Essays on Forecasting Financial and Economic Time Series

Mansur, Mohaimen

Essays on Forecasting Financial and Economic Time Series

by

Mohaimen Mansur

A thesis submitted for the Degree of

Doctor of Philosophy (Ph.D.) in Economics

School of Economics and Finance

Queen Mary

University of London

March 2014

AUTHOR'S DECLARATION

I wish to declare that no part of this doctoral thesis, titled as "Essays on Forecasting Financial and Economic Time Series", and submitted to the University of London in pursuance of the degree of Doctor of Philosophy (Ph.D.) in Economics, has been presented to any University for any degree.

Signed

Mohaimen Mansur

EXTENDED ABSTRACT

This thesis comprises three main chapters focusing on a number of issues related to forecasting economic and financial time series.

Chapter 2 contains a detailed empirical study comparing forecast performance of a number of popular term structure models in predicting the UK yield curve. Several questions are addressed and investigated, such as whether macroeconomic information helps in forecasting yields and whether predicting performance of models change over time. We find evidence of significant time-variation in forecast accuracy of competing models, particularly during the recent financial crisis period.

Chapter 3 explores density forecasts of the yield curve which, unlike the point forecasts, provide a full account of possible uncertainties surrounding the forecasts. We contribute by evaluating predictive performance of the recently developed stochastic-volatility arbitrage-free Nelson-Siegel models of Christensen et al. (2010). The one-month-ahead predictive densities of the models appear to be inferior compared to those of their constant-volatility counterparts. The advantage of modelling time-varying volatilities becomes evident only when forecasting interest rates at longer horizons.

Chapter 3 deals with a more general problem of forecasting time series under structural change and long memory noise. Presence of long memory in

the data is often easily confused with structural change. Wrongly accounting for one when the other is present may lead to serious forecast failure. In our search for a forecast method that can perform reliably in presence of both features we extend the recent work of Giraitis et al. (2013). A forecast strategy with data-dependent discounting is adopted and typical robust-to-structural-change methods such as rolling window regression, forecast averaging and exponentially weighted moving average methods are exploited. We provide detailed theoretical analyses of forecast optimality by considering certain types of structural changes and various degrees of long range dependence in noise. An extensive Monte Carlo study and empirical application to many UK time series ensure usefulness of adaptive forecast methods.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisors, Professor Andrea Carriero and Professor Liudas Giraitis, for their prudent guidance throughout my doctoral studies. Special thanks are due to Professor Liudas Giraitis for providing detailed technical advice and to Professor George Kapetanios for offering valuable suggestions on computational aspects. I would also like to thank my parents and friends for believing in me and encouraging me in pursuing this scholarly endeavour. Last but not least, I am grateful to my beloved wife Seheli who had been a constant source of inspiration and support during many difficult times. Without her by my side this dissertation would not have been completed in time.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Time series forecasting is an integral part of economics and finance. Reliable predictions help making informed decisions and consequently, forecasting has found important roles in a wide variety of economic and financial activities which include monetary and fiscal policy designing, business planning, financial asset and risk management, among others. Ability to produce accurate forecasts is valued as one of the principal attributes of a good dynamic model, but it is hardly a guaranteed property. It is not unlikely that sophisticated models which fit historical data extremely well often fail to outperform naive models in terms of out-of-sample forecast performance. Several issues have been held responsible in the literature (see, e.g., Egorov et al. (2006)). A prominent one is the so called 'data snooping' which says that parametrically rich models often overfit past data without capturing the true data generating process. Parameter proliferation can again cause substantial estimation uncertainty which deteriorates forecasts. Presence of structural change in time series data can inflict parameter instability and therefore, have pernicious effect on out-of-sample forecasting performance of a model which previously exhibited good in-sample fit. Therefore, it may not be ideal to judge a time series model on the basis of its in-sample performance alone. Careful assessment of out-of-sample forecasts can often insure against data snooping problem and help select the best data-explaining model.

The field of economic forecasting is vast and expanding rapidly. The objective of this thesis is to shed light only on a few very specific topics in this respect. The focus is two-fold. The first part explicitly deals with certain aspects of forecasting the term structure of interest rates and takes empirical approaches to investigate them. The second part looks at a wider problem of forecasting time series under structural change and long memory persistence and combines theoretical justifications with empirical evidence.

We begin the first part by conducting a comprehensive empirical study which focuses on forecasting the term structure of UK interest rates. There are several papers which attempt to model the UK yield curve and capture its historical variable dynamics (e.g., Lildholdt et al. (2007), Bianchi et al. (2009)). However, forecasting the future movements of the UK yield curve has received little attention in the literature. The contribution of this chapter is explicitly to this end and it exploits two very popular classes of yield curve models in predicting the future course of the bond yields, namely the Nelson-Siegel models and no-arbitrage affine models. The chapter attempts to answer several interesting questions related to forecasting bond yields. First, it investigates whether incorporating macroeconomic information helps in improving predictive performance of models that, otherwise, would rely only on cross-sectional information of the yield curve. There is historical evidence of advantageous effects of adding economic fundamentals on interest rate prediction (see Ang and Piazzesi (2003), Diebold et al. (2006), Moench (2008), among others). Most of the studies, however, predominantly use the US yield curves data and it is our interest to investigate whether documented findings hold also for the UK. Second, the chapter examines whether predictive performance of models change over time. Following the onset of the recent financial crisis in late 2007 the UK yield curve has undergone dramatic changes. The short rates plummeted to their historical lows and the spread between the short and long rates widened substantially. It is, therefore, unlikely that a single term structure model will consistently produce the best forecasts of the UK yield curve. Following recent literature we argue that a global measure summarising model performance over a long forecast period may be misleading when comparing competing models. We evaluate quality of forecasts across sub-periods and apply the fluctuation test of Giacomini and Rossi (2010) to detect significant time-variation in predictability of different models.

While the focus of the first chapter in part one is purely point forecasts of the conditional mean of bond yields, a second chapter contributes by exploring density forecasts of the yield curve. The point forecasts are often considered to be of limited value in the sense that they do not account for the uncertainty associated with future forecasts. A more desirable course of action is constructing density forecasts which predict the entire distribution of future yields and acknowledge possible uncertainties. Density forecasts allow computation of higher order moments such as variance, skewness and kurtosis, knowledge of which play important role in risk management and derivative pricing. Unfortunately, empirical literature on probability forecast of the yield curve is narrow. To our knowledge the only contribution is Egorov et al. (2006) who evaluate performance of widely popular no-arbitrage affine term structure models (ATSMs) in forecasting the conditional predictive density of bond yields. Provided that these authors document unsatisfactory density forecasts from ATSMs, a better alternative forecasting model is to be sought for. We contribute by computing and evaluating out-of-sample forecasts of a recently developed class of models namely the stochastic-volatility arbitrage-free Nelson-Siegel models of Christensen et al. (2010). While the models are parallel to the ATSM in terms of theoretical consistency they enjoy several advantages over the ATSM. They are more parsimonious, easier to estimate and offer meaningful interpretations for latent state variables. These models have been found to fit the conditional mean and variance of US yields reasonably well (Christensen et al. (2010)), but, to our knowledge, they have not been evaluated for forecast accuracy. We take this opportunity to compare several variants of the models in terms of their predictive performance. The questions which are of particular interest are whether enforcing no-arbitrage restriction and/or modelling time-varying volatility improve predictive performance of Nelson-Siegel models. We resort to various metrics used in recent literature such as probability integral transforms, coverage rates, log predictive density scores in order to extensively assess the quality of calibrated density forecasts.

In the second part we divert our concentration from the specific problem of forecasting interest rates to a more general problem of forecasting time series under structural changes. In economics and finance literature structural change is considered as a common phenomenon (see Stock and Watson (1996)) and often regarded as the principal cause of forecast failures (see Hendry (2000)). This chapter addresses the broad problem of making reliable real-time forecasts of a time series in the presence of ongoing structural changes and focuses on specific cases where the scenario is further complicated by noises which are contaminated with long range dependence.

A major source of motivation behind conducting the research is an ongoing argument about possible spurious relationship between long memory and structural change (see, e.g., Diebold and Inoue (2001), Gourieroux and Jasiak (2001), Granger and Hyung (2004)). It is being increasingly evident from econometrics literature that the presence of long memory in the data can be easily confused with structural change. Wrongly accounting for one when the other is present or acknowledging only one when both are present may lead to serious forecast failure. Given that it is often difficult to distinguish between the two, it is desirable to establish forecast methods that are robust to structural change and also appropriately account for long memory persistence. Our contribution is specifically to this end.

We approach the problem by exploiting an adaptive forecast strategy that is recently advocated in Giraitis et al. (2013). This strategy is attractive in many ways. First, it computes forecasts of a time series in a simplistic framework of weighted average of past data and avoids complicated modelling of structural breaks and consequent estimation of associated parameters. Second, it utilises forecast methods which essentially function by downweighting historical data. Such a class of methods includes rolling window forecasts, forecast averaging across estimation windows, exponentially weighted moving averages, among others and are largely considered to be robust to historical and recent structural changes. Third, unlike most models which are designed to tackle specific types of structural changes, predominantly breaks, the adaptive method makes minimal structural assumptions and is applicable to various forms of structural changes ranging from breaks to smooth and cyclical trends.

By considering short-memory noise processes Giraitis et al. (2013) develop an in-sample cross-validation based technique to tune the downweighting parameter and theoretically prove that such data-selected discount rate minimises mean squared forecast error (MSFE) asymptotically. They confirm empirical usefulness of the methods by conducting a simulation study and applying it to a large number of US time series. We extend their work by introducing a more complex yet realistic forecasting environment where structural change in a dynamic model is accompanied by noises with long memory persistence. We consider a number of specific types of structural changes and provide detailed theoretical justification for asymptotic optimality of forecasts based on the proposed methods. An extensive Monte Carlo study follows to illustrate small sample performance of data-tuned robust methods. Finally, we take the methods to real data and examine their effectiveness by forecasting a number of UK macroeconomic and financial time series.

The thesis is organised in four chapters. Chapter 1 addresses the research questions that are investigated in this dissertation and provides a brief summary of each chapter.

Chapter 2 reviews in details two well-known class of term structure models – the no-arbitrage affine models and the Nelson-Siegel models, describes how forecasts of future yields can be generated using the models and how macroeconomic information can be incorporated. Using monthly zero-coupon bond yields of the UK, the predictive performance of the competing models, both with and without macroeconomic information, are then compared. Detailed robustness checks are conducted in order to check for time variation in models' quality of forecasts.

Chapter 3 presents a survey of variants of Nelson-Siegel type yield curve models – ranging from standard constant volatility atheoretical models to arbitrage-free stochastic volatility models. We discuss how these models can be exploited to produce density forecasts of interest rates. The models are then evaluated in terms of their calibrated predictive density when applied in forecasting the US yield curve. A number of criteria have been used to scrutinise quality of density forecasts and detailed results of out-of-sample forecast exercise are reported.

Chapter 4 concentrates on a general problem of forecasting time series in presence of structural change and long range persistence. A simple adaptive forecast strategy which predicts by downweighting older data and relies on data-dependent selection of a tuning parameter is revised and discussed. Theoretical justifications for asymptotic optimality of such forecasts are presented by considering processes with long memory noises. An extensive Monte Carlo study confirms good small sample performance of data-tuned discounting. A detailed empirical forecast exercise using many UK time series show that the methods are also practically useful.

Chapter 2

Forecasting the Term Structure of UK Interest Rates

2.1 Introduction

The yield curve, often known as the term structure of interest rates, is one of the most widely studied topics in both finance and economics. It explicitly looks at the relationship between different maturity periods of bonds (commonly government bonds which are free of default-risk) and interest rates (yields) earned on them. Knowledge of the yield curve can be imperative for researchers, policy makers and market participants for various reasons. Interest rate models are the building blocks of fundamental financial activities such as pricing assets, managing portfolios and hedging financial risks. The yield curve has important implications in economics too. It has been documented in the literature as one of the most reliable and consistent predictors of recession, inflation and output growth (see Estrella and Mishkin (1998) and Stock and Watson (2003)). Its role as a successful economic indicator can be effectively exploited by central banks in formulating monetary policy decisions. Debt policies can also be benefited from the term structure of interest rates as it can help in deciding maturity lengths of newly issued bonds. It is not surprising, therefore, that the yield curve has attracted a lot of interest from researchers around the world and enormous amount of efforts has been devoted to modelling and forecasting the term structure of interest rates. Despite significant progress in the modelling aspect the task of accurately predicting the yield curve, however, remains to be a challenging endeavour even today. Many sophisticated yield curve models which show impressive fit to historical data fail to produce forecasts better than mere 'no-change' forecasts.

Academic literature has evolved mainly around two classes of term structure models, namely the affine term structure models and the Nelson-Siegel models. Early development comes through the hands of affine models (Vasicek (1977), Cox et al. (1985) and Duffie and Kan (1996)). These models express yields as linear functions of a small number of state variables which are extracted from the cross-section of a spectrum of yields with different maturities. An attractive feature of affine models is that they impose, by construction, cross-equation restrictions to rule out risk-free arbitrage opportunities. Given that most bond markets are well-organised and highly liquid, any arbitrage opportunity is expected to be traded away by the market participants. By ensuring absence of arbitrage affine models, thus, comply with a desirable theoretical requirement. De Jong (2000) and Dai and Singleton (2000) confirm the models' good ability to fit the yield curve in-sample, but Duffee (2002) reports rather disappointing forecasting performance out-of-sample. An excellent revision of affine term structure models is provided in Piazzesi (2010).

A second class of models proposed by Nelson and Siegel (1987) takes a more statistical approach to approximating the yield curve in a factor model framework by imposing a particular exponential structure on the factor loadings. They show that despite having a concrete theoretical foundation their parsimonious three-factor model is able to replicate various shapes of the yield curve such as monotonic, concave, S-shaped etc. Several more flexible extensions of the model are later proposed by including additional factors (see Svensson (1994), Björk and Christensen (1999)). Diebold and Li (2006) introduce dynamics in the original static versions of the model. By using monthly U.S. government bond yields they show that the dynamic Nelson-Siegel model not only fits the data very well in-sample but also produces accurate forecasts, particularly over long forecast horizons. De Pooter (2007) compares predictive performance of several multi-factor variants of the Nelson-Siegel model and finds that a more flexible four-factor model with two distinct slope factors outperforms its two- and three-factor counterparts in terms of better in-sample fit and out-of-sample forecasting. The Nelson-Siegel model has long been criticised for not being arbitrage-free by design until Christensen et al. (2011) develop a no-arbitrage version of the model. They demonstrate that imposing freedom of arbitrage improves forecasts of dynamic Nelson-Siegel models when predicting the U.S. yield curve.

The early and most basic versions of the term structure models have been developed purely from financial interests and have long been studied without considering any possible relationships of bond yields with wider economy. The models typically employ only a small set of unobserved factors, often linked to physical attributes of the yield curve such as its level, slope and curvature, to explain movements along the term structure of interest rates and ignore potential economic forces that could drive such movements. According to the expectations theory of the term structure of interest rates long rates are expected values of the risk-adjusted future short rates. Short rates themselves are popular policy instruments and are usually controlled by central banks and other monetary authorities in response to change in macroeconomic conditions.[1] It is, therefore, reasonable to infer that the entire yield curve responds to macroeconomic shocks. Nonetheless, it is not more than a decade ago that researchers have begun to exploit macroeconomic information in modelling and forecasting the yield curve. Ang and Piazzesi (2003) incorporate two macroeconomic factors, one related to inflation and the other to real activity, in a standard no-arbitrage affine framework and constrain macro-yields relationship to be unidirectional by allowing yields to depend on macro factors and not vice versa. They find that macroeconomic information not only explains a healthy proportion of movements in the yield curve, particularly in the short and medium rates, but also significantly improves historically documented poor forecasting performance of these classes of models. However, the forecast gains over the random walk is small and evaluated only at one-month horizon. Diebold et al. (2006) analyse a more generic bidirectional interactions between the yield curve and economy by augmenting three macroeconomic variables (representing policy instrument, inflation and real activity) with the latent Nelson-Siegel factors. They report a number of important findings: there are strong correlations between inflation and the level factor and also between real activity and the slope factor and there is strong evidence of causality from macro variables towards yield curve movements but a weaker evidence of a reverse influence. Their focus, however, is entirely on modelling the yield curve dynamics in-sample rather than out-of-sample forecasting. A group of researchers adopts more structural approaches to explaining yield curve movements where they combine macroeconomic models to no-arbitrage yield curve models. Important contributions come from Wu (2006), Hördahl et al. (2006) and Rudebusch and Wu (2008). Amongst these studies only Hördahl et al. (2006) test their model's predictive ability with

---

[1] The analogy dates back to Taylor (1993) who proposes a monetary policy rule to determine how much the central banks should change the nominal interest rate in response to changes in inflation and output.

a pseudo out-of-sample forecast exercise and report its superior performance when forecasting German zero-coupon bond yields.

Recently another strand of literature has emerged motivated by an argument that central banks today set policy rate depending on a large set of macroeconomic variables rather than only a few key indicators (Bernake and Boivin (2003)). This would mean that a shock in one or more of the variables in the set would affect the movements in the short end of the yield curve and such effects would then feed into the long end. Several papers explore modelling and forecasting the U.S. yield curve in a data-rich environment by exploiting the so called dynamic factor models and extracting a number of factors from a large panel of macroeconomic time series. Favero et al. (2012), Moench (2008) and De Pooter et al.(2010) using either Nelson-Siegel or no-arbitrage affine models or using both models show that considering a broader set of economic information is beneficial for forecasting yields.

The principal objective of this thesis chapter is providing an empirical comparison of state-of-the-art models of term structure of interest rates based on their performance in forecasting the UK yield curve. Although the literature on yield curve modelling is rich with contributions of researchers around the world, relatively few papers explore dynamic behaviour of the UK yield curve with major contributions coming from the central bank researchers. Lildholdt et al. (2007), using an affine macro-factor model, investigate historical fluctuations in the UK yield curve and attribute the movements at the short-end of the yield curve to change in monetary policy and that at the long-end to changes in inflation target. Kaminska (2012) augments a structural framework with a standard no-arbitrage affine term structure model to argue that risk premia is driven by structural macroeconomic shocks rather than by non-structural risk components. Bianchi et al. (2009) adopted time-varying parameters and stochastic volatilities in a standard Nelson-Siegel model with an aim to modelling interaction between interest rates and macroeconomy and find that the bilateral relationship changes substantially across different policy regimes. While all these studies confirm material changes in dynamic properties of the UK yield curve and their proposed models' reasonably good fit to historical data, they remain silent about the models' ability to forecast future yields. This essay contributes by exploring dynamic behaviour of the UK yield curve from an out-of-sample forecasting perspective.

In its build up the chapter addresses several interesting issues related to yield forecasting. A crucial one is possible time-variation in predictive performance of term structure models. The time path of the UK yield curve has undergone significant changes. For example, following the adoption of inflation targeting policy in 1992 there is a large decline in its volatility. A latest example is the recent financial crisis period of 2008-2010 which sees short-term interest rates to plummet to their historical low and spread between short and long rates to become extremely wide. It is, therefore, likely that a previously well-performing model may forecast poorly during this changed environment while a previously ill-performing model generates accurate forecasts. We shed light on this by evaluating effects of instability on quality of model forecasts. Following recent literature, we also assess the role of macroeconomic information in predicting yields. Like the bond yields many macroeconomic indicators of the UK went through dramatic change during the 2008-2010 crisis, signalling a recession. We contribute by investigating whether variations in macroeconomic fundamentals can help in forecasting recent changes in yield dynamics. To summarise, the objective of the paper is three-fold: first and foremost, compare forecasting performance of models widely used in literature in predicting UK nominal zero-coupon bond yields. Second, determine whether the forecasting performance of models evolve over time. And finally, assess potential forecasting benefit, if any, from incorporating macroeconomic variables in standard model frameworks.

The findings of the paper can be summerised as follows: i) forecasting ability of models vary significantly over time, ii) models without macroeconomic information forecast well during periods when interest rates are relatively stable, iii) an affine model augmented with macro-factors consistently produces accurate forecasts during the recent crisis period of near-zero short rates and wide spread and iv) choice of expanding-window and rolling-window forecasting schemes can affect forecasting performance of certain models.

The chapter is organised as follows. Section 2.2 provides a comprehensive description of the term structure models used for forecasting. Besides presenting the econometric frameworks we explain how the models accommodate macroeconomic information and how they are estimated. Section 2.3 introduces and summarises empirical data containing both yields and macroeconomic variables for the UK. Results of in-sample fit of the models are presented in section 2.4. Section 2.5 discusses the forecast methodology and evaluation of forecasts generated by competing models. Section 2.6 concludes.

2.2 Term Structure Models

We compare predictive performance of two classes of multi-factor term structure models that are the most popular among academicians as well as practitioners. The first class, the Nelson-Siegel model, functions by imposing a specific exponential structure on the loadings of the factors and has been developed with an aim to capture various shapes of a typical yield curve. The other class, the affine term structure model, is more founded on financial theory and builds on cross-equation no-arbitrage restrictions. Variants of both models where some macroeconomic variables are entertained as yield-explaining factors are also considered. We also include simple unrestricted linear dynamics such as AR(1) on yield levels and random walk to serve as benchmarks. In what follows we discuss in details the frameworks of the models, how macroeconomic information is incorporated and how the models are estimated.

2.2.1 The Dynamic Nelson-Siegel Model

The Yields-Only Model

Nelson and Siegel (1987) has proposed modelling the yield curve in an exponential components framework using a mathematical approximating function (a Laguerre polynomial) and showed that their parametrically parsimonious specification can provide enough smoothness and flexibility to capture varieties of yield curve shapes. If y_t^m denotes the continuously compounded yield-to-maturity of an m-period bond at time t, the functional form linking the yield curve to maturities can be written as

$$y(m) = \beta_1 + \beta_2 \frac{1 - e^{-\lambda m}}{\lambda m} + \beta_3 \left(\frac{1 - e^{-\lambda m}}{\lambda m} - e^{-\lambda m} \right), \tag{2.2.1}$$

where β_1, β_2, β_3 and λ are the parameters of the model. λ determines the rate of exponential decay of the loadings on $\beta_2$and β_3. Although the Nelson-Siegel model does not inherently enforce no-arbitrage assumption it certainly complies with several desirable properties of zero-coupon yields. For example, as maturity m tends to zero, the yield-to-maturity reduces to instantaneous short rate, r and when m increases indefinitely, the yield-to-maturity converges to β_1, a constant.

The original Nelson-Siegel model is static in definition and is meant to fit the cross section of yields of different maturities at a particular point in time. Diebold and Li (2006) extend it to a dynamic factor model:

$$y_t(m) = \beta_{1t} + \beta_{2t} \frac{1 - e^{-\lambda m}}{\lambda m} + \beta_{3t} \left(\frac{1 - e^{-\lambda m}}{\lambda m} - e^{-\lambda m} \right), \tag{2.2.2}$$

where β_{1t}, β_{2t} and β_{3t} are now time-varying latent factors with associated loadings of 1, $\frac{1-e^{-\lambda m}}{\lambda m}$ and $\frac{1-e^{-\lambda m}}{\lambda m} - e^{-\lambda m}$. Depending on different limiting behaviours of the loadings Diebold and Li (2006) interprete β_{1t}, β_{2t} and β_{3t} as the level, the slope and the curvature of the yield curve, respectively. For example, the loading on β_{1t} is constant at 1 for all maturities. Consequently, changes in β_{1t} affect all yields uniformly and thus control the level of the curve. Loadings on β_{2t} goes to one as $m \to 0$ but decays fast to zero as $m \to \infty$. Shocks in β_{2t}, therefore, primarily affect the short end of the yield curve and thus induce variations in yield spreads. Finally, the loadings on β_{3t} converges to zero as $m \to 0$ and $m \to \infty$ but is concave in m. Accordingly, shocks in β_{3t} have a dominant effect on yields with mid-term maturities and consequently, on curvature of the entire yield curve. Diebold and Li (2006) consider factor independence by allowing separate AR(1) dynamics for each factor and a more general case of correlated factors by allowing also a single VAR(1) process.

