

[8] Similar to Stickel (1992), a standard deviation less than 0.25 is arbitrarily set to 0.25 to mitigate small denominators. Our results are not affected by this operation.

Equations (2) and (4) measure the abnormal return associated with the different groups of analysts' forecast revisions. Equation (3) and (5) measures the proportion of abnormal return explained by each group of analysts' forecast revisions. The size variable is a proxy for the differences in firms' information environment[9] but also for foreign investors' ownership since they tend to concentrate their investments on high-capitalization liquid firms.

5.2. Results for the Impact of Forecast Revisions on Security Prices

Table 6 reports the mean cumulative abnormal return during the forecast release period. The price reaction depends on the size of the revision. The cumulative abnormal returns display important standard deviations and consequently only the stock returns associated with the bottom 50% sub-sample display statistically significant price reactions. Conversely, other revisions do not impact on prices. This is consistent with Stickel (1992, 1995) who documents a non-linear relation between forecast revisions and price reactions. Therefore, the regressions are restricted to revisions of a given magnitude.

Results for the cross-sectional regressions (2) and (3) are reported in table 7. First, panel A results indicate that, following downward revisions, there is no difference in the average size of the stock price reaction across groups. On the other hand, following large upward revisions (top 10%), the average cumulative abnormal return is significantly smaller for local analysts than for expatriate ones. The same is true for foreign analysts but the regression coefficient is only marginally significant. Second, results reported in panel B indicate that the stock price reaction following analysts forecast revisions is only significant for expatriate analysts large upward revisions. Unfortunately, the null hypothesis of equality across coefficients cannot be rejected by the F-tests presented in columns 7 through 9.

Table 6. Stock price reactions following forecast surprises

	All FSUR		Bottom 10%	Bottom 50%		Top 50%	Top 10%
Mean (%)	-0.07	**	-0.16	-0.12	**	-0.02	-0.17
Standard deviation (%)	4.39		4.73	4.52		4.27	4.42
N	16699		1670	8352		8347	1670

This table reports some descriptive statistics about the cumulative abnormal returns (CARs) following forecasts' revisions. Cumulative abnormal returns are computed as the difference between the buy-and-hold return for the firm's common stock and the value-weighted Datastream country index during the forecast release period (days 0 and 1). The column All FSUR reports statistics on CARs for all forecast surprise level. Bottom 10% reports CARs for forecast surprises located in the top 10% of the distribution. Bottom 50% reports statistics for CAR's located in the bottom 50% of the distribution. In the column Top 50%, statistics are reported for CAR's located in the top 50% of the distribution. Top 10% reports statistics for CAR's located in the top 10% of the distribution.

***, **, * denote significance at the 1%, 5%, and 10% levels, respectively.

[9] Stickel (1995), among others, reports that buy and sell recommendations induces a greater price reaction for smaller companies than for larger ones.

Table 7. Stock price reactions following analyst forecast revisions

Panel A: $CAR_{jt} = \beta_0 + \beta_1 FSUR_{ijt} + \beta_2 LOC_i + \beta_3 FOR_i + \beta_4 LNSIZE_{jt} + \varepsilon_{jt}$

FSUR Cutoff	β_0	β_1	β_2	β_3	β_4	N
Bottom 10%	-0.46	0.09	-0.24	0.02	0.08	1670
	(-0.60)	(0.63)	(-0.77)	(0.05)	(0.88)	
Bottom 50%	-0.59 *	0.07	0.07	0.17	0.06	8352
	(-1.91)	(1.03)	(0.51)	(1.17)	(1.55)	
Top 50%	0.03	-0.04	-0.20	-0.27 **	0.02	8347
	(0.09)	(-0.56)	(-1.57)	(-1.97)	0.58	
Top 10%	1.07	0.20	-0.60 **	-0.52 *	-0.16 *	1670
	(1.35)	(1.42)	(-2.13)	(-1.68)	(-1.73)	

Panel B: $CAR_{jt} = \beta_0 + \beta_1 LOC_i \times FSUR_{ijt} + \beta_2 FOR_i \times FSUR_{ijt} + \beta_3 EXPAT_i \times FSUR_{ijt} + \beta_4 LNSIZE_{jt} + \varepsilon_{jt}$

FSUR Cutoff	β_0	β_1	β_2	β_3	β_4	$\beta_1 = \beta_2$	$\beta_1 = \beta_3$	$\beta_2 = \beta_3$	N
Bottom 10%	-0.56	0.13	0.05	0.06	0.08	0.40	0.25	0.01	1670
	(-0.76)	(0.86)	(0.30)	(0.35)	(0.87)				
Bottom 50%	-0.50 *	0.13	0.13	0.03	0.06	0.99	0.56	0.02	8352
	(-1.73)	(1.52)	(1.52)	(0.25)	(1.55)				
Top 50%	-0.14	-0.11	-0.03	0.12	0.02	0.42	2.89 *	1.04	8347
	(-0.51)	(-1.25)	(-0.29)	(0.98)	(0.53)				
Top 10%	0.64	0.13	0.22	0.35 **	-0.16	0.46	2.24	0.65	1670
	(0.84)	(0.81)	(1.36)	(1.99)	(-1.74)				

This table presents the coefficients obtained by regressing the cumulative abnormal returns following forecast revisions on the magnitude of the revision, firm size, and dummy variables indicating analysts' status. Revisions are dated within the firm's current fiscal year over the 1993-1999 period. CAR_{jt} is the cumulative abnormal return to security i during the release period (days 0 and +1). $FSUR_{ijt}$ is the forecast surprise following analyst i's revision at date t. $LNSIZE_{jt}$ is the natural logarithm of the market value (in USD) of common stock at fiscal year end. LOC_i is a dummy variable that takes a value of 1 if analyst i is employed by a local brokerage house and 0 otherwise. FOR_i is a dummy variable that takes a value of 1 if analyst i is employed by a foreign brokerage house without local residence and 0 otherwise. $EXPAT_i$ is a dummy variable that takes a value of 1 if analyst i is employed by a foreign brokerage house with local residence and 0 otherwise. All coefficients are multiplied by 100. T-statistics are based on White (1980). For each regression the adjusted R^2 are less than 0.01.

***, **, * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 8 reports the results for the cross-sectional regressions (4) and (5). First, as reported in panel A, the average cumulative return does not differ between leader and follower analysts. Second, as indicated by panel B results, there is a significant market reaction following foreign and expatriate leaders large upward revisions (top 10%). The F-tests indicate that the regression coefficients associated to foreign leaders' revisions are significantly higher than those associated to local leaders and followers' revisions.

Table 8. Stock price reactions following leaders and followers forecast revisions

Panel A: $CAR_{jt} = \beta_0 + \beta_1 FSUR_{ijt} + \beta_2 LOCLEAD_{ij} + \beta_3 FORLEAD_{ij} + \beta_4 EXPATLEAD_{ij} + \beta_5 LNSIZE_{jt} + \varepsilon_{jt}$

FSUR Cutoff	β_0	β_1	β_2	β_3	β_4	β_5	N
Bottom 10%	-0.56	0.09	-0.63	0.52	-0.61	0.09	1670
	(-0.76)	(0.64)	(-1.19)	(0.86)	(-0.59)	(0.92)	
Bottom 50%	-0.48	0.07	-0.24	-0.03	-0.30	0.06	8352
	(-1.66)	(1.05)	(-1.15)	(-0.10)	(-0.72)	(1.57)	
Top 50%	-0.17	-0.03	0.17	-0.33	0.22	0.02	8347
	(-0.62)	(-0.52)	(0.78)	(-1.17)	(0.64)	(0.64)	
Top 10%	0.60	0.20	-0.18	0.78	0.97	-0.16 *	1670
	(0.79)	(1.38)	(-0.40)	(1.24)	(1.23)	(-1.72)	

Panel B: $CAR_{jt} = \beta_0 + \beta_1 LOCLEAD_{ij} \times FSUR_{ijt} + \beta_2 FORLEAD_{ij} \times FSUR_{ijt} + \beta_3 EXPATLEAD_{ij} \times FSUR_{ijt} + \beta_4 FOL_{ij} \times FSUR_{ijt} + \beta_5 LNSIZE_{jt} + \varepsilon_{jt}$

FSUR Cutoff	β_0	β_1	β_2	β_3	β_4	β_5	$\beta_1 = \beta_2$	$\beta_1 = \beta_3$	$\beta_1 = \beta_4$	$\beta_2 = \beta_3$	$\beta_2 = \beta_4$	$\beta_3 = \beta_4$	N
Bottom 10%	-0.55	0.24	0.04	0.46	0.07	0.08	0.26	0.23	0.43	0.77	0.01	1.04	1670
	(-0.75)	(0.76)	(0.14)	(1.05)	(0.51)	(0.86)							
Bottom 50%	-0.50 *	0.26	0.12	0.42	0.05	0.06	0.19	0.14	0.93	0.50	0.07	1.19	8352
	(-1.73)	(1.03)	(0.46)	(1.04)	(0.68)	(1.56)							
Top 50%	-0.14	-0.11	0.33	0.34	-0.05	0.02	1.41	1.32	0.05	0.00	1.64	1.48	8347
	(-0.53)	(-0.45)	(1.11)	(1.26)	(-0.78)	(0.55)							
Top 10%	0.64	0.10	0.88 ***	0.60 **	0.17	-0.16 *	3.60 *	1.35	0.08	0.35	4.41 **	1.44	1670
	(0.85)	(0.37)	(2.80)	(1.96)	(1.18)	(-1.73)							

This table presents the coefficients obtained by regressing the cumulative abnormal returns following forecast revisions on the magnitude of the revision, firm size, and dummy variables indicating analysts' status. Revisions are dated within the firm's current fiscal year over the 1993-1999 period. CAR_{jt} is the cumulative abnormal return to security i during the release period (days 0 and +1). $FSUR_{ijt}$ is the forecast surprise following analyst i's revision at date t. $LNSIZE_{jt}$ is the natural logarithm of the market value (in USD) of common stock at fiscal year end. $LOCLEAD_{ij}$ is a dummy variable that takes a value of 1 if analyst i is a local leader and 0 otherwise. $FORLEAD_{ij}$ is a dummy variable that takes a value of 1 if analyst i is a foreign leader and 0 otherwise. $EXPATLEAD_{ij}$ is a dummy variable that takes a value of 1 if analyst i is an expatriate leader and 0 otherwise. FOL_{ij} is a dummy variable that takes a value of 1 if analyst i is a follower and 0 otherwise. All coefficients are multiplied by 100. T-statistics are based on White (1980). For each regression the adjusted R^2 are less than 0.01.

***, **, * denote significance at the 1%, 5%, and 10% levels, respectively.

Overall, this section shows that there are almost no significant differences in the incremental information contained in financial analysts forecasts revisions. However, the market seems to consider the forecasts issued by local and, to some extent, by expatriate leaders as being more informative than those issued by other analysts. This is consistent with the view that foreign leaders' revisions have a greater information content than other analysts' revisions.

6. Conclusions

Foreign financial analysts' EPS forecasts are more timely than expatriate and local analysts' forecasts. Building on CDL methodology, 1301 leader analysts are identified. Out of these leaders, 444 are foreign. This is significantly greater than the proportion of foreign analysts' forecasts in the sample. Conversely, analysts with local residence display a significant tendency to follow the "crowd".

In terms of forecast accuracy, analysts working for foreign brokerage houses are better at predicting firms' EPS than local analysts. Surprisingly, we detect no significant differences in forecast accuracy for companies with quoted ADRs. This may indicate that foreign and expatriate analysts' superior performance vanishes for companies with richer information environment.

Finally, stock prices react positively to upward forecast revisions released by foreign and expatriate leader analysts. The coefficient associated to foreign leaders forecast surprises is significantly greater than that associated to follower forecast surprises. It is also marginally greater than the coefficient associated to local leaders forecast surprises.

We see that foreign analysts outperform their local peers across all our performance measures. This suggests that residence does not give local financial analysts an advantage relative to their foreign counterparts. The difference between foreign and expatriate analysts' performance is less evident. Foreign analysts outperform their expatriate peers for one out of three performance measures. This suggests that agency problems, due to tighter investment banking relationships between resident analysts' firms and local companies, are not influencing financial analysts' objectivity on Latin American markets. Overall, our results are consistent with better information and greater sophistication on the part of analysts employed by foreign brokerage houses. This superiority may be linked to the superior resources available to analysts who work for important international brokerage houses, to the better international expertise of these analysts, or to their greater talent.

The present results are consistent with a better information on the part of foreign investors. Foreigners' portfolio profits on emerging markets, such as those documented by Seasholes (2000), may be driven by the better ability of foreign analysts at analyzing firms' situation for their clients. However, further research is needed to understand which category of investors (foreign or domestic) trade around foreign and local analysts' revisions. Finally, the practical implication of this investigation is that investors should rely more heavily on foreign financial analysts' forecasts than on local ones when they invest in Latin American markets.

Appendix

Leader-to-Follower Ratio

The Leader-to-Follower Ratio (LFR) for a particular analyst *a* who provides forecasts for firm *j* is expressed a follows:

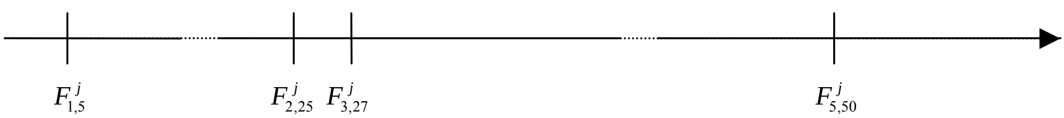
$$LFR_a^j = \frac{T_{0,a}^j}{T_{1,a}^j}, \tag{6}$$

where T_0 and T_1 are respectively the cumulative lead- and follow time for the *K* forecasts made by analyst *a* on firm *j* during a particular fiscal year. They are defined as follows:

$$T_{0,a}^j = \sum_{k=1}^{K}\sum_{h=1}^{H} t_{jhk}^0 \tag{7}$$

$$T_{1,a}^j = \sum_{k=1}^{K}\sum_{h=1}^{H} t_{jhk}^1 \tag{8}$$

t_{jmk}^0 (t_{jmk}^1) denotes the number of days by which forecast *h* precedes (follows) the *k*-th forecast may by analyst *a* for firm *j*. *H* is the number of forecasts made by other analysts that precede and follow the release of the *k*-th forecast of analyst *a*. The above figure provides an illustration of the idea underlying the LFR ratio. The forecast issued by analyst *a* for firm *j* at date *t* is denoted as $F_{a,t}^j$.



From this example, we see that analyst 2 issues a forecast on day 25. The preceding forecast was issued 20 days before on day 5. The following forecast is released soon afterward, on day 27. Taking into account only these one preceding and one following forecasts, analyst 2's LFR ratio is $\frac{20}{2} = 10$. Analyst 2 would therefore classified as a leader. To the same extent, analyst 3 is a follower analyst. He issues a forecast right after analysts 2 and no one free-rides on its forecast since the next to issue a forecast is analyst 4, only 23 days later. Its LFR would then be $\frac{1}{23} \cong 0.04$.

.

Acknowledgements

The authors thanks Michel Dubois, Christophe Pérignon, René Stulz, Ernst-Ludvig Von Thadden, and participants at FAME doctoral workshop, 5th Conference of the Swiss Society for Financial Market Research, 2002 FMA European meeting, 2002 EFMA meeting, and 2002 FMA US meeting for their helpful comments.

The authors acknowledge the contribution of Thomson Financial for providing earnings per share forecast data, available through the Institutional Brokers Estimate System. This data has been provided as part of a broad academic program to encourage earnings expectations research.

The authors aknowledge the financial support of the Swiss National Science Fondation (grant nr. 1214-065220).

References

Brennan, M.J. and H.H. Cao, 1997, International portfolio investment flows, *Journal of Finance* **52**, 1851-1880.

Brown, L.D., 1997, Analyst forecasting errors: Additional evidence, *Financial Analysts Journal,* November/December, 81-88.

Chang, C., 2002, *Information footholds: Expatriate analysts in an emerging market,* Working Paper, Haas School of Business, U.C. Berkeley.

Choe, H., Kho, B.C. and R.M. Stulz, 2000, *Do domestic investors have more valuable information about individual stocks than foreign investors?, Working paper*, Seoul National University and Ohio State University.

Cooper, R.A., Day, T.E. and C.M. Lewis, 2001, Following the leader: A study of individual analysts' earnings forecasts, *Journal of Financial Economics* **61**, 383-416.

Coval, J.D. and T.J. Moskowitz, 2001, The geography of investment: Informed trading and asset prices, *Journal of Political Economy* **109**, 811-841.

Grinblatt, M. and M. Keloharju, 2000, The investment behaviour and performance of various investor types: A study of Finland's unique data set, *Journal of Financial Economics* **55**, 43-67.

Harvey, C.R., 1995, Predictable risk and returns in emerging markets, *Review of Financial Studies* **8**, 773-816.

Hope, O., 2003, *Accounting policy disclosures and analysts' forecasts,* Contemporary Accounting Research, Forthcoming.

Kang, S., O'Brien, J. and K. Sivaramakrishnan, 1994, Analysts' interim earnings forecasts: Evidence on the forecasting process, *Journal of Accounting Research* **32**, 103-112.

Kang, J. and R.M. Stulz, 1997, Why is there a home bias? An analysis of foreign portfolio equity ownership in Japan, *Journal of Financial Economics* **46**, 2-28.

Kini, O., Shehzad, M., Rebello, M. and A. Venkateswaran, 2003, *On the determinants of international analyst research coverage,* Working Paper, Robinson College of Business, Georgia State University.