Diebold et al. (2006) identify that the latent nature of the factors allows the dynamic Nelson-Siegel model to be represented in a state-space system, a framework which can explicitly handle time series models with unobserved variables in a unified methodology. The measurement equation is formulated simply by adding maturity-specific stochastic error terms $\varepsilon_t(m)$ on the right-hand side of the yield equation (2.2.2). In matrix notations this can be written as

$$\mathbf{Y}_t = \mathbf{\Lambda}(\lambda)\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_{t,} \qquad t = 1, 2, ..., T, \tag{2.2.3}$$

where $\mathbf{Y}_t = [y_t(m_1), ..., y_t(m_N)]'$ is the vector of yields, $\boldsymbol{\beta}_t = [\beta_{1t}, \beta_{2t}, \beta_{3t}]'$ is the vector of latent factors and $\boldsymbol{\varepsilon}_t = [\varepsilon_t(m_1), ..., \varepsilon_t(m_N)]'$ is the vector of measurement errors. These errors can arise from various sources such as methods of yield extraction, mistakes in data entry, lack of synchrony in sampled data etc. $\mathbf{\Lambda}(\lambda)$ is a 3×3 matrix of factor loadings with its (i, j)-th element given by

$$\Lambda_{ij}(\lambda) = \begin{cases} 1, & j = 1 \\ (1 - e^{-\lambda m_i})/\lambda m_i, & j = 2 \\ \frac{1-e^{-\lambda m_i}}{\lambda m_i} - e^{-\lambda m_i}, & j = 3. \end{cases}$$

The state equation defines a VAR(1) factor dynamics which essentially nests an AR(1) process:

$$\boldsymbol{\beta}_t = \boldsymbol{\mu} + \Phi \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \backsim N(0, \Sigma_\eta), \qquad t = 1, 2, ..., T, \tag{2.2.4}$$

where $\boldsymbol{\mu}$ is a 3×1 vector of factor means, Φ is a 3×3 matrix of VAR coefficients, $\boldsymbol{\eta}_t$ is a 3×1 vector of disturbances and Σ_η is the variance-covariance matrix of disturbances. Note that for an AR(1) dynamics Φ and Σ_η are assumed to be diagonal. A final assumption is that the measurement- and state-equation disturbances are orthogonal to each other, i.e.,

$$\begin{bmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\eta}_t \end{bmatrix} \sim NID\left(\begin{bmatrix} 0_{N\times1} \\ 0_{3\times1} \end{bmatrix}, \begin{bmatrix} \Sigma_\varepsilon & \\ & \Sigma_\eta \end{bmatrix} \right). \tag{2.2.5}$$

The Yields-Macro Model

When extending the yields-only model to incorporate macroeconomic information we closely follow Diebold et al. (2006) and include three variables, namely official bank rate (BR_t), annual CPI Inflation (INF_t) and unemployment rate (U_t) as measures of policy instrument, inflation rate and economic activity, respectively.[2] The macroeconomic variables enter the set of state variables alongside the latent factors. A new state-space system is constructed with

[2] According to Diebold et al. (2006) this is the minimal set of indicators that can explain fundamental macroeconomic dynamics. Note, however, that while they use manufacturing capacity utilisation of the US as one of the variables, due to unavailability of such a variable for the UK we exploit unemployment rate to represent real activity in the economy.

equations (2.2.3)-(2.2.5) substituted by

$$\mathbf{Y}_t = \mathbf{\Lambda}(\lambda)\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \qquad t = 1, 2, ..., T, \tag{2.2.6}$$

$$\mathbf{f}_t = \boldsymbol{\mu} + \Phi\mathbf{f}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \backsim N(0, \Sigma_\eta), \quad t = 1, 2, ..., T, \tag{2.2.7}$$

$$\begin{bmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\eta}_t \end{bmatrix} \sim NID\left(\begin{bmatrix} 0_{N\times 1} \\ 0_{6\times 1} \end{bmatrix}, \begin{bmatrix} \Sigma_\epsilon & \\ & \Sigma_\eta \end{bmatrix} \right), \tag{2.2.8}$$

where $\mathbf{f}_t = [BR_t, INF_t, U_t, \beta_{1t}, \beta_{2t}, \beta_{3t}]'$ with dimensions of $\boldsymbol{\mu}$, Φ, $\boldsymbol{\eta}_t$, $\mathbf{\Lambda}$ and Σ_η are increased accordingly to account for additional three macroeconomic variables.[3] We maintain the assumption of bidirectional causality between yields and macroeconomic variables by allowing Φ and Σ_η to be full matrices.

Estimation

For estimating the yields-only model we use both a two-step approach following Diebold and Li (2006) and a one-step approach following Diebold et al. (2006).

The two-step approach requires λ to be fixed which allows us to compute the maturity-specific factor loadings in $\mathbf{\Lambda}(\lambda)$.[4] At each period t the measurement equation (2.2.3) then reduces to a linear regression model and OLS is applied in the first step to obtain period by period estimates of $\boldsymbol{\beta}_t = [\beta_{1t}, \beta_{2t}, \beta_{3t}]'$ using the cross-section of the spectrum of yields. In step two, we specify the dynamics of the latent factors by fitting the transition equation (2.2.4) and estimating related parameters. We label these latent-factor models, estimated in two stages, as $NS2\_AR$ and $NS2\_VAR$ depending on whether an AR(1)

---

[3] $\boldsymbol{\mu}$ is now a 6×1 vector, Φ is a 6×6 matrix and Λ is an $N \times 6$ matrix. We adopt the same parsimonious representation as in Diebold et al. (2006) and restrict the three left most columns of Λ to contain only zeros meaning that the macroeconomic variables have no direct influence on yields and that three latent factors are sufficient to explain interest rate dynamics.

[4] Following Diebold and Li (2006) we set $\lambda = 0.0609$ which maximizes the loading on β_3 at a 30-month maturity.

or a VAR(1) process has been considered for describing the evolution of the factors.

The one step approach uses the state-space system of equations (2.2.3)-(2.2.5) and applies a Kalman filter-induced maximum likelihood method to estimate all the parameters of the state and the measurement equations simultaneously, namely $\boldsymbol{\mu}$, Φ, Σ_η, $\boldsymbol{\Lambda}(\lambda)$ and Σ_ε. Thus, λ is not fixed and estimated together with other parameters. We consider a VAR(1) dynamics for the factors by assuming that both Φ and Σ_η are full matrices. We refer to this one-step-estimated yields-only model as $NS1$.

Finally, for the model with macroeconomic variables we only apply the one-step maximum likelihood approach to estimate the models in the state-space system defined by equations (2.2.6) - (2.2.8) and denote this model as NS_M.

2.2.2 Affine Term Structure Model

The Yields-Only Model

While the structure imposed on factor loadings in a Nelson-Siegel model is based on a convenient mathematical function used primarily to ensure smoothness across yields, such a structure in an affine model is founded on some cross-equation restrictions adopted for enforcing a theoretically desirable requirement of freedom of arbitrage. Whether imposing such restrictions helps in forecasting yields is still a debatable issue, but they can certainly help in pricing bonds and other financial instruments and estimating yields of unobserved intermediate maturities in a financially consistent manner. Affine models have been developed and have long been studied in a continuous-time environment with specifying a diffusion process for latent factor dynamics (Duffie and Kan (1996)). In our analysis we, however, adopt a discrete-time version of the models proposed in Ang and Piazzesi (2003). In particular, we assume a zero-mean Gaussian VAR(1) dynamics for a set of K latent factors, $F = [l_1, l_2, .., l_K]'$ which drive the movements of the yield curve:

$$F_t = \Psi F_{t-1} + u_t, \tag{2.2.9}$$

where, $u_t \backsim N(0, \Sigma\Sigma')$ with Σ being a lower triangular Cholesky matrix and Ψ is a $K \times K$ matrix of coefficients which govern the dynamics. The short rate is assumed to be an affine function of the factors:

$$r_t = \delta_0 + \delta_1^{'} F_t, \tag{2.2.10}$$

where δ_0 is a scalar and δ_1 is a $K \times 1$ vector. The nominal pricing kernel which is assumed to price all assets in the economy, is modelled as[5]

$$M_{t+1} = \exp(-r_t - 0.5\lambda_t^{'}\lambda_t - \lambda_t^{'} u_{t+1}), \tag{2.2.11}$$

where λ_t are market prices of risk which are assumed to be affine in the underlying state variables and depend only on contemporaneous observations of the model factors. The risk pricing equation, therefore, takes the form

$$\lambda_t = \lambda_0 + \lambda_1 F_t, \tag{2.2.12}$$

where λ_0 is a $K \times 1$ vector and λ_1 is a $K \times K$ matrix. In an arbitrage-free market, the price of an m-months to maturity zero-coupon bond in period t must equal the expected discounted value of the price of an $(m-1)$-months to maturity bond in period $t+1$. This leads to the recursive pricing formula:

$$P_t^m = E_t[M_{t+1} P_{t+1}^{m-1}], \tag{2.2.13}$$

where the expectation is taken under the risk-neutral measure. The bond prices are then exponential linear functions of the state vector:

$$P_t^m = \exp(A_m + B_m^{'} F_t), \tag{2.2.14}$$

[5] The pricing kernel is also known as the stochastic discount factor.

where the scalar A_m and the K-dimensional coefficient vector B_m depend on the time-to-maturiy m. Ang and Piazzesi (2003) show that imposing no-arbitrage results in the following recursive equations for bond pricing coefficients:

$$A_m = A_{m-1} + B_{m-1}^{'}(\mu - \Sigma\lambda_0) + 0.5B_m^{\prime}\Sigma\Sigma^{'}B_m - \delta_0, \tag{2.2.15}$$

$$B_m^{'} = B_{m-1}^{'}\left(\Psi - \Sigma\lambda_1\right) - \delta_1^{'}, \tag{2.2.16}$$

with $A_1 = -\delta_0$ and $B_1 = -\delta_1$. [6]

The continuously compounded yield on an m-period zero-coupon bond is then given by

$$y_t(m) = -\frac{\log p_t^m}{m} \tag{2.2.17}$$

$$= a_m + b_m^{'} F_t, \tag{2.2.18}$$

where $a_m = -A_m/m$ and $b_m = -B_m^{\prime}/m$. Thus, the yields are also affine functions of the state variables F_t.

The affine term structure model, analogously to its Nelson-Siegel counterpart, can be easily cast into a compact state-space framework. The transition and the measurement equations of the system can be summarised as

$$F_t = \Psi F_{t-1} + u_t, \tag{2.2.19}$$

$$Y_t = A + BF_t + v_t, \tag{2.2.20}$$

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} \backsim IIDN\left(\begin{bmatrix} 0_{N\times 1} \\ 0_{K\times 1} \end{bmatrix}, \begin{bmatrix} \Sigma_u & 0 \\ 0 & \Sigma_v \end{bmatrix}\right), \tag{2.2.21}$$

where $Y_t = [y_t(m_1), y_t(m_2), ..., y_t(m_N)]^{\prime}$ is a vector of all N yields at hand, $A = [a_{m_1}, a_{m_2}, ..., a_{m_N}]^{'}$ is a $N \times 1$ vector and $B = [b_{m_1}, b_{m_2}, ..., b_{m_N}]^{'}$ is a

[6] See Ang and Piazzesi (2003) or Moench (2008) for detailed derivation of these recursive formulae.

$N \times K$ matrix of bond pricing coefficients defined by equations (2.2.15) and (2.2.16). v_t is a vector of i.i.d. Gaussian measurement errors with variance-covariance matrix Σ_v and $\Sigma_u = \Sigma\Sigma^{'}$. We refer to the yields-only no-arbitrage affine model as $ATSM$.

Following common practice we set $K = 3$ to adopt a three-factor model where the latent factors are interpreted as level, slope and curvature of the yield curve. Dai and Singleton (2000) show that in a VAR setting linear transformation and rotation of unobserved factors achieve observationally equivalent yields. Following Ang and Piazzesi (2003) we identify the factors in the Gaussian specification by adopting the a simple normalisation where we assume that Ψ is lower-triangular and Σ_u is an identity matrix I. Then, δ_0 is the unconditional mean of the observed short-rate (which we approximate by the 3-month treasury bill rate). Similar to the the Nelson-Siegel models we assume Σ_v to be diagonal.

The Yields-Macro Model

Adding macroeconomic information in affine models is as straightforward as in Nelson-Siegel models. We follow Ang and Piazzesi (2003) to first collect time series information on a number of variables related to inflation and economic activity and standardise each series to have zero mean and unit variance.[7] Then, we extract the first principal components from each group, namely 'inflation' and 'economic activity' and denote them as M_1 and M_2, respectively. We restrict the macro factors to follow a simple VAR(1) process in order to keep the model parsimonious in terms of parameters.[8] The state-space representation of the model with macro factors then consists of the following

[7] We describe all the macroeconomic variables in section 2.3.

[8] Ang and Piazzesi (2003), however, adopted a VAR(12) dynamics for the macro factors. But unlike their study the principal goal of our work is forecasting and so parametric parsimony of a dynamic model is desirable.

equations:

$$Z_t = \Psi Z_{t-1} + u_t, \tag{2.2.22}$$

$$Y_t = A + BZ_t + v_t, \tag{2.2.23}$$

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} \backsim IIDN \left(\begin{bmatrix} 0_{q\times 1} \\ 0_{(K+2)\times 1} \end{bmatrix}, \begin{bmatrix} \Sigma_u & 0 \\ 0 & \Sigma_v \end{bmatrix} \right), \tag{2.2.24}$$

where $Z_t = [M_1, M_2, F_t^{'}]'$ and dimensions of Ψ, u_t and Σ_u are increased accordingly to adjust for additional observed macroeconomic factors. Following Ang and Piazzesi (2003) we impose independence between the latent and macro factors.[9] This implies that for a model with three latent factors the lower-left 3×2 corner and the upper-right 2×3 corner of Ψ in the state-space system (2.2.22)-(2.2.23) contain only zeros. The lower-right 3×3 corner is restricted to be lower-triangular to match desirable canonical representation of latent factors. We refer to the macro-added affine model as $ATSM\_M$.

Estimation

Both the yields-only model, $ATSM$ and the yields-macro model, $ATSM\_M$ are estimated in one-step by using the associated state-space representations and applying the Kalman-filter induced maximum likelihood estimation process. This allows simultaneous estimation of all parameters and extraction of latent yield factors in a unified framework.

[9] This strong assumption has two drawbacks. First, it defies historical evidence that the term structure predicts macroeconomic dynamics and the general logic that policy rate affects inflation and economic activity (see Ang and Piazzesi (2003)).

The Benchmark and a Simple Competitor

Random Walk

Since time series of bond yields show high persistence, a driftless random walk (RW) process is often found to capture yield dynamics very well. The yields are modelled as

$$y_t(m) = y_{t-1}(m) + \sigma(m)\zeta_t(m), \quad \zeta_t(m) \sim N(0,1). \tag{2.2.25}$$

The model implies that interest rates are not predictable and any forecast is taken to be the last available observation. It is well-documented in the empirical literature that standard term structure models struggle to outperform the naive random walk forecasts (e.g., Duffee (2002), Ang and Piazzesi (2003) and Diebold and Li (2006)). Therefore, we use the 'no change' forecast as the benchmark against which we compare the forecasts of all other competing models.

AR(1) on Yield Levels

The term structure literature often finds simple and parsimonious models to produce more accurate forecasts than sophisticated models. We, therefore, present an univariate AR(1) on yield levels which often produces good yield forecasts. The AR(1) dynamics is given by

$$y_t(m) = c(m) + \gamma(m)y_{t-1}(m) + \sigma(m)\zeta_t(m), \quad \zeta_t(m) \sim N(0,1), \tag{2.2.26}$$

where $c(m)$, $\gamma(m)$ and $\sigma(m)$ are scalar parameters.

We denote this model as AR and treat it like a second benchmark.

2.3 Data Description

The data set for our empirical analysis consists of monthly UK nominal zero-coupon bond yields from January 1989 to November 2010. We use end-month spot interest rates reported at fixed maturities of 3, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months. Yields at maturities of one year and longer have been downloaded from the Bank of England website which publish them in daily frequency.[10] These yields have been derived from UK government bond (gilt) prices and General Collateral (GC) repo rates by applying a spline based estimation technique.[11] The shortest yield of 3-month maturity is, however, proxied by the 3-month treasury bill rate which has been downloaded using Datastream. We do so because of absence of repo market and consequent irregular availability of yields at very short maturities (less than one year) before March, 1997.

Figure 2.A.1 plots a subset of the sample yields over time. In the beginning of the sample the UK experience an inverted yield curve with higher yields at shorter maturities. Levels of yields are in general high ranging between 9% to 13%. During the period 1989-1999 the yields remain more or less volatile. First, there is a sharp increase in all the yields in the year 1989 and they reach their highest levels in a year time. This period is then followed by a longer period of plummeting rates which lasts till the beginning of 1994. Moderate amount of fluctuations remains till the end of 1999. 2000-2007 is a period of a more stable and flatter yield curve with yields at different maturities staying close to each other. Following the start of the recent global financial crisis in the third quarter of 2007 the short yields drop abruptly to their historical low in reaction to a lowered official bank rate by the Bank of England. The longest rate does not, however, decline as much creating a very wide spread between the short and long end of the yield curve.

Descriptive statistics of yields are presented in **Table 2.B.1**. The average yield curve is downward sloping over maturities of 3 months to 1 year, but

---

[10] The Bank of England website is www.bankofengland.co.uk.

[11] See Anderson and Sleath (2001) for detailed technical description of the estimation method.

is upward sloping from then onwards. The standard deviations and sample autocorrelations suggest that volatility decreases with maturity and long rates are more persistent than the short rates. The shapes of the mean and median yield curve is very different: while the short end of the average curve is downward sloping, it is upward sloping for the median curve and while mid-to-long end of the average curve is upward sloping, it is downward sloping for median curve. These facts, together with reported non-zero skewness and excess kurtosis, imply that yields are probably non-normal. Yields of different maturities are highly correlated and the closer the maturities are to each other the stronger is the relation. The weakest correlation of 87% is observed between the 3-month treasury bill rate and 10-year yield.

For macroeconomic information we include the official bank rate as a measure of policy instrument, a group of inflation related variables namely consumer's price index (CPI), producer's price index (PPI) and retailer's price index (RPI) and a group of economic activity related variables namely unemployment rate, the claimant count rate, the growth in employment and annual growth in index of production (IOP). Time series data on the official bank rate are collected from the website of Bank of England, while the remaining are downloaded from the website of Office for National Statistics[12]. The sample time period for each macroeconomic series coincides with that for the yield data.

Figure 2.A.2 shows time series plots of three key macroeconomic variables - the official bank rate, CPI inflation and unemployment rate. For each series we can identify three different time periods with distinctive patterns. During early to mid 90s all the three curves show noticeable movements and reach their maximums for the sample. The bank rate falls substantially after rising to the highest level of 15%, inflation rate fluctuates between 6% to 8% after an initial increase and then fall sharply to 2%, and unemployment rate rises and

[12] www.statistics.gov.uk

exceeds 10% and is then followed by a gradual decrease. Mid 90's to the third quarter of the year 2007 marks the period of stability. The bank rate stays close to 5%, inflation rate remains within the BOE's target rate of 2% and unemployment rate slowly drops to 5% and is maintained at that level. The global financial crisis that starts on September, 2007 breaks the stability and the macroeconomic series begin to show some aberrant behaviours following the crisis. The bank rate is adjusted and held fixed at its historical lowest level of 0.5%, inflation becomes volatile with sharp fluctuations between 1% and 5%, and unemployment rises by almost 2% to 7%.

2.4 In-Sample Performance

Although our principal interest is systematic evaluation of forecast ability of competing term structure models, we begin our empirical analysis by assessing how well the models fit the cross-section of the UK yield curve. We use root mean squared error (RMSE) of fitted residuals, a standard and widely used evaluation criterion, for measuring the goodness of model fit. Provided that both Nelson-Siegel and affine term structure models have good empirical records of estimating bond yields we expect good in-sample performance from all the models. Results presented in **Table 2.B.2** confirm this. Small values of RMSE, expressed in percentages, indicate that overall the models fit the data well. Noticeable difference in RMSE values of the $NS2\_AR$ and $NS1$ models implies that the quality of data-fit of the Nelson-Siegel models varies depending on whether factors are assumed to be independent or correlated, if λ parameter is treated to be fixed or estimated freely and/or whether model is estimated in one or two steps. Identical fit for the $NS1$ and NS_M models indicates that there is no significant benefit from incorporating macroeconomic information in Nelson-Siegel models.[13] Adding macroeconomic fundamentals, however, clearly improves the fit of affine term structure models. The macro-yields model, $ATSM\_M$ reports RMSEs which are consistently lower than those of its yields-only counterpart, $ATSM$. Gains are, however, small. Overall, the Nelson-Siegel models, $NS1$ and NS_M provide better fit than the independent factor specification, $NS2\_AR$ and the two affine models, except for the 3-month yield and yields with maturities longer than seven years.

It is well recognised in term structure literature that a small number of factors that can be distilled from the cross-section of yields are sufficient to describe variations in the entire yield curve (e.g., see Litterman and Scheinkman (1991), Bliss (1997)). A factor model with three latent factors is the most commonly used specification and the factors are often linked to three attributes of the yield curve - namely level, slope and curvature (see Litterman and Scheinkman (1991) and Diebold and Li (2006)). Another way of evaluating how well our three-factor Nelson-Siegel and affine models fit the UK yields is to look at the agreement between time series plots of actual level, slope and curvature of the sample yield curve and model extracted factors. The plots are presented in **Figure 2.A.3**.[14] We follow Diebold and Li (2006) in defining the true level, slope and curvature as the 10-year yield, the difference between the 10-year and 3-month yield, and twice the 2-year yield minus the sum of 3-month and 10-year yield, respectively. Close agreement between empirical and model generated factors indicates good approximation to yield dynamics by both the Nelson-Siegel and affine models.

[13] RMSE values of the NS1 and NS_M models are not exactly identical, but differences are too small to be observed in percentage values expressed in three decimal points.

[14] Since factors generated by different models operate at different levels we standardise them for a fare comparison.

2.5 Out-of-Sample Forecasting

In this section we evaluate the predictive performance of competing forecasting models. Alongside standard yields-only versions we use variants of these models which incorporate macroeconomic variables. The models are applied to generate forecasts of future government bond yields of the UK and compared in terms of forecast accuracy. We investigate a number of issues: whether additional macroeconomic information helps in out-of-sample forecasting and whether predictive ability of models changes over time. In what follows we sequentially describe the forecast design and how the forecasts are constructed, and finally, discuss the results of the forecasting exercise.

2.5.1 Forecast Procedure

We generate model forecasts by adopting separate recursive and rolling-window estimation procedures. The two schemes differ in terms of how much of the past information is used for making a forecast in the future. Under the recursive scheme, often called an expanding-window scheme, parameters of a model are estimated using all data available at each forecast origin (the point in time when a forecast is made). Under a rolling-window scheme, however, a data-window of fixed size is rolled over the full sample to update the parameter estimates and generate future forecasts at each forecast origin. For each strategy we use the most up-to-date information available at the time when a forecast is made. Rossi (2012) argues that choice of expanding or rolling window estimation may play an important role in forecasting in presence of structural breaks. Rolling estimation is expected to forecast better in case of large and recurrent breaks while recursive estimation is expected to be advantageous when breaks are small or absent. Although we do not specifically test for breaks in our sample yields series, it is our interest to investigate how application of rolling and recursive schemes affects forecast accuracy of our competing models, particularly during the period of unusually low short rates following the recent financial crisis.

We construct 1-, 6- and 12-month-ahead forecasts over the period 2001:1-2010:11. We decide to keep the forecast period and consequently, the number of forecasts fixed irrespective of the length of forecast horizon.[15] Such a design allows us to directly compare the quality of forecasts made at different horizons but requires that we adjust estimation sample according to the length of the forecast horizon. Thus, when generating 1-month-ahead forecasts using the recursive scheme the initial estimation window is set to 1989:1-2000:12 so that the first forecast is made at 2001:1 and the last estimation window is set to 1989:1-2010:10 so that the last forecast is made at 2010:11. For longer horizons we reduce the estimation sample in order to maintain a fixed number of 119 out-of-sample forecasts. For example, the first estimation window for 12-month-ahead prediction is 1989:1-2000:1 so that the first forecast is made at 2001:1 and the last estimation window is 1989:1-2009:11 so that the last forecast is made at 2010:11. Under rolling-window estimation scheme we generate forecasts over the same forecast period 2001:1-2010:11. For each forecast horizon h we keep the size of the estimation window fixed at 133 observations. This, however, requires that the initial estimation period is different for different h. For example, when forecasting 1-month-ahead the initial estimation window is 1989:12-2000:12 so that the forecast is made at 2001:1 and the last estimation window is 1999:10-2010:10 so that the last forecast is made at 2010:11. When forecasting 12-month-ahead, however, the first estimation window is 1989:1-2000:1 so that the first forecast is made at 2001:1 and the last estimation window is 1998:11-2009:11 so that the last forecast is made

[15] This is important because one of our objectives is to compare forecasts at various horizons made over the financial crisis period which constitute the last 35 months of the sample.

at 2010:11. When predicting yields multi-period ahead, we construct iterated forecasts where the one-period-ahead model is iterated forward to make forecasts at longer horizons.

2.5.2 Construction of Forecasts

We describe below how out-of-sample forecasts are computed for all the reported models including the benchmark. As mentioned earlier, we use an iterative approach to generate multi-step ahead forecasts.

Nelson-Siegel Models

Once the parameter vector $\{\boldsymbol{\mu}, \Phi, \Sigma_\eta, \Sigma_\varepsilon, \lambda\}$ of the state-space system is estimated at each time point t, an h-step-ahead forecast of the state vector is computed by iterating forward the estimated state equations (2.2.4) and (2.2.7):

$$\widehat{\boldsymbol{\beta}}_{t+h|t} = \left(\sum_{i=0}^{h-1} \widehat{\Phi}^i\right) \widehat{\boldsymbol{\mu}} + \widehat{\Phi}^h \widehat{\boldsymbol{\beta}}_t, \tag{2.5.1}$$

$$\widehat{\mathbf{f}}_{t+h|t} = \left(\sum_{i=0}^{h-1} \widehat{\Phi}^i\right) \widehat{\boldsymbol{\mu}} + \widehat{\Phi}^h \widehat{\mathbf{f}}_t. \tag{2.5.2}$$

Equation (2.5.1) corresponds to the yields-only models, $NS2\_AR$, $NS2\_VAR$ and $NS1$, and equation (2.5.2) corresponds to the yields-macro model, NS_M. The h-step-ahead yield forecasts are then made by substituting the factor forecasts in the estimated measurement equations:

$$\widehat{Y}_{t+h} = \boldsymbol{\Lambda}(\widehat{\lambda}) \widehat{\boldsymbol{\beta}}_{t+h|t}, \tag{2.5.3}$$

$$\widehat{Y}_{t+h} = \boldsymbol{\Lambda}(\widehat{\lambda}) \widehat{\mathbf{f}}_{t+h|t}. \tag{2.5.4}$$

Note that for the model $NS2\_AR$, which is estimated in two steps, $\widehat{\lambda} = \lambda = 0.0609$ is fixed.

Affine Term Structure Models

Similar to the Nelson-Siegel models, when the parameter vector $\{\Psi, \delta_1, \lambda_0, \lambda_1, \Sigma_v\}$ of the affine models is estimated at any forecast origin, the h-step-ahead forecasts of state variables are computed by iterating forward the estimated state equations (2.2.19) and (2.2.22):

$$\widehat{F}_{t+h|t} = \widehat{\Psi}^h \widehat{F}_t, \tag{2.5.5}$$

$$\widehat{Z}_{t+h|t} = \widehat{\Psi}^h \widehat{Z}_t. \tag{2.5.6}$$

Equation (2.5.5) corresponds to the yields-only model, $ATSM$ and equation (2.5.6) corresponds to the yields-macro model, $ATSM\_M$.

Estimated state and measurement equation parameters are placed in recursive pricing equations (2.2.15) and (2.2.16) to obtain estimates of A and B. Finally, the h-step-ahead yield forecasts are then made by substituting the factors by their forecasts in the estimated measurement equations:

$$\widehat{Y}_{t+h} = \widehat{A} + \widehat{B}\widehat{F}_{t+h|t}, \tag{2.5.7}$$

$$\widehat{Y}_{t+h} = \widehat{A} + \widehat{B}\widehat{Z}_{t+h|t}. \tag{2.5.8}$$

Random Walk

The benchmark forecast is simply the 'last available observation' or 'no-change' forecast, i.e.,

$$\widehat{y}_{t+h}(m_i) = y_t(m_i).$$

AR(1) on Yield Levels

We use the OLS to estimate at each forecast origin the parameters c and γ of the AR(1) process (2.2.26) and then construct the h-step-ahead forecasts of yields with maturity m_i as

$$\widehat{y}_{t+h}(m_i) = \left(\sum_{i=0}^{h-1} \widehat{\gamma}(m_i)\right) \widehat{c}(m_i) + \widehat{\gamma}^h(m_i) y_t.$$

2.5.3 Forecast Evaluation Criteria and Tests for Model Comparison

We evaluate the quality of forecasts by computing root mean squared forecast error (RMSFE). It is a popular measure of predictive performance of time series models and is symmetric in nature as it penalises negative and positive errors equally. The smaller the value of RMSFE the better is the forecast accuracy of a model. Let $\widehat{y}^j_{t+h}(m)$ denote the h-month-ahead forecast of an m-maturity yield $y_{t+h}(m)$ made by model j. Then the associated RMSFE is defined by

$$RMSFE^j_{m,h} = \sqrt{\frac{1}{T_n} \sum \left(\widehat{y}^j_{t+h}(m) - y_{t+h}(m)\right)^2},$$

where the sum is computed over total number of forecasts, T_n.

We compare the predictive performance of models by reporting RMSFEs relative to the benchmark which is the random walk. The relative root mean squared forecast error (RRMSFE) is computed as:

$RRMSFE^j_{m,h} = \frac{RMSFE^j_{m,h}}{RMSFE^{RW}_{m,h}}$

A value of $RRMSFE^j_{m,h}$ smaller than one indicates that model j forecasts better than the random walk.

In order to assess statistical significance of any forecast gain or loss relative to the benchmark we apply the unconditional version of Giacomini and White (2006) test of forecast comparison. The null hypothesis of the test is that of equal predictive performance (measured in terms of loss functions such as squared errors) of two competing models. One major advantage of the test is that it can be effectively applied to forecasts based on both nested and non-nested models.[16] The test is robust to choice of estimation procedures (e.g., Bayesian and fully-, semi- and non-parametric methods) but requires that the size of estimation sample be finite. We could, therefore, apply the test only for the rolling-window based forecasting exercise. For the expanding-window estimation scheme we only report the RRMSFE without commenting on statistical significance of any predictive gain.

Giacomini and Rossi (2010) argue that in presence of instability standard tests of forecast accuracy are not insightful and propose a more appropriate fluctuation test for testing statistical significance of evolving relative forecast performance of two competing models. During our forecast period 2001:1 - 2010:11 the UK yield curve shows changing behaviours: a period of stability is followed by some erratic characteristics during the financial crisis, such as dramatic fall of short rates and pronounced widening of spread. We, therefore, use the fluctuation test to evaluate significance of possible time variation in predictive ability of a forecast model relative to random walk, the benchmark. The test statistic is calculated as standardised difference between the $MSFE^{j}_{m,h}$ and $MSFE^{RW}_{m,h}$ computed over a rolling-window of 35-months. Negative values of the test statistic imply that the model under consideration is better than the random walk. Giacomini and Rossi (2010) report critical values of the test. Since the test statistic is equivalent to that of Giacomini and White (2006) test it can also be used only for the rolling-window forecasts.

2.5.4 Forecast Results

Results of model performance for the entire forecasting period of 2001:1-2010:11 are presented in **Table 2.B.3.** The table has two horizontal panels,

[16] This is particularly important for our study since many of the forecasting models nest the random walk benchmark. Alternative tests such as Diebold and Mariano (1995) test are not suitable for comparing nested models, and therefore, avoided.

panel (a) reporting results for the recursive forecasts and panel (b) reporting results for the rolling-window forecasts. Each panel is again divided into three vertical blocks containing results for three different forecast horizons: 1-month, 6-month and 12-month. The second row in each block reports five selected maturities ranging from 3 months to 10 years, while the remaining rows report the root mean squared forecast errors relative to random walk, $RRMSFE$. A value of $RRMSFE$ smaller than one implies that the model under consideration outperforms the random walk benchmark. The bold number under each maturity represents the minimum $RRMSFE$ identifying the best forecasting model. An asterisk indicates that the predictive gain over random walk is statistically significant under Giacomini and White (2006) test ('*' and '**' indicate significance at 10% and 5% levels, respectively).

In order to check temporal robustness of our forecast results we further conduct a stability check by evaluating predictive performance over two non-overlapping sub-samples. The first subsample consists of the first seven years of the forecast period, during which the interest rates are relatively less volatile and the spread between the long and short rates are particularly narrow. The second sub-sample span the last 35 months of the forecast period and is marked by near-zero short term yields and wide spread. $RRMSFE$ results for the two sub-periods are reported in **Tables 2.B.4 - 2.B.5**. These tables are structured and interpreted the same way as the full-sample table **Table 2.B.3**. We discuss the results of the forecasting exercise in details below.

Full Sample Results

Sample 2001:1 - 2010:11

From panel (a) of **Table 2.B.3** we find that the 1-month-ahead forecast results under the recursive scheme are very mixed. No single model stands out to be the best predictor by outperforming the random walk across all or many maturities. Most of the models, however, forecast the shortest rate of 3-month maturity very well and no-arbitrage affine models yield the largest $RRMSFE$ reductions of about 8%. Poor forecasting performance relative to the benchmark does not improve at longer horizons except that when forecasting 12-month-ahead the two specifications of the Nelson-Siegel yields-only models, denoted as $NS2\_AR$ and $NS1$, predict the 3-month and 12-month yield more accurately than the 'last observation' benchmark. The AR(1) on yields level forecasts the short end of the yield curve reasonably well. Models that incorporate additional macroeconomic information generate inferior forecasts compared to their yields-only counterparts. Overall, the random walk shows the best predictive ability and any gain over it is small.

Results of the rolling-window scheme, presented in panel (b), show a more clear pattern in the relative forecasting performance of individual models. At 1-month horizon, the no-arbitrage affine term structure model with macroeconomic factors, namely the $ATSM\_M$, consistently beats the random walk for all yields, except for that of 10-year maturity. This result is very much in line with the findings of Ang and Piazzesi (2003) who report impressive 1-month ahead predictive performance of a similar model when forecasting the US zero-coupon bond yields. The advantage of the $ATSM\_M$ model over other competing models does not, however, sustain over longer forecast horizons. At 6-month horizon, the Nelson-Siegel yields-only model with an AR(1) factor dynamics, the $NS2\_AR$ model shows the most consistent forecast accuracy. Most of the gains are, however, not statistically significant. The 12-month-ahead forecasts reveal a similar story as has been established under the recursive scheme: the yields-only models produce superior forecasts of short yields and adding macroeconomic fundamentals appears to be disadvantageous for forecasting at longer horizons. One major distinction between the results of rolling- and expanding-window exercises is that of a relatively much worse performance of the NS_M model in the former. When forecasting 12-month-ahead using a recursive scheme, penalties relative to benchmark random walk range between 11%-20%. But with forecasting using rolling-windows such costs lie within a much higher range of 50%-140% and increase with maturity. These findings possibly imply that the heavily parameterised yields-macro Nelson-Siegel model may require accounting for information of distant past for its stable estimation and generation of reasonable forecasts.

Subsample Results

Sample 2001:1 - 2007:12

During this relatively long subsample the UK yield curves remain stable and yields with different maturities stay close to each other. Panel (a) of **Table 2.B.4** contains results for the recursive forecasting scheme. Like the full-sample exercise $RRMSFE$ result at 1-month horizon is inconclusive and does not guide towards a clear preference for a best model. However, as the forecast horizon increases, a definitive pattern emerges. Apparently, over 6- and 12-month horizons the $NS2\_AR$ model systematically beats all its competitors across almost all reported maturities. This result resembles the overwhelming long-horizon predictive accuracy of the $NS2\_AR$ model observed in Diebold and Li (2006) while forecasting US yields. At longer horizons the $NS2\_VAR$ model fares reasonably well against the benchmark, but the $NS1$ model performs poorly. This finding implies that during the period of stable interest rates a VAR(1) factor-dynamics is helping in forecasting only when the Nelson-Siegel model is estimated in two steps and/or λ parameter is held fixed at 0.0609. The no-arbitrage affine models render worse predictive ability than the random walk benchmark. Finally, incorporating macroeconomic information deteriorates forecast accuracy of yields-only models, as is evident from large $RRMSFE$s of the macro-yields models NS_M and $ATSM\_M$.

Most of the findings under recursive forecasting also hold for the rolling-window exercise. For example, the $NS2\_AR$ model maintains supreme predictive ability and dominates alternative models. Its gains against the benchmark random walk are larger than in recursive case and are mostly statistically significant, particularly at longer forecast horizons. For 12-month-ahead forecasts the margins of significant outperformance is as high as 16% to 26%. In general, the impact of using macroeconomic variables in forecasting is negative except for the fact that they improve predictive ability of the NS_M model when forecasting the longest yield of 10-year maturity. This is interesting considering the fact that the NS_M model performs miserably on the whole sample. Results of rolling-window scheme that are noticeably different from those of recursive exercise include relatively better and statistically significant 1-month-ahead forecast of the yields-only model $ATSM$ and worse performance of the $NS2\_VAR$ model in the former.

Sample 2008:1 - 2010:11

This 35-month subperiod accommodates the recent financial crisis experienced by the UK and not surprisingly records some irregular behaviour on the part of the yield curve. During this period the short rates decline sharply by about 4.5% from 5% to 0.5% and the spreads between the long and the short rates widen substantially. Thus, this subsample offers us a platform to investigate whether the crisis has any impact on the relative forecasting power of the competing term structure models. We first analyse panel (a) of **Table 2.B.5** that reports the expanding-window results. The sole domination of the $NS2\_AR$ model observed during the pre-crisis period vanishes. It fails to outperform the benchmark random walk, except for only the 3-month rate. At the longest forecast horizon of 12-month, the NS_M model forecasts the 3- and 12-month rate better than the benchmark, the $NS2\_AR$ and the $NS2\_VAR$, but it is outperformed by the $NS1$ model which is the most accurate. This suggests that improvement in the predictive power of the Nelson-Siegel models is attributable to estimating them in one-step within a state-space framework, rather than to taking macroeconomic information into account. The macroeconomic fundamentals, however, clearly helps the no-arbitrage affine models in forecasting the short-term yields. The $ATSM\_M$ model predicts the 3- and 12-month rates most accurately at the 6-month forecast horizon and is able to surpass random walk and the yields-only affine model $ATSM$ in forecasting the same rates at 12-month horizon. A simple AR(1) model also shows reasonably good predictive performance.

The rolling-window results, reported in panel (b) of **Table 2.B.5**, render convincing evidence of strikingly superior forecasting accuracy of the $ATSM\_M$ model. For 1-month-ahead forecasts it comfortably outperforms random walk and other competitors across all maturities except 10-year. Gains over random walk are large and range between 9% to 18%. This dominant feat is carried also to longer forecast horizons. At 12-month horizon the $ATSM\_M$ model achieves highly significant $RRMSFE$ reduction of 21% and 34% over the benchmark for yields with 3-year and 5-year maturities, respectively. The yields-only model $ATSM$ also does well, but adding macroeconomic factors seems to play an important role in improving forecast accuracy on affine models during the financial crisis. Overall, no-arbitrage affine models forecast better than their Nelson-Siegel counterparts. Similar to the findings as in the recursive scheme, disappointing predictive performance of the $NS2\_AR$ model is evident. Models with VAR(1) factor dynamics, such as the $NS2\_VAR$ and $NS1$ show occasional good forecasting accuracy, particularly at the short end of the yield curve. The model that suffers the most in predicting yields is the NS_M model, reporting at times $RMSFE$ values which are 2 to 3 times larger than those of random walk. This performance of the NS_M is relatively much worse than its rolling-window performance in the pre-crisis period. Interestingly, its yields-only counterpart, the $NS1$ model, shows a reverse pattern by forecasting consistently better in the crisis period than in the pre-crisis period, particularly at longer horizons. These observations imply that structural change in the macroeconomic variables and yields during the crisis period may have played a role in the forecast failure of the heavily parameterised model NS_M. However, a comparison between rolling-window and recursive forecasts confirms that the inferior rolling-window performance is potentially due to accounting for recent data and discarding older information.

Testing Time-varying Predictive Ability

It is evident from the sub-sample analysis that the models perform differently across subperiods. A clear example is the $NS2\_AR$ model which dominates other competing models in the first sub-sample, but forecasts miserably in the second subsample. Another example is the $ATSM\_M$ model which demonstrates an exact opposite forecast pattern by yielding unsatisfactory predictive performance in the first subsample and producing superior forecasts in the second. As we have discussed earlier, we further investigate possible time-variation in forecast accuracy of models by exploiting the fluctuation test of Giacomini and Rossi (2010). The test tracks relative accuracy of two competing models over time and monitors for significant deviations in performance. **Figures 2.A.4 - 2.A.6** plot the evolution of the fluctuation test statistics with their critical values for 1-month, 6-month and 12-month forecast horizons, respectively. Each figure has two panels. Panel (a) reports the results for the 12-month yield and panel (b) reports the results for a relatively longer 5-year yield.

Figure 2.A.4 shows that in the first few years of the out-of-sample period only the $AR(1)$ model produces forecasts of the 12-month yields that are significantly worse than those of the random walk. Otherwise, for yields of bonds with maturities of both 12 and 60 months, forecasts of competing models are not significantly different from random walk forecasts until mid 2007. Since then the models' forecasting power, however, starts to change dramatically.

The only model that consistently and significantly outperforms the random walk during the post-2007 financial crisis period is the $ATSM\_M$ model. All other models show traces of inferior forecasting ability after some time. For many of them predictive performance becomes significantly worse than the benchmark at certain points, e.g., $NS2\_AR$, $ATSM$ and NS_M when forecasting the 12-month yield and NS_M, $ATSM$ and AR when forecasting the 60-month yield.

Results of fluctuation tests for the 6-month-ahead forecasts are presented in two panels of **Figure 2.A.5**. Variable predictive performance of the models, more prominent for the longer maturity yields, can be observed from the very beginning of the forecast period. The most remarkable feature is the early superiority of the $NS2\_AR$ model in contrast to relatively much poorer performance of alternative models compared to the random walk benchmark. However, the advantage of the $NS2\_AR$ model soon dies out and the performance of other models improve over time. Over the period of mid 2004 to mid 2007 forecast accuracy of all models is very similar to that of random walk. But on the verge of the crisis in 2007, the paths of relative performance of the forecasting models start to diverge. Although only the $ATSM$ model forecasts the 12-month interest rate significantly better than the random walk, the $ATSM\_M$ and $NS1$ models also do reasonably well. For a longer 60-month yield, the $ATSM\_M$ model comfortably outperforms all its competitors over the latest period of interest rate anomaly, at times with significant gains over the random walk. The NS_M model, as for 1-month-ahead forecasts, displays significant forecast failure at the end of the sample. The performance of the $NS2\_AR$ model becomes increasingly disappointing over time.

Figure 2.A.6 looks at time-variation in 12-month-ahead forecasts and reports many similar patterns as those observed at 6-month horizon. Three periods of distinctive features can clearly be identified: a first period of early but fading superiority of the $NS2\_AR$ model, a second period of relatively equal

and stable forecast performance of competing models and finally, a third period of unusually low short rates and high spread. During the latter period which accommodates the crisis the $NS2\_AR$ and NS_M models produce significantly inferior forecasts, the affine models generate significantly more accurate short rate predictions than the benchmark, and the $ATSM\_M$ model stands out to be the most dominating predictor and shows significant advantage of incorporating macroeconomic information.

Explaining Variations in Model Performance During the Crisis

One of our major findings is apparent change in relative forecasting performance of competing models during the recent financial crisis. Unquestionably, the crisis induced some unusual behaviour in the UK yield curve which was relatively stable for a substantially long period prior to the crisis. In response to Band of England's lowering of the official rate in the early 2009 the short yields plummeted towards zero and the gap between short and long rates widened. Forecasts of many models under study deteriorated relative to the benchmark random walk, possibly because of failure to account for structural changes in the yield curve dynamics and consequent parameter instability inflicted by the crisis.

Careful comparisons can shed further light on potential reasons behind variations in model-specific predictive performance during the crisis. The $NS2\_AR$ model which forecasted the UK yield curve exceptionally well during 2001-2004 conceded large errors during the crisis affected period of 2008-2010. Its poor performance relative to the $NS1$ model in the latter period may be explained by two factors. First, using an AR(1) dynamics for latent yield curve factors which impose factor independence is too restrictive and predicting the changed dynamics of interest rates in crisis requires a more flexible VAR(1) specification which allows interactions among all the factors. Second, fixing the lambda parameter which determines the decay in the loadings of slope and curvature factors appears to be similarly restrictive and estimating it along with other model parameters is desirable. Inferior performance of the NS_M model compared to its yields-only counterparts during the crisis possibly implies that the drastic change in yield dynamics caused adverse parameter instability in the model which already suffers from parameter proliferation arising from addition of macroeconomic variables. The affine term structure models, $ATSM$ and $ATSM\_M$, performed better than the Nelson-Siegel models, particularly at the longest forecast horizon, indicating possibly that imposition of no-arbitrage restrictions became important in forecasting UK yields during the crisis. Overall, the term structure of interest rates of the UK became harder to predict and the random walk benchmark predicting no change appeared to be difficult to outperform. The only model that consistently showed improved forecasting power is the $ATSM\_M$ model. This indicates that incorporating observed macroeconomic factors can benefit the arbitrage-free affine models in explaining and predicting the crisis period anomaly in bond yields.

2.6 Concluding Remarks

Forecasting of the term structure of UK interest rates has received little attention in the finance or economics literature. This paper, using UK nominal zero-coupon bond yield data, attempts to shed light on this important, yet overlooked issue. In its build up this essay tries to answer a number of important questions related to yield curve forecasting. First, it searches for a well-performing forecasting model for the UK yield curve by providing a survey of a number of popular term structure models and comparing them in terms of forecast accuracy. Second, it examines whether adding macroeconomic information in the latent factor models improves predictive power. Third, it undertakes both a recursive and a rolling-window forecast exercise to determine whether the length of information history used in estimating dynamic models play an important role in generating accurate forecasts. Finally and most importantly, it investigates whether the predictive performance of the yield curve models changes over time. In order to investigate such changing behaviour the full forecast period is divided into two subperiods: one records fairly stable yield movements and the other accommodates the recent financial crisis period reporting abnormal yield dynamics. The models are then evaluated across the two subsamples in order to check for consistency in forecasting performance. Instead of relying solely on a time-invariant global measure of predictive ability, we make inference by following the entire time path of the models' local relative performance and testing for possible significant time variation.