Lang, M.H., Lins, K.V. and D. Miller, 2003, ADRs, analysts, and accuracy: Does cross listing in the U.S. improve a firm's information environment and increase market value, *Journal of Accounting Research,* Forthcoming.

Lin, H. and M.F. McNichols, 1998, Underwriting relationships, analysts' earnings forecasts and investment recommendations, *Journal of Accounting and Economics* **25**, 101-127.

Malloy, C.J., 2003, The geography of equity analysis, Journal of Finance, Forthcoming.

Michealy, R. and K.L. Womack, 1999, Conflict of interest and the credibility of underwriter analyst recommendations, *Review of Financial Studies* **12**, 653-686.

Orpurt, S.F., 2003, *Local asymmetric information advantages: International evidence from analysts' European firm earnings forecasts,* Working Paper, Graduate School of Business, University of Chicago.

Seasholes, M., 2000, *Smart foreign traders in emerging markets,* Working paper, Harvard Business School, Cambridge, M.A.

Stickel, S., 1992, Reputation and performance among security analysts, *Journal of Finance* **47**, 1811-1836.

Stickel, S., 1995, The anatomy of the performance of buy and sell recommendations, *Financial Analysts Journal,* September/October, 25-39.

White, H., 1980, A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica* **48**, 817-838.

In: Economic Forecasting
Editor: Alan T. Molnar, pp. 213-227

ISBN: 978-1-60741-068-3

Chapter 9

MODELING AND FORECASTING INCOME TAX REVENUE: THE CASE OF UZBEKISTAN

Marat Ibragimov[1], Rustam Ibragimov[2,*]
and Nishanbay Sirajiddinov[1]
[1]Center for Economic Research, Tashkent, Uzbekistan
[2]Department of Economics, Harvard University, Cambridge, MA, USA

Abstract

Income tax revenue crucially depends on the wage distribution across and within the industries. However, many transition economies present a challenge for a sound econometric analysis due to data unavailability. The paper presents an approach to modeling and forecasting income tax revenues in an economy under missing data on individual wages within the industries. We consider the situations where only the aggregate industry-level data and sample observations for a few industries are available. Using the example of the Uzbek economy in 1995-2005, we show how the econometric analysis of wage distributions and the implied tax revenues can be conducted in such settings. One of the main conclusions of the paper is that the distributions of wages and the implied tax revenues in the economy are well approximated by Gamma distributions with semi-heavy tails that decay slower than those of Gaussian variables.

JEL classification: C13, D31, H24, P20

Keywords: Income tax; Wage distribution; Inequality

---

* E-mail address: ribragim@fas.harvard.edu (Corresponding author: Rustam Ibragimov). Phone: (617) 496-4795; Fax: (617) 495-7730. Department of Economics, Harvard University, Littauer Center, 1805 Cambridge St., Cambridge, MA 02138.

1. Objectives of the Paper

This paper discusses an approach to econometric analysis and forecasting for income tax revenue in situations where the data on wages and income are largely unavailable. Naturally, income tax revenue crucially depends on the wage distribution in the economy. However, many transition economies present a challenge for a sound econometric analysis due to data unavailability. This is typically the case for employee-level data, including wages and income.

We show that, in certain cases, one can make inference on wage distribution and the implied distribution of income tax revenues even in absence of microeconomic level data for individual households. We discuss an approach that allows one to estimate the form and parameters of the wage distribution in the economy using the official data on average wages across the industries. We then use these results to obtain estimates for the distribution of tax revenues. One of the main conclusions of the paper is that Gamma densities provide appropriate models for wage distributions. These approximations imply that the distributions of wages and the implied tax revenues have semi-heavy tails that decay slower than those of Gaussian variables.

The analysis presented in the paper is based on the official data on average wages across the industries and a sample of data for employees' wages in one of the industries, construction. The analysis thus uses the existing minimum of the available information. The estimates presented are, therefore, some of the only results that are possible to obtain in such settings.

2. Inference on Wage Distribution under Missing Household Income Data

Tables A1 and A2 in the appendix provide government data on the number of employees and the average wages across the industries of the Uzbek economy. Table 1 below provides a summary of the implied inequality measures for the wage distribution in the Uzbek economy. The inequality measures summarized in the table and their properties are discussed in detail, among others, in Ch. 13 in Marshall and Olkin (1979).

Table 1. Statistical characteristics of wage distribution in the Uzbek economy in 1995, 2000 and 2005

	1995	2000	2005
Minimal wage, soum	250	1320	9400
Average wage, $\overline{w}$, soum	1057	11225.1	85865.0
Gini coefficient[1]	0.235	0.233	0.252
The number of employees with wages less than $\overline{w}$	60%	43%	42%
Minimal majority[2]	68%	67,4%	67,5%

[1] The Gini coefficient is computed under the assumption that each industry of the economy is considered as an individual unit regardless of the number of employees in the industry. In other words, here, the Gini coefficient characterizes the inequality in distribution of the average wage across the industries.

Table 2 presents Gini coefficients for the wage distribution in the Uzbek economy computed under the assumption that wages are uniformly distributed within each industry.

Table 2. Gini coefficients for wages in Uzbekistan under the assumption of uniform wage distribution within the industries

	1995	2000	2005
Gini coefficient	0.161	0.253	0.262

The Gini coefficient values reported in Tables 1 and 2 are considerably small comparing to other Newly Independent States. These values do not reflect the true inequality in wage distribution in the total economy.

The inequality in the total economy reflects both the disparities among the industries and inequalities in distributions within them. Thus, taking into account the wage inequality within the industries would lead to an increase in the calculated Gini coefficients and other inequality measures considered in Table 1.

As an example, we estimate the wage distribution within the construction industry. This wage distribution is estimated using a sample of wages in 1995 for 256 employees in the industry available to us (see Table A3 and Figure A1 in the appendix). The Gini coefficient for the construction industry estimated using the sample is 0.29.

Consider a random variable (r.v.) *X* with truncated normal density

$$f(x) = \begin{cases} \dfrac{1}{A\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}, & if \;\; x > a; \\ 0, \;\; if \;\; x \le a. & \end{cases} \tag{1}$$

where the truncation level *a*=250 (soum) is the minimal wage in 1995 and $A = \frac{1}{\sigma\sqrt{2\pi}} \int_a^{\infty} e^{-(x-\mu)^2/(2\sigma^2)} dx = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$ is a normalizing constant.

If *a* is known, one may use the relations for the mean *E*[*X*] and the variance *Var*[*X*] of *X* given by

$$\begin{cases} E[X] = \mu + \sigma B \\ Var[X] = \sigma^2 (1 - B^2 + \dfrac{a-\mu}{\sigma} B) \end{cases} \tag{2}$$

[2] In the context of political science, the minimal majority is the smallest number of individuals controlling a majority of legislature. This index can also be used as an inequality measure and has a simple expression in terms of the Lorenz curve characterizing the inequality among subjects: if wages $w_1, \ldots, w_n$ determine the Lorenz curve *h*, then the minimal majority is $h^{-1}(0.5)$ (see Ch. 13 in Marshall and Olkin, 1979). Thus, in Table 1, minimal majority equals to the proportion of the employees who have 50% of the total income (wage fund).

where $B = \dfrac{\varphi\left(\dfrac{a-\mu}{\sigma}\right)}{A}$, to estimate μ and σ by the method of moments (see Section 19.3-4 in Korn and Korn, 1968, and Section 10.1 in Johnson, Kotz and Balakrishnan, 1994).

The p-value for the Kolmogorov-Smirnov test of the null hypothesis that wage in the construction industry follows the above distribution equals to α=0.21. Thus, the null hypothesis is not rejected at the $\alpha \le 0.21$ significance level (see Figure A2 in the appendix).

The coefficient of variation for the fitted truncated normal distribution is $CV = \dfrac{\sqrt{Var[X]}}{|E[X]|} = 0.51$.

To estimate the wage distribution in the whole economy we assume that, in each industry, wage has a truncated normal distribution with the same coefficient of variation CV=0.51 as in construction, that is, the parameters $E[X]$ and $Var[X]$ satisfy $\sqrt{Var[X]} = 0.51E[X]$. The wage distribution for the whole economy is a mixture of wage distributions within the industries. The cdf of wage in the whole economy is determined as

$$F(x) = \frac{N_1}{N} F_1(x) + \ldots + \frac{N_m}{N} F_m(x)\,, \tag{3}$$

where $N=N_1+\ldots+N_m$ is the total number of employees in the economy, N_j is the number of employees in the ith industry, and $F_j(x)$ is the cdf of wage in the jth industry, j=1, ..., m.

Figures A3-A5 in the appendix provide the cdf's for wages in the Uzbek economy in 1995, 2000 and 2005. The calculations are based on the above assumption of equal coefficients of variation for wage distributions within the industries. Evidently, this assumption may be lifted if, in addition to the data on the average wages in each of the industries, one also has the data on the wage variances within the industries.

Table 3 provides the means, variances and Gini coefficients for the fitted wage distribution in the Uzbek economy in 1995, 2000 and 2005. Columns 5 and 6 of Table A1 and columns 5, 6, 10 and 11 of Table A2 in the appendix provide the parameters μ and σ for truncated normal distributions (1) and the distributions' means $E[X]$ and variances $Var[X]$ calculated using (2) for each of the industries in 1995, 2000 and 2005.

Table 3. Statistical characteristics of wage distribution in the Uzbek economy in 1995, 2000 and 2005

	1995	**2000**	**2005**
Mean	1307	19973	136667
Standard deviation	808	12968	89396
Gini coefficient	0.321	0.356	0.359

The interpretation of Gini coefficients presented in Tables 1-3 can be described as follows.

1. The coefficients in Table 1 correspond to the case where one measures the inequality among the industries using known values of average wages in them. This equals to the true inequality index for the whole economy if wage is equally distributed within each of the industries and, in addition, the number of employees in the industries is the same.
2. The estimates in Table 2 correspond to the case where the average wages and the number of employees in the industries are known. Wage distributions within the industries are assumed to be uniform.
3. Table 3 provides the estimates of the wage inequality in the whole economy where the inequality in wage distributions within the industries is taken into account.

One should note that the Gini coefficient values in Table 1 provide lower bounds for the true Gini coefficients for the whole economy. The inequality indices in Table 3 may be regarded as the most appropriate ones. In addition, they are the closest to the true inequality measures under some natural general assumptions (see the discussion and related results in Ibragimov and Walden, 2007).

According to the results in this section, the inequality in wage distribution over the whole economy is relatively small, even when the inequalities within the industries are taken into account. The situation is similar to that in Russia in the first half of 90's where one observed "the power of institutional features in the wage settings that tended to dominate the redistributive effects transmitted through high inflation and decentralization in wage settings" (Commander, McHale and Yemtsov, 1995, p. 165) and relative stability of the low inequality among the industries.

3. Income Tax Revenue Distribution and Forecasting

We follow the following two criteria in approximating the wage distribution in the economy:

1. The distribution law should be widely known and parsimonious with a relatively small number of parameters;
2. The measure of approximation should have a clear interpretation.

Among two parameter distributions, the best fit appeared to be provided by gamma distribution with cdf

$$F_{gamma}(x;k;\theta) = \int_0^x f(u;k;\theta)du,$$

where $f(x;k;\theta) = x^{k-1}\dfrac{e^{-x/\theta}}{\theta^k \Gamma(k)}$, $x > 0$, $k, \theta > 0$ is the pdf of the distribution. The gamma cdf $F_{gamma}(x;k;\theta)$ has tails that asymptotically decay as $x^{\alpha}\exp(-\beta x)$ as $x \to \infty$. Thus, the tails of the distribution belong to the class of semi-heavy tails that decay slower than those of Gaussian distribution but faster than any power law tails.[3]

The shape and scale parameters *k* and *θ* are estimated by the method of moments using the relations

$$\begin{cases} E[X] = k\theta \\ Var[X] = k\theta^2 \end{cases}$$

for the mean *E*[*X*] and variance *Var*[*X*] of a gamma r.v. *X* with the pdf $f(x;k;\theta)$.

Table 4 provides a summary of the implied approximation for the wage cdf *F* estimated in (3) in the case where the measure $\Delta_{gamma} = \max|F(x) - F_{gamma}(x;k;\theta)|$ is used as a measure of approximation. For comparison, Table 4 also presents the values $\Delta_{normal} = \max|F(x) - F_{normal}(x;k;\theta)|$ for the normal distribution with the parameters *m*=*E*[*X*] and $\sigma = \sqrt{Var[X]}$ (see also Figures A3-A5 in the appendix).

Table 4. Gamma distribution approximation results.

	1995	**2000**	**2005**
E[*X*]	1307.43	19973.28	136666.91
Var[*X*]	653450.74	168164070.26	7991717405.72
k	2.615916	2.372278605	2.337150
θ	499.798148	8419.450581	58475.876093
Δ_{gamma}	0.0496	0.0237	0.02533
Δ_{normal}	0.0977	0.0808	0.0792

As is seen from the table, the gamma distribution provides better approximation to *F*(*x*) than does the normal distribution.

[3] Such tails were considered among others, by Barndorff-Nielsen (1997) and Barndorff-Nielsen and Shephard (2001) in the context of applications of Normal Inverse Gaussian distributions and their extensions and by Malevergne, Pisarenko and Sornette (2005) who consider the fitting by stretched exponential and related distributions. Together with widely used power law distributions applied in many works (see, for instance, the discussion in Loretan and Phillips, 1994, Gabaix, Gopikrishnan, Plerou and Stanley, 2003, Ibragimov, 2005, Rachev, Menn and Fabozzi, 2005, and references therein), semi-heavy tailed distributions were reported to provide good fit for a number of economic and financial time series.

The results obtained can be applied to forecast the income tax revenue using the tax rate and wage data. We describe the approach to forecasting using the parameters $E[X]$ and $Var[X]$ for 2005.

Denote by T_{total} the yearly income tax revenue in the economy, in million soum, and by w_{min} the minimal monthly wage, in soum. Let $T(x)$ denote the monthly income tax revenue from one employee with wage x.

The tax rates in Uzbekistan in 2005 are provided in Table A`4 in the appendix.

The function relating the income tax $T(x)$ to the wage x has the following form:

$$T(x) = \begin{cases} t_1 x, \ 0 < x \le w_1, \\ t_1 w_1 + t_2(x - w_1), \ \ w_1 < x \le w_2, \\ \dots \\ t_1 w_1 + t_2(w_2 - w_1) + \dots + t_n(x - w_{n-1}), \ \ x > w_{n-1} \end{cases}, \qquad (4)$$

where x is the wage, t_i are the tax rates for different wage levels, i=1, ..., n; $[w_{\min}; w_1]$, $[w_1; w_2]$, ...,

$[w_{n-1}; \infty)$ is the partition of the set of possible wage values $[w_{min};\infty)$ into strata that correspond to the different tax rates (see Table A4 in the appendix). Thus function (4) has the following form for the Uzbek economy in 2005:

$$T(x) = \begin{cases} 0.13x, \ 0 < x \le 5w_{\min}, \\ -0.4w_{\min} + 0.21x, \ \ 5w_{\min} < x \le 10w_{\min}, \\ -1.3w_{\min} + 0.3x, \ \ x > 10w_{\min}, \end{cases}$$

where w_{min}=9400 soum.

As before, we assume that the wage level X has the gamma distribution with the cdf $F_{gamma}(x; k; \theta)$ and the pdf $f(x; k; \theta)$. Therefore, the cdf of the r.v. $t=T(x)$ is given by $F_{gamma}(T^{-1}(t); k; \theta)$ and the pdf of $t=T(x)$ is $f(T^{-1}(t); k; \theta)$ (see Figure A6 in the appendix).

According to the official site of the Ministry of Finance of the Republic of Uzbekistan[4], the income tax revenue in the Uzbek economy in 2005 was equal to T_{actual}=465641.1 million soum.

To compare income tax revenues predicted by the above results with the actual values, we make the following assumptions:

1. The monthly income tax revenue equals to $1/12$ of yearly tax revenue;
2. The income tax revenue from each employee equals to $1/N$ of the income tax revenue over the whole economy, where N is the number of employees included in calculation of the tax base. The monthly income tax revenue from each employee is the r.v. $T(x)$ with the cdf $F_{gamma}(T^{-1}(t);k;\theta)$.

[4] http://www.mf.uz/

Table 5 presents the estimates for the tail probabilities of yearly income tax revenues over the whole economy that hold under these assumptions.

Starting with 2000, the average wage for agriculture has not been calculated or reported by the State Committee for Statistics (Goskomstat) of the Republic of Uzbekistan due to the absence of data for the private sector (see Tables A1 and A2 in the appendix). Therefore, the income tax revenues from agriculture seem to be negligible in 2000 and 2005. Nevertheless, we present the estimates for the tail probabilities of the income tax revenue for both the cases where N equals to the total number of employees in the economy and where N does not include the employees in agriculture.

Table 5. Tail probabilities of the total income tax revenues in 2005

N, тыс.чел.	$P(T_{total} > T_{actual}) = P(T_{total} > 465641.1$ million soum)
10196.3: Total number of employees in all the industries of the economy	0.9496
7110.6: Total number of employees in all industries except agriculture	0.8988

The results in Table 5 suggest very high probability of the total tax revenue in the model being not less than the actual value of 465641.1 million soum in 2005. This probability is estimated to be about 0.9 if agriculture is excluded in calculation of the tax base, and to be about 0.95 if the estimate is calculated using the total number of employees in the economy.

4. Conclusion

The approach discussed in the paper allows one to obtain estimates of the income tax revenues via gamma distribution approximations using data on average wages in the economy and wage variances. At the same time, the calculations are based on several strict assumptions, in particular, those listed below

1. Within each industry, wage distribution is truncated normal;
2. Wage variances in all the industries are the same;
3. Wage in the whole economy follows a gamma distribution.