A first-hand idea from the full sample results is that it is difficult to outperform simple 'no-change' forecasts. A subsample analysis, however, reveals that the forecasting power of models vary significantly across subperiods. Both the recursive and the rolling-window exercises confirm that the yields-only models perform better during periods when the term structure of interest rates shows a stable pattern, such as the period 2001-2007 during which UK interests rates were less volatile and spread between long and short rates is relatively narrow. In particular, a Nelson-Siegel yields-only model with AR(1) factor dynamics renders superior predictive ability against all competing models. During the recent crisis period of 2008-2010 the interest rates, however, exhibit some aberrant characteristics, such as sharp decline in short rates and pronounced widening of the spread. It is evident from the rolling-window scheme that only a no-arbitrage affine model with macroeconomic information can predict these unusual behaviours of yields most accurately and most consistently. A test of time varying relative performance further corroborates this finding. One additional finding is that the choice of forecasting designs may substantially affect forecasting ability of certain models. One clear example is the Nelson-Siegel model with macro variables which display reasonably good predictive ability under the recursive scheme but thoroughly disappointing performance under the rolling-window scheme.

There is scope for further research. It is worth investigating how alternative empirically successful forecasting models such as the no-arbitrage Nelson-Siegel model of Christensen et al. (2011), the factor-augmented VAR model of Moench (2008) and the structural model of Hördahl et al. (2006) fare in predicting the UK yield curves. Evidence of substantial time variation in forecasting performance of models and failure of a single model to consistently produce accurate forecasts imply that forecast combinations across models may potentially improve forecasts of the term structure of UK interest rates. This is, however, beyond the scope of this study.

2.A Appendix A: Figures

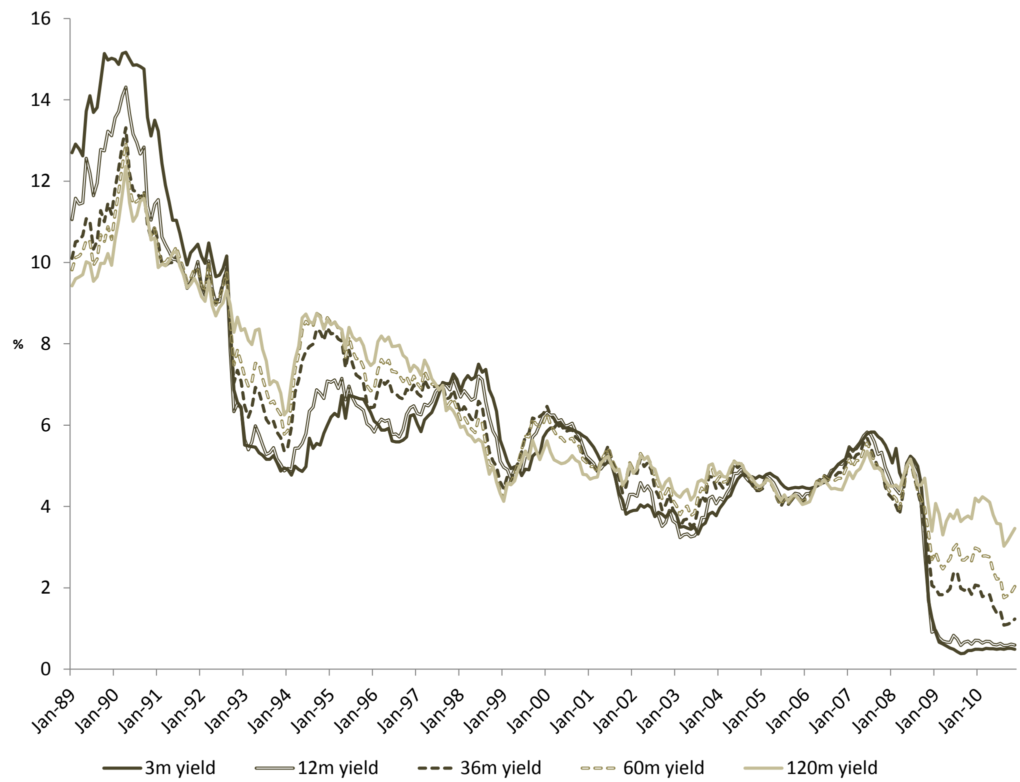
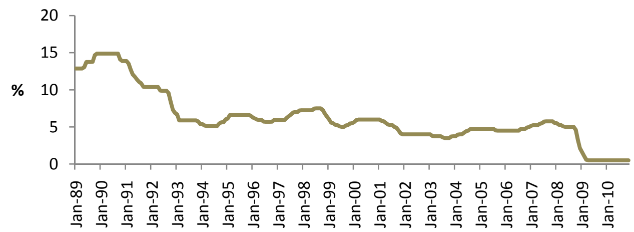
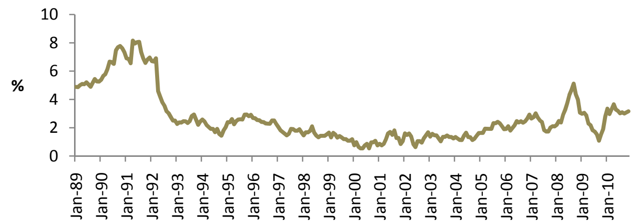


Figure 2.A.1: The UK yield curves

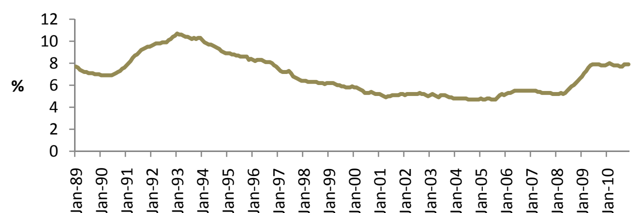
Note: Time series plots of the end-month zero-coupon bond yields for the UK. The sample period is January 1989 - November 2010 and selected maturities are 3, 12, 36, 60 and 120 months.



(a) Official Bank Rate



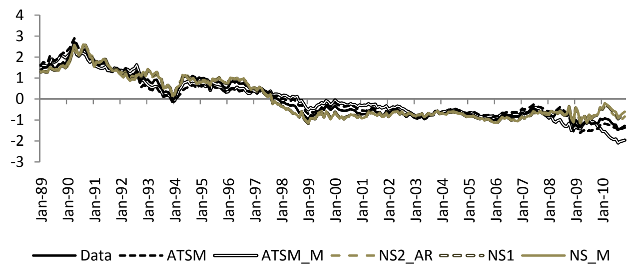
(b) CPI Inflation



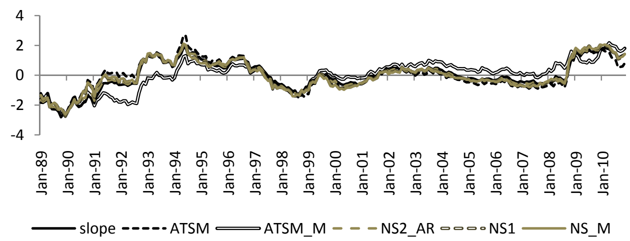
(c) Unemployment Rate

Figure 2.A.2: Time series plots of selected UK macroeconomic variables

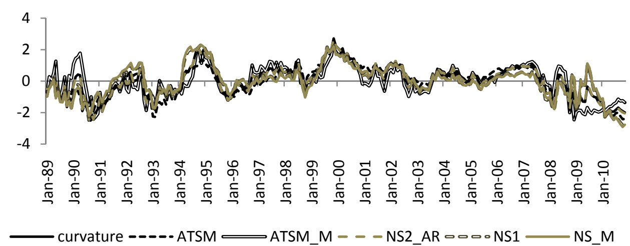
Note: The figure shows time series plots of the official bank rate, CPI inflation and unemployment rate, each expressed in percentages. The sample period is January 1989 - November 2010.



(a) Level



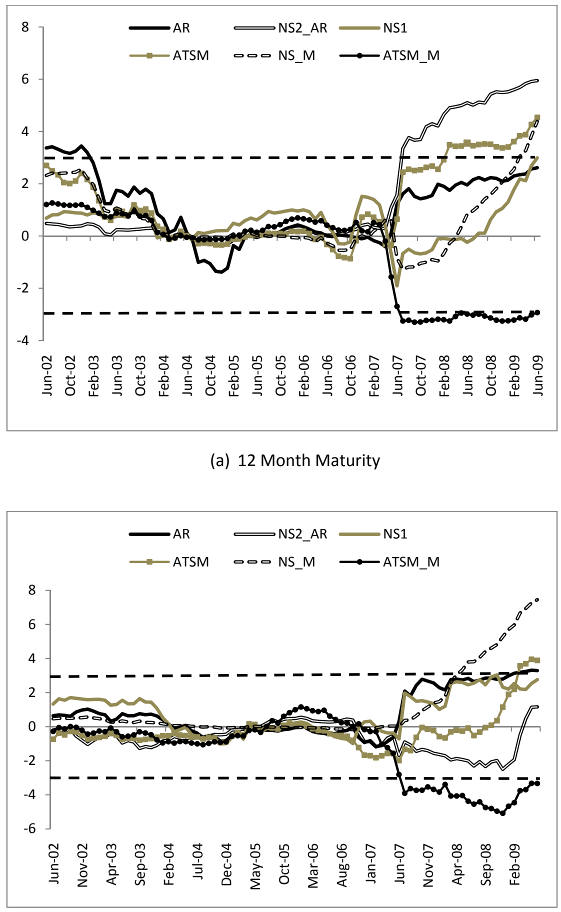
(b) Slope



(c) Curvature

Figure 2.A.3: Data-based vs model-based level, slope and curvature

Note: The data-based level, slope and curvature are defined as the 10-year yield, the difference between the 10-year and 3-month yield and twice the 2-year yield minus the sum of 3-month and 10-year yield, respectively. Data and model-based factors are all standardised for convenience of comparison.

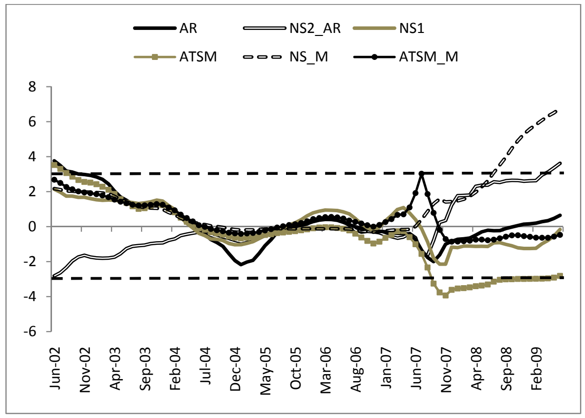


(a) 12 Month Maturity

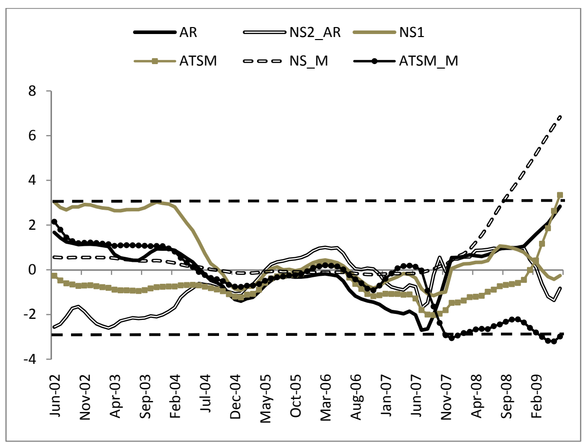
(b) 60 Month Maturity

Figure 2.A.4: Time path of fluctuation test statistic of Giacomini and Rossi (2010) (1-month-ahead forecasts)

Note: The fluctuation test statistic is calculated as the standardised difference between the MSFEs of a competing model and the random walk benchmark. A negative value of the statistic implies that the corresponding model is better than random walk. The horizontal dashed lines indicate two-sided critical values of the statistic.



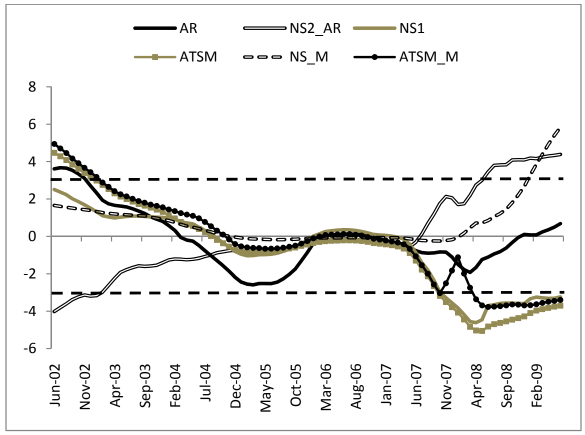
(a) 12 Month Maturity



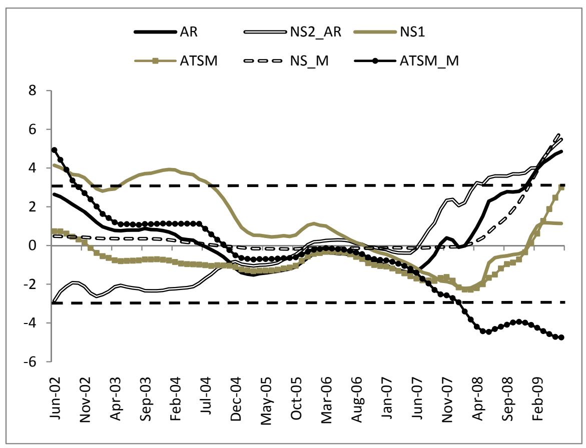
(b) 60 Month Maturity

Figure 2.A.5: Time path of fluctuation test statistic of Giacomini and Rossi (2010) (6-month-ahead forecasts)

Note: The fluctuation test statistic is calculated as the standardised difference between the MSFEs of a competing model and the random walk benchmark. A negative value of the statistic implies that the corresponding model is better than random walk. The horizontal dashed lines indicate two-sided critical values of the statistic.



(a) 12 Month Maturity



(a) 60 Month Maturity

Figure 2.A.6: Time path of fluctuation test statistic of Giacomini and Rossi (2010) (12-month-ahead forecasts)

Note: The fluctuation test statistic is calculated as the standardised difference between the MSFEs of a competing model and the random walk benchmark. A negative value of the statistic implies that the corresponding model is better than random walk. The horizontal dashed lines indicate two-sided critical values of the statistic.

2.B Appendix B: Tables

Table 2.B.1: Descriptive statistics, the UK yield curves

Maturity (months)	Mean	Median	Std dev	Skew	Kurt	Min	Max	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
3	6.064	5.420	3.442	1.027	1.060	0.380	15.170	0.984	0.687	0.218
12	5.884	5.436	2.986	0.717	0.751	0.579	14.311	0.982	0.684	0.258
15	5.897	5.445	2.910	0.668	0.645	0.609	14.158	0.982	0.688	0.277
18	5.918	5.444	2.842	0.628	0.534	0.652	14.005	0.981	0.692	0.298
21	5.943	5.477	2.782	0.599	0.427	0.707	13.859	0.981	0.697	0.318
24	5.969	5.473	2.729	0.578	0.328	0.771	13.725	0.981	0.703	0.337
30	6.019	5.515	2.643	0.555	0.152	0.918	13.494	0.981	0.714	0.370
36	6.065	5.566	2.577	0.551	0.009	1.080	13.315	0.981	0.725	0.397
48	6.141	5.595	2.484	0.574	-0.200	1.422	13.075	0.982	0.747	0.438
60	6.199	5.529	2.421	0.613	-0.340	1.757	12.932	0.983	0.767	0.468
72	6.242	5.418	2.374	0.650	-0.441	2.070	12.828	0.983	0.783	0.491
84	6.272	5.350	2.334	0.680	-0.521	2.353	12.733	0.984	0.797	0.510
96	6.292	5.280	2.298	0.700	-0.593	2.606	12.629	0.985	0.809	0.525
108	6.302	5.216	2.263	0.710	-0.664	2.828	12.508	0.985	0.819	0.539
120	6.306	5.153	2.229	0.711	-0.736	3.022	12.368	0.986	0.827	0.550

Notes: The table reports summary statistics for the end-month UK nominal zero-coupon bond yields. The yields are annualised and expressed in percentages. The sample period is January 1989 - November 2010. The last three columns report sample autocorrelations at displacements of 1, 12 and 30 months.

Table 2.B.2: In-sample fit of models (RMSE of yield curve residuals)

Maturity (months)	NS2_AR	NS1	NS_M	ATSM	ATSM_M
3	0.101	0.483	0.483	0.066	0.064
12	0.105	0.051	0.051	0.085	0.072
15	0.076	0.017	0.017	0.057	0.051
18	0.049	0.000	0.000	0.035	0.035
21	0.028	0.006	0.006	0.024	0.023
24	0.019	0.007	0.007	0.026	0.018
30	0.039	0.000	0.000	0.039	0.026
36	0.057	0.010	0.010	0.049	0.038
48	0.072	0.025	0.025	0.055	0.049
60	0.069	0.029	0.029	0.051	0.047
72	0.054	0.020	0.020	0.041	0.036
84	0.031	0.000	0.000	0.028	0.021
96	0.016	0.030	0.030	0.021	0.014
108	0.049	0.067	0.067	0.037	0.031
120	0.093	0.110	0.110	0.065	0.056

Notes: The table reports root mean squared errors of model fitted residuals expressed in percentages. The NS2_AR refers to a Nelson-Siegel yields-only model with an AR(1) factor dynamics estimated in two steps, the NS1 refers to a Nelson-Siegel yields-only model with a VAR(1) dynamics estimated in one step, the NS_M refers to a Nelson-Siegel yields-macro model, the ATSM refers to a no-arbitrage yields-only affine model and ATSM_M refers to a no-arbitrage macro-augmented affine model.

Table 2.B.3 Relative root mean squared forecast errors (RRMSFEs) with random walk forecasts as the benchmark, Jan 2001-Nov 2010

	1-month horizon					6-month horizon					12-month horizon				
Maturity	3m	12m	36m	60m	120m	3m	12m	36m	60m	120m	3m	12m	36m	60m	120m
(a) Recursive															
AR	0.994	1.001	1.002	1.002	1.002	0.981	**0.993**	1.005	1.005	1.004	0.967	0.987	1.013	1.013	1.004
NS2_AR	0.971	1.144	1.030	1.066	1.029	0.993	1.029	1.051	1.079	1.036	**0.965**	0.997	1.025	1.034	1.022
NS2_VAR	0.945	1.052	**0.992**	1.003	1.039	**0.974**	1.028	1.094	1.102	1.033	1.000	1.136	1.265	1.222	1.087
NS1	1.596	**0.996**	1.003	1.001	1.094	1.065	1.001	1.009	1.007	1.055	0.968	**0.981**	1.027	1.029	1.168
ATSM	**0.920**	1.079	1.025	1.046	1.017	0.982	1.035	1.083	1.062	**0.954**	0.971	1.073	1.147	1.099	1.005
NS_M	1.615	1.012	1.017	1.014	1.094	1.181	1.090	1.081	1.055	1.067	1.129	1.119	1.138	1.105	1.196
ATSM_M	0.921	1.016	1.138	1.077	1.076	0.987	1.024	1.103	1.095	0.984	1.155	1.144	1.184	1.187	1.305
(b) Rolling															
AR	1.015	1.027	1.021	1.020	1.010	1.010	1.021	1.018	1.029	1.033	0.992	1.014	1.045	1.071	1.068
NS2_AR	0.961	1.119	1.012	1.012	**0.970**	0.980	1.002	**0.982**	**0.962**	**0.872****	0.955*	0.995	1.032	1.032	0.994
NS2_VAR	0.907	1.096	0.999	1.012	1.009	0.957	1.017	1.037	1.035	0.991	**0.882**	0.981	1.107	1.146	1.216
NS1	1.574	1.056	1.051	1.041	1.054	1.042	1.028	1.060	1.062	0.999	0.894	**0.968**	1.115	1.176	1.246
ATSM	**0.862**	1.082	**0.991**	1.029	1.020	**0.952**	1.001	1.025	1.035	0.939	0.923	0.997	1.053	1.043	**0.946**
NS_M	1.414	1.210	1.617	1.659	1.213	1.288	1.613	2.178	2.329	1.961	1.492	1.781	2.299	2.387	2.102
ATSM_M	0.934	**0.952**	**0.984**	**0.949**	1.191	1.085	1.072	1.014	0.964	1.146	1.017	1.026	1.007	**0.976**	1.327

Notes: A value of RRMSFE below one indicates better forecast relative to the random walk. Bold entries along each maturity highlight the minimum relative RMSFEs (best forecasts). '*' and '**' mark gains that are significant at 10% and 5% according to Giacomini and White (2006) test. The test is applied only to rolling-window forecasts.

Table 2.B.4 Relative root mean squared forecast errors (RRMSFEs) with random walk forecasts as the benchmark, Jan 2001-Dec 2007

	1-month horizon					6-month horizon					12-month horizon				
Maturity	3m	12m	36m	60m	120m	3m	12m	36m	60m	120m	3m	12m	36m	60m	120m
Recursive															
AR	0.998	**0.999**	1.002	1.002	1.004	0.999	0.994	1.008	1.011	1.017	0.992	0.987	1.018	1.023	1.027
NS2_AR	0.979	1.042	1.013	1.010	1.043	**0.979**	**0.964**	**0.940**	**0.961**	**0.978**	**0.926**	**0.861**	**0.825**	**0.856**	0.947
NS2_VAR	1.101	1.019	1.009	0.995	1.047	1.089	1.039	0.994	0.981	0.994	1.031	1.011	0.950	0.901	**0.923**
NS1	2.206	1.020	**0.999**	0.991	1.073	1.281	1.096	1.020	1.018	1.059	1.322	1.177	1.060	1.052	1.177
ATSM	0.932	1.05	1.016	1.004	1.019	1.029	1.109	1.040	1.008	0.983	1.147	1.174	1.069	1.013	1.025
NS_M	1.958	1.042	1.004	**0.989**	1.077	1.406	1.288	1.112	1.040	1.076	1.621	1.475	1.192	1.065	1.157
ATSM_M	0.920	1.019	1.090	1.016	1.032	1.467	1.237	1.038	1.005	1.016	1.813	1.580	1.319	1.293	1.481
Rolling															
AR	1.063	1.038	1.010	1.003	**0.997**	1.059	1.074	1.027	1.007	0.985	0.970	1.034	1.039	1.018	0.984
NS2_AR	0.917	1.025	1.006	0.998	1.047	**0.809***	**0.872***	**0.903***	**0.927**	**0.957**	**0.738***	**0.748****	**0.794****	**0.841****	0.948
NS2_VAR	0.984	1.051	1.004	0.997	1.070	0.990	1.082	1.077	1.090	1.149	1.032	1.109	1.136	1.170	1.359
NS1	2.004	1.035	1.020	1.017	1.081	1.107	1.125	1.130	1.160	1.180	1.102	1.153	1.222	1.292	1.460
ATSM	**0.884***	1.069	**0.990**	**0.977**	1.037	1.083	1.126	0.985	0.959	0.993	1.215	1.186	1.010	0.961	1.026
NS_M	1.903	1.172	1.106	1.069	1.105	1.542	1.601	1.387	1.258	1.065	1.671	1.649	1.388	1.188	**0.934**
ATSM_M	1.267	1.130	1.045	0.997	1.099	1.412	1.338	1.155	1.104	1.174	1.467	1.397	1.274	1.254	1.499

Notes: A value of RRMSFE below one indicates better forecast relative to the random walk. Bold entries along each maturity highlight the minimum relative RMSFEs (best forecasts). '*' and '**' mark gains that are significant at 10% and 5% according to Giacomini and White (2006) test. The test is applied only to rolling-window forecasts.