Each of these assumptions is very strong. The assumptions are motivated only by the absence of sample data across all the industries. Even if such studies are conducted in transition economies, their results are often inaccessible. On the other hand, the approach to wage distribution modeling and forecasting the implied income tax revenues discussed in the paper is based on the existing minimum of the available data. The estimates presented are, therefore, some of the only results that are possible to obtain in such circumstances.

Appendix

Table A1. Wage distribution in the industries of Uzbek economy in 1995

Industries	Number of employees, *N*, thousands	Average wage, *E*[*X*], soum	Variance, *Var*[*X*]	μ	σ
1	2	3	4	5	6
1.Heavy industry	854.9	1529	608072.4	1264	973.14
2. Horticulture	378.1	808	169809.9	342.32	655.49
3. Animal husbandry	84.3	620	99982.44	-522.74	723.05
4. Forestry	5.3	694	125273.5	-3.9443	659.67
5.Transportation	259.5	1427	529649.2	1158	919.95
6.Communications	43.4	1521	601726	1255.8	968.94
7.Construction	320.5	1655	712420.4	1392.4	1039.9
8. Construction-related services	13.9	1915	953845.2	1651.4	1180.1
9. Trade	194.4	556	80406.27	-1726.4	888.4
10. Public catering	46.9	528	72511.72	-1733.8	860.71
11.Computer services	2.7	1716	765905	1453.7	1072.6
12. Housing and utilities	116.9	1064	294458.2	747.27	743.15
13. Health	469.6	665	115022.7	-153.27	674.25
14.Education	858.5	577	86594.83	-1317.8	840.35
15.Arts and culture	62.4	770	154213.3	251.76	650.92
16.Science	41.6	1209	380183.2	920.57	810.42
17.Insurance and pensions	38	1789	832455.5	1526.7	1111.8
18.Administration	87	1041	281865.4	717.77	733.17
19.Other industries	252.1	949	234246.3	590.78	696.16

The values μ and σ are the parameters of the truncated normal distribution (1) fitted to the industry data, and *Var*[*X*] is the variance of the distribution calculated using (2).

Source: The State Committee for Statistics (Goskomstat) of the Republic of Uzbekistan (Columns 1-3) and the authors' calculations (Columns 4-6).

Table A2. Wage distribution in the industries of Uzbek economy in 2000 and 2005

Industries	2000					2005				
1	2	3	4	5	6	7	8	9	10	11
	N	$E[X]$	$Var[X]$	μ	σ	N	$E[X]$	$Var[X]$	μ	σ
1.Heavy industry	1145	21861.5	124308074	19970	12774	1145	145364.2	5496104461	132430.0	85171.0
2. Construction	676	18796.6	91896342	17064	11054	676	140114.8	5106320438	127460.0	82226.0
3. Transportation	313.7	18569.3	89687023	16848	10926	367.97	134430.6	4700419575	122060.0	79038.0
4. Communications	68.3	25081.1	163619392	23016	14584	80.1	135686.9	4788680620	123250.0	79742.0
5. Trade and public catering	754	9815.6	25059452	8427.6	6070.5	754	70209.0	1282111887	60321.0	43398.0
6. Health	587	8177.2	17392138	6780.9	5193	587	48411.0	609576842	37913.0	31924.0
7. Education	1054	9317.3	22579624	7932.6	5801.2	1274	55454.2	799851274	45420.0	35524.0
8. Arts and culture	61.02	9591.9	23930226	8205.8	5949.4	73.75	57977.2	874287115	48027.0	36846.0
9. Science	30.98	16053.3	67029991	14453	9518.6	37.445	90070.1	2110090877	83956.0	53187.0
10. Housing and utilities	251	13265.4	45769964	11784	7966.4	251	84703.9	1866150447	74494.0	51332.0
11. Credits and insurance	52	24834.8	160420567	22783	14446	52	198770.0	1.0276E+10	182950.0	115210.0
12. Other industries	904.3	30779.5	246412885	28397	17793	1812.4	200156.0	1.042E+10	184260.0	115990.0

The notations used are the same as in Table A1.

Source: The State Committee for Statistics (Goskomstat) of the Republic of Uzbekistan (Columns 1-3, 7, 8) and the authors' calculations (Columns 4-6, 9-11).

Note: Starting with 2000, the average wage for agriculture has not been calculated by the Goskomstat due to the absence of data for the private sector.

Table A3. Wage distribution in the construction industry

***W*, %**	***n*, %**		***W*, %**	***n*, %**		***W*, %**	***n*, %**
0-4	1,56		74-79	3,13		148-153	1,56
4-9	1,17		79-83	2,73		153-157	0,39
9-13	3,13		83-87	1,95		157-161	1,17
13-17	1,56		87-92	2,73		161-166	0,78
17-22	1,56		92-96	3,52		166-170	1,56
22-26	1,17		96-100	4,69		170175	0,78
26-31	1,17		100-105	3,13		175-179	0,78
31-35	1,56		105-109	4,30		179-183	1,56
35-39	2,34		109-113	2,34		183-188	1,17
39-44	3,13		113-118	3,52		188-192	1,17
44-48	0,39		118-122	5,08		192-196	0,39
48-52	1,56		122-127	3,13		196-205	0,39
52-57	3,52		127-131	3,13		205-218	0,78
57-61	1,95		131-135	2,73		218-236	0,39
61-65	2,34		135-140	3,52		236-246	0,39
65-70	1,95		140-144	1,95		>246	0,39
70-74	1,95		144-148	2,73			

Based on a sample of 256 employees. *W*,% denotes the average wage in percent to the average wage in the sample; *n*,% is the percentage of employees in the sample who receive the wage within the indicated ranges.

Table A4. Income tax rates in Uzbekistan in 2005

Range of the income, *I*	**Corresponding tax amount, *T(I)***
$I \le 5w_{\min}$	$T(I) = 0.13I$
$5w_{\min} < I \le 10w_{\min}$	$T(5w_{\min}) + 0.21(I - 5w_{\min})$
$I > 10w_{\min}$	$T(10w_{\min}) + 0.3(I - 10w_{\min})$

The value *I* denotes the income to date from the beginning of 2005; $w_{\min}$ denotes the minimal wage in 2005.

Source: State Tax Committee of the Republic of Uzbekistan

Figure A1. Wage frequency distribution in the construction industry.

Based on a sample of 256 employees. *Wage,%* denotes the average wage in percent to the average wage in the sample. *Number of workers,%* is the percentage of employees in the sample who receive the wage within the indicated ranges.

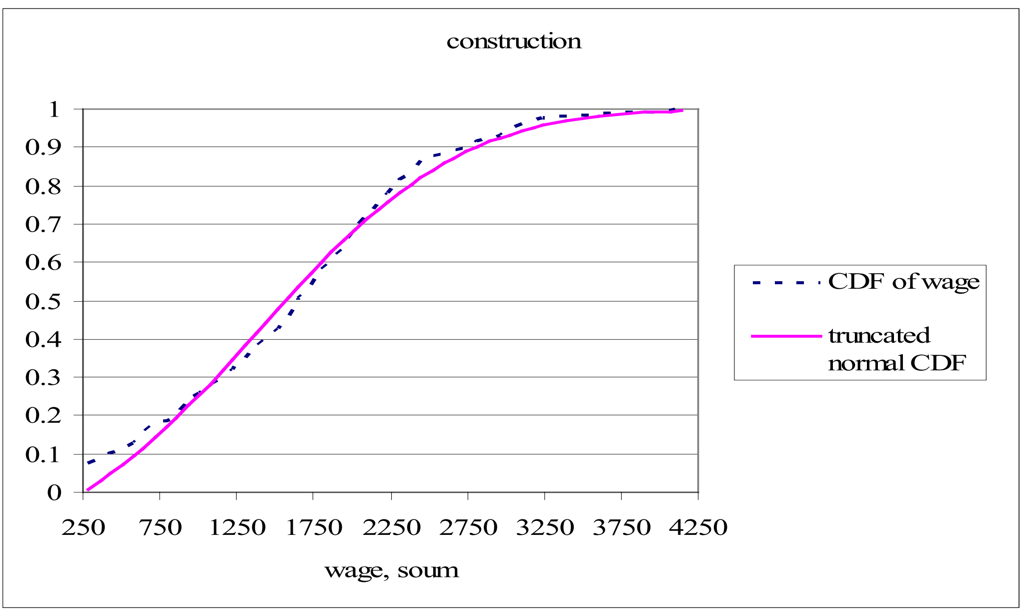


Figure A2. The cdf of wage in the construction industry and the cdf of the fitted truncated normal distribution.

Figure A3. The cdf $F(x)$ for the wage distribution in the Uzbek economy in 1995 and the cdf's of the fitted gamma and normal distributions.

Figure A4. The cdf $F(x)$ for the wage distribution in the Uzbek economy in 2000 and the cdf's of the fitted gamma and truncated normal distributions.

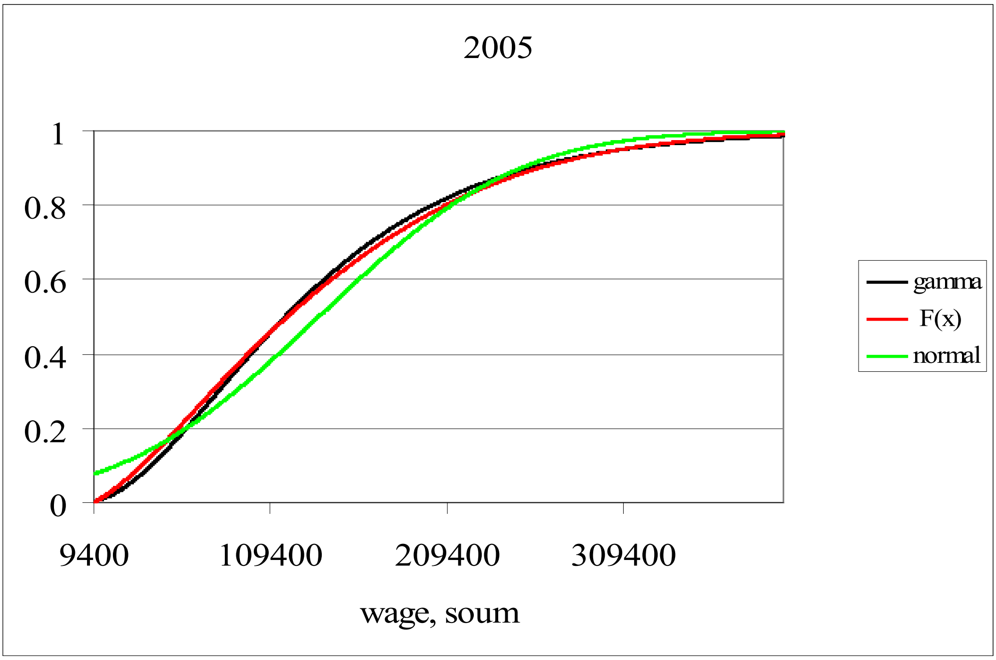


Figure A5. The cdf *F*(*x*) for the wage distribution in the Uzbek economy in 2005 and the cdf's of the fitted gamma and truncated normal distributions.

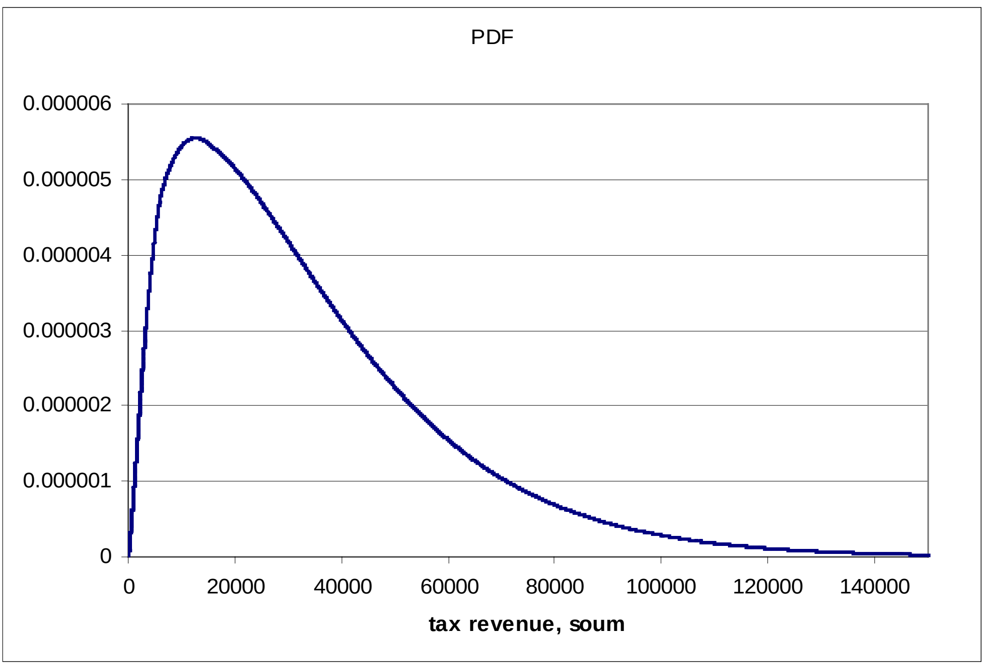


Figure A6. The pdf of the estimated income tax revenue from one employee in 2005.

References

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and the modeling of stock returns. *Scandinavian Journal of Statistics* **24**, 1-13.

Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **63**, 167-241.

Commander, S., McHale, J. and Yemtsov, R. (1995). *Russia.* In S. Commander and F. Coricelli, eds "*Unemployment, Structural Transformation and Labor Market in the Eastern Europe and Russia",* EDI Development Studies, World Bank, Washington, DC. , pp. 147-191.

Gabaix, X., Gopikrishnan, P., Plerou, V. and Stanley, H. E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature* **423**, 267-270.

Ibragimov R. and Walden, J. (2007). Value at risk under dependence and heavy-tailedness: Models with common shocks. *Working paper, Harvard University and University of California at Berkeley*.

Ibragimov, R. (2005), *New majorization theory in economics and martingale convergence results in econometrics*, Ph.D. dissertation, Yale University.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous univariate distributions*. Vol. 1, Second edition. John Wiley, New York.

Korn, G. A. and Korn, T. M. (1968). *Mathematical handbook for scientists and engineers.* Second edition, McGraw-Hill, New York.

Loretan, M. and Phillips, P. C. B. (1994). Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance* **1**, 211-248.

Malevergne, Y., Pisarenko V. and Sornette D. (2005). Empirical distributions of stock returns: between the stretched exponential and the power law? *Quantitative Finance* **5,** 379-401.

Marshall, A. W. & Olkin, I. (1979), *Inequalities: theory of majorization and its applications*, Academic Press, New York.

Rachev, S. T., Menn, C. and Fabozzi, F. J. (2005). *Fat-tailed and skewed asset return distributions: Implications for risk management, portfolio selection, and option pricing*, Wiley, Hoboken, NJ.

In: Economic Forecasting
Editor: Alan T. Molnar, pp. 229-247

ISBN 978-1-60741-068-3

Chapter 10

FORECASTING THE UNCONDITIONAL AND CONDITIONAL KURTOSIS OF THE ASSET RETURNS DISTRIBUTION*

Trino Manuel Ñíguez[1,†]***, Javier Perote***[2] ***and Antonio Rubia***[3]
[1]University of Westminster
[2]Rey Juan Carlos University
[3]University of Alicante

Abstract

This paper analyzes the out-of-sample ability of different parametric and semi-parametric GARCH-type models to forecast the conditional variance and the conditional and unconditional kurtosis of three types of financial assets (stock index, exchange rate and Treasury Note). For this purpose, we consider the Gaussian and Student-t GARCH models by Bollerslev (1986, 1987), and two different time-varying conditional kurtosis GARCH models based on the Student-t and a transformed Gram-Charlier density.

Key words: Gram-Charlier densities; Financial data; High-order moments; Out-of-sample forecasting.

JEL classification: C16, G1.

1. Introduction

The literature related to financial econometrics and asset pricing has shown that the conditional distribution of high-frequency returns exhibits stylized features that include excess of kurtosis, negative skewness, and temporal persistence in conditional moments. Remarkably, time dependency may be a characteristic that not only is present in the dynamics of the

*Financial support from the Spanish Ministry of Education through project SEJ2006-06104/ECON is gratefully acknowledged.

†E-mail address: T.M.Niguez@wmin.ac.uk. Phone: +44 (0) 22 7911 5000. Fax: +44 (0) 22 7911 5839. Department of Economics and Quantitative Methods, Westminster Business School, University of Westminster, 35 Marylebone Road, NW1 5LS London, UK.

expected return and the conditional variance, but also in higher-order moments $E_t\left(r_{t+1}^s\right)$, $s \geq 3$; see, Nelson (1996). Among these, the conditional skewness and kurtosis (related to the third- and fourth-order conditional central moments, respectively) are particularly relevant for their implications in risk management, asset pricing, and optimal portfolio selection, as pointed out by Chunhachinda, Dandapani, Hamid and Prakash (1997), Harvey and Siddique (2000), Christie-David and Chaudhry (2001) and Schmidt (2002). For instance, rational investors concerned with the non-Gaussian properties of returns would be averse to negative skewness and high kurtosis. As a result, the composition of their optimal portfolio would change (everything else being equal) whenever they expect changes in any of those characteristics. In this regard, Fang and Lai (1997) have reported empirical evidence of positive risk premiums for conditional skewness and conditional kurtosis in the US market. All this has given rise to a string of recent articles which have gone beyond the traditional modelling and forecasting of the conditional volatility to also focus on the time-varying properties of higher-order moments. The models proposed in this literature include both parametric (see, among others, Hansen 1994, Dueker 1997, Harvey and Siddique 1999, and Brooks, Burke, Heravi and Persand 2005) and semi-parametric approaches, such as entropy distributions (Rockinger and Jondeau 2002), and Gram-Charlier densities (León, Rubio and Serna 2005).

The econometric modelling of high-order moments attempts to exploit time dependency to improve the forecasts which are typically needed in financial applications. In this paper, we analyze the out-of-sample ability of different parametric and semi-parametric GARCH-type models to forecast the conditional variance as well as the conditional and unconditional kurtosis of several classes of financial assets. We do not focus on asymmetric distributions (*i.e.*, we do not consider skewness in this paper) so that we can specifically isolate the gains from modelling kurtosis, which is widely considered as the most representative non-Gaussian stylized feature of financial data. We compare four different approaches in our study with increasing degree of complexity, going from the standard GARCH model to more sophisticated specifications. The starting point in our analysis is the simple Gaussian GARCH model by Bollerslev (1986), which implies the same degree of constant conditional kurtosis as the Normal distribution. Second, we consider the straightforward generalization of this model suggested by Bollerslev (1987) which, on the basis of the conditional Student-t distribution, is able to capture the underlying conditional kurtosis in the data, still assuming constant kurtosis. Next, we consider a further generalization, the so-called Student-t GARCHK model, suggested in Brooks *et al.* (2005). This is intended to fit the time-varying dynamics of the conditional variance and the kurtosis separately via a Student-t distribution with a degrees of freedom parameter that is allowed to vary over time. Finally, we consider a restricted version of the semi-parametric GARCH model with time-varying conditional kurtosis proposed in León et al. (2005) as an alternative to the Student-t GARCHK model. The semi-parametric approach relies upon a Gram-Charlier type polynomial expansion so that the resulting probability density function is flexible enough to approximate any unknown density, without imposing any assumption on the underlying conditional distribution.

The main questions we try to solve refer to $i)$ whether conditional kurtosis models are able to yield better out-of-sample forecasts, and $ii)$ which conditional kurtosis approach (parametric or semi-parametric) is more appropriate for applied purposes. These are ultimately empirical questions that we shall address statistically in this paper by means of an out-of-sample forecasting analysis. In particular, we compute one-day ahead forecasts of the conditional variances and the conditional and unconditional kurtosis implied by the different models, and compare their forecasting ability in terms of the Mean Square Error (MSE) loss-function. Patton (2006) has recently argued that only few of the volatility loss functions are not affected by the choice of the proxy used, showing that the MSE loss-function is robust. The same arguments may hold when analyzing higher-order conditional moments for which the true values are not directly observable and must be proxied with sampling error. We use the procedure in Diebold and Mariano (1995) to test statistically whether the differences observed across the different GARCH models are truly significant.

The remainder of this paper is organized as follows. Section 2 describes the conditional models that we use to fit and forecast conditional variance and kurtosis. Section 3 discusses the main features of the empirical analysis. Finally, Section 4 summarizes and concludes.

2. Modelling Forecasting Conditional Variance and Higher-Order Moments

Let us first introduce the basic data generating process and the general notation used throughout the paper. We observe a sample of (daily) asset prices from which we compute the series of returns $r_t = 100 \log\left(P_t/P_{t-1}\right), \ t = 1, ..., T$. We assume that the returns follow the dynamics,

$$\begin{aligned} r_t &= E_{t-1}(r_t) + \varepsilon_t, \\ \varepsilon_t &= h_t^{1/2}\eta_t; \end{aligned} \tag{1}$$

where the conditional expectation $E_{t-1}(\cdot) = E\left(\cdot|I_{t-1}\right)$ is taken on the observable set of information available up to time $t-1$, denoted as I_{t-1}. The set of random innovations η_t are conditionally distributed according to certain density function $f\left(\eta_t|I_{t-1}\right)$ that satisfies $E_{t-1}\left(\eta_t\right) = 0$ and $E_{t-1}\left(\eta_t^2\right) = 1$, with $E\left(\eta_t^s\right) < \infty$ for some $s > 2$.

The expected return in the model is given by $E_{t-1}(r_t)$. We shall use a simple AR(1) model, $E_{t-1}(r_t) = c + \rho r_{t-1}$, to filter out any predictable component in the conditional mean of the series. The conditional variance of the process is given by $h_t = E_{t-1}(\varepsilon_t^2)$, which is the main object of interest in many papers that specifically focus on volatility modelling. In this regard, one of the most widely used models is the GARCH(1,1) process of Bollerslev (1986), which assumes a linear functional form,

$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1} \tag{2}$$

with the parameter restrictions $\omega > 0,\ \alpha, \beta \geq 0$ ensuring almost sure positiveness in the conditional variance process. The additional restriction $\alpha+\beta < 1$ is sufficient and necessary for $E\left(\varepsilon_t^2\right) < \infty$, whereas the existence of higher-order moments imply further restrictions on the driving parameters (α, β) as well as the existence of suitable moments of η_t (e.g., the unconditional fourth-order moment is well defined when $\kappa_\eta\alpha^2 + \beta^2 + 2\alpha\beta < 1$, with κ_η denoting the kurtosis of η_t).

The enormous success of the GARCH(1,1) model strives in its appealing interpretation, large degree of statistical parsimony, and computational tractability. The main GARCH equation describes the conditional variance forecast h_t as a weighted average of a constant term (long-run variance), ω, the previous variance forecast, h_{t-1}, and a proxy for the conditional variance given the information which was not available when the previous forecast was made (related to new information arrivals), ε_{t-1}^2. As a result, the model is able to capture the main stylized feature in the conditional variance (namely, clustering and persistence) by resorting to a small number of parameters in a fairly simple representation, from which one-step and multi-step forecasts can easily be obtained.[1] Further generalizations that conform the broad GARCH family arise by readily extending this basic structure (for instance, towards including leverage and other non-linear effects), and/or by considering different assumptions on the conditional distribution $f(\eta_t|I_{t-1})$. We shall discuss in more detail the basic model and several of its extensions intended to capture excess of kurtosis and time dependence in higher-order moments in the following subsections.

2.1. GARCH Modelling

2.1.1. Gaussian GARCH

The simplest approach in GARCH-type modelling is the Gaussian GARCH(1,1) model of Bollerslev (1986). In addition, to the basic data generating process (1)-(2), it is assumed that the conditional shocks $\{\eta_t\}$ are independent and identically normally distributed with mean 0 and variance 1, *i.e.*, it is imposed the particularly strong restriction

$$\eta_t \sim iid\mathcal{N}(0,1), \tag{3}$$

or $f(\eta_t|I_{t-1}) = f(\eta_t) = (2\pi)^{-1/2}\exp\left(-\eta_t^2/2\right)$. The model is fully specified with this assumption, and the relevant parameters $\xi_0 := (c,\rho,\omega,\alpha,\beta)'$ can then be estimated from the sample by Maximum Likelihood [ML henceforth]. Under conventional assumptions on the pre-sample observations which do not play any relevant role when the sample is large enough, the log-likelihood function of the t-th observation, after dropping a constant term, can be written as:

$$l_t(\xi_0) = -\frac{1}{2}\log h_t - \frac{\varepsilon_t^2}{2h_t}; \quad t = 1,...,T, \tag{4}$$

Since the information matrix related to the two sets of parameters involved (conditional mean and conditional variance) is block-diagonal, the respective parameter vectors, say $\xi_{0m} := (c,\rho)'$ and $\xi_{0v} := (\omega,\alpha,\beta)'$, can be estimated separately. We shall proceed in this way, computing first the demeaned series $\hat{\varepsilon}_t = r_t - \hat{c}_T - \hat{\rho}_T r_{t-1}$, and then estimating the remaining parameters given $\{\hat{\varepsilon}_t\}$.[2]

[1] The empirical analysis in Hansen and Lunde (2005) makes an out-of-sample comparison of over 300 different volatility models using daily exchange rate data. They find that none of these models is able to provide a significantly better forecast than the GARCH(1,1) model.

[2] The orthogonality condition $E\left(\partial l_t(\xi_0)/\partial\xi_{0m}\xi_{0v}\right) = 0$ holds for all the models considered in this paper. It is usual to estimate the parameters in the AR(1) model by Least Squares, whereas the parameters related to the conditional variance (and higher-order moments) must be estimated by ML.

Conditional normality is a fairly restrictive assumption which is widely accepted not to hold in the majority of applications involving real financial data. This fact is observed even when the data are sampled on a relatively low frequency basis which implies a high degree of aggregation. Fortunately, it is also widely accepted that the Normal assumption does not play a critical role when the main purpose is model fitting and/or volatility forecasting and, in fact, there are both computational and statistical reasons that have supported the wide use of the Gaussian GARCH in applied settings. First, the Gaussian log-likelihood function, $\mathcal{L}(\xi_0) = \sum_{t=1}^{T} l_t(\xi_0)$, is very tractable and typically does not pose any computational problems in order to be optimized numerically – hence, the GARCH model is directly implemented in most statistical packages, and can be estimated even with a spreadsheet. Second, and more importantly, the resultant estimation, namely $\hat{\xi}_T = \arg\max_{\xi_0} \mathcal{L}(\xi_0)$, is known to be $\sqrt{T}-$consistent and asymptotically normally distributed under certain regularity conditions even if η_t are not really Gaussian distributed (in this case $\hat{\xi}_T$ is referred to as the quasi maximum likelihood estimator [QML]); see Weiss (1986), Bollerslev and Wooldridge (1992), Lee and Hansen (1994), and Newey and Steigerwald (1997). These properties ensure tractability and accuracy for many applications in which the main aim is to obtain consistent estimates of ξ_0, and/or forecasts of the conditional variance process, which are simply determined as $\hat{h}_{T+s} = \hat{\omega}_T + \hat{\alpha}_T \hat{\varepsilon}_T^2 + \hat{\beta}_T \hat{h}_{T-1}$ for $s = 1$, and $\hat{h}_{T+s} = \hat{\omega}_T + \left(\hat{\alpha}_T + \hat{\beta}_T\right) \hat{h}_{T+s-1}$ for $s > 1$.

2.1.2. Student-*t* GARCH

The Normal assumption may be convenient, but it turns out to be too restrictive for applications on risk management and asset pricing, because these require the conditional density of η_t, and not just volatility estimations.[3] The failure of the Normal assumption is mainly due to the large degree of kurtosis that is typically observable in real data, which in turn is related to the magnitude and the frequency of extreme values that characterize almost any financial time series. Although the unconditional distribution implied by the Gaussian GARCH(1,1) model is leptokurtic (Bollerslev, 1986), often this model cannot generate large enough values to match the range which is observed in practice owing to limitations in its statistical properties; see Carnero, Peña and Ruiz (2004) for a discussion on this topic. Furthermore, the empirical distribution of the estimates $\hat{\eta}_t$, given the ML estimates $\hat{\xi}_T$, also suggest an excess of conditional kurtosis over the theoretical level which is implied by the Normal distribution.[4] Overall, the empirical evidence largely supports the existence of strong leptokurtosis in both the unconditional and conditional distributions of returns, thereby suggesting model misspecification in the Gaussian GARCH approach.

This observation motivated further extensions aiming to capture extreme movements through heavy-tailed distributions. A very simple, yet useful extension, was early suggested by Bollerslev (1987), who proposed a transformed Student t-distribution with v degrees of

[3] A leading example is the Value at Risk methodology. The percentiles of η_t, together with the forecasts of the future conditional variance, jointly determine the maximum expected loss of an asset at a certain significance level.

[4] For instance, the standardized residuals of the Gaussian GARCH model studied in Section 3 below have a kurtosis of nearly 6 in the in-sample period considered for the S&P index. The Jarque-Bera test for normality (JB=444.56) rejects the hypothesis of normality. Similar results are obtained for the remaining time-series.

freedom to accommodate the excess of kurtosis, *i.e.,*

$$\eta_t \sim iidt_v\left(0,1\right), \tag{5}$$

where the degrees of freedom parameter, v, is directly characterized by the shape of the underlying distribution, and can be estimated by ML from the available data (subject to the restriction $v > 2$ so that the variance process is well defined). Apart from a constant term, the relevant log-likelihood function is given by,

$$l_t\left(\xi_1\right) = \log\left(\frac{\Gamma\left(\left(\nu+1\right)/2\right)\right]}{\Gamma(\nu/2)}\right) - \frac{1}{2}\log\left(\left(v-2\right)h_t\right) - \frac{v+1}{2}\log\left[1+\frac{\varepsilon_t^2}{\left(v-2\right)h_t}\right], \tag{6}$$

with $\xi_1 = \left(v, \xi_0'\right)'$, and $\Gamma\left(\cdot\right)$ denoting the Gamma function. When $1/\hat{v}_T \to 0$, the conditional distribution approaches a Normal distribution and the Gaussian restriction may be acceptable. However, for small values such that $1/\hat{v}_T > 0$, the empirical distribution has fatter tails than the corresponding Normal distribution. For many empirical applications related to risk-management, such as Value at Risk, the Student-t GARCH model tends to provide a superior performance over the Gaussian GARCH model; see, for instance, Alexander (1998).

2.2. Further Approaches: Modelling Higher-Order Conditional Moments

2.2.1. The Student-t GARCHK Model

The Student-t GARCH model provides further flexibility to capture constant unconditional leptokurtosis. Obviously, there is no prior reason to believe that higher-order conditional moments should remain unchanged, other than for model simplicity and computational tractability. Consequently, Brooks et al., (2005) proposed a further extension of this model, the so-called Student-t GARCHK, by allowing the possibility of heterogeneity in the conditional distribution $f\left(\eta_t|I_{t-1}\right)$ due to time-varying kurtosis.

Considering the basic GARCH model, the key assumption now is that $\{\eta_t\}$ are conditionally distributed according to a Student-t distribution with a time-varying number of degrees of freedom, say v_t, which evolves independently of the dynamics followed by the conditional variance. In particular, the characteristic restriction is given by

$$f\left(\eta_t|I_{t-1}\right) \sim t_{v_t}\left(0,1\right); \;\; \nu_t = \frac{2\left(2k_t-3\right)}{k_t-3}, \tag{7}$$

where k_t is the conditional kurtosis of the process at time t. In the same spirit of the structural GARCH modelling, an autoregressive moving average process is used to capture the dynamics of the conditional kurtosis:

$$k_t = \kappa + \delta\left(\frac{\varepsilon_{t-1}^4}{h_{t-1}^2}\right) + \theta k_{t-1}. \tag{8}$$

As in equation (2), the parameter restrictions $\kappa > 0$, $\delta, \theta \geq 0$, are sufficient for ensuring positiveness in the resultant process. Note that k_t arises as a weighted combination of a long-run constant value, the previous kurtosis forecast, and a term with updated information of the conditional kurtosis as proxied by $\left(\varepsilon_{t-1}^2/h_{t-1}\right)^2$. Under the restriction $\delta = \theta = 0$, the model reduces to the constant kurtosis model studied in the previous subsection, which suggests an easy way to statistically test for the suitability of the time-varying specification. It is important to remark that the conditional variance process, h_t, and the conditional kurtosis process, k_t, are not contemporaneously functionally related, so they may be parameterized individually as desired by using different specifications than those discussed above, for instance, introducing nonlinearities or dependence upon other variables. The log-likelihood of the model, apart from a constant term, is given by,

$$l_t\left(\xi_2\right) = \log\left(\frac{\Gamma[(\nu_t+1)/2]}{\Gamma(\nu_t/2)}\right) - \frac{1}{2}\log\left((\nu_t-2)h_t\right) - \frac{\nu_t+1}{2}\log\left[1+\frac{\varepsilon_t^2}{h_t(\nu_t-2)}\right], \tag{9}$$

with $\xi_2 = \left(\kappa, \delta, \theta, \xi_0'\right)'$, and $v_t > 4$ to ensure the existence of the first fourth-order moments. The similitudes between (9) are (6) are obvious, since the time-varying kurtosis generalizes the constant kurtosis model by simply allowing time variability in the degrees of freedom parameter.

The empirical in-sample evidence discussed in Brooks et al., (2005, Section 3) for several US and UK equities and bonds supports the hypothesis of heterogeneity in the conditional kurtosis, largely outperforming the specification with constant kurtosis.

2.2.2. The Gram-Charlier GARCHK Model

Let us start this section by recalling the dynamics of the conditional variance-kurtosis models which have been discussed thus far:

$$\begin{aligned} r_t &= E_{t-1}(r_t) + \varepsilon_t;\ \varepsilon_t = h_t^{1/2}\eta_t, \\ h_t &= \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}, \\ k_t &= \kappa + \delta\left(\frac{\varepsilon_{t-1}^4}{h_{t-1}^2}\right) + \theta k_{t-1}, \end{aligned} \tag{10}$$

given the set of unknown parameter $\xi_2 = \left(\kappa, \delta, \theta, \xi_0'\right)'$. Instead of imposing a particular assumption on the conditional distribution of η_t as we did in the previous sections (Normal, Student-t, or Student-t with time-varying degrees of freedom), we may use a Gram-Charlier type of expansion to fit semi-parametrically the unknown density function $f\left(\eta_t|I_{t-1}\right)$. This is the central point discussed in the model proposed in León et al. (2005), which we summarize below.