Table 2.B.5 Relative root mean squared forecast errors (RRMSFEs) with random walk forecasts as the benchmark, Jan 2008-Nov 2010

	1-month horizon					6-month horizon					12-month horizon				
Maturity	3m	12m	36m	60m	120m	3m	12m	36m	60m	120m	3m	12m	36m	60m	120m
Recursive															
AR	0.993	1.001	1.003	1.002	1.000	0.977	0.993	1.003	1.002	0.993	0.961	0.986	1.010	1.005	**0.973**
NS2_AR	0.970	1.205	1.048	1.128	1.017	0.996	1.047	1.109	1.154	1.083	0.975	1.038	1.134	1.162	1.110
NS2_VAR	**0.904**	1.073	**0.973**	1.012	1.033	0.948	1.024	1.147	1.179	1.066	0.993	1.174	1.430	1.436	1.263
NS1	1.414	**0.980**	1.007	1.012	1.110	1.013	0.971	1.002	1.000	1.050	**0.856**	**0.907**	1.006	1.010	1.157
ATSM	0.917	1.098	1.035	1.093	1.016	0.972	1.011	1.107	1.097	**0.928**	0.921	1.037	1.195	1.167	0.979
NS_M	1.523	0.993	1.032	1.043	1.108	1.127	1.023	1.063	1.066	1.059	0.965	0.972	1.101	1.136	1.243
ATSM_M	0.922	1.014	1.192	1.145	1.109	**0.851**	**0.952**	1.138	1.154	0.955	0.915	0.956	1.089	1.092	1.042
Rolling															
AR	1.003	1.020	1.033	1.039	1.020	1.000	1.006	1.012	1.044	1.072	0.998	1.007	1.049	1.113	1.165
NS2_AR	0.971	1.176	1.018	1.028	**0.907**	1.013	1.038	1.025	0.985	**0.790**	1.003	1.064	1.159	1.167	1.049
NS2_VAR	0.888	1.124	0.993	1.029	0.960	0.950	0.997	1.014	0.995	0.830	0.839	0.935	1.087	1.125	1.010
NS1	1.454	1.068	1.086	1.069	1.033	1.027	0.997	1.018	0.989	0.810	0.834	0.897	1.040	1.070	0.912
ATSM	0.856	1.090	0.992	1.087	1.006	**0.921**	**0.961**	1.047	1.084	0.890	**0.833**	0.925	1.080	1.107	0.835
NS_M	1.271	1.234	2.056	2.157	1.291	1.227	1.616	2.518	2.844	2.490	1.443	1.823	2.728	3.045	2.977
ATSM_M	**0.837**	**0.817**	**0.910**	**0.889**	1.258	1.003	0.980	**0.925**	**0.855**	1.122	0.866	**0.869**	**0.789****	**0.659****	1.072

Notes: A value of RRMSFE below one indicates better forecast relative to the random walk. Bold entries along each maturity highlight the minimum relative RMSFEs (best forecasts). '*' and '**' mark gains that are significant at 10% and 5% according to Giacomini and White (2006) test. The test is applied only to rolling-window forecasts.

Chapter 3

Density Forecasts of Bond Yields: Evaluating Arbitrage-free Nelson-Siegel Models with Stochastic Volatilities

3.1 Introduction

We continue empirical research on yield curve forecasting in Chapter 3. While Chapter 2 assesses the predictive performance of the term structure models in terms of point forecasts of bond yields this chapter focuses on evaluating models on the basis of density forecasts which provide a full distribution of the future predicted yields.

For this work we exploit several variants of a well-known class of yield curve models - the Nelson-Siegel models. These models enjoy a number of attractive properties which made them very popular among researchers, both in academia and central banks. The Nelson-Siegel model is parsimonious in terms of parameters to be estimated, they are easy to estimate and tract and they are flexible enough to capture many possible shapes of an yield curve. Since its development by Nelson and Siegel (1987), the original model has gone through significant refinements. Diebold and Li (2006) made the original static model dynamic in order to capture evolution of bond yields over time. Christensen et al. (2011) proposed an arbitrage-free version of the model and made it theoretically more sound and competitive. Christensen et al. (2010) went one step further by incorporating stochastic volatility to model dynamics of interest rate fluctuations. The main objective of this chapter is to compare different specifications of the Nelson-Siegel models, simple to complex, in terms of their ability to generate forecasts of bond yields out-of-sample. Particular emphasis has been given to evaluation of density forecasts of the yield curve, as opposed to point forecasts.

Why should we care about density forecasts of bond yields? A natural motivation comes from the importance of forecasting predictive density in general. A point forecast provides a single future value for a variable of interest and is easy to compute and interpret. However, a criticism of it is that it does not take into account the uncertainty surrounding the prediction. Density forecasts provide a detailed description of such uncertainties as they are essentially estimates of the complete probability distribution of the possible future values of the variable of interest. This is particularly helpful for policy makers who can incorporate forecast uncertainties in their policy decisions. It has a become a common practice for most of the central banks around the world (e.g. Bank of England, Bank of Canada, Norges Bank etc.) to routinely issue predictive distributions for many key economic indicators such as inflation, GDP, policy rate etc. Density forecasts has found even more frequent applications in finance where risk or uncertainty plays a crucial role. The most prominent use is in rapidly growing financial risk management industry where a full account of the predictive density of future portfolio returns helps to track certain features of the distribution such as value-at-risk which are often used as measures of risk exposure. J.P. Morgan, Reuters, Bloomberg routinely publish density forecasts of key measures of portfolio risk.

The density forecast of the term structure of interest rates has received little attention in finance literature. To our knowledge the only contributions came from Hong and Li (2005) and Egorov et al. (2006) who proposed non-parametric tests for evaluating density forecasts and applied them to compare a number of affine term structure models (ATSMs) when forecasting the joint conditional probability density of bond yields. They found unsatisfactory density forecasts from ATSMs in continuous time which is reminiscent of Duffee (2002) who found similar disappointing performance of discrete-time ATSMs in terms of point forecasts using the US yields. Performance of the ATSM in fitting conditional volatility of yields is not good enough either (Collin-Dufresne et al. (2009)).[1] Moreover, it is well-documented in literature that in general, it is difficult to estimate ATSMs, particularly its prices of risk parameters (see Duffee (2011)). Such empirical failure of ATSM prompts search for an alternative and more competitive model. The result is the recent addition of arbitrage-free Nelson-Siegel models with stochastic volatilities proposed by Christensen et al. (2010). The models combine several attractive properties which are important from both theoretical and empirical perspectives. For example they account for time variation in yield curve volatility and are, therefore, more flexible than the constant volatility Nelson-Siegel models of Diebold and Li (2006) and Christensen et al. (2011). There are existing Nelson-Siegel

[1] Jacobs and Karoui (2009), however, argue that ability of affine term sturcture models to capture conditional volatility of the yield curve is sensitive to choice of sample periods.

models which account for time-varying volatility in bond yields. Koopman et al. (2010) find improved fit by adding a common GARCH-type volatility factor that drives volatility of the entire cross-section of the yield curve. Hautsch and Yang (2012) allow the level, slope and curvature factors to induce stochastic volatility in the model by introducing three additional state variables. They also confirm benefits from including stochastic volatility in terms of better in-sample performance and reduced forecast uncertainty. However, none of these models are arbitrage-free by nature. The stochastic-volatility no-arbitrage models of Christensen et al. (2010) have an advantage over these models in that they are corrected for risk-free arbitrage opportunities and are, therefore, theoretically more sound. Using daily US and UK interest rates Christensen et al. (2010) show that the models provide good in-sample fit and can explain substantial proportion of stochastic volatility observed in the data. However, to our knowledge forecasting performance of these models is not tested. This chapter systematically evaluates forecasting performance of these models, particularly their ability in calibrating predictive densities of bond yield out-of-sample and investigates to what extent they satisfy the quest for a model that produce good density forecasts of interest rates. The questions which are of particular interest are whether enforcing no-arbitrage restriction and/or modelling time-varying volatility improve predictive performance of Nelson-Siegel models.[2] We employ various metrics used in literature such as probability integral transforms, coverage rates, log predictive density scores in order to assess the quality of density forecasts.

The chapter is structured as follows: section 3.2 provides a detailed description of the competing Nelson-Siegel models with and without stochastic

[2] There is, however, considerable amount of debate in the literature about whether no-arbitrage restrictions help in forecasting yields. Duffee (2011) and Joslin et al. (2011) argue that predictions of pricing factors are independent of imposition of cross-sectional restrictions. Nonetheless, Christensen et. al (2011) find improvements in forecasts of US bond yields from making Nelson-Siegel models arbitrage-free.

volatilities and their estimation procedure. Section 3.3 describes the yield data used for empirical analysis and evaluates the models' ability to fit the data in-sample. Section 3.4 provides detailed comparison of models in terms of both point and density forecasts of bond yields. Section 3.5 concludes.

3.2 Forecasting Models

In this section we describe different specifications of the Nelson-Siegel term structure models which will be evaluated on the basis of their ability to forecast bond yields. The models differ in terms of two basic features: whether they adopt no-arbitrage restrictions or not and whether volatility of yields is assumed constant or modelled as time-varying. As explained in the introduction, we follow Christensen et al. (2010) to explicitly consider time-varying volatilities that are generated only by latent factors that are extracted from cross-section of the yield curve. The stochastic volatility specifications differ depending on how many of the factors drive the volatility.

3.2.1 Standard Dynamic Nelson-Siegel Model

The model has been already introduced in Section 2.2 of the previous chapter. It is a very popular dynamic term structure model and is a result of temporal extension by Diebold and Li (2006) of the original static model of Nelson and Siegel (1987). We denote the three time-varying latent factors, level, slope and curvature by L_t, S_t and C_t, respectively. The yields are then expressed in a dynamic exponential factor-model framework:

$$y_t(m) = L_t + S_t \frac{1 - e^{-\lambda m}}{\lambda m} + C_t \left(\frac{1 - e^{-\lambda m}}{\lambda m} - e^{-\lambda m} \right). \tag{3.2.1}$$

The DNS model has been widely applied and analysed for modelling yields and its success in forecasting has been well-documented in the literature. Diebold and Li (2006) report its superior predictive ability by assuming an AR(1) factor dynamics and using U.S. government bond yields. Despite its good empirical behaviour, the DNS model has been heavily criticised for not imposing no-arbitrage restrictions which lie in the heart of financial theory of asset pricing. This simple model assumes that the implied volatility in bond yields is constant or time-invariant. We use the DNS as the benchmark and compare its forecast performance against more sophisticated Nelson-Siegel counterparts which are arbitrage-free and/or modelled with time-varying volatility.

3.2.2 Arbitrage-free Nelson-Siegel with Constant Volatility

Motivated by the importance of making the DNS model more consistent with finance theory, Christensen et al. (2011) propose an arbitrage-free version of the dynamic Nelson-Siegel model with constant volatility (denoted hereby as $AFNS_0$). The $AFNS_0$ model is derived in a continuous-time affine diffusion environment as described in Duffie and Kan (1996). Following Christensen et al. (2010) we first define the general affine process which encompasses the $AFNS_0$ and also the arbitrage-free Nelson-Siegel models with stochastic volatilities used in this study.

Let X be an N-dimensional vector of state variables. Then under no-arbitrage affine framework X_t follows a Markovian diffusion process and its dynamics under risk neutral measure Q can be defined by the following stochastic differential equation (SDE):

$$dX_t = K^Q(\theta^Q - X_t)dt + \Sigma S(X_t)dZ_t^Q, \tag{3.2.2}$$

where Z^Q is a vector of N independent standard Brownian motions, $\theta^Q \in$

$\mathbf{R}^N$ is the mean vector of the process, $K^Q \in \mathbf{R}^{N\times N}$ determines the speed of mean reversion and $\Sigma \in \mathbf{R}^{N\times N}$ is the constant component of the volatility of the process. The state-dependent $(N \times N)$ diagonal matrix $S(X_t)$ introduces conditional heteroskedasticity or time-varying volatility in the bond yields and its i-th diagonal element given by

$$\sqrt{\alpha^i + \beta_1^i X_t^1 + ... + \beta_n^i X_t^N}.$$

The short rate (instantaneous risk-free rate) is modelled as an affine function of the underlying state variables

$$r_t = \delta_0 + {\delta_1}' X_t,$$

where $\delta_0 \in \mathbf{R}$ and $\delta_1 \in \mathbf{R}^n$ are bounded and continuous functions.

If we define $P_t(m)$ as the time t price of a \$1 zero-coupon bond which is maturing at time $t + m$, then $P_t(m)$ can be expressed as exponential affine functions of the state variables (Duffie and Kan (1996)):

$$P_t(m) = \exp(B(m)' X_t + A(m)),$$

where the pricing coefficients $B(m)$ and $A(m)$ solve the following system of Ricatti ordinary differential equations (ODEs):

$$\frac{dB(m)}{dm} = -\delta_1 - (K^Q)' B(m) + \frac{1}{2} \sum_{j=1}^{N} (\Sigma' B(m) B(m)' \Sigma)_{j,j} (\beta^j)', \quad B(0) = 0, \tag{3.2.3}$$

$$\frac{dA(m)}{dm} = -\delta_0 + B(m)' K^Q \theta^Q + \frac{1}{2} \sum_{j=1}^{N} (\Sigma' B(m) B(m)' \Sigma)_{j,j} \alpha^j, \quad A(0) = 0. \tag{3.2.4}$$

Functional relationship between the yield and the price of an m-period zero-coupon bond then implies that the expression for the yield reduces to an affine functions of X_t:

$$y_t(m) = -\frac{1}{m} \log P(X_t, m) = -\frac{B(m)'}{m} X_t - \frac{A(m)}{m}. \tag{3.2.5}$$

Finally, the model is fully described with a risk price specification which translates its risk-free dynamics to dynamics under real world (or historic) P-measure. Following Christensen et al. (2010) we use the extended affine risk premium specification of Cheridito et al. (2007) for all the models except for one where we are restricted to use the essentially affine risk premium structure of Duffee (2002). The risk premium Γ_t in extended specification is given by

$$\Gamma_t = S^{-1}(X_t)\gamma^0 + S^{-1}(X_t)\gamma^1 X_t,$$

and the one in essentially affine specification is given by

$$\Gamma_t = S(X_t)\gamma^0 + S^{-1}(X_t)\gamma^1 X_t,$$

where $\gamma^0 \in \mathbf{R}^N$and $\gamma^1 \in \mathbf{R}^{NxN}$contain unrestricted parameters. The relationship between real-world yield curve dynamics under the P-measure and risk-neutral dynamics under the Q-measure is given by

$$dZ_t^Q = dZ_t^P + \Gamma_t dt. \tag{3.2.6}$$

The dynamics of the state vector under the P-measure is obtained by subtracting the term $\Sigma S(X_t)\Gamma_t$ from the SDE of Q-dynamics in (3.2.2) and replacing dZ_t^Q by dZ_t^P. A general expression for the P-dynamics can then be given by

$$dX_t = K^P(\theta^P - X_t)dt + \Sigma S(X_t)dZ_t^P \tag{3.2.7}$$

In the constant volatility AFNS model ($AFNS_0$) of Christensen et al. (2011), the instantaneous risk-free rate is explicitly defined as the sum of the first two latent factors:

$$r_t = X_t^1 + X_t^2. \tag{3.2.8}$$

The vector of state variables $X_t = (X_t^1, X_t^2, X_t^3)$ follows a Gaussian (as opposed to square-root) diffusion process in the sense that there is no conditional heteroskedasticity in the volatility of the yield factors and the volatility is driven by a constant Σ.[3] Without loss of generality we can set $S(X_t)$ to an identity matrix and define the Q-dynamics of X_t by the following system of linear SDEs:

$$dX_t = K^Q(\theta^Q - X_t)dt + \Sigma dZ_t^Q.$$

Christensen et al. (2011) have showed that if the mean reversion matrix K^Q has the following particular specification

$$K^Q = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix}, \tag{3.2.9}$$

then the recursive bond pricing coefficients in (3.2.3) are solved as

$$\begin{aligned} B^1(m) &= -m \\ B^2(m) &= -\left(\frac{1-e^{-\lambda m}}{\lambda}\right) \\ B^3(m) &= me^{-\lambda m} - \left(\frac{1-e^{-\lambda m}}{\lambda}\right) \end{aligned} \tag{3.2.10}$$

and the yields-factors relationship can be written as

$$y_t(m) = X_t^1 + \left(\frac{1-e^{-\lambda m}}{\lambda m}\right) X_t^2 + \left(\frac{1-e^{-\lambda m}}{\lambda m} - e^{-\lambda m}\right) X_t^3 - \frac{A(m)}{m},$$

which preserves the same loadings as the original model of Nelson and Siegel (1987) for level, slope and curvature. There is, however, an additional yield-adjustment term which is time invariant and is a function of maturity only.

The extended affine risk premium in the Gaussian framework implies

$$\Gamma_t = \gamma^0 + \gamma^1 X_t$$

[3] See Piazzesi (2010) for the distinction between a Gaussian and a square-root process.

which together with measure change equation (3.2.6) imply that the P-dynamics of the state vector X_t is given by

$$dX_t = K^P(\theta^P - X_t)dt + \Sigma dZ_t^P.$$

Christensen et al. (2011) further showed that these models can only be identified if the elements of θ^Q, the mean vector under the Q-measure, are zero and the volatility matrix Σ is no more than a triangular matrix. Under the assumption of independence of yield factors Σ is diagonal and the P-dynamics takes the form:

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = \begin{pmatrix} \kappa_{11}^P & 0 & 0 \\ 0 & \kappa_{22}^P & 0 \\ 0 & 0 & \kappa_{33}^P \end{pmatrix} \left[\begin{pmatrix} \theta_1^P \\ \theta_2^P \\ \theta_3^P \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \right] dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} \begin{pmatrix} dZ_t^{1,P} \\ dZ_t^{2,P} \\ dZ_t^{3,P} \end{pmatrix}.$$

The arbitrage-free version of Nelson-Siegel model $AFNS_0$ has been reported to provide better yield forecasts than the already successful dynamic Nelson-Siegel model DNS of Diebold and Li (2006) which does not correct for arbitrage opportunities (Christensen et al. (2011)).

3.2.3 AFNS with Stochastic Volatility

Time varying volatility is a key feature of bond yield data. Therefore, assuming constant volatility, as considered in the DNS and $AFNS_0$ models, limits flexibility of yield curve models. Christensen et al. (2010) extend $AFNS_0$ to incorporate stochastic volatility where the volatility is spanned entirely by latent yield curve factors. The principal objective of this paper is to investigate whether incorporating stochastic volatility can further improve predictive power of the AFNS models. We review all the models described in Christensen et al. (2010). Throughout all the models the short rate is modelled as in equation (3.2.8) and the particular structure of k^Qmean reversion matrix in (3.2.9) is maintained to match closely the desirable Nelson-Siegel factor loading structure in the zero-coupon bond yield function. The models mainly differ in terms of how many and which of the factors drive stochastic volatility. Consequently, they have different specifications for the state-dependent stochastic volatility inducing matrix, $S(X_t)$ which affects dynamics of factors under the risk-neutral and physical measures. We define parsimonious versions of the models where the three factors move independently of each other.

AFNS with One Stochastic Volatility Factor

There are two feasible AFNS stochastic volatility specifications that allow just one factor to exhibit stochastic volatility - in one the volatility is induced by level (denoted as $AFNS_1\_L$) and in the other by curvature (denoted as $AFNS_1\_C$).[4]

The Q-dynamics of the state vector X_t in a correctly identified $AFNS_1\_L$ requires that in equation (3.2.2) we set

$$S(X_t) = \begin{pmatrix} \sqrt{X_t^1} & 0 & 0 \\ 0 & \sqrt{1+\beta_{21}X_t^1} & 0 \\ 0 & 0 & \sqrt{1+\beta_{31}X_t^1} \end{pmatrix}, \ \theta_2^Q = \theta_3^Q = 0 \text{ and}$$

[4] Christensen et al. (2010) show that a model with slope as a stochastic volatility factor is not admissible because of the particular stucture of K^Q in (3.2.9).

$k_{11}^{Q} = \varepsilon$, where ε is a small positive number. As is observable, volatility in slope and curvature is influenced by the level factor X_t^1 and the volatility sensitivity parameter $\beta_{21}$and β_{31} measure the extent of such influence. The level factor follows a square-root process.

For the factor loadings in the zero-coupon bond prices, $B^1(m)$ is the solution to

$$\begin{aligned}\frac{dB^1(m)}{dm} &= -1 - \varepsilon B^1(m) + \frac{1}{2}\sigma_{11}^2 (B^1(m))^2 + \frac{1}{2}\beta_{21}\sigma_{22}^2 (B^2(m))^2 \\ &\quad + \frac{1}{2}\beta_{31}\sigma_{33}^2 (B^3(m))^2,\end{aligned}$$

while $B^2(m)$and $B^3(m)$ keep the original Nelson-Siegel expressions as in (3.2.10).

The yield-adjustment term, $A(m)$ solves the following ODE:

$$\frac{dA(m)}{dm} = B(m)^{'} K^Q \theta^Q + \frac{1}{2}\sigma_{22}^2 B^2(m)^2 + \frac{1}{2}\sigma_{33}^2 B^3(m)^2.$$

For the $AFNS_1\_L$ model the extended affine risk premium is not viable (see Christensen et al. (2010)) and we adopt the essentially affine risk premium structure of Duffee (2002). The P-dynamics for the independent-factor specification is given by

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = \begin{pmatrix} \kappa_{11}^P & 0 & 0 \\ 0 & \kappa_{22}^P & 0 \\ 0 & 0 & \kappa_{33}^P \end{pmatrix} \left[\begin{pmatrix} \theta_1^P \\ \theta_2^P \\ \theta_3^P \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \right] dt$$

$$+ \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} \begin{pmatrix} \sqrt{X_t^1} & 0 & 0 \\ 0 & \sqrt{1+\beta_{21}X_t^1} & 0 \\ 0 & 0 & \sqrt{1+\beta_{31}X_t^1} \end{pmatrix} \begin{pmatrix} dZ_t^{1,P} \\ dZ_t^{2,P} \\ dZ_t^{3,P} \end{pmatrix}.$$

Following Christensen et al. (2010) we implement a required restriction on the mean parameter θ_1^P which is $\theta_1^P = \frac{\varepsilon.\theta_1^Q}{\kappa_{11}^P}$.

The model is completed by specifying the Feller conditions which ensure that the square-root process X_t^1 never attains zero. The conditions are implemented as $\kappa_{11}^P \theta_1^P > 0$ and $\varepsilon.\theta_1^Q > 0$.