Under certain regularity conditions, any probability density function (pdf henceforth) can be expanded in an infinite series of derivatives of the standard Normal density, $\phi(\eta_t)$, as follows,

$$f\left(\eta_t|I_{t-1}\right) = \phi(\eta_t)\sum_{s=0}^{\infty} d_{st}H_s(\eta_t), \tag{11}$$

where $H_s(\eta_t)$ is the s^{th} order Hermite polynomial defined in terms of the s^{th} order derivative of the Gaussian pdf:

$$\frac{d^s\phi(\eta_t)}{d\eta_t^s} = (-1)^s\phi(\eta_t)H_s(\eta_t). \tag{12}$$

For applied purposes, the infinite expansion is not operative and has to be truncated. Thus, considering the finite expansion (approximation) of $f(\eta_t|I_{t-1})$ in (11) with a truncation factor up to the fourth-order moment, we obtain:

$$f(\eta_t|I_{t-1}) \simeq \phi(\eta_t)\left[1 + \frac{s_t}{3!}(\eta_t^3 - 3\eta_t) + \frac{k_t - 3}{4!}(\eta_t^4 - 6\eta_t^2 + 3)\right] = \phi(\eta_t)\psi(\eta_t), \tag{13}$$

where the polynomial $\psi(\eta_t)$ is defined implicitly, and the terms s_t and k_t correspond to the conditional skewness and kurtosis, respectively. Note that the resulting approximation, $\phi(\eta_t)\psi(\eta_t)$, is characterized by the underlying dynamics of the conditional moments up to the fourth-order moment and the set of unknown parameters ξ_2. We do not overload unnecessarily the notation by remarking the latter feature as this is completely clear at this point.

Since the approximation based on a finite polynomial expansion of $f(\eta_t|I_{t-1})$ implies certain amount of truncation error, the right-hand side of (13) cannot be seen as a proper density function. The main reason is that $\phi(\eta_t)\psi(\eta_t)$ is not ensured to be almost surely positive uniformly on the parameter space of ξ_2. This unappealing feature does not only suppose a major shortcoming from a theoretical viewpoint, but also may cause the failure of the ML estimation in empirical settings. León et al. (2005) propose a solution building on the same methodology as Gallant and Nychka (1987), Gallant and Tauchen (1989) and Gallant Nychka and Fenton (1996). In essence, they achieved a well-defined pdf by first using a simple positive transformation of $\psi(\eta_t)$ that ensures almost-surely positiveness (namely, squaring $\psi(\eta_t)$, although other transformation in similar spirit are possible as well), and then re-normalizing the resulting function by a suitable scaling factor such that the resulting function intregates up to one. More specifically, given the normalizing factor,

$$\Delta_t = \int_{-\infty}^{\infty} \phi(\eta_t)\psi^2(\eta_t)\, d\eta_t = 1 + \frac{s_t^2}{3!} + \frac{(k_t - 3)^2}{4!} \tag{14}$$

the transformed Gram-Charlier probability density function, denoted $f^*(\eta_t|I_{t-1})$, can readily be written as

$$f^*(\eta_t|I_{t-1}) = \left(\frac{1}{\Delta_t}\right)\phi(\eta_t)\psi^2(\eta_t). \tag{15}$$

Note that the Hermite polynomial that characterize $\psi(\eta_t)$ convey information about the empirical degree of conditional moments, and so does $f^*(\eta_t|I_{t-1})$, from which ξ_2 can be identified from the observable data. However, the terms s_t and k_t no longer admit the interpretation of conditional moments, and further adjustments to forecast the conditional moments given $f^*(\eta_t|I_{t-1})$ are necessary; see Section 3.2.2 for further details. Since we are restricting ourselves to symmetric conditional distributions, we set $s_t = 0$ in (13) for all t, and denote as $\bar{\psi}(\eta_t)$ the restricted version of the model. Hence, the normalizing factor reduces accordingly to $\bar{\Delta}_t = 1 + (k_t - 3)^2/4!$, and the corresponding log-likelihood function, apart from a constant term, is given by

$$l_t(\xi_2) = -\frac{1}{2}\ln h_t - \frac{\varepsilon_t^2}{2h_t} - \ln\bar{\Delta}_t + \ln\left[\bar{\psi}^2(\eta_t)\right]. \quad (16)$$

It is worth remarking at this point the similitudes between this function and the Gaussian log-likelihood function (4) used in the basic GARCH model: The Gram-Charlier log-likelihood function simply adds two adjustment terms to the latter in order to capture the non-Gaussian features of the data, in our case, conditional kurtosis dynamics. In fact, the Gaussian likelihood (4) is nested as a particular case by setting constant kurtosis $(\delta = \theta = 0)$, equal to that of the Normal distribution (*i.e.*, $k_t = \kappa = 3$). As in Brooks et al. (2005), the empirical results in León et al. (2005) indicate a significant presence of time-variability in the higher-order moments which support the suitability of conditional kurtosis.

3. Empirical Analysis

3.1. The Data

The data used in this study are daily returns (scaled by a factor of 100) of the S&P500 index (SP), the GBP(£)/US Dollar($) exchange rate (FX), and the 10 years Treasury Notes (TN). The series are sampled over the period June 9, 1993 to June 8, 2008 for a total of $T = 3,912$ observations obtained from Datastream. Table 1 displays some descriptive information for the total sample. As expected, stock returns are much more volatile (as measured by the unconditional volatility) than the other series. The unconditional distribution of any of these series shows clearly non-Gaussian features, such as a (mild) skewness in the case of the SP and FX series, and a remarked excess of kurtosis over the Normal distribution due to outliers in the three time series considered. The Jarque-Bera tests for normality are easily rejected, particularly in the case of the stock index time-series.[5] The analysis of dependence through the Ljung-Box portmanteau test statists shows some form of weak dependence in the level of the returns, and a strong, persistent correlation in higher-order moments.

3.2. Modelling and Out-of-Sample Forecasting

We split the total sample into an in-sample period to estimate the models, and an out-the-sample window to make a total of $N = 500$ one-step predictions of the conditional variance and kurtosis by means of a rolling-window procedure. To assess the ability of the different GARCH models involved, we need a time-varying measure of the actual conditional moments. In both cases, the main problem for addressing forecasting ability is that the true conditional variance and kurtosis are not observable and have to be approached by means of statistical proxies which often can only provide a crude measure.

[5]The Jarque-Bera test for normality uses the test statistic $JB = T\left[\frac{s^2}{3!} + \frac{(k-3)^2}{4!}\right]$, where s and k are the sample skewness and kurtosis, respectively. The test is asymptotically distributed as $\chi^2(2)$, and is rejected for non-zero values of the sampled skewness and/or excess of kurtosis.

Table 1. Descriptive statistics for daily returns

Statistic	SP	FX	TN
Sample		9/06/1993 - 8/06/2008	
Observations		3913	
Mean	0.0561	-0.0067	-0.0093
Median	0.1186	0.0000	0.0000
Maximum	16.107	3.4233	6.1278
Minimum	-15.419	-4.2211	-5.1238
St. Dev.	1.8624	0.5112	1.1807
Skewness	-0.1033	-0.0131	0.3568
Kurtosis	11.01	5.8469	5.4564
Jarque-Bera	10470 (0.00)	1324 (0.00)	1066.8 (0.00)
Ljung-Box Q(1)-r_t	17.29 (0.00)	16.65 (0.00)	4.0521 (0.04)
Ljung-Box Q(20)-r_t^2	1047.1 (0.00)	229.4 (0.00)	1322.2 (0.00)
Ljung-Box Q(20)-r_t^3	93.50 (0.00)	399.7 (0.00)	73.28 (0.00)
Ljung-Box Q(20)-r_t^4	120.7 (0.00)	459.7 (0.00)	297.1 (0.00)
LR: $(\kappa, \delta, \theta) = 0$	146.9 (0.00)	154.4 (0.00)	153.2 (0.00)
LR: $(\delta, \theta) = 0$	77.70 (0.00)	9.43 (0.00)	14.1 (0.00)

The Jarque-Bera normality test is asymptotically distributed as a $\chi^2(2)$ under the null of normality, the Ljung-Box is asymptotically distributed as a $\chi^2(\varsigma)$, ς being the autocorrelation order, the Likelihood Ratio test (LR) is asymptotically distributed as a $\chi^2(q)$ being q the number of restrictions under the null, (asymptotic p-values in parenthesis).The critical values of $\chi^2(1)$, $\chi^2(2)$, $\chi^2(3)$ and $\chi^2(20)$ are 3.84, 5.99, 7.81, 31.41, at 5% level, respectively.

In the context of volatility forecasting, the empirical proxies considered in most papers are based on measurable transformations of the absolute-valued unexpected returns $|\varepsilon_t|$, most frequently ε_t^2. Following this literature, we shall consider ε_{T+1}^2 as a proxy for variance in this paper.[6] Although there exists an agreement (at least empirically) on how conditional variance could be proxied, to the best of our knowledge there is no obvious guidance on how to approach conditional kurtosis. Given that the estimation bias may be more significant when considering higher-order moments, and that the choice of the proxy necessarily conditions the results, we consider different proxies for the conditional kurtosis. In particular, we take the sample kurtosis in the m days immediately following the last in-sample observation, namely

$$\bar{k}_{T+1,m} = \left(\frac{\frac{1}{m}\sum_{j=1}^{m} \left(\varepsilon_{T+j} - \bar{\varepsilon}_m\right)^4}{\left[\frac{1}{m}\sum_{j=1}^{m} \left(\varepsilon_{T+j} - \bar{\varepsilon}_m\right)^2\right]^2} \right); \quad \bar{\varepsilon}_m = \sum_{j=1}^{m} \varepsilon_{T+j}/m, \tag{17}$$

[6] The availability of intraday data has motivated the use of a new strand of proxies for volatility which provide a more accurate measure based on realized volatility.

with $m = \{5, 50, 500\}$.[7] The choice of m here seeks a compromise between the tautological notion of *conditional* and the statistical problems related to measurement errors in the relevant statistic when using a few number of observations. As $m \to 1$, the proxy is more erratic and extremely noisy, and can severely be influenced by a few large observations, whereas the largest window in our analysis ($m = 500$ observations) is related to the out-of-sample unconditional kurtosis.

We consider the MSE as a loss function for any of the conditional variance and kurtosis forecasts series, *i.e.*, we compute the statistic $N^{-1}\sum_{i=1}^{N}\left(\hat{\pi}_{T+1,i} - \bar{\pi}_{T+1,i}\right)^2$ for each model and time-series, where $\hat{\pi}_{T+1,i}$ is the i-th prediction (either conditional variance or kurtosis) and $\bar{\pi}_{T+1,i}$ the proxy for the actual value. The forecasting performance is compared in statistical terms by means of the test proposed by Diebold and Mariano (1995). This test assumes no differences between the loss functions of two alternative models under the null hypothesis. The null is rejected for large values of the statistic $DM = \bar{x}/\sqrt{2\pi f_x\left(w=0\right)/N}$, where $\bar{x}$ denotes the sample mean of the differences in the forecasting errors of the two alternative models, $f_x\left(w=0\right)$ is the spectral density function of the forecasting error differences evaluated at the zero frequency (long-run variance), and N is the total number of forecasts. The statistic is asymptotically distributed as a standard Normal random variable under the null.

3.2.1. In-sample Analysis

The slight (positive) autocorrelation pattern in the conditional mean of the returns is filtered out by fitting an AR(1) process estimated by Least-Squares. Given the demeaned series, $\hat{\varepsilon}_t$, all the conditional variance-kurtosis models are then estimated by optimizing the corresponding log-likelihood functions using the Newton-Raphson method, and initializing the conditional variance and kurtosis dynamics with values equal to the corresponding unconditional moments. Convergence in the optimization process to the global extremes is easily obtained in the case of the simplest Gaussian and Student-t GARCH models. Similarly, the estimation of the Student-t and Gram-Charlier GARCHK models is not computationally troublesome providing that the starting values are chosen properly.[8] The main results from the estimation in the in-sample period are displayed in Table 2 below.

The estimates of the conditional variance for the three series show the usual degree of high persistence and low sensitivity to shocks which is commonly observed in daily asset returns. Persistence is related to the magnitude of the coefficient $\hat{\alpha}_T + \hat{\beta}_T$, which tends to be slightly smaller than unity, while sensitivity to new information arrivals is measured through $\hat{\alpha}_T$, which takes small values empirically. Owing to the large degree of unconditional kurtosis in the data, the Student-t GARCH model determines a degrees of freedom parameter around 6 for all the series. This result confirms that extreme observations in real data are much more likely to occur in relation to the Normal distribution. Assuming that the true distribution is a Student-t, higher-order moments larger than 6 would not be well-defined. The models that allow for time-varying kurtosis reject the hypothesis of constant kurtosis, since the restriction $\delta = \theta = 0$ is easily rejected by a standard Likelihood Ratio

[7] We also used othe values for m, noting no qualitative difference with the results reported in the main text.

[8] Also, in order to avoid convergence to local extremes, the optimization routine is monitored using a grid of different starting values. Normal convergence is obtained in all the cases.

Table 2. GARCH in-sample estimation results

		GARCH-n	GARCH-t	GARCHK-t	GARCHK-GC
		Panel 1: SP			
Mean equation	μ	0.056 (1.73)			
	ϕ	0.083 (5.26)			
Variance equation	ω	0.118 (3.05)	0.107 (2.76)	0.107 (2.96)	0.113 (3.49)
	α	0.159 (5.36)	0.139 (4.94)	0.155 (5.66)	0.155 (6.01)
	β	0.809 (22.7)	0.832 (23.3)	0.819 (24.2)	0.807 (25.9)
Kurtosis equation	κ			0.801 (1.49)	1.887 (4.12)
	δ			0.024 (1.07)	0.008 (0.64)
	θ			0.842 (8.87)	0.448 (3.39)
DoF	ν		6.202 (9.60)	[5.977]	
AIC		3.731	3.704	3.681	3.687
		Panel 2: FX			
Mean equation	μ	-0.007 (-0.89)			
	ϕ	-0.065 (-4.08)			
Variance equation	ω	0.007 (1.96)	0.003 (1.89)	0.004 (1.83)	0.003 (2.22)
	α	0.036 (3.30)	0.031 (3.96)	0.032 (3.59)	0.038 (5.83)
	β	0.937 (44.1)	0.967 (80.4)	0.955 (69.6)	0.944 (93.9)
Kurtosis equation	κ			2.883 (1.63)	2.684 (2.66)
	δ			0.049 (0.28)	0.008 (5.82)
	θ			0.599 (2.61)	0.223 (0.77)*
DoF	ν		5.319 (10.7)	[5.344]	
AIC		1.478	1.424	1.422	1.432
		Panel 3: TN			
Mean equation	μ	-0.0091 (-0.46)			
	ϕ	0.0322 (2.01)			
Variance equation	ω	0.012 (2.45)	0.007 (2.11)	0.007 (2.13)	0.009 (2.49)
	α	0.039 (5.16)	0.038 (5.71)	0.041 (4.87)	0.037 (5.73)
	β	0.951 (96.8)	0.958 (126)	0.954 (104)	0.953 (117)
Kurtosis equation	κ			4.197 (2.11)	0.671 (3.73)
	δ			0.665 (1.38)	0.001 (0.28)
	θ			0.216 (0.72)	0.806 (15.8)
DoF	ν		5.685 (10.3)	[4.695]	
AIC		3.009	2.969	2.956	2.963

Estimation results (robust QML t-statistics in brackets) for the Gaussian GARCH (GARCH-n), the Student-t GARCH (GARCH-t), the Student-t GARCHK (GARCHK-t) and the Gram-Charlier GARCHK models (GARCHK-GC). AIC denotes the Akaike Information Criterion statistic. The row DoF shows the estimated degrees of freedom parameter in the Student-t distribution under the GARCH-t model, $\widehat{\nu}_T$, and the unconditional kurtosis implied by the estimated parameters, $\kappa/\left(1-\delta-\theta\right)$, in case of the GARCHK-t model.

test in the three time series considered. Overall, the empirical evidence we observe perfectly agrees with the results in Brooks et al. (2005) and León et al. (2005), showing that extending GARCH models toward accounting for time-varying kurtosis leads to a better in-sample fitting.

There are two further interesting features that arise when comparing the results observed across the different types of estimation techniques involving time-varying kurtosis, and the different classes of financial assets considered. First, whereas the estimates of the GARCH equation remain virtually unaltered given the different models, the estimates of the driving parameters of the conditional kurtosis, say $\xi_{2k} = (\kappa, \delta, \theta)'$, reveal very different dynamics depending on whether a Student-t or a transformed Gram-Charlier densities is used. In particular, the dynamics of the conditional kurtosis tend to be much more persistent under the assumption of the Student-t distribution, whereas the parameter related to the arrivals of new information, δ, tend to be not significant under the Gram-Charlier fitting. This feature shows that the driving parameters of the kurtosis are particularly sensitive to the model assumptions that must capture the actual tail-behavior of the underlying distribution. Second, the estimated dynamics of the conditional kurtosis of the Treasure Notes time series differ remarkably from those estimated for the SP and the FX series. This evidence is in sharp contrast to the dynamics followed by the variance, which tend to show the same type of pattern across fairly different classes of financial assets.

Both features seem to suggest that, whereas the dynamics of the conditional variance can always be characterized by 'stylized features' (a small estimated α coefficient, and high persistence as measured by the estimated term $\alpha+\beta$), the dynamics of the conditional kurtosis may exhibit a much more idiosyncratic behavior and vary across the class of financial asset and sample period considered. This empirical observation should not be very surprising, since the dynamics of the conditional kurtosis are strongly related to the likelihood and the magnitude of extreme observations (*i.e.*, outliers), which in turn are known to show a large degree of heterogeneity and irregular behavior. The main implication is that, whereas GARCH models tend to yield similar estimation outcomes regardless the financial time-series and the market considered, conditional-kurtosis modelling may yield quite different results depending on the asset considered, and the relevant assumption about the tail-behaviour of the conditional distribution.

3.2.2. Out-of-sample Analysis

One-step forecasts of conditional variance from the Gaussian GARCH, Student-t GARCH, and Student-t GARCHK models are easily obtained as

$$\widehat{h}_{T+1} = \widehat{E}_T(h_{T+1}) = \widehat{\omega}_T + \widehat{\alpha}_T \hat{\varepsilon}_T^2 + \widehat{\beta}_T \widehat{h}_T. \tag{18}$$

For the Gram-Charlier GARCHK model, the forecast $\widehat{h}_{T+1}$ is obtained as

$$\widehat{h}_{T+1} = \left(\widehat{\omega}_T + \widehat{\alpha}_T \hat{\varepsilon}_T^2 + \widehat{\beta}_T \widehat{h}_T\right) \left[\frac{1+216\widehat{d}_4^2}{1+24\widehat{d}_4^2}\right]; \;\; \widehat{d}_4^2 = \frac{\widehat{k}_{T+1}-3}{4!}. \tag{19}$$

For the Gaussian GARCH model, the kurtosis is constant and equals 3, whilst for the Student-t GARCH model the kurtosis forecast is given by $\widehat{k}_{T+1} = 3(\widehat{\nu}_T - 2)/(\widehat{\nu}_T - 4)$, being $\widehat{\nu}_T$ the degrees of freedom parameter estimated in the in-sample period. The forecasts of the conditional kurtosis of the Student-t GARCH models are simply given by

$$\widehat{k}_{T+1} = \widehat{E}_T(k_{T+1}) = \widehat{\tau}_T + \widehat{\delta}_T \left(\frac{\hat{\varepsilon}_T^2}{\hat{h}_T} \right)^2 + \widehat{\theta}_T \widehat{k}_T \tag{20}$$

while the Gram-Charlier GARCHK determines a conditional forecast given by:

$$\widehat{k}_{T+1} = \frac{(3 + 2952\widehat{d}_4^2 + 12\widehat{d}_4^2)(1 + 24\widehat{d}_4^2)}{(1 + 216\widehat{d}_4^2)^2} \tag{21}$$

The MSEs for the GARCH models used to forecast conditional variance and kurtosis are presented in Table 3, while Table 4 shows the p-values related to the Diebold-Mariano test (a value smaller than 0.05 implies that the model with smallest MSE in the comparison yields a significant improvement at the 95% confidence level). Some comments follow. We note that the differences in the MSE loss functions for volatility forecasting are not generally significant across the different GARCH models considered in all the series analyzed. This is not surprising, since Gaussian GARCH forecasts are known to be accurate in mean from the property of consistency (discussed in Section 2), and because the dynamics of the conditional kurtosis are modelled independently of the dynamics followed by the conditional variance, we should expect no interaction between them. Therefore, considering further dynamics in the higher-order moments, or allowing for excess of kurtosis in the conditional distribution, would hardly improve empirically the on-average accuracy of the variance forecasts made by a simple Gaussian GARCH model.

In relation to (un)conditional kurtosis, the main results of our analysis are the following. First, the model that yields better out-of-sample forecast of the unconditional kurtosis given the proxy $\bar{k}_{T+1}^{(500)}$, consistently across the three series considered, is the Gram-Charlier GARCHK model. Owing to its semi-parametric nature, the Gram-Charlier type modelling does not rely upon a specific assumption on the underlying distribution of the data, which provides robustness against potential departures over the parametric models (which, on the other hand, would be consistent and more efficient under correct specification). As we have seen from the empirical results for the conditional variance, robustness turns out to be a precious property when making predictions, and of course this property also applies when considering higher-order moments. Second, and related to the previous consideration, we observe that the Gram-Charlier GARCHK model largely overperforms the Student-t GARCHK model in forecasting conditional kurtosis, as proxied for small values of m in $\bar{k}_{T+1}^{(m)}$. Overall, these findings suggest that the assumption of a Student-t distribution with time-varying degrees of freedom may not be appropriate for applied purposes related to conditional-kurtosis forecasting.

Finally, we can only observe mixed and somewhat inconclusive evidence regarding the empirical importance of modelling time-varying kurtosis, since only in the case of the TN time series there seems to be statistical improvements over the simplest Gaussian GARCH model, and only when using the Gram-Charlier GARCHK specification. There are several reasons that may explain, at least partially, the seeming failure of the conditional-kurtosis GARCH models in the SP and FX time series. First, the presence of measurement errors in the proxy considered $\bar{k}_{T+1}^{(m)}$ may end up playing a significant role in the MSE loss-function (particularly as $m \to 1$) given the particularities of the time-series involved, and leading to distorted empirical conclusions. Second, although the conditional-kurtosis GARCH models may provide a better fit in the in-sample period, this does not necessarily imply that these models have to improve the out-of-sample forecast performance. There are two different reasons supporting this statement, both of them being rooted in the high degree of heterogeneity and idiosyncratic behavior of the (conditional) kurtosis. On the one hand, Korkie, Sivakumar and Turtle (2006) have argued that the persistence in the higher-order moments of financial returns may be a statistical artifact related to variance spillovers, so there would not be any gain from forecasting these dynamics. If this pervasive effect exists, it may be more important for some variables than for others, as we have documented statistical gains from modelling kurtosis in the case of the TN time series. On the other hand, even if the conditional kurtosis does really change over time, its dynamics are necessarily linked to the particularities of the data generating process that drives extreme observations and irregular outliers. This feature brings up further statistical concerns, because the time-varying kurtosis process has to be characterized empirically by finite-sample ML estimates that are strongly conditioned by the assumption on the underlying distribution (as we have seen in the previous section) and which, furthermore, may suffer from important biases related to the occurrence and magnitude of outliers in the in-sample period. A few irregular, large enough outliers are perfectly able to strongly bias the ML estimates of the conditional kurtosis in the attempt to provide the best possible in-sample fit, given the underlying assumption that determines the theoretical likelihood and magnitude of extreme observations, but at the logical cost of poorly forecasting on-average the out-of-sample dynamics in which such extreme observations do not occur.

Table 3. Out-of-sample volatility and kurtosis MSE forecasting performance

	A: Gaussian GARCH				C: Student-t GARCHK			
		$\bar{k}_{T+1}^{(m)}$				$\bar{k}_{T+1}^{(m)}$		
	h_{T+1}	5	50	500	h_{T+1}	5	50	500
SP	37.88	5.627	0.221	2.177	37.81	14.08	7.662	1.698
FX	0.133	22.17	1.266	0.771	0.133	34.74	19.31	10.75
TN	0.719	5.321	0.717	5.847	0.717	31.06	24.46	13.19

	B: Student-t GARCH				D: Gram-Charlier GARCHK			
		$\bar{k}_{T+1}^{(m)}$				$\bar{k}_{T+1}^{(m)}$		
	h_{T+1}	5	50	500	h_{T+1}	5	50	500
SP	37.53	14.60	8.201	1.937	37.97	6.305	0.149	0.643
FX	0.133	34.64	19.07	10.56	0.133	21.90	1.229	0.011
TN	0.715	17.97	10.52	1.653	0.717	6.184	0.552	3.017

This table shows the Mean Square Error (MSE) for the one-step ahead conditional variance and kurtosis forecasts from the different GARCH models used in the analysis. The proxies considered for the kurtosis, $\bar{k}_{T+1,m}$, are estimated from the sample kurtosis of the first $m = 5$, 50 and 500 days immediately following the last day in the in-sample window.

Table 4. Diebold and Mariano statistics

	GARCH-t				GARCHK-t				GARCHK-GC			
		$\bar{k}^{(m)}_{T+1}$				$\bar{k}^{(m)}_{T+1}$				$\bar{k}^{(m)}_{T+1}$		
	h_{T+1}	5	50	500	h_{T+1}	5	50	500	h_{T+1}	5	50	500
SP												
GARCH-n	0.05	0.00	0.00	0.00	0.32	0.00	0.00	0.00	0.22	0.00	0.00	0.00
GARCH-t					0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
GARCHK-t									0.07	0.00	0.00	0.00
FX												
GARCH-n	0.28	0.00	0.00	0.00	0.21	0.00	0.00	0.00	0.06	0.29	0.00	0.00
GARCH-t					0.34	0.13	0.09	0.08	0.49	0.00	0.00	0.00
GARCHK-t									0.44	0.00	0.00	0.00
TN												
GARCH-n	0.03	0.00	0.00	0.00	0.09	0.00	0.00	0.07	0.01	0.00	0.00	0.00
GARCH-t					0.13	0.02	0.01	0.01	0.09	0.00	0.00	0.00
GARCHK-t									0.35	0.00	0.00	0.02

This table reports the results of the DM test for the difference of the MSE loss function from the GARCH models under analysis (see notation in Table 2). The entries are DM test p-values for the predictive ability of the model in the row versus the model in the column.

4. Concluding Remarks

Several papers have argued that the kurtosis of returns may exhibit clusters and time dependency similar to the characteristic patterns which are observable in the proxies of conditional variance. The modelling of the conditional third- and fourth-order moments tend to improve the in-sample goodness of fit over the simplest GARCH models that assume constant higher-order moments. The main aim of this paper is to provide better insight on whether accounting for time-varying kurtosis is valuable for out-of-sample forecasting of both conditional variance and conditional kurtosis, and which procedure (among several of the parametric and semi-parametric alternatives that have been suggested in the literature) is better suited for empirical purposes.

As in the previous literature, our empirical results on three different classes of financial assets confirm that the semi-parametric Gram-Charlier and the Student-t GARCHK models that allow for time-varying kurtosis provide a better in-sample goodness-of-fit over the constant-kurtosis GARCH models, with the parametric Student-t distribution slightly overperforming the Gram-Charlier type distribution. For forecasting purposes, the best procedure to forecast the unconditional kurtosis seems to be the Gram-Charlier GARCHK model, as the semi-parametric nature of this approach provides robust properties against model misspecification which may ruin the out-of-sample forecasting ability of the model. Similarly, this methodology largely overperforms the Student-t distribution with time-varying degrees of freedom parameter in forecasting conditional kurtosis, which overall suggests that the semi-parametric approximation may be better indicated in practice. Unfortunately, the Gram-Charlier GARCHK model does not always achieve a significant success in beating the forecasts made by the simplest Gaussian GARCH model, at least given the proxies for conditional kurtosis considered in this paper. The lack of conclusive results for some of the time series analyzed may be, at least partially, a statistical artifact due to the sizeable measurement errors in the proxies used in the analysis. However, it is also possible that the empirical success of forecasting higher-order moments strongly depends on the class of asset and the sample period considered, given that the data generating process of the conditional kurtosis does not seem to exhibit the same degree of parameter uniformity as, for instance, the conditional variance does: whereas we always observe the same sort of stylized features in the GARCH-type estimates of the conditional variance, the conditional kurtosis exhibits a large degree of idiosyncratic behavior. Hence, the econometric modelling allowing for time-varying kurtosis may not generally necessarily enhance the out-of-sample forecasting performance of the models, even if in-sample results seem to suggest the opposite. More research on the empirical role of the dynamics of the conditional kurtosis seems deserved.

References

[1] lexander CO. 1998. Volatility and correlation: methods, models and applications. In: *Risk Management and Analysis: Measuring and Modelling Financial Risk* (C.O. Alexander, Ed.) Wiley.

[2] ollerslev T. 1986. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31**, 307-327.

[3] ollerslev T. 1987. A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* **69**, 542-547.

[4] ollerslev T, Wooldridge JM. 1992. Quasi-Maximum likelihood estimation and inference in dynamics models with time-varying covariances. *Econometrics Reviews* **11**, 143-172.

[5] arnero MA, Peña D, Ruiz E. 2007. Effects of outliers on the identification and estimation of GARCH models. *Journal of Time Series Analysis* **4**, 471-497.

[6] hristie-David R, Chaudhry M. 2001. Coskewness and cokurtosis in futures markets. *Journal of Empirical Finance* **8**, 55-81.

[7] hunhachinda P, Dandapani K, Hamid S, Prakash A. 1997. Portfolio selection and skewness: evidence from international stock markets. *Journal of Banking and Finance* **21**, 143–167.

[8] iebold FX, Mariano R. 1995. Comparing predictive accuracy. *Journal of Business and Economics Statistics* **13**, 253-264.

[9] ueker MJ. 1997. Markov switching in GARCH processes and mean-reverting stock-market volatility. *Journal of Business and Economic Statistics* **15**, 26-34.

[10] ang H, Lai TY. 1997. Co-kurtosis and capital asset pricing. *The Financial Review* **32**, 293-307.

[11] allant AR, Nychka DW. 1987. Seminonparametric maximum likelihood estimation. *Econometrica* **55**, 363-390.

[12] allant AR, Tauchen G. 1989. Seminonparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications. *Econometrica* **57**, 5, 1091-1120.

[13] allant AR, Nychka DW, Fenton V. 1996. Convergence rates of SNP density estimators. *Econometrica* **64**, 719-727.

[14] ansen BE. 1994. Autoregressive conditional density estimation. *International Economic Review* **35**, 705-730.

[15] ansen PR, Lunde A. 2005. A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* **20**, 873-889.

[16] arvey CR, Siddique A. 1999. Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis* **34**, 465–487.

[17] arvey CR, Siddique A. 2000. Conditional Skewness in Asset Pricing Tests. *Journal of Finance* **55**, 1263-1295.

[18] ondeau E, Rockinger M. 2000. Conditional volatility, skewness and kurtosis: existence and persistence. *Working Paper* **77**. Banque de France & *HEC School of Management*.

[19] ondeau E, Rockinger M. 2001. Gram-Charlier Densities. *Journal of Economic Dynamics and Control*, **25**, 1457-1483.

[20] orkie B, Sivakumar R, Turtle HJ. 2006. Variance spillover and skewness in financial asset returns. *The Financial Review* **41**, 139-156.

[21] ee SW, Hansen BE. 1994. Asymptotic properties of the maximum likelihood estimator and test of the stability of parameters of the GARCH and IGARCH models. *Econometric Theory* **10**, 29-52.

[22] eón A, Rubio G, Serna G. 2005. Autoregressive conditional volatility, skewness and kurtosis. *Quarterly Review of Economics and Finance* **45**, 599-618.

[23] elson D. 1996. Asymptotic filtering theory for multivariate ARCH models. *Journal of Econometrics* **71**, 1–47.

[24] ewey W, Steigerwald D. 1997. Asymptotic bias for quai-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica* **65**, 587-599.

[25] atton A. 2006. Volatility forecast comparison using imperfect volatility proxies. *Research Paper* **175**. Quantitative Finance Research Centre. University of Technology Sydney.

[26] argan JD. 1976. Econometric estimators and the Edgeworth approximation. *Econometrica* **44**, 421-448.

[27] chmidt R. 2002. Tail dependence for elliptically contoured distributions. *Mathematical Methods of Operations Research* **55** (2), 301-327.

[28] eiss AA. 1986. Asymptotic theory of ARCH models: Estimation and testing. *Econometric Theory* **2**, 107-131.

In: Economic Forecasting
Editor: Alan T. Molnar, pp. 249-262
ISBN: 978-1-60741-068-3

Chapter 11

TRANSPORTING TURKISH EXAM TAKERS: A NEW USE FOR AN OLD MODEL

***Nurhan Davutyan*[a] *and Mert C. Demir*[b]**
Marmara University-Goztepe Campus,
Istanbul, Turkey

Abstract

This paper argues that the transportation model of linear programming can be used to administer the Public Personnel Language Exam of Turkey in many different locations instead of just one, as is the current practice. It shows the resulting system to be much less costly. Furthermore, once the decision about number of locations is made, the resulting system can be managed either in a centralized or decentralized manner. A mixed mode of management is outlined, some historical perspectives on the genesis of the transportation model are offered and some ideas regarding the reasons for the current wasteful practices are presented. The possibility of applying the same policy reform in other MENA (Middle East and North Africa) countries is discussed in brief.

1. Introduction

Most optimization textbooks, without going into their history and evolution, present the transportation model as a well-known technique and link it to early work by the 1975 Nobel laureate Tjalling Koopmans prior to and during WWII. See for instance Denardo[1]. However, its roots go much deeper. Thompson and Thore[2] explain that mathematician Gaspard Mongé first formulated the model in the 1790s. While working on the construction of military fortifications for Napoleon, he had encountered a "cut and fill" problem, that of moving piles

[a] E-mail address: ndavutyan@marmara.edu.tr, dnurhan@hotmail.com. Nurhan Davutyan, Professor, Department of Industrial Engineering, Marmara University-Goztepe Campus 34722, Istanbul, Turkey (Corresponding author)
[b] E-mail address: mert.c.demir @gmail.com. Mert Demir: Director, Pensions Monitoring Center, Kozyatağı Istanbul/Turkey
[1] Denardo, E. 2002. *The Science of Decision-Making*, John Wiley, New York, p107.
[2] Thompson, G. and Thore, S. 1992. *Computational Economics*, Scientific Press, San Francisco, pp 6-7.

of dirt from locations where it was not needed to locations where it was needed. He stated that problem – without solving it – in mathematical form as a kind of transportation problem.

For our purposes, it is important to stress the following point: Such a formulation requires recognizing the "work/effort" of the soldiers, who would presumably do the moving, as something valuable and worth optimizing. In our modern world – at least in the Western Hemisphere – where almost everything is explicitly priced, this point may appear self-evident but it is worth emphasizing. Moreover we should recall that unpaid labor of all sorts – known as corvée – was quite common prior to 1789, in Ancient Regime, France. In fact, we may conjecture that ideas regarding civil liberties ushered in at that time, the abolition of corvée and Mongé's formulation of the problem were interrelated. The life and career of Gaspard Mongé, an ardent libertarian, support this view, Ball[3]. Schumpeter[4] gives numerous examples of these kinds of related phenomena whereby the emergence of their "necessary conditions" spurs discoveries and innovations. In anticipation of making a point in our next section, we hope Turkey's accession talks with the European Union will play a triggering role, leading to developments comparable to the 18th century abolition of corvée in France.