The Q-dynamics of the state vector X_t in a correctly identified $AFNS_1\_C$ model (where curvature is the sole driver of stochastic volatility) requires that in equation (3.2.2) we set

$$S(X_t) = \begin{pmatrix} \sqrt{1+\beta_{13}X_t^3} & 0 & 0 \\ 0 & \sqrt{1+\beta_{23}X_t^3} & 0 \\ 0 & 0 & \sqrt{X_t^3} \end{pmatrix} \text{ and } \theta_1^Q = \theta_2^Q = 0.$$

In this model class the first two factor loadings $B^1(m)$ and $B^2(m)$ are identical to those in (3.2.10), while $B^3(m)$ is the solution to:

$$\begin{aligned} \frac{dB^3(m)}{dm} &= \lambda B^2(m) - \lambda B^3(m) + \frac{1}{2}\sigma_{33}^2(B^3(m))^2 + \frac{1}{2}\beta_{13}\sigma_{11}^2(B^1(m))^2 \\ &\quad + \frac{1}{2}\beta_{23}\sigma_{22}^2(B^2(m))^2. \end{aligned} \tag{3.2.11}$$

The yield-adjustment term, $A(m)$ solves the following ODE:

$$\frac{dA(m)}{dm} = B(m)^{'} K^Q\theta^Q + \frac{1}{2}\sigma_{11}^2(B^1(m))^2 + \frac{1}{2}\sigma_{22}^2(B^2(m))^2. \tag{3.2.12}$$

We estimate this model using the extended affine risk premium specifications and the independent-factor P-dynamics is given by

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = \begin{pmatrix} \kappa_{11}^P & 0 & 0 \\ 0 & \kappa_{22}^P & 0 \\ 0 & 0 & \kappa_{33}^P \end{pmatrix} \left[\begin{pmatrix} \theta_1^P \\ \theta_2^P \\ \theta_3^P \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \right] dt$$

$$+ \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} \begin{pmatrix} \sqrt{1+\beta_{13}X_t^3} & 0 & 0 \\ 0 & \sqrt{1+\beta_{23}X_t^3} & 0 \\ 0 & 0 & \sqrt{X_t^3} \end{pmatrix} \begin{pmatrix} dZ_t^{1,P} \\ dZ_t^{2,P} \\ dZ_t^{3,P} \end{pmatrix}.$$

The Feller condition requires $\kappa_{33}^P\theta_3^P > \frac{1}{2}\sigma_{33}^2$ and $\lambda\theta_3^Q > \frac{1}{2}\sigma_{33}^2$.

AFNS with Two Stochastic Volatility Factors

There are two feasible models under specifications where volatility is dictated by two factors. One is where level and curvature together exhibit stochastic volatility (denoted as $AFNS_2 - LC$) and the other is where slope and curvature together drive stochastic volatility (denoted as $AFNS_2 - SC$).[5] We find difficulty in estimating the $AFNS_2 - SC$ model for the particular data set we are using and therefore, exclude it from our analysis.

The Q-dynamics of a correctly identified $AFNS_2\_LC$ requires that in equation (3.2.2) we set

$$S(X_t) = \begin{pmatrix} \sqrt{X_t^1} & 0 & 0 \\ 0 & \sqrt{1+\beta_{21}X_t^1+\beta_{23}X_t^3} & 0 \\ 0 & 0 & \sqrt{X_t^3} \end{pmatrix}, \ \theta_2^Q = 0 \text{ and } k_{11}^Q = \varepsilon,$$

where ε is a small positive number.

The factor loadings $B^1(m)$ and $B^3(m)$ of the zero-coupon bond price function are unique solutions to the following set of ODEs:

$$\frac{dB^1(m)}{dm} = -1 - \varepsilon B^1(m) + \frac{1}{2}\sigma_{11}^2(B^1(m))^2 + \frac{1}{2}\beta_{21}\sigma_{22}^2(B^2(m))^2,$$

$$\frac{dB^3(m)}{dm} = \lambda B^2(m) - \lambda B^3(m) + \frac{1}{2}\sigma_{33}^2(B^3(m))^2 + \frac{1}{2}\beta_{23}\sigma_{22}^2(B^2(m))^2.$$

$B^2(m)$ remains the same as in (3.2.10). Hence, X_t^2 preserves its role as a slope factor. The $A(m)$-function is the solution to:

$$\frac{dA(m)}{dm} = B(m)^{'} K^Q\theta^Q + \frac{1}{2}\sigma_{22}^2B^2(m)^2.$$

[5] See Christensen et al. (2010) for detailed reasoning of why level and slope together cannot enter the model as drivers of stochastic volatility.

Using the extended affine risk premium structure, the independent-factor P-dynamics is given by

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = \begin{pmatrix} \kappa_{11}^P & 0 & 0 \\ 0 & \kappa_{22}^P & 0 \\ 0 & 0 & \kappa_{33}^P \end{pmatrix} \left[\begin{pmatrix} \theta_1^P \\ \theta_2^P \\ \theta_3^P \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \right] dt$$

$$+ \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} \begin{pmatrix} \sqrt{X_t^1} & 0 & 0 \\ 0 & \sqrt{1 + \beta_{21} X_t^1 + \beta_{23} X_t^3} & 0 \\ 0 & 0 & \sqrt{X_t^3} \end{pmatrix} \begin{pmatrix} dZ_t^{1,P} \\ dZ_t^{2,P} \\ dZ_t^{3,P} \end{pmatrix}.$$

For the level factor, the condition $\varepsilon.\theta_1^Q = \kappa_{11}^P \theta_1^P$ must be satisfied. Feller conditions are given by $\kappa_{33}^P \theta_3^P > \frac{1}{2}\sigma_{33}^2$ and $\lambda\theta_3^Q > \frac{1}{2}\sigma_{33}^2$.

AFNS with Three Stochastic Volatility Factors

In the last specification all three factors exhibit stochastic volatility (denoted by $AFNS_3$). The Q-dynamics of a correctly identified $AFNS_3$ requires that in equation (3.2.2) we set

$$S(X_t) = \begin{pmatrix} \sqrt{X_t^1} & 0 & 0 \\ 0 & \sqrt{X_t^2} & 0 \\ 0 & 0 & \sqrt{X_t^3} \end{pmatrix}$$ and $k_{11}^Q = \varepsilon$, where ε is a small positive number.

In this model class, the factor loadings in the zero-coupon bond price function are given by the unique solutions to

$$\frac{dB^1(m)}{dm} = -1 - \varepsilon B^1(m) + \frac{1}{2}\sigma_{11}^2 (B^1(m))^2,$$

$$\frac{dB^2(m)}{dm} = -1 - \lambda B^2(m) + \frac{1}{2}\sigma_{22}^2 (B^2(m))^2,$$

$$\frac{dB^3(m)}{dm} = \lambda B^2(m) - \lambda B^3(m) + \frac{1}{2}\sigma_{33}^2(B^3(m))^2,$$

and the yield-adjustment term $A(t,T)$ is given by the solution to:

$$\frac{dA(m)}{dm} = B(m)^{'} K^Q \theta^Q.$$

Applying the extended affine risk premium specification, the independent-factor P-dynamics is given by

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = \begin{pmatrix} \kappa_{11}^P & 0 & 0 \\ 0 & \kappa_{22}^P & 0 \\ 0 & 0 & \kappa_{33}^P \end{pmatrix} \left[\begin{pmatrix} \theta_1^P \\ \theta_2^P \\ \theta_3^P \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \right] dt$$
$$+ \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} \begin{pmatrix} \sqrt{X_t^1} & 0 & 0 \\ 0 & \sqrt{X_t^2} & 0 \\ 0 & 0 & \sqrt{X_t^3} \end{pmatrix} \begin{pmatrix} dZ_t^{1,P} \\ dZ_t^{2,P} \\ dZ_t^{3,P} \end{pmatrix}.$$

For X_t^1, the constraint $\varepsilon.\theta_1^Q = \kappa_{11}^P \theta_1^P$ must be satisfied. The Feller conditions which must be satisfied are:

$\kappa_{22}^P \theta_2^P > \frac{1}{2}\sigma_{22}^2$; $\lambda\theta_2^Q - \lambda\theta_3^Q > \frac{1}{2}\sigma_{22}^2$; $\kappa_{33}^P \theta_3^P > \frac{1}{2}\sigma_{33}^2$ and $\lambda\theta_3^Q > \frac{1}{2}\sigma_{33}^2$.

3.2.4 Estimation Framework

All the Nelson-Siegel specifications under this study can be conveniently represented in state-space frameworks. For estimation of such unobserved factor dynamic models we use a standard maximum likelihood technique which use Kalman filter for extraction of latent yields.[6] We start by writing the transition equations for different specifications. For the benchmark DNS model, the

[6] Diebold et al. (2006), Christensen et al. (2011) and Christensen et al. (2010) use such Kalman filter induced maximum likelihood method for estimating DNS, $AFNS_0$ and $AFNS$ with stochastic volatility, respectively.

transition equation is a VAR(1) dynamics for the state vector $X_t = \{L_t, S_t, C_t\}$:

$$X_t = (I - \phi)\mu + \phi X_{t-1} + \eta_t, \quad \eta_t \sim N(0, \Omega). \tag{3.2.13}$$

Specification of state dynamics of arbitrage-free Nelson-Siegel models requires defining the first two conditional moments of the latent factors under the P-measure. The expression for the conditional mean vector is the same for both the constant volatility and stochastic volatility cases:

$$E^P[X_T|X_t] = (I - \exp(-K^P(T-t)))\theta^P + \exp(-K^P(T-t))X_t. \tag{3.2.14}$$

However, the definitions of conditional variance matrices are different. The one for $AFNS_0$ is time-invariant and is given by

$$Q = V_0^P[X_T|X_t] = \int_0^{T-t} \exp(-K^P u)\Sigma'\Sigma \exp(-(K^P)'u)du. \tag{3.2.15}$$

The conditional variance matrix for $AFNS$ with stochastic volatilities are state-dependent and can be computed as

$$\begin{aligned} Q_T(X_t) &= V_1^P[X_T|X_t] = \int_t^T \exp(-K^P(T-u))\Sigma S(E^P[X_u|X_t]) \\ &\quad \times S(E^P[X_u|X_t])'\Sigma \exp(-(K^P)'(T-u))du. \end{aligned} \tag{3.2.16}$$

The state equation is then defined as a discrete version of the continuous-time P-dynamics of the latent factors:

$$X_t = (I - \exp(-K^P\triangle t))\theta^P + \exp(-K^P\triangle t)X_{t-1} + \eta_t, \tag{3.2.17}$$

where $\triangle t$ is the time between observations. For $AFNS_0$ $\eta_t \sim N(0, Q)$ and for stochastic volatility $AFNS$ $\eta_t \sim N(0, Q_t(X_{t-1}))$ with $Q_t(X_{t-1}) = V^P[X_t|X_{t-1}]$ given by equation (3.2.16). The measurement equation is obtained by adding stochastic disturbance terms to equation (3.2.5) where yields are expressed as deterministic linear functions of state variables:

$$y_t(m_i) = -\frac{1}{\tau}B(m_i)'X_t - \frac{1}{\tau}A(m_i) + \varepsilon_t(m_i), \quad i = 1, 2, ..., N.$$

The above equation can be more compactly written in matrix notation as:

$$Y_t = \overline{A} + \overline{B}X_t + \varepsilon_t,$$

where Y_t is a vector of N observed yields and ε_t is an $N \times 1$ vector of independent and identically distributed Gaussian white noise measurement errors, i.e., $\varepsilon_t \sim i.i.d.N(0, R)$. $\overline{A}$ is the vector of yield-adjustment terms and $\overline{B}$ is the loading matrix defined respectively as[7]:

$$\overline{A} = \begin{bmatrix} -\frac{A(m_1)}{m_1} \\ -\frac{A(m_2)}{m_2} \\ \vdots \\ -\frac{A(m_N)}{m_N} \end{bmatrix} \quad \text{and } \overline{B} = \begin{bmatrix} -\frac{B(m_1)}{m_1} \\ -\frac{B(m_2)}{m_2} \\ \vdots \\ -\frac{B(m_N)}{m_N} \end{bmatrix} \quad .$$

Note that for DNS there is no yield-adjustment term and hence, $\overline{A} = 0$. The measurement disturbance covariance matrix R is assumed to be diagonal. Measurement and transition disturbances are assumed to be orthogonal to each other.

We briefly describe the Kalman filter that operates in two recursion steps - a prediction step and an updating step. Let $X_{t-1|t-1}$ denote an update of the state vector that has been obtained at period $t-1$ using information up to $t-1$and let $P_{t|-1t-1}$ be its mean square error matrix. Then forecasts for the next period t are obtained in the prediction step as

$$X_{t|t-1} = a + bX_{t-1|t-1},$$

where $a = (I - \exp(-K^P \triangle t))\theta^P$ and $b = \exp(-K^P \triangle t)$[8] and

$$P_{t|t-1} = bP_{t|-1|t-1}b' + Q_t(X_{t-1|t-1}),$$

[7] Elements of $\overline{A}$ and $\overline{B}$ for $AFNS$ are solutions of ordinary differential equations and we use Matlab funtion ode45 to numerically solve them.

[8] In the empirical excercise the matrix exponents are evaluted in Matlab which uses Padé approximation.

where $Q(.)$ is computed by the conditional variance formula (3.2.16).

The yield prediction error and its variance are obtained as

$$v_{t|t-1} = Y_t - \overline{A} - \overline{B}X_{t|t-1},$$

and

$$F_{t|t-1} = \overline{B}P_{t|t-1}\overline{B}' + R.$$

In the update step at time t, the prediction $X_{t|t-1}$ made at time $t-1$ is improved by using additional information contained in Y_t:

$$X_{t|t} = X_{t|t-1} + P_{t|t-1}\overline{B}'F_{t|t-1}^{-1}v_{t|t-1},$$

$$P_{t|t} = P_{t|t-1} + P_{t|t-1}\overline{B}'F_{t|t-1}^{-1}\overline{B}P_{t|t-1}.$$

The unknown parameters of the state-space model are estimated by maximizing the log likelihood given by

$$l(Y_1, Y_2, ..., Y_N; \Psi) = -\frac{NT}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln\left|F_{t|t-1}\right| - \frac{1}{2}\sum_{t=1}^{T}v_{t|t-1}'F_{t|t-1}v_{t|t-1}.$$

The Kalman filter is initialised at the unconditional mean and covariance matrix [9]

$$\widehat{X_o} = \theta^P \text{ and } \widehat{\Sigma_0} = \int_0^\infty e^{-K^P s}\Sigma S(\theta^P)S(\theta^P)'\Sigma' e^{-(K^P)' s}ds.$$

Finally, the standard deviations of the estimated parameters are calculated as

$$\Sigma(\widehat{\Psi}) = \frac{1}{T}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \log l_t(\widehat{\Psi})}{\partial(\widehat{\Psi})}\frac{\partial \log l_t(\widehat{\Psi})'}{\partial(\widehat{\Psi})}\right]^{-1},$$

where $\widehat{\Psi}$ denotes the optimal parameter set.

There are, however, caveats of using the above Kalman filter based maximum likelihood technique for $AFNS$ with stochastic volatilities. The discrete

[9] The conditional and unconditional moments in the estimation are calculated following Fackler (2000) who extends the analytic solutions provided in Fisher and Giles (1996).

state equation (3.2.17) assumes that the distribution of that the state variables be Gaussian which is unlikely because of introduction of stochastic volatilities in the models . Under the assumption of normality, the Kalman filter induced maximum likelihood estimation is only quasi-maximum. Moreover, in spite of forcing the parameter sets to satisfy Feller and other non-negativity conditions the discretisation (3.2.17) can drive the square-root processes to negative territory. If this happens we replace the negative value by zero following the literature (e.g., see Duffee (1999), Christensen et al. (2010)).

3.3 Data and In-Sample Fit

For our empirical analysis we opt to use data which have been analysed in Christensen et al. (2011). The data set consists of monthly U.S. zero-coupon bond yields from January 1987 to December 2002. The yields are end-of-month and reported at sixteen different maturities: 3, 6, 9, 12, 18, 24, 36, 48, 60, 84, 96, 108, 120, 180, 240 and 360 months. We find this particular data set attractive for our forecasting exercise for several reasons. First, the yields are Fama-Bliss unsmoothed yields and therefore, represents the true raw yields better than those extracted by smoothing methods such as interpolating with Nelson-Siegel type functions or fitting splines.[10] Second, the data cover the cross-section of the yield curve reasonably well as it includes yields with very short maturity (e.g., 3 months) to very long maturity (e.g., 30 years). Third, application of stochastic-volatility-Nelson-Siegel models on this particular data allows us to compare some of our results directly with those of Christensen et

[10] Construction of true yields by fitting a Nelson-Siegel type function may provide an unfair advantage to any Nelson-Siegel yield curve model for out-of-sample forecasting. Since all our competing models belong to the Nelson-Siegel class, this is not desirable. Yields estimated by alternative smoothing techniques such as spline-fittng are often considered to be distorted to some extent.

al. (2011) obtained with constant volatility models.[11]

Figure 3.A.1 plots time series of yields over the entire sample period of January 1987 to December 2002. It is evident that the level, slope and curvature of the yield curve vary substantially over time depicting many different shapes - upward sloping, flat, inverted and so on.

Table 3.B.1 presents summary statistics of the sample yields and confirms a number of stylised features of a typical yield curve. The average yield curve, represented by means of yields of reported maturities, slopes upward. The rear end of the yield curve, however, tilts downwards with average yield of the 30-year bond smaller than that of the 20-year bond.[12] Decreasing standard deviations for longer yields imply that the short end of the yield curve is more volatile than the long end. The sample autocorrelations reveal that all the yields are very persistent and that persistence increases with maturity. Autocorrelations of longer yields are high even at lags of two years.

All the Nelson-Siegel models under study exploit three latent factors which have unique definitions: level, slope and curvature.[13] Following Christensen et al. (2010) we perform a principal component analysis in order to investigate the appropriateness of use of such three-factor models to our data. Results are summarised in **Table 3.B.2.** First three principal components explain about 99.8% of the total variation in the yield curve and their loadings on different yields shed light on their nature. With negative loadings of somewhat similar size on all the yields the first principal component acts like the level of the yield curve. Any change in it would affect all yields all most equally and in the same

[11] Our forecast design, however, is different from the setup of Christensen et al. (2011). Therefore, conclusions drawn from forecast results of DNS and $AFNS_0$ are similar, but not readily comparable.

[12] Litterman et al. (1991) relates this reduction of yields on longer maturities to the convexity of discount factor which prices bonds. They argue that the latter increases with increased volatility.

[13] The fact that three factors are sufficient to model the term structure of interest rates is well documented in literature. See Litterman and Scheinkman (1991), for example.

direction. With loadings of opposite signs for shorter and longer maturities the second principal component captures the slope of the yield curve. Shocks in it determines the steepness and flatness of the yield curve. Finally, the third principal component has negative loadings on the two ends of the yield curve and positive loadings in the middle suggesting that the component controls the curvature of the yield curve. Overall, the results justify application of three-factor Nelson-Siegel models to the data.

We estimate all the six models over the full sample period of January, 1987 to December, 2002. The parameter estimates of the benchmark model DNS are presented in **Table 3.B.3** and those of the $AFNS$ models are presented in **Table 3.B.4**. We replicate results of estimation of the DNS and the $AFNS_0$ models, as presented in Christensen et al. (2011), with minor discrepancies. For the $AFNS$ models with stochastic volatilities the patterns that can be identified from estimated values of the parameters are very similar to that reported in Christensen et al. (2010) which uses daily U.S. yields. Whether a factor is inducing stochastic volatility or not largely contributes to the variations in the estimated mean parameters in θ^P and estimated volatility parameters in σ across maturities. The signs of elements of θ^P are predominantly determined by the fact that any volatility-generating factor has to be non-negative under a square-root process. Since different factors operate at different scales estimated values of factor means are also substantially different. Factors generating volatility have higher estimated σ values and if a factor does not produce stochastic volatility its associated estimated σ is very close to the corresponding estimate in the $AFNS_0$ model. There is substantial variation in K^P matrix across models. However, for all models the level factor is the most persistent while the curvature being the least persistent. The estimated β volatility sensitivity parameters suggest that the level factor induces substantial stochastic volatility in both slope and curvature but curvature hardly contributes in generating volatility in the level. The estimated values of λ lie within a range of 0.49-0.82.

Following common practice we assess goodness of in-sample fit of different models by comparing root-mean-squared-errors (RMSE) of fitted yields. Reports presented in **Table 3.B.5** show that performance of models are mixed. No single model provides the best fit to the entire cross-section of the yield curve. A comparison of the DNS and the $AFNS_0$ models implies that imposition of no-arbitrage restriction deteriorates the fit to the short-end of the yield curve, particularly to yields with 3- and 6-month maturities. However, introduction of stochastic volatility through the level factor, as modelled by the $AFNS_1 - L$ model, outweighs much of these losses. RMSE values of the models $AFNS_1 - L$ and $AFNS_2 - LC$ are similar across all yields but they are different from those of $AFNS_0$. This probably suggests that when both level and curvature are allowed to induce volatility, most of the in-sample volatility is accounted for by the level factor and it alone can generate sufficient amount of stochastic volatility to produce cross-sectional fit which is substantially different from that of the constant volatility counterpart. RMSE values of the models $AFNS_0$ and $AFNS_1 - C$ are somewhat different for the first few short-maturity yields, but similar otherwise. This indicates that volatility which is generated through curvature alone affects only the short end of the yield curve. Interestingly, variations in RMSEs of the models $AFNS_0$, $AFNS_1 - L$ and $AFNS_3$ imply that all the three factors together produce stochastic volatility which affects the in-sample fit of data differently than do volatility generated by level factor alone or by constant volatility. An additional noteworthy observation is that the models $AFNS_0$ and $AFNS_1 - C$ fit the longest end of the curve (i.e. the 30-year yield) very well but provide poor fit to 15- and 20-year yields. The other time-varying volatility models $AFNS_1 - L$, $AFNS_2 - LC$ and $AFNS_3$ models deliver an opposite performance by fitting the 15- and 20-year yields relatively well and fitting the 30-year yield miserably. Overall, the $AFNS_2 - LC$ model, where both level and curvature are the drivers of stochastic volatility, is the most consistent in capturing the yield curve dynamics well as it produces the maximum number of minimum RMSEs.

3.4 Out-of-Sample Forecasting

The main focus of this chapter is systematic evaluation of predictive performance of Nelson-Siegel term structure models out-of-sample. Particular interests lie in investigating whether accounting for no-arbitrage restrictions together with time-varying volatility improves forecast accuracy of models and also whether performance of models in terms of point forecasts of conditional mean of yields is consistent with their ability to generate conditional predictive densities of yields. The competing models which we have discussed earlier range from non-arbitrage-free constant volatility benchmark to arbitrage-free stochastic volatility specifications. In what follows we explain how forecasts, both point and probability, are generated using each model and analyse results of both types of forecasts in details.

3.4.1 Point Forecasts

Forecast Design and Construction of Forecasts

We construct 1-, 3-, 6- and 12-month-ahead forecasts of the U.S. yield curve using all the Nelson-Siegel specifications described in Section 2. We estimate and forecast using a rolling-window sample. The first estimation sample is January, 1987 to January, 1996; the next is February, 1987 to February, 1996 and so on. The last estimation sample for the 1-step-ahead forecast ends in November, 2002 and 83 forecasts are generated altogether. For 3-, 6- and 12-month horizons, final estimation samples end in September, 2002 (81 forecasts), June, 2002 (78 forecasts) and December, 2001 (72 forecasts), respectively. For multi-step ahead forecasts we use iterated forecasts, as opposed to direct forecasts.[14] Forecasts of yields are made it two steps: first h-step-ahead forecasts of factors are produced and then these are used to predict yields h steps forward.