2. Setting

Twice a year, in May and November, the Student Selection and Placement Center (SSPC), a governmental agency in Turkey, administers a language exam known as the Public Personnel Language Exam or PPLE by its acronym. The exam takes place in Ankara, the capital. Anyone in Turkey who wants to prove her/his proficiency officially, in a language other than Turkish has to take this exam once every five years. The career paths and salaries of the individuals involved are linked to their exam scores.

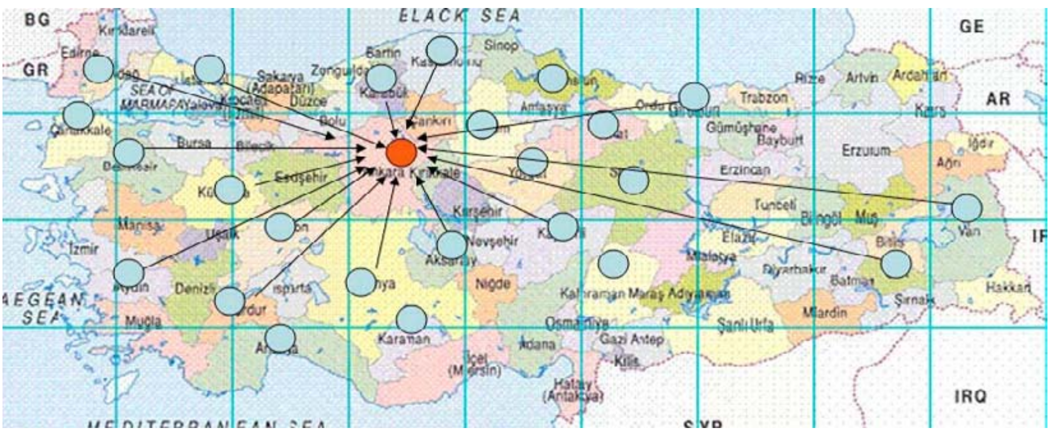


Figure 1. Current Situation: All Examinees Must Travel to the Capital.

---

[3] Ball, R. W. W. 1908. *A Short Account of the History of Mathematics*, 4th ed., www.maths.tcd.ie/pub/HistMath/People/Monge, transcribed by D.R. Wilkins.

[4] Schumpeter, J A 1951, 1986. *History of Economic Analysis*, Oxford University Press, (Schumpeter Elizabeth B ed.)

Not surprisingly, thousands of people converge to Ankara to take it (see Figure 1). After a three-hour exam, they return to their hometowns, which in most cases are hundreds/thousands of kilometers away. With only a change of dates, a similar exam is administered again in the spring and fall, this time exclusively for university personnel. As further discussed in our last section, a cautious guess of the yearly magnitude involved is one hundred thousand individuals. The people involved are essentially white collar workers from both the private and public sector: administrators, teachers, engineers, doctors and other health personnel etc.

The purpose of this centralized and illiberal examination policy, like so many others, is to accelerate modernization; in this particular case, by encouraging white collar workers to learn European languages. However, the means chosen – involving quasi-compulsory travel– is distinctly "unmodern" and discussed in numerous works on political history. For example, Atabaki and Zurcher[5]:

> ' . . The rights of the individual and his relationship with the state were of marginal rather than central significance in the eyes of Middle Eastern modernizers, and critical reason and individual autonomy seemed to have little relevance. The main reason for such discrepancy lay in the fact that the development of modern European societies was synchronized with and benefited from the age of European colonialism and imperialism and wars against the Orient. Modernization in the Middle East was a defensive reaction.'
> (Italics added).

In short, what strikes us today as "unmodern" and, as we will demonstrate, "wasteful", was and is part of an "authoritarian modernization" drive. For example, there is a practice that consists of requiring people about once every five years, to stay at home, on the last Sunday of November for the purpose of being physically counted. This way of conducting the population census -instead of using statistical sampling- is clearly wasteful. According to Kish[6], a number of countries in Africa follow a similar practice, on the grounds that it enhances social/national cohesion.

In this particular instance, at least at its inception, the centralized exam fulfilled a twofold purpose: reinforcement of the nation building effort by having people, many of them provincial decision-makers, go to Ankara periodically and thereby strengthening the view that everything of importance is done at the capital city and the prevention of fraud especially in areas where sentiments of local patriotism is strong. We think that a country like Turkey, which is beginning accession talks with the European Union, should have enough self-confidence to put aside the first rationale and to recognize the counter productive potential of the second. Lejour and de Mooij[7] provide an overall assessment of Turkey's accession to the EU. Interestingly, they conclude, "*if* Turkey would succeed in *reforming its domestic institutions* in response to the EU-membership, consumption per capita in Turkey could rise by an *additional* 9%" (*Italics added*). The two examples of waste (the exam and the census practice) we provide – there are others – support this view.

---

[5] Atabaki, T. and Zurcher, E. J. 2004). Men of Order: Authoritarian Modernization Under Ataturk and Reza *Shah*, IB Tauris, London and New York. pp 1-12.

[6] Kish, L. 2003. Selected Papers (Kalton G and Heeringa S, Eds.), John Wiley, New York.

[7] Lejour, A. M. and De Mooij, R. A. 2005. "Turkish Delight: Does Turkey's Accession to the EU Bring Economic Benefits," *Kyklos*, Vol. 58: pp. 87-120.

Moreover as discussed in the next and last sections, the choice of exam locations can easily exclude regions where incidence of malfeasance is deemed likely to occur. Thus, we hope accession talks with the EU will provide the necessary impetus to rationalize every aspect of public life, thereby eliminating sundry wasteful practices, including the one we are discussing. In addition, given the similarities of the modernization experience in numerous MENA countries, similar considerations may apply as well, even in the absence of a formal EU integration process.

3. Costs Involved

We now turn to the more narrow “economic” aspects of our problem. Let CPS stand for the “cost of the present situation” which can be represented as follows:

$$CPS = CT + CW + CA \tag{3.1}$$

In (3.1), CT represents total travel and lodging expenses, CW, total opportunity cost of wasted time, and CA, total cost of deaths and various injuries caused directly and indirectly by PPLE induced travel. These include traffic accidents resulting from extra travel as well as various untreated injuries due to the relevant health personnel being on the roads rather than on duty. This last point can be important for rural areas suffering from an acute shortage of health care personnel.

Admittedly, to estimate the second and third items requires some ingenuity. As matters stand, CPS is the sum of private costs born by all PPLE goers plus untreated injury costs. Abstracting from the latter and using the terminology of Section 4, note that CPS equals “Number of examgoers times ci,6”, where ci,6 stands for the per person cost of going from various provinces to Ankara. Thus, no collective entity bears that cost and that is one reason for its existence. Hence, CPS is a good example of the distinction between private versus social cost notions. For future reference let CPSactual stand for the particular current magnitude of CPS.

4. Centralized Solution

It may be argued that the Bergson-Samuelson-Stiglitz approach that ‘assumes a single social-welfare-maximizing principal’ (i.e.Hegel’s philosopher-king) who acts like ‘an omnipotent, omniscient and benevolent dictator’ might apply in our case, Dixit[8]. Thus, suppose the SSPC decides to play the role of Hegel’s philosopher-king. The obvious solution would be to administer the exam in each of the 81 provinces of Turkey. This would drive CPS down to zero. However, partly legitimate fears of fraud prevent this solution. This paper proposes to administer the exam in suitably chosen provinces where the exam’s integrity can be safeguarded. (In Turkey, each province carries the name of its capital city. Thus, we will use the terms city and province interchangeably.)

---

[8] Dixit, A. 1996. *The Making of Economic Policy*, Munich Lectures in Economics, MIT Press, Cambridge, MA. p68.

In our last section, we use the LP dual values to outline a method of choosing such locations. The rest of this section will argue that the problem at hand is a variant of the well-known transportation problem of linear programming.

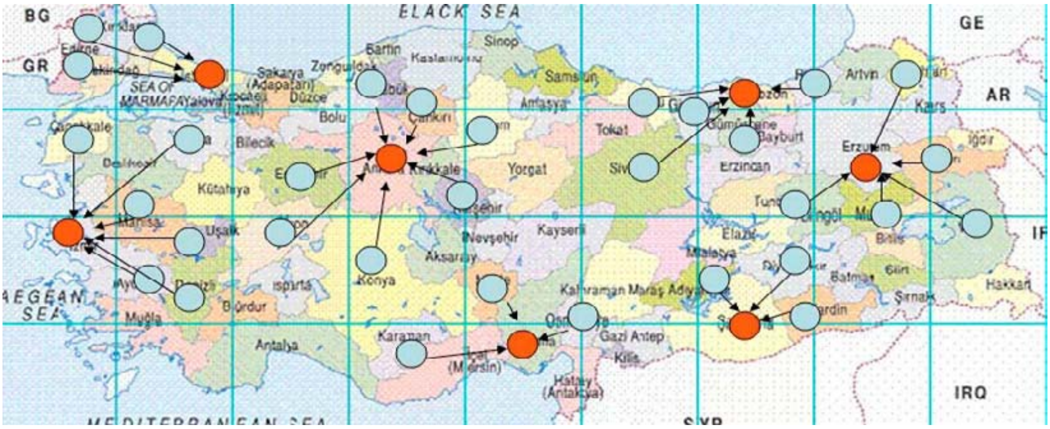


Figure 2. Illustrative Solution: More Than One Exam Location.

Let *i* index over the m provinces of Turkey (currently m=81), and *j* index over exam locations, $j = 1 \ldots n$. Figure 2 illustrates this situation for seven arbitrarily chosen exam sites. Let a_i be the number of exam takers from province *i,* and c_{ij} denote the per person cost of going from province *i* to exam location *j* (Consistent with Equation (3.1), each such c_{ij} is composed of three parts: travel and lodging, wasted time, and accident cost. Difficulties of measurement, especially for the last one, not withstanding, they clearly exist both at individual and social levels). Note that when a province also serves as an exam site the relevant cost is zero, i.e. $c_{ii} = 0$. Let x_{ij} represent the number of people from province *i* going to exam location *j*.

The optimization problem faced by the SSPC can be formulated as in 4.1 – 4.3.

$$\text{Min } Z_{cent} = \sum(i,j)\, c_{ij}\, x_{ij} \tag{4.1}$$

$$s.t. \sum_{j=1}^{n} x_{ij} \geq a_i\,,\ i = 1 \ldots m \tag{4.2}$$

$$x_{ij} \geq 0 \tag{4.3}$$

Verbally, (4.2) states the following: the total number of exam takers sent from each province to all exam sites is at least the number of exam takers from that province. The above problem can be easily solved using standard procedures. For further reference, let z^*_{cent} refers to the cost magnitude generated by the optimal program.

Now we state our assumption, which operationalizes the solution procedure.

Assumption 4.1. Let j^* be an exam site and $I(j^*)$ be the subset of the provinces such that $c^*_{ij} = \min_j \{c_{ij}\}$, then $\sum_{i \in I(j^*)} a_i \leq b_{j^*}$, where b_{j^*} is the number of slots at exam site j^*.

This assumption states that a center has enough examination slots for all examinees that would find it closest to them. Since it is known that bi ≥ ai for all i, and there is considerable freedom in choosing exam sites, we believe this assumption is quite reasonable.

Suppose the closest exam location to a certain province I is J. Also let *x*I1… x*Ij… x*In* refer to that portion of the optimal solution pertaining to province I. If Assumption 4.1 holds for all exam sites, then all magnitudes except *x*IJ* will be zero and *x*IJ* will be equal to *aI*. The intuitive reason should be obvious: the least costly way of handling the exam takers from province I is to send all of them to J. Since each exam site is capable of handling all examinees in its vicinity, the choices of examinees can be thought of as independent decisions. This is achieved via Dantzig-Wolfe decomposition theorem, Dantzig[9]. Total decomposition is possible since the binding constraint (in this case the capacity limitation) is vacuous. This is also seen by the dual problem given in Equations 5.1-5.3.

$$\text{Max } W_{\text{cent}} = \Sigma_i\, a_i y_i \tag{5.1}$$

$$s.t.\ y_i \le c_{ij}, j = 1\ldots n, i = 1\ldots m \tag{5.2}$$

$$y_i \ge 0 \tag{5.3}$$

Equation (5.2) states that the dual variables, yi , for a non-exam site *i* are set to the minimum travel cost among all exam sites *j*. For exam sites (*i=j*) however, yi , reduces to 0 since c_{ii} =0.

5. Kaldorian Improvement

Since the Pareto improvement notion is too restrictive (if one single person loses, change is excluded even when the whole rest of society benefits) Kaldor proposed an alternative criterion, Baumol[10]. The criterion used is, if the overall gains of those who prefer the status quo post exceed the overall losses of those who prefer the status quo ante, then go ahead and implement the change.

In our case, the change in question is to rationalize the PPLE, which means to administer the exam in many different locations instead of just one. Note that once the decision to switch is taken, the new situation can be managed either in a centralized or decentralized fashion. Thus, the terms centralized/decentralized are used in their economic design or mathematical programming sense, and not in their political science sense. In addition, a decentralized mode of management does not mean the “disappearance” of the center. Under such a mode, the center merely cedes the right to choose among exam locations to the examinees. The center still prepares the exam questions, keeps records and provides general supervision.

We note the multi-location exam can be administered using the Internet as the medium through which the questions are delivered and the answers are collected and graded. This will drastically reduce the number of individuals who physically handle the questions and

---

[9] Dantzig, G. B. 1963. *Linear Programming and Extensions*, Princeton University Press, New Jersey.
[10] Baumol, W. 1977. *Economic Theory and Operations Analysis*, 4th ed., Prentice Hall, New Jersey, pp 527-9.

answers. As a result, the likelihood of fraud will decrease considerably. This is the method used in several EU countries to administer various proficiency exams. Geoffrion and Krishnan[11] provide an overview of the complementarities between operations research and the Internet.

Proposition I. CPSactual > z^*cent > 0 for $n \neq m$

This states that so long as m (# of provinces) > n (# of exam sites), z**cent* > 0, and obviously when $m = n = 81$, $z^*cent = 0$.

Proof. CPSactual is the cost with only one exam site, namely Ankara. Also *z*cent* decreases with each additional exam site. Note that each such decreasing total cost has 3 components: CT +CW +CA. CT, namely the cost of travelling and lodging for examgoers, is a benefit item for hoteliers and bus operators; hence, the gain to examgoers is cancelled by the loss of hotel and bus operators. Note that this does not require any particular assumption, apart from the natural one of the analyst's neutrality between these two groups. On the other hand, the other two are deadweight losses since nobody in society gains from their presence. Thus, when each one of these latter cost items diminishes with each additional exam site, the reductions represent a net social gain. This establishes the validity of our conjecture regarding Kaldorian improvement.

As for the second part, if the exam takes place in all provinces, there will be neither congestion and wasted time, nor any transportation cost; hence all three components of *z*cent* will be 0.

6. Decentralized Solution

Let us now assume that the examinees in each province *i* act independently. We further assume that x^k_{ij} denotes the action of the kth individual from province *i.* It can be argued, when given the option of choosing among n exam locations, the kth individual will face the following general linear programming problem (there are *ai* of these problems, one for each examinee, in province *i*):

$$\text{Min } Z_i^k = \Sigma_{(i,j)}\ c_{ij}x_{ij}^k \qquad (6.1)$$

$$\text{s.t. } \sum_{j=1}^{n} x_{ij}^k = 1,\ i=1....m \qquad (6.2)$$

$$x_{ij}^k >= 0 \qquad (6.3)$$

Let $x^k{}_{*I1} \ldots x^k{}_{*IJ} \ldots . x^k{}_{*In}$ refer to the optimal values for an arbitrary person k in region I. Intuitively it should be clear that all values except $x^k{}_{*IJ}$ will be zero and $x^k{}_{*IJ}$ will be 1,

[11] Geoffrion, A. and Krishnan, R. 2001. "Prospects for Operations Research in the E-Business Era," *Interfaces*, Vol. 31, No. 2: pp. 6-36.

because the least costly way for an examgoer in province I to take the exam is in the nearest exam location J.

Proposition II. $\sum_{k=1}^{a_i} x^{*\,k}_{ij} = x^*_{ij} = a_i$ for all i and j.

Proof. Note that (6.2), when summed up over *k* gives (4.2). The groups in each region are non-overlapping. Hence, the equivalence follows. Since this holds for any solution satisfying (6.2) and (4.2), it also holds for the optimum one.

Each examinee (irrespective of province) will have to solve the linear problem (6.1)-(6.3). Let

$$Z^*_{decent} = \sum_{i=1}^{m} \sum_{k=1}^{a_i} Z^{*k}_{i}$$

be the optimum magnitude obtained when all individuals' optimum costs are totalled for all provinces.

Proposition III. $Z^*_{cent} = Z^*_{decent}$

Proof. Noting that the costs in (4.1) and (6.1) are the same and using Proposition 2, the result follows.

Finally, a mixed mode of management is also possible. It can be achieved as follows: the examinees have the right to select their exam site and actually do so. Thus each one of them solves her/his linear problem. Let the resulting outcome be called the decentralized actual solution. The exam goers are required to inform the center of their province of origin – and of nothing else! Based on this knowledge, that is the *ai's* and the costs (the *cij*'s) that it estimates, the center solves the LP problem of (4.1)-(4.3). Let the result be called the centralized notional solution. Propositions 2 and 3 show these two solutions would essentially coincide. Occasionally, the center may use this procedure to verify the overall integrity of the system.

7. Numerical Illustration

7.1. Parameters

This section calculates the savings –in the form of reduced costs– to be achieved by implementing the ideas outlined in our paper. Throughout this exercise, we strive for realism.

Firstly, we do not include the costs of administering the exam itself. The reason is straightforward: we are comparing scenarios where the number of exam locations is gradually increased from one (only Ankara) to eighty-one, namely in every province. However, in each case, the language examination takes place and thus the administrative costs are incurred.

Secondly, we assume the number of examtakers to be the same irrespective of the number of exam locations. In reality, we can expect that number to increase somewhat as candidates have more places to choose from. The situation bears some similarity to the "discouraged worker effect" phenomenon. This involves unemployed people not looking for work when they reckon their chances are slim. To that extent measured unemployment falls short of true unemployment. By a similar logic if the cost of traveling to Ankara is prohibitive for some would be examtakers, then the actual number of examtakers that we estimate and use in our scenarios has a downward bias. As a result, the benefit figures we come up with need to be revised upward.

Third, we take one hundred thousand individuals to be a realistic estimate for the number of examinees. We arrived at this figure via the following logic. According to SSPC's website, during May 2005, 40,000 individuals took the PPLE exam. The corresponding figures for the equivalent language examinations of university and medical personnel are 23,000 and 5,800 respectively. Since these are for the first half of that year, this sum (roughly) of 69,000 should be multiplied by two. Choosing to err on the side of caution, we ended up with 100,000 as our yearly estimate.

We allocated this total among our 81 provinces in accordance with the population share of the province in the 2000 census. For example, Adana whose population was 1.85 millions according to that census –the total being 67.8 million– has a share of 0.0273. Thus, under the status quo, 2730 people from Adana must travel to Ankara for exam purposes on a yearly basis.

Fourth, we take the yearly income of an examinee to grow at 3% per year starting with $8,600. The justifications for these figures are as follows: a) according to SSI (State Statistical Institute) the Turkish per capita, GDP is around $6,800. We posit the typical examinee to be an above average earner and come up with a round starting salary of $8,600. b) We believe a real 3% growth rate to be plausible, especially in view of our time horizon of 30 years. That is we assume each examinee to have a working life of 30 years. We take 5% as the time discount factor. These parameters yield roughly $188,000 as the present value of the typical person's future income stream. For lack of any better alternative, we use that figure as the "measure of a life's worth" or "cost of one's death". Consistent with these income figures we use an hourly wage of $ 4.5. Multiplying this number with the relevant travel duration yields an estimate for the cost of wasted time.

Finally, we use $1,000 as the cost of one non-fatal injury. Since we could not find an acceptable shortcut to estimate the number of lives saved due to "less/no travel" by medical personnel, we had to exclude them from our analysis.

We take the likelihood of fatality to be:

$$3941 \text{ persons} / 175{,}236{,}000{,}000 \text{ person-kms.} = 2.252 \text{ E}{-08}.$$

According to information provided on the Highways Directorate website, the numerator represents the number of fatalities during 2000, and the denominator measures total traffic flow during that same year. A similar analysis, using information from the same source, yields 6.622 E –07 as the likelihood of a non-fatal injury. Multiplying the first number with the distances involved and our "value of life" measure, we estimate the per person cost of fatalities. For non-fatal injuries, we repeat the procedure using the second number.

7.2. Choosing Exam Sites

As demonstrated by Proposition 1, it is possible to drive the costs, caused by the current "only Ankara" policy, down to zero. This requires holding the exam in every one of the 81 Turkish provinces. Here we outline a dual price based method of choosing the most beneficial exam sites sequentially: starting with the current situation (Ankara is the sole exam location) we solve the dual (5.1-5.3). The dual price for each city is the cost (consisting of cost of wasted time, cost of fatalities and cost of non-fatal injuries) which is directly proportional to its distance to the nearest exam site. (These shadow prices are listed as Col. 2 of Appendix 1). Hence, among the non-exam cities, the one with the highest product of "number of exam takers and shadow price" is the next one to be included in the set of exam sites. (See Col. 3 of Appendix 1). Recalling that the product in question represents the best lower bound we have for the amount of savings to be achieved by including the city among exam locations, it follows this is a reasonable approximation for identifying the next best candidate. Iterating in this manner, we add another city to the set of exam locations at each stage. The resulting total costs are depicted in Figure 3 as a function of the number of exam locations. As shown in that figure, starting from roughly $3.5 million, yearly waste decreases drastically with the first 15 cities. This cuts annual waste by around $3 million or about 86% of the total.

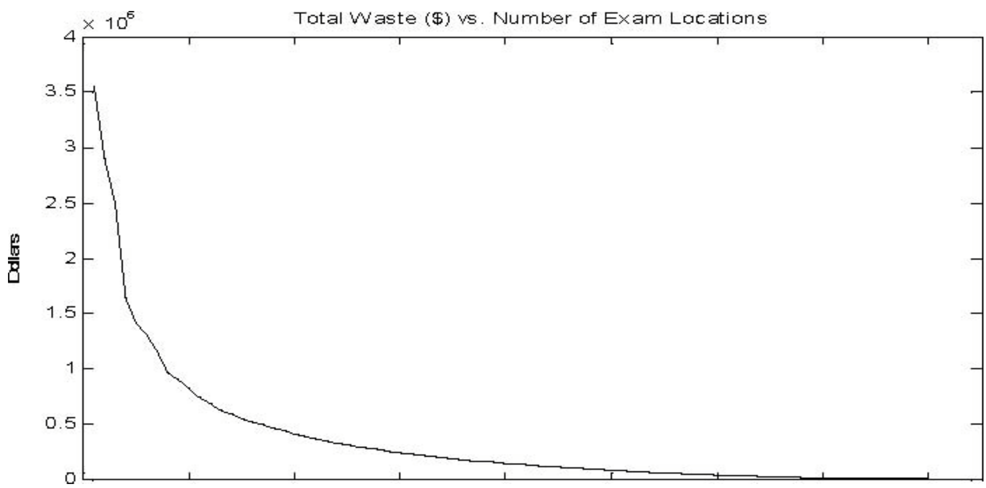


Figure 3. Total Waste as a Function of Exam Locations Number of cities.

With the exception of Diyarbakir, Sanliurfa and Van, these 15 cities providing the bulk of the gains happen to be situated in the most developed parts of the country, where individualism prevails and thus chances of fraud are negligible. For illustrative purposes, we give the first 7 cities and the resulting costs in Table 1. Appendix 1 contains the marginal (Col. 4) and total savings (Col. 5) involved for each of the 81 provinces as well as an index linking its license plate number –used as numerical identifier in Table 1– to its name.

A few remarks regarding the actual as distinct from the computational process of choosing exam sites are in order. As discussed previously, in the tradition bound parts of the country where local allegiances and tribal loyalties abound, fraud is possible.

For instance,during a presentation of this paper at Middle East Technical University, a faculty member who is a scion of a prominent Diyarbakır family openly stated that in his hometown “the exam’s integrity could not be guaranteed”. We hasten to add in 2000 –the year of the presentation– the possibilities of the Internet were not fully understood. Thus our colleague envisioned a theft prone process whereby printed material would be transported by minor functionaries back and forth between Ankara and Diyarbakır.

Table 1. Travel Related Waste as a Function of the Number of Exam Locations

Number of Exam Cities (license plate numbers are used as identifier)	Cost of Wasted Time (Mill $)	Cost of traffic accident related non-fatal injuries ($)	Cost of traffic accident related fatalities ($)	Total (Mill $)
1 (6)	3.312	34,123	2.089e5	3.556
2 (6,34)	2.702	27,834	1.704e5	2.900
3 (6,34,35)	2.330	23,998	1.469e5	2.501
4 (6,34,35,21)	1.529	15,746	96,387	1.641
5 (6,34,35,21,63)	1.316	13,553	82,961	1.412
6 (6,34,35,21,63,7)	1.217	12,536	76,741	1.306
7 (6,34,35,21,63,7,33)	1.081	11,130	68,131	1.160

Therefore, the choice of exam sites has to be the result of balancing the savings to be achieved against that possibility. Thus, it has to be a political decision, but hopefully not a difficult one since the most beneficial potential exam sites are also the ones with the least likelihood of malfeasance. Moreover, if the Internet is chosen as the medium through which the exam is administered, the possibility of fraud and theft will be minimized for the reasons given previously. Thus, we can easily visualize a gradual process of increasing the number of exam sites by stages. One can start with provinces where the current Internet infrastructure is capable of handling the extra traffic. In the developed parts of the country where the bulk of the savings occur, the existing infrastructure is robust and reliable. Over time, the system can be extended to the whole country.

As for actual implementation, we believe the ideas outlined in the special issue of Interfaces (March April 2001), devoted to web applications of operational research, are relevant. In particular, Keskinocak and Sridhar[12] discuss the quantitative aspects of using the Internet for logistics management; whereas Cohen et al[13] give numerous real life examples of web based information processing and decision making support systems.

8. Conclusions

We have demonstrated that by administering the language exam in many locations instead of just one, the current policy, nonnegligible savings will be achieved. We have also indicated how the proposed multicenter policy can actually be implemented. Finally, we presented some numerical estimates for the gains involved.

Finally yet importantly, we feel the necessity to stress the following point. Some policy makers may consider the yearly savings involved (around $ 3.5 million) as too small to warrant interest. We beg to disagree. The promotion and nurturing of a culture of

[12] Keskinocak, P. and Sridhar, T. 2001. “Quantitative Analysis for Internet Enabled Supply Chains,” *Interfaces*, Vol. 31, No. 2: pp. 70-89.

[13] Cohen, M. D., Kelly, and Medaglia, A 2001. “Decision Support with Web-Enabled Software,” *Interfaces*, Vol. 31, No. 2: pp.109-129.

optimization is paramount for accelerating modernization and growth. Waste avoidance by governmental authorities, whatever the monetary magnitude involved, is an integral part of that process. In addition, the overall benefits of such a cultural transformation exceed the purely pecuniary ones; in this particular instance even a belated recognition of the examinees' value of time will impact their morale positively.

Appendix 1

Rank	Shadow Price($)	Shadow Price*ai ($)	Marginal Saving ($)	Total Saving ($)	License Plate Number	Province Name
1					6	Ankara
2	31.33	462,880	655,200	655,200	34	Istanbul
3	38.99	193,800	399,800	1,055,000	35	Izmir
4	63.55	127,740	859,900	1,914,900	21	Diyarbakir
5	56.58	120,460	228,500	2,143,400	63	Sanliurfa
6	32.36	82,069	105,900	2,249,300	7	Antalya
7	28.57	69,588	146,600	2,395,900	33	Icel
8	43.06	61,916	186,540	2,582,440	61	Trabzon
9	17.80	57,555	68,190	2,650,630	42	Konya
10	16.77	52,549	80,610	2,731,240	16	Bursa
11	23.88	42,568	79,300	2,810,540	55	Samsun
12	21.80	34,102	54,360	2,864,900	38	Kayseri
13	26.15	33,840	63,340	2,928,240	65	Van
14	17.87	33,044	38,740	2,966,980	31	Hatay
15	20.70	28,608	42,320	3,009,300	25	Erzurum
16	17.39	21,892	24,910	3,034,210	44	Malatya
17	15.94	19,988	27,660	3,061,870	20	Denizli
18	15.94	19,462	26,750	3,088,620	60	Tokat
19	15.11	18,103	24,760	3,113,380	3	Afyon
20	9.52	18,054	28,020	3,141,400	27	Gaziantep
21	18.49	16,791	23,470	3,164,870	67	Zonguldak
22	10.35	16,426	19,310	3,184,180	10	Baliskesir
23	11.39	14,903	19,870	3,204,050	52	Ordu
24	7.66	13,626	22,730	3,226,780	41	Kocaeli
25	13.39	13,480	18,240	3,245,020	66	Yozgat
26	4.76	12,988	14,630	3,259,650	1	Adana
27	8.63	12,093	15,950	3,275,600	9	Aydin
28	10.21	10,631	12,180	3,287,780	26	Eskisehir
29	15.66	10,479	15,210	3,302,990	49	Mus
30	19.18	9,994	12,300	3,315,290	73	Sirnak
31	12.77	9,957	11,320	3,326,610	4	Agri
32	14.42	9,893	10,060	3,336,670	17	Canakkale
33	16.70	9,251	9,800	3,346,470	37	Kastamonu
34	15.39	9,140	14,010	3,360,480	22	Edirne
35	9.04	8,316	8,320	3,368,800	59	Tekirdag
36	7.38	8,225	8,220	3,377,020	58	Sivas

Appendix 1. Continued

Rank	Shadow Price($)	Shadow Price*ai ($)	Marginal Saving ($)	Total Saving ($)	License Plate Number	Province Name
37	5.52	8,159	8,160	3,385,180	46	Kahramanmaras
38	10.14	7,689	9,570	3,394,750	32	Isparta
39	6.90	7,280	7,280	3,402,030	48	Mugla
40	7.52	6,920	6,920	3,408,950	2	Adiyaman
41	6.62	6,889	6,890	3,415,840	47	Mardin
42	14.21	6,809	9,660	3,425,500	36	Kars
43	7.25	6,383	7,200	3,432,700	19	Corum
44	13.25	6,187	7,180	3,439,880	24	Erzincan
45	9.18	6,185	6,370	3,446,250	72	Batman
46	10.07	5,883	6,970	3,453,220	68	Aksaray
47	6.97	5,854	5,852	3,459,072	23	Elazig
48	5.38	5,215	5,215	3,464,287	43	Kutahya
49	14.01	4,889	4,889	3,469,176	30	Hakkari
50	2.48	4,618	4,618	3,473,794	45	Manisa
51	8.35	4,283	4,283	3,478,077	51	Nigde
52	10.42	4,157	7,000	3,485,077	14	Bolu
53	6.00	4,064	4,064	3,489,141	80	Osmaniye
54	14.28	4,042	4,042	3,493,183	8	Artvin
55	11.59	3,860	3,860	3,497,043	57	Sinop
56	7.87	3,736	3,736	3,500,779	64	Usak
57	6.35	3,422	3,422	3,504,201	5	Amasya
58	5.87	3,361	3,361	3,507,562	13	Bitlis
59	9.25	3,070	3,163	3,510,725	78	Karabuk
60	7.66	3,056	3,056	3,513,781	18	Cankiri
61	5.24	2,968	2,969	3,516,750	71	Kirikkale
62	7.94	2,968	2,967	3,519,717	12	Bingol
63	2.55	2,847	2,847	3,522,564	54	Sakarya
64	5.24	2,832	2,832	3,525,396	53	Rize
65	7.87	2,824	2,824	3,528,220	70	Karaman
66	7.52	2,805	2,805	3,531,025	40	Kirsehir
67	6.14	2,389	2,389	3,533,414	56	Siirt
68	9.59	2,388	2,388	3,535,802	76	Igdir
69	5.18	2,365	2,365	3,538,167	50	Nevsehir
70	3.04	2,347	2,347	3,540,514	28	Giresun
71	4.28	2,071	2,071	3,542,585	39	Kirklareli
72	6.90	1,904	2,381	3,544,966	29	Gumushane
73	5.52	1,584	1,585	3,546,551	11	Bilecik
74	5.80	1,577	1,577	3,548,127	74	Bartin
75	3.11	1,438	1,438	3,549,565	81	Duzce
76	3.45	1,308	1,308	3,550,872	15	Burdur
77	8.97	1,238	1,238	3,552,110	62	Tunceli
78	6.14	1,210	1,210	3,553,320	75	Ardahan
79	4.49	1,117	1,117	3,554,437	77	Yalova
80	5.24	755	755	3,555,192	69	Bayburt
81	2.42	408	408	3,555,600	79	Kilis

References

Atabaki, T. and Zurcher, E. J. 2004). *Men of Order: Authoritarian Modernization Under Ataturk and Reza Shah*, IB Tauris, London and New York. pp 1-12.

Ball, R. W. W. 1908. *A Short Account of the History of Mathematics*, 4th ed., www.maths.tcd.ie/pub/HistMath/People/Monge, transcribed by D.R. Wilkins.

Baumol, W. 1977. *EconomicTheory and Operations Analysis*, 4th ed., Prentice Hall, New Jersey, pp 527-9.

Cohen, M. D., Kelly, C. and Medaglia, A 2001. "Decision Support with Web-Enabled Software," *Interfaces*, Vol. 31, No. 2: pp.109-129.

Dantzig, G. B. 1963. *Linear Programming and Extensions*, Princeton University Press, New Jersey.

Denardo, E. 2002. *The Science of Decision-Making*, John Wiley, New York, p107.

Dixit, A. 1996. *The Making of Economic Policy, Munich Lectures in Economics*, MIT Press, Cambridge, MA. p68.

Geoffrion, A. and Krishnan, R. 2001. "Prospects for Operations Research in the E-Business Era," *Interfaces,* Vol. 31, No. 2: pp. 6-36.

Keskinocak, P. and Sridhar, T. 2001. "Quantitative Analysis for Internet Enabled Supply Chains," *Interfaces*, Vol. 31, No. 2: pp. 70-89.

Kish, L. 2003. Selected Papers (Kalton G and Heeringa S, Eds.), John Wiley, New York.

Lejour, A. M. and De Mooij, R. A. 2005. "Turkish Delight: Does Turkey's Accession to the EU Bring Economic Benefits," *Kyklos,* Vol. 58: pp. 87-120.

Schumpeter, J A 1951, 1986. *History of Economic Analysis*, Oxford University Press, (Schumpeter E B ed.)

Thompson, G. and Thore, S. 1992. *Computational Economics*, Scientific Press, San Francisco, pp 6-7.

INDEX

A

B

C

D

E

F

G

H

I

J

K

L

M

Q

R

S

T

U

V

W

Y