An h-step-ahead forecast of an m-maturity yield which is made at time t using the DNS model is given by

$$\widehat{y}_{t+h|t}(m) = \widehat{L}_{t+h|t} + \widehat{S}_{t+h|t}\left(\frac{1-e^{-\lambda m}}{\lambda m}\right) + \widehat{C}_{t+h|t}\left(\frac{1-e^{-\lambda m}}{\lambda m} - e^{-\lambda m}\right). \quad (3.4.1)$$

Defining the state vector as $X_t = (L_t, S_t, C_t)$, an h-step-ahead forecast $\widehat{X}_{t+h|t}$ is computed as

$$\widehat{X}_{t+h|t} = \left(\sum_{i=0}^{h-1}\phi^i\right)(I-\phi)\mu + \phi^h X_t. \quad (3.4.2)$$

For the $AFNS$ specifications, an h-step-ahead forecast of an m-maturity yield is constructed as

$$\widehat{y}_{t+h|t}(m) = \widehat{X}^1_{t+h|t} + \widehat{X}^2_{t+h|t}\left(\frac{1-e^{-\lambda m}}{\lambda m}\right) + \widehat{X}^3_{t+h|t}\left(\frac{1-e^{-\lambda m}}{\lambda m} - e^{-\lambda m}\right) - \frac{A(m)}{m}, \quad (3.4.3)$$

where, $\widehat{X}_{t+h|t}$ is an h-step-ahead forecast of the state vector $X_t = (X_t^1, X_t^2, X_t^3)$ and is given by

$$\widehat{X}_{t+h|t} = (I - \exp(-K^P h))\theta^P + \exp(-K^P h)X_t. \quad (3.4.4)$$

Forecast Evaluation

We use root mean squared forecast error (RMSFE) as the criterion for evaluating the accuracy of yield curve models and compare their predictive performance in terms of RMSFE relative to the benchmark model DNS. For an h-step-ahead forecast of m-maturity yield produced from model j, we compute

$$RMSFE_j^h(m) = \sqrt{n_h^{-1}\sum(y_{t+h}(m) - \widehat{y}^{(j)}_{t+h|t}(m))^2}$$

[14] See Marcellino et al. (2006) for a discussion on direct and iterated forecasts.

and define the relative RMSFE as

$$RRMSFE_j^h(m) = RMSFE_j^h(m)/RMSFE_{DNS}^h(m).$$

A value of RRMSFE smaller than one implies that the corresponding model forecasts better than the benchmark.

Tables 3.B.6 and **3.B.7** present mean forecast errors (MFEs) of models in basis points. A negative value indicates over prediction while a positive value implies under prediction. We test the null hypothesis of zero MFE against alternative hypotheses of negative and positive biases and mark significance at 1%, 5% and 10% level.

MFE values are predominantly negative indicating that the model-generated forecasts of many yields are, on average, too high. Average forecast errors for the constant volatility models DNS and $AFNS_0$ are consistently negative across yields with maturities of 10 years and less. Values for the $AFNS_3$ model are negative all through except for 1-month-ahead forecasts of the shortest yield with 3-month maturity. For the models $AFNS_1{-}L$ and $AFNS_2{-}LC$ prediction biases are more mixed in sign. For example, 1-month-ahead MFEs of these two models are positive for yields with 3-month, 7-year and 20-year maturities but negative for the rest of the maturities. Over the shortest forecast horizon of one month, the benchmark model DNS turns out to be the worst forecasting model in the sense that it is the only model that reports MFEs that are significantly different from zero for the entire cross-section of the yields. The absolute forecast bias of the DNS model are also the highest for all yields below 10-year maturity. Relatively better performance of the $AFNS$ models implies that imposition of no-arbitrage restriction helps in forecasting yield levels, at least up to moderately long maturity. Overall, the $AFNS$ models forecast yields with 3-month to 7-year maturities well at 1-month horizon with the highest average prediction error not exceeding 8 basis points in absolute value. Forecasts of the 10-year yield are, however, significantly high for all the models with mean forecast errors being between -10 to -19 basis points. Finally, forecast performance of competing models are mixed for the longest two yields: while the $AFNS_1-L$ and $AFNS_2-LC$ models predict the 20-year yield with forecast biases of only 2 basis points, they provide considerably large and significantly non-zero biases of around 75 basis points for the 30-year yield. Performance of the $AFNS_0$ and $AFNS_1-C$ models are opposite: while average biases are staggering 50 basis points when forecasting the 20-year yield they are merely 6 basis points when forecasting the 30-year yield. This extreme forecast behaviour of the models at the long end of the yield curve is not surprising considering similar in-sample performance of the models. The 1-month-ahead forecasts of the 30-year yield that are generated by the $AFNS_3$ model are highly biased with mean forecast errors exceeding 100 basis points.

Most of the findings for 1-month-ahead forecasts are also preserved at longer horizons. On average, the benchmark DNS model continues to be the most biased model while the $AFNS$ models where the level factor is one of the drivers of volatility generate the least forecast errors except at the longest maturity.

The root-mean-square-forecast-errors (RMSFEs) which is a broader criterion of evaluating predictive performance have been reported in **Table 3.B.8** and **Table 3.B.9**. The first row of each panel of the table reports RMSFE of the benchmark model DNS. The remaining rows present RRMSFEs which are ratios of RMSFE of each model relative to the DNS, as defined above. Any value below one means that the corresponding model forecasts better than the benchmark. We test whether any gain or loss against the benchmark DNS is significant by applying Giacomini and White (2006) test. The null hypothesis of the test is that of equal predictive ability for a model and the DNS benchmark. Statistical significance at 1%, 5% and 10% is reported.

All the $AFNS$ models fare very well against the constant volatility benchmark, at least for yields that are maturing in ten years and less. At the 1-month horizon, models show varied predictive ability across different yields and it is difficult to identify a single best and most consistent predictor. The most accurate forecasts are shared among stochastic volatility models $AFNS_1 - L$, $AFNS_1 - C$ and $AFNS_2 - LC$. The gains, are, however, small except for the longest two maturities with the maximum gain not exceeding 6%. Extreme and contrasting forecast performance of the models for the 20- and 30-year yields, which is evident in MFE analysis, is also reflected in the reported RRMSFEs. The models $AFNS_1 - L$, $AFNS_2 - LC$ and $AFNS_3$ predict the 20-year yield with gains of around 40% over the DNS benchmark, but they are convincingly outperformed by the benchmark when forecasting the 30-year yield. The margin of prediction loss is more than 200% for the $AFNS_3$ model. The models $AFNS_0$ and $AFNS_1 - C$ lose 20% against the benchmark when forecasting the 20-year yield but gain a same percentage in forecasting the 30-year yield.

At longer forecast horizons of three months and six months the $AFNS_2 - LC$ model appears to be the best predictive model by outperforming all the competitors across most of the yields. However, forecast gains are rarely significant. For instance, among all the 3-month-ahead predictions, only the $AFNS_1 - C$'s forecast of the 3-month yield and the $AFNS_1 - L$ and $AFNS_2 - LC$'s forecasts of the 10-year and 20-year yields are significantly better than those of the DNS. At 12-month horizon the $AFNS_3$ forecasts many of the yields most accurately.

3.4.2 Density Forecasts

As we discussed earlier, point forecasts do not provide any description of uncertainties associated with forecasts and therefore, capture only a partial/incomplete account of the predictive power of a time series model. A much broader picture of forecast performance is available through density forecasts which estimate the entire probability distribution of possible forecasts at each point in time in the future. Calibration of density forecasts of the yield curve which usually exhibit considerable amount of time-varying fluctuations requires sufficient account of stochastic volatility in the underlying models. Diebold and Rudebusch (2013) rightly emphasise "if interest centers on interval or density forecasts of yields or yield factors, then stochastic volatility is of direct and intrinsic interest and cannot be ignored." One of the principal interests of this paper lies in investigating how stochastic volatility AFNS models fare against their constant volatility counterparts in portraying predictive densities of yields.

We use Monte Carlo simulation to produce density forecasts of yields. The basic approach involves generating alternative outcomes (forecasts) artificially and approximating the distribution of forecasts each point in time in the forecast period. Note that we do not take account of parameter estimation uncertainty and directly use point estimates of state-space parameters in the simulation. The simulation process can be more methodically described in the following steps:

1. Estimate a model using all the information up to time t. Let $\widetilde{\Psi}_t$ denote the vector of point estimates of unknown parameters and $\widetilde{X}_t$ denote the vector of latent factors extracted by Kalman filter in the estimation.

2. For $h = 1$ draw one-step-ahead forecast of state equation disturbance vector, η_{t+1} from an appropriate multivariate normal distribution specific to a model. Note that for DNS $\eta_{t+1} \sim MN(0, \Omega)$, for $AFNS_0$ $\eta_{t+1} \sim MN(0, Q)$ and for $AFNS$ with stochastic volatility $\eta_{t+1} \sim MN(0, Q(\widetilde{X}_t))$, where $Q(\widetilde{X}_t)$ is given by (3.2.16).

3. At time t generate 1-step-ahead prediction of factors using the state equation as:

$$\widehat{X}_{t+1|t} = a(\widetilde{\Psi}_t) + b(\widetilde{\Psi}_t)\widetilde{X}_t + \eta_{t+1}, \tag{3.4.5}$$

where a and b are appropriately computed for DNS and $AFNS$ specifications according to definitions given in the estimation.

4. Multi-step-ahead forecasts of latent state vector are simulated by drawing h-step-ahead factor forecast errors η_{t+h} from the model-specific multivariate normal distributions stated in step 2 and then iterating forward the state equation for $h = 2, 3, ..., h_{\max}$:

$$\widehat{X}_{t+h|t} = a(\widetilde{\Psi}_t) + b(\widetilde{\Psi}_t)\widetilde{X}_{t+h-1|t} + \eta_{t+h}. \tag{3.4.6}$$

Note that for $AFNS$ with time-varying volatilities variance of η_{t+h} is $Q(\widetilde{X}_{t+h-1|t})$ which is state-dependent.

5. Finally, realisations of h-step-ahead forecasts of yields are approximated by drawing the measurement equation disturbance vector ε_{t+h} from a multivariate normal distribution with mean 0 and variance $\widehat{R}$, where $\widehat{R} \in \widetilde{\Psi}_t$ and inserting them in the yield measurement equation along with the factor forecasts:

$$\widehat{Y}_{t+h|t} = A(\widetilde{\Psi}_t) + B(\widetilde{\Psi}_t)\widehat{X}_{t+h|t} + \varepsilon_{t+h}. \tag{3.4.7}$$

Forecasts for the period $t+2$ are obtained on the basis of estimates of parameters which use information up to time $t+1$ and then repeating all the steps described above. This process is continued until forecasts for the last period T are generated. One can generate many artificial paths for h-step-ahead forecasts of factors and yields by drawing the transition and measurement equation errors many times accordingly. At each forecast origin, predictive density is numerically approximated from the simulated replications.

Interval Forecast

One of the simplest ways of evaluating density forecasts is interval forecast or coverage rate. The idea is to estimate the time path of forecast intervals with certain coverage probability α and then compute proportion of true realisations that fall inside the interval. If the forecasting model is correctly specified about $\alpha\%$ of the realised values are expected to fall within the interval. This is very similar to back-testing in risk management where forecasted losses from estimated value-at-risk (VaR) are evaluated by looking back at past and checking how many times actual losses exceed the VaR limit. Giordani and Villani (2010) and Clark (2011) have used coverage rates as means of assessing macroeconomic density forecasts. Following Clark (2011) we construct 70% intervals by computing the 15th and 85th percentiles from the calibrated predictive densities.

Table 3.B.10 reports coverage rates generated by competing forecast models for a selection of maturities. An accurately constructed interval should contain about 70% of the real-time yields observed over the forecast period. A coverage rate of more than 70% implies that, on average, for a given sample, the predictive density is too wide and a rate below 70% means it is too narrow.

The 1-month-ahead predictive intervals of all models are extremely wide for yields with maturities below one year. Coverage rates for the 3-month yield are in the range of 92%-98% and those for the 6-month yield lie within 87%-95%. Performance of models in terms of matching the nominal coverage rate of 70% are mixed at the shortest forecast horizon. The $AFNS_1 - C$ and the $AFNS_0$ models provide the best forecast intervals for yields of bonds with maturities of two years or less, the benchmark DNS model predicts 3- and 5-year yields very well with coverage probabilities of 68%-72% and the $AFNS_1 - L$ and the $AFNS_2 - LC$ models calibrate the interval most accurately for longer yields with 10-year and 20-year maturities. Interesting patterns can be identified in coverage probabilities reported for different models, at least for all yields with maturities of twenty years and less. The pair, $AFNS_1 - L$ and $AFNS_2 - LC$, consistently produces intervals that are too wide with true yields falling inside the intervals much more frequently than the desired nominal rate of 70%. For the rest of the models, including the constant volatility specifications, coverage probabilities are decreasing with years-to-maturity. Calibrated forecast intervals of the arbitrage-free specifications $AFNS_0$, $AFNS_1-C$ and $AFNS_3$ become imprecisely narrow at the long end of the yield curve. For instance, coverage probabilities of these models are only 51%-53% for the 20-year yield. Inferior 1-month-ahead predictive performance of the $AFNS_1 - L$, $AFNS_2 - LC$ and $AFNS_3$ models, as observed in point forecasts of the 30-year yield, is reflected also in interval forecasts. For this longest interest rate coverage probabilities of the models $AFNS_1 - L$ and $AFNS_2 - LC$ are only 31% while that for $AFNS_3$ is just 11%. Overall, at 1-month horizon the benchmark model DNS reports coverage rates that are, on average, most consistently close to the true rate. Thus, it is evident from the sample that imposing no-arbitrage restriction and/or incorporating stochastic volatility have deteriorated calibration of 1-month-ahead predictive densities.

Coverage probabilities of forecasting models change at longer horizons. $AFNS$ models where level is one of the drivers of stochastic volatility show improved calibration of forecast intervals. At 6-month and 12-month horizons the $AFNS_1 - L$ and $AFNS_2 - LC$ models provide coverage rates that are closest to 70% for most of the yields. 12-month-ahead coverage probabilities of all models are very low for yields of two years maturity and below.

Probability Integral Transforms

Probability integral transforms (PITs) provide an informal but useful qualitative approach to assessing accuracy of density forecasts. The PIT of a realization y_t with respect to density forecast $p_{t-1}(y_t)$ can be defined as

$$z_t = \int_{-\infty}^{y_t} p_t(u)du.$$

For a correctly specified forecast density the PIT series should be i.i.d. uniform variates in the interval [0,1]. The idea of evaluating distributional assumption using PIT was first proposed in Rosenblatt (1952) and later used in Diebold et al. (1998) for assessing optimality of predictive density. Diebold et al. (1998) used a number of simple visual assessment techniques - such as, histograms for checking uniformity and correlograms of generalized residuals for checking independence of PITs.

For yields of selected maturities and forecast horizons of one month, **Figures 3.A.2-3.A.7** present PIT historgrams obtained as decile counts of PIT transforms. If the 1-month-ahead density forecasts are optimal the histograms would be flat (with heights of 8.3 per bin) to confirm that PITs are i.i.d. U[0,1]. Results show that forecasts of all the models are, in general, poor as they suffer material departures from uniformity. Departures are severe for shorter yields, particularly for 3-month and 6-month yields with distributions of PITs looking more like normal distributions. The constant volatility models DNS and $AFNS_0$ predicts the 1-year and the 3-year yields better than their stochastic volatility counterparts with relatively flatter PIT histograms. The $AFNS_1 - L$ and $AFNS_2 - LC$ models, however, produce superior density forecasts for yields with longer maturities, particularly 5, 7 and 20-year maturities. Predictive densities for the 10-year and the 30-year yields are far from convincing irrespective of model types.

Normal Transforms of PIT

The quality of models' forecast densities can also be visually investigated by plotting the normalized forecast error over time. The normalized forecast error at time t is defined as $u_t = \Phi^{-1}(z_t)$, where z_t denotes the PIT of one-step-ahead forecast errors and Φ^{-1} is the inverse of standard normal cumulative density function. Independence and uniformity of the PIT series would then mean that normalized forecast errors are independently distributed as standard normal which should be the case if a model is correctly specified.

Time series plots of normalised errors over time are presented in **Figures 3.A.8-3.A.13.** Clear distinctions among models' performances at the long end of the yield curve, as already identified from PIT historgrams, are also evident in the time series plots. However, since forecasts of yields generated by the competing models are very similar, plot of normalised errors are not very informative. Therefore, we resort to more formal evaluation of density forecast through statistical tests proposed in Berkowitz (2001).[15] The tests are based on the following AR(1) dynamics for the normalised errors:

$$u_t - \mu_u = \rho_u(u_{t-1} - \mu_u) + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_u^2). \tag{3.4.8}$$

Evaluating if $u_t \sim iidN(0,1)$ is then equivalent to testing the null hypothesis $H_0 : \mu_u = 0,\ \rho_u = 0$ and $\sigma_u^2 = 1$. The likelihood ratio test statistic is given by[16]

$$LR = -2(l(0,1,0) - l(\widehat{\mu}_u, \widehat{\sigma}_u^2, \widehat{\rho}_u)),$$

where the hats denote the estimated values. Under H_0, LR is distributed as χ^2 with 3 degrees of freedom, one for each of the three restrictions. A test of only standard normality of normalised errors can be constructed by specifying the null as $H_0' : \mu_u = 0$ and $\sigma_u^2 = 1$ and computing the test statistic as

$$LR = -2(l(0,1,\widehat{\rho}_u) - l(\widehat{\mu}_u, \widehat{\sigma}_u^2, \widehat{\rho}_u)),$$

which follows a χ^2-distribution with 2 degrees of freedom.

We report results of both tests along with a number of other metrics in **Table 3.B.11** and **Table 3.B.12.** The first panel reports the means of the normalised errors with t-statistics for testing the null of a zero mean. The second panel reports the variances of the normalised errors with t-statistics for

[15] Betkowitz (2001) document better power for tests based on normality of normalised errors than tests based on uniformity of PITs.

[16] See Berkowitz (2001) for an expression of the exact log-likelihood function associated with the AR(1) process.

testing the null that the variance equals one. The third panel reports the AR(1) coefficients estimated by regressing errors on a constant and first lags along with t-statistics for testing the null that the AR(1) coefficient is zero. The fourth and the final panel reports results of the Berkowitz's (2001) likelihood ratio tests with p-values of the joint test of independence and normality and p-values of the test of normality with zero mean and unit variance only (in parentheses).

Results of means of normalised errors resemble results of MFE for point forecast. The benchmark model DNS does poorly by reporting average normalised PITs which are significantly different from zero for most of the yields. Mean errors from the $AFNS_0$ model are, however, consistently closer to zero than those from the DNS model, at least for yields of 10-year maturity or less and in most cases they are significantly non-zero. For example, average of normalised errors from the DNS model are -0.140, -0.191 and -0.302 at maturities of three months, six months and five years and they are all statistically significant. The reported numbers for the $AFNS_0$ model on the same set of yields are -0.055, -0.107 and -0.253 which are not different from zero. The stochastic volatility models $AFNS_1 - L$, $AFNS_2 - LC$ and $AFNS_3$ fair well except at the longest yield of thirty years.

Reported variances of all the models are considerably and significantly lower than one for the shortest yield with 3-month maturity. The benchmark DNS is the most consistent in matching the unit variance. Variances of its errors are not significantly different from one for yields with maturities of one year and more. The $AFNS_0$ model reports variances which are, in general, much higher than those of the DNS model and for most of the longer yields variances are significantly higher than unity with large t-statistics. For instance, the error variance for the DNS model is 1.025 for the 7-year yield with a t-statistic of 0.205 and the error variance for the $AFNS_0$ model is 1.846 for the same yield with a t-statistic of 2.900. The stochastic volatility models

$AFNS_1 - L$ and $AFNS_2 - LC$ report variances which are consistently lower than one and also lower than those reported by the constant volatility counterparts DNS and $AFNS_0$ across all yields except for the one with 20-year maturity. For yields with two years to twenty years maturities, the variances are, however, not statistically different from unity. The stochastic volatility model $AFNS_3$ where all the three factors drive volatility provides variances which are even higher than those of $AFNS_0$ for a number of yields.

There is little evidence of independence of normalised errors. For the constant volatility models DNS and $AFNS_0$, the estimated AR(1) coefficients of error dynamics are significantly different from zero for all the yields except the one with 10-year maturity. The stochastic volatility models $AFNS_1 - L$, $AFNS_2 - LC$ and $AFNS_3$, however, perform slightly better by not showing significant autocorrelation for multiple yields, e.g., yields with maturities of 3, 5 and 7 years. For the 5-year yield reported AR(1) coefficients of DNS, $AFNS_0$, $AFNS_1 - L$, $AFNS_2 - LC$ and $AFNS_3$ are 0.248, 0.241, 0.226, 0.224 and 0.211, respectively and associated t-statistics for testing the null of no serial autocorrelations are 2.088, 2.137, 1.648, 1.557 and 1.340, respectively.

The above metrics look at the distributional and independence requirements individually and therefore, provide only a first hand idea about the quality of the 1-month-ahead density forecasts of yields. The requirements can jointly and therefore, more appropriately be tested by Berkowitz (2001) tests. P-values for jointly testing the H_0 of independence and standard normality of normalised errors reveal that in general the models fail the overall test. The $AFNS_2 - LC$ model is the least bad as it survives the test for two of the yields, the 5-year yield with a p-value of 0.064 and the 7-year yield with a p-value of 0.140. Two other models, the $AFNS_1 - L$ and the benchmark DNS show significantly good calibration of one-step-ahead forecast density for only the 7-year yield. P-values of a second test of just normality with zero mean and unit variance (reported in parentheses) shed light on whether any failure is driven by autocorrelation. When only matching the standard normal distribution in terms of first two moments (mean and variance) is of concern, some patterns can be observed among models' performance. The $AFNS_0$, $AFNS_1 - C$ and $AFNS_3$ do well on yields of medium maturities of one year to three years; the $AFNS_1 - L$ and $AFNS_2 - LC$ pass the test on yields of longer maturities of five, seven and twenty years and the benchmark DNS match the distribution well on yields of one year to seven years maturity.

Log Predictive Density Scores

The accuracy of density forecasts can be most broadly summarised and evaluated using log predictive density scores (LPDS). We follow the quadratic formula of Adolfson et al. (2005) where the log predictive score of h-step ahead predictive density at forecast origin t is defined as

$$S_t(y_{t+h}) = -2 \log p_t(y_{t+h}),$$

where, $p_t(y_{t+h})$ is the forecast density of N-dimensional vector of yields. Since we assumed $p_t(y_{t+h})$ to be multivariate normal, the LPDS can be expressed as

$$S_t(y_{t+h}) = n \log(2\pi) + \log \left| V_{t+h|t} \right| + (y_{t+h} - \overline{y}_{t+h|t})' V_{t+h|t}^{-1} (y_{t+h} - \overline{y}_{t+h|t}),$$

where $\overline{y}_{t+h|t}$ and $V_{t+h|t}$ are the mean and covariance matrix of h-step-ahead forecast distribution, being at time t. An average LPDS over the hold-out sample is defined as

$$S(h) = N_h^{-1} \sum S_t(y_{t+h})$$

where, N_h is the number of h-step ahead forecasts. The lower the score the better is the predictive ability of a model.

Table 3.B.13 reports the log predictive density scores of the forecasting models. Since yields of different maturities move together scores are similar across yields. Results of 1-month-ahead forecasts are mostly inconclusive.

For 3-month to 5-year yields, the model $AFNS_1 - C$, where curvature factor is the only driver of stochastic volatility, generates the lowest scores and therefore, the most accurate forecast density. The constant volatility models DNS and $AFNS_0$ are equally competitive. The pair of models, $AFNS_1 - L$ and $AFNS_2 - LC$, in which stochastic volatility is primarily induced by the level factor, produces inferior probability forecasts compared to their constant volatility counterparts except for the 20-year yield. Poor scores of the models $AFNS_1 - L$, $AFNS_2 - LC$ and $AFNS_3$ for the longest yield with 30-year maturity resemble similar disappointing predictive performance of the models in terms of RRMSFEs and coverage rates. The $AFNS_3$ model where all the three factors are responsible for generating time varying volatility turns out to be the worst model by scoring the highest across most of the maturities.

The log predictive scores increase with forecast horizon implying that predictions are less accurate for longer horizons. Multi-step-ahead scores reveal that some of the $AFNS$ models start to show improved predictive densities. At 3-month horizon the $AFNS_1 - L$ model provides the minimum scores for yields maturing in two to twenty years. Its scores are closely matched by those of the $AFNS_2 - LC$ model. At 6-month and 12-month horizons these two models produce scores that are consistently lower than those of the constant volatility models DNS and $AFNS_0$ across the entire cross-section of the yield curve except for the longest yield with 30-year maturity. Interestingly, a comparison between the two constant volatility models indicates that log predictive scores of the $AFNS_0$ are consistently higher than those of the DNS benchmark at forecast horizons of three months and more. This together with previous results imply that adopting no-arbitrage restriction appears to deteriorate prediction of forecast density at longer horizons, but the loss is more than compensated by gains through considering time-varying volatility, particularly that driven by the level factor. Even the $AFNS_3$ model which performs badly at 1-month horizon consistently outperforms the $AFNS_0$ at

12-month horizon, but fails to beat the DNS at many maturities.

Differences in log predictive scores of a model and that of the benchmark DNS are reported in **Table 3.B.14** and **Table 3.B.15.** We identify significant superior and inferior performance against the benchmark by using Giacomini and White (2006) test of equal predictive ability. The loss functions used in the tests are log predictive scores. Negative differences indicate that the calibrated predictive density of a model is better than that of the DNS benchmark while positive differences mean worse approximation on the part of the model. At 1-month-horizon the models $AFNS_0$ and $AFNS_1 - C$ have significant gains over the DNS at the short end of the curve. The models $AFNS_1 - L$ and $AFNS_2 - LC$ generate density forecasts that are worse than those of the benchmark for all yields except the 20-year yield where they gain significantly. It is difficult to find significance at longer horizons. The only significant gains come from the models $AFNS_1 - L$ and $AFNS_2 - LC$ when forecasting the 20-year yield at 3-month horizon. The 3- and 6-month-ahead density forecasts of these two models are, however, significantly inferior to forecasts of the DNS benchmark for the longest yield of 30-year maturity.

3.5 Conclusion

Current and past literature on term structure of interest rates heavily focus on comparing models in terms of point forecasts of either the mean or the variance of the yields. These are often of limited values as they do not account for uncertainties surrounding a prediction. Density forecasts which provide a full description of predictive densities of a model are more attractive and desirable. Apart from common moments they allow for computation of a range of uncertainty related measures such as quantiles. There is, however, an ongoing search for a term structure model which can calibrate distribution of future yields reasonably well. Following the failure of the well-known affine models in this context, as documented in Egorov et al. (2006), it is time to look for alternative models. Arbitrage-free Nelson-Siegel models with stochastic volatilities, proposed in Christensen et al. (2010), are certainly worthy candidates. They adopt several properties which are attractive from both theoretical and empirical perspectives. In particular, they are parametrically parsimonious, easily estimable and tractable, arbitrage-free and modeller of time-varying volatility of yields. This chapter extensively analyses and evaluates forecast performance of different specifications of arbitrage-free Nelson-Siegel models, both in terms of point forecasts and density forecasts.

Results of point forecasts fare well in favour of arbitrage-free Nelson-Siegel Models. Using the same data set of U.S. bond yields but a different forecast design we find similar conclusions as in Christensen et al. (2011)[17]: adopting no-arbitrage restrictions helps in improving forecast accuracy of dynamic Nelson-Siegel Models. Further predictive gains can be achieved by modelling stochastic volatility, particularly from forecasting at longer horizons. The simple benchmark DNS which does not rectify for risk-free arbitrage opportunities and time-varying volatilities consistently reports mean forecast errors which are higher than those of the $AFNS$ counterparts and root mean square errors which are relatively higher across all forecast horizons. Models where level is one of the drivers of stochastic volatility, in particular, show superior predictive ability. While the $AFNS_2 - LC$ model has clear advantage in 3- and 6-month-ahead forecasts, the $AFNS_3$ model, where all the three factors generate stochastic volatility, provides the most competitive forecast at the longest horizon of twelve months. There is, however, a caveat of using $AFNS$ with stochastic volatility to our sample; they fail miserably in fitting and forecasting the yield with the longest maturity of 30 years. This is not surprising

[17] We opt to use a rolling-window rather than an expanding-window scheme. We also curtail the estimation sample to have a longer forecast period.

considering the fact that these models have been found to show little flexibility in capturing any tilt at the longest end of the yield curve in-sample. Very similar in-sample fit and out-of-sample forecast performance of $AFNS_0$ and $AFNS_1 - C$ imply that curvature alone accounts for insignificant proportion of time-varying volatility.

Implications of findings for density forecasts are somewhat different. We use a number of metrics and tests to evaluate predictive densities of yields generated by the competing models. Results of a joint test of independence, zero mean and unit variance of normalised errors reveal that all models' prediction of 1-month-ahead forecast distribution is far from satisfactory. However, when serial correlation is ignored and just matching the first two moments of the standard normal distribution is considered, the constant volatility models DNS and $AFNS_0$ show good performance for medium-long maturity yields and stochastic volatility models $AFNS_1 - L$ and $AFNS_2 - LC$ exhibit good performance for a few long maturity yields. Density forecasts of short rates with maturities below one year are inferior for all the Nelson-Siegel models. These findings are also supported by visual representations of PIT histograms. When models are compared in terms of forecast intervals, all models are found to produce very wide intervals for short yields. Nonetheless, constant volatility models fare better than their stochastic volatility counterparts in matching the true coverage probability for 1-month-ahead forecasts. Calibration of predictive intervals of stochastic volatility models, however, improves with longer forecast horizons. Evaluation of density forecasts in terms of log predictive scores leads to similar findings: better predictive densities for constant volatility models and $AFNS_1{-}C$ at one-month horizon and more competitive density forecasts for the $AFNS$ models with stochastic volatilities at longer forecast horizons. Gains over the benchmark DNS are, however, rarely statistically significant. Disappointing performance of stochastic volatility $AFNS$ models for the longest yield of 30 year maturity is confirmed by all means of density forecast evaluation.

Overall, there are evidence of benefits from accounting for time-varying volatility which is spanned by three latent factors of arbitrage-free Nelson-Siegel models. But the models' generated joint predictive densities of yields are far from convincing. Further investigation is required to check if incorporation of unspanned stochastic volatility, as advocated by Collin-Dufresne et al. (2009), can improve density forecasts of such models.

3.A Appendix C: Figures

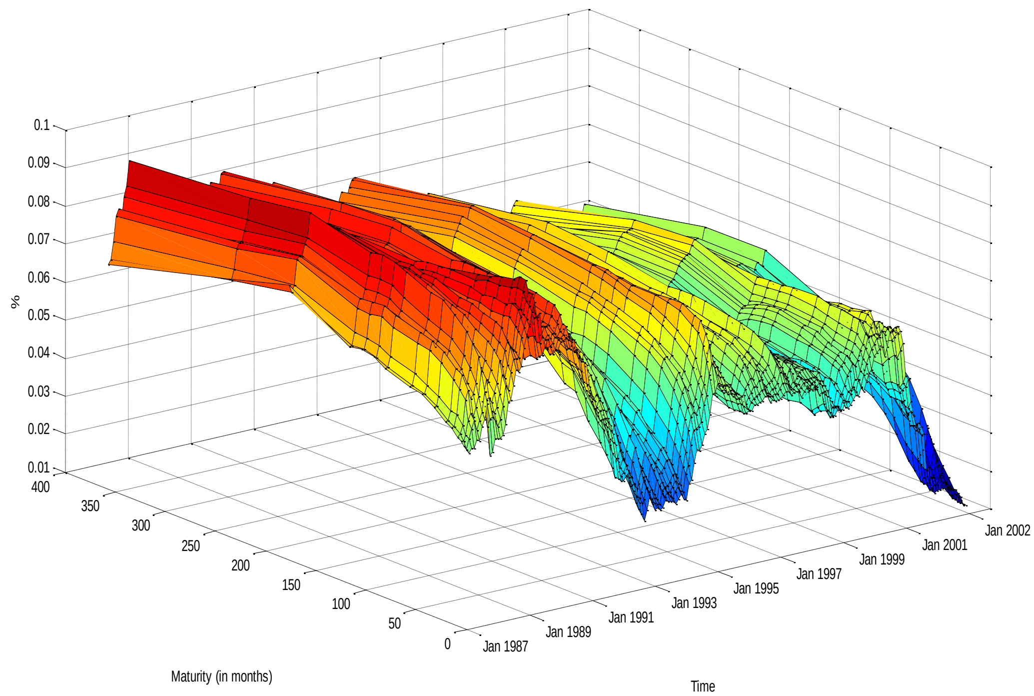


Figure 3.A.1: The US yield curves

Note: The data are Fama-Bliss unsmoothed zero-coupon bond yields for the period January 1987 - December 2002.

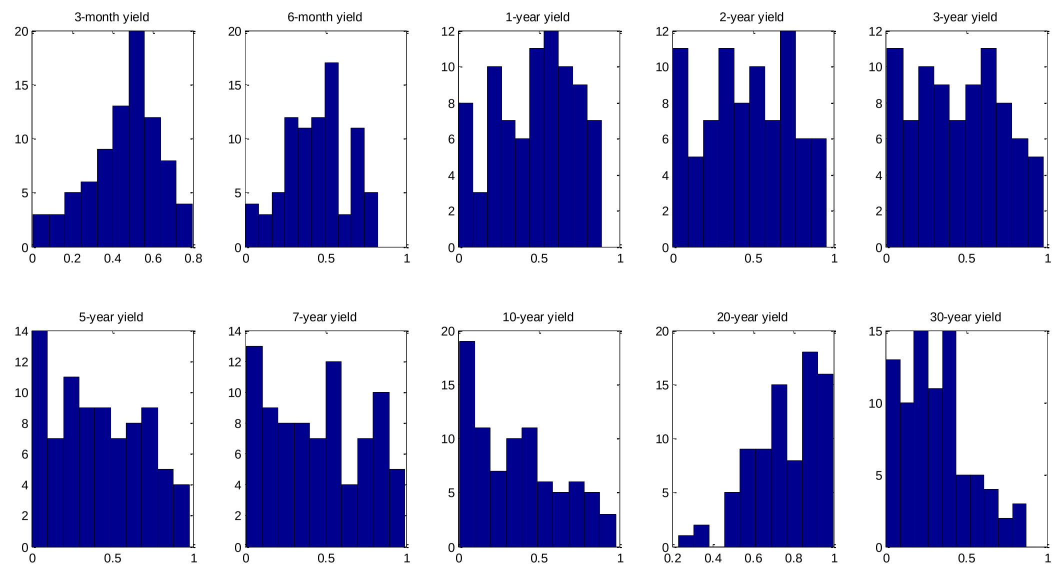


Figure 3.A.2: PIT histograms generated by the DNS model

Note: The histograms are decile counts of the probability integral transforms based on 1-month-ahead forecasts of bond yields with selected maturities. The forecast period is February 1996 to December 2002.

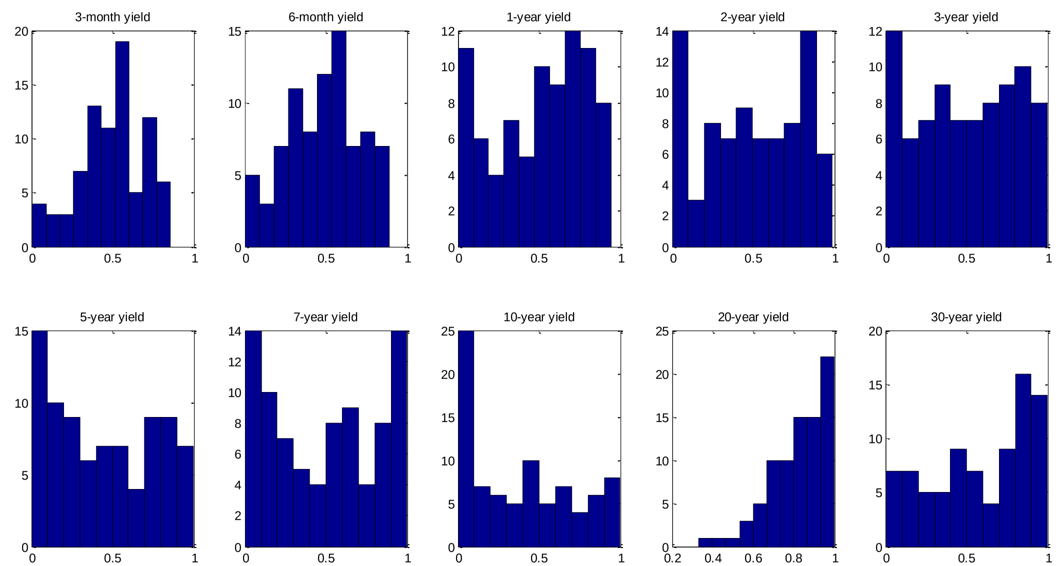


Figure 3.A.3: PIT histograms generated by the $AFNS_0$ model

Note: The histograms are decile counts of the probability integral transforms based on 1-month-ahead forecasts of bond yields with selected maturities. The forecast period is February 1996 to December 2002.

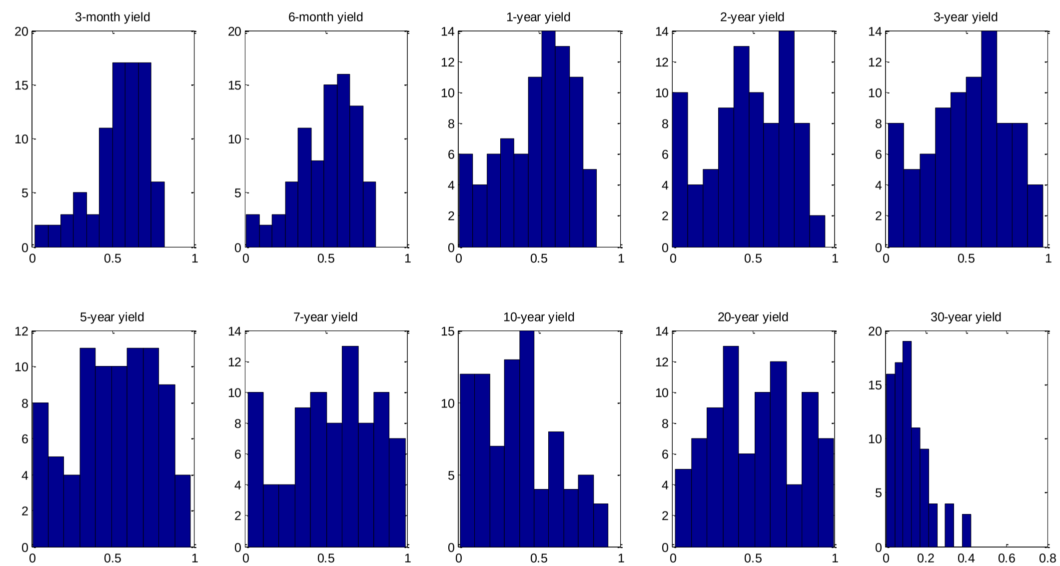


Figure 3.A.4: PIT histograms generated by the $AFNS_1 - L$ model

Note: The histograms are decile counts of the probability integral transforms based on 1-month-ahead forecasts of bond yields with selected maturities. The forecast period is February 1996 to December 2002.

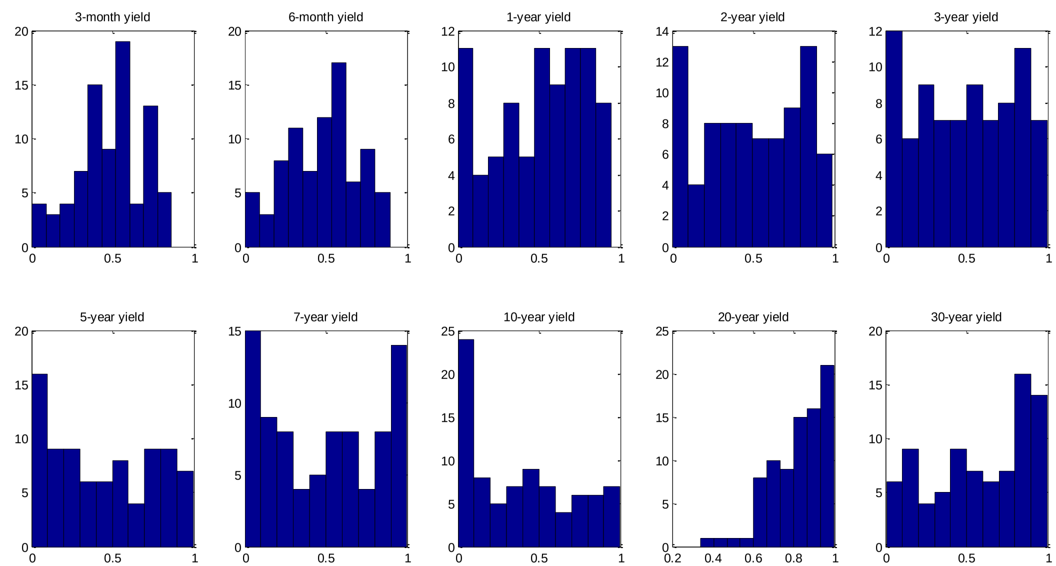


Figure 3.A.5: PIT histograms generated by the $AFNS_1 - C$ model

Note: The histograms are decile counts of the probability integral transforms based on 1-month-ahead forecasts of bond yields with selected maturities. The forecast period is February 1996 to December 2002.

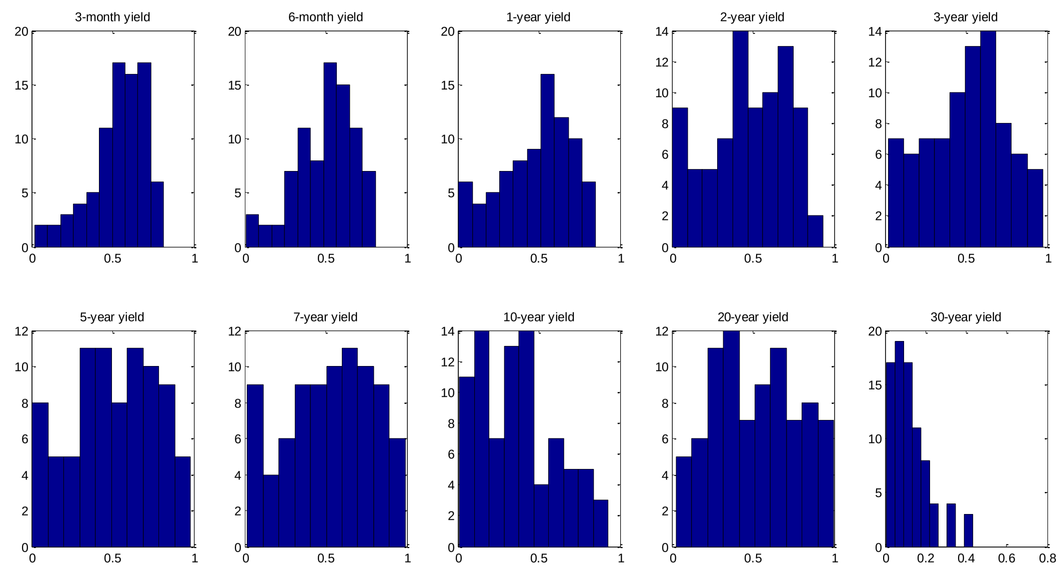


Figure 3.A.6: PIT histograms generated by the $AFNS_2 - LC$ model

Note: The histograms are decile counts of the probability integral transforms based on 1-month-ahead forecasts of bond yields with selected maturities. The forecast period is February 1996 to December 2002.

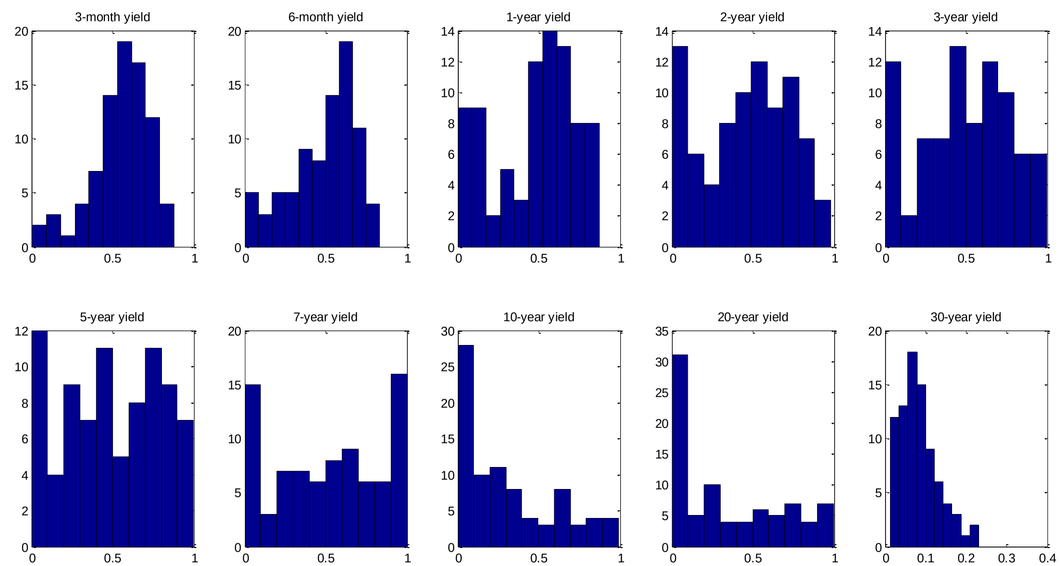


Figure 3.A.7: PIT histograms generated by the $AFNS_3$ model

Note: The histograms are decile counts of the probability integral transforms based on 1-month-ahead forecasts of bond yields with selected maturities. The forecast period is February 1996 to December 2002.

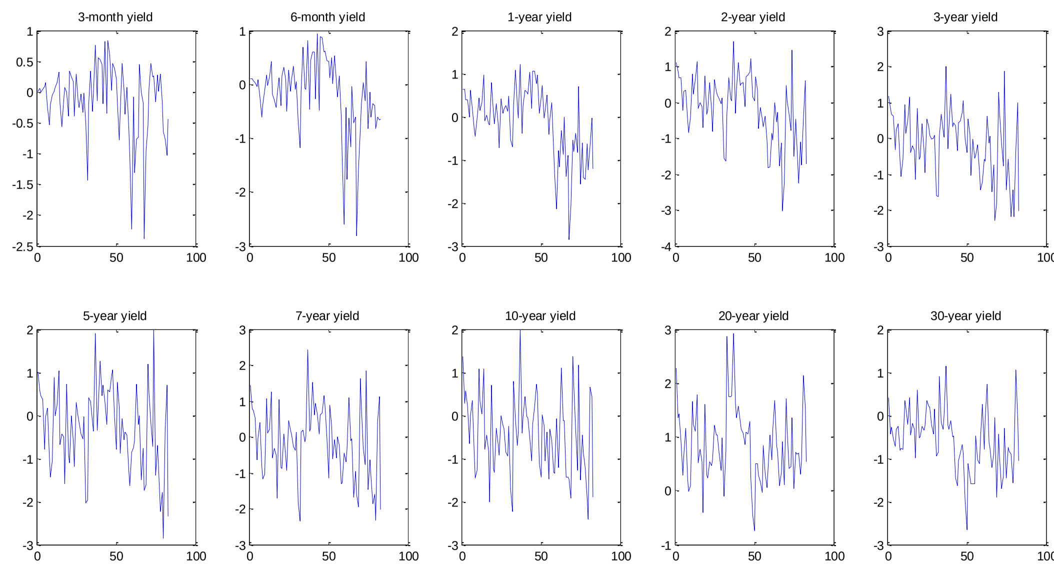


Figure 3.A.8: Normalised forecast errors generated by the DNS model

Note: The normalised forecast errors are defined as $\Phi^{-1}(z_t)$ where z_t denotes the probability integral transforms of 1-month-ahead forecast errors and Φ^{-1} is the inverse of the standard normal cumulative density function. The forecast period is February 1996 to December 2002.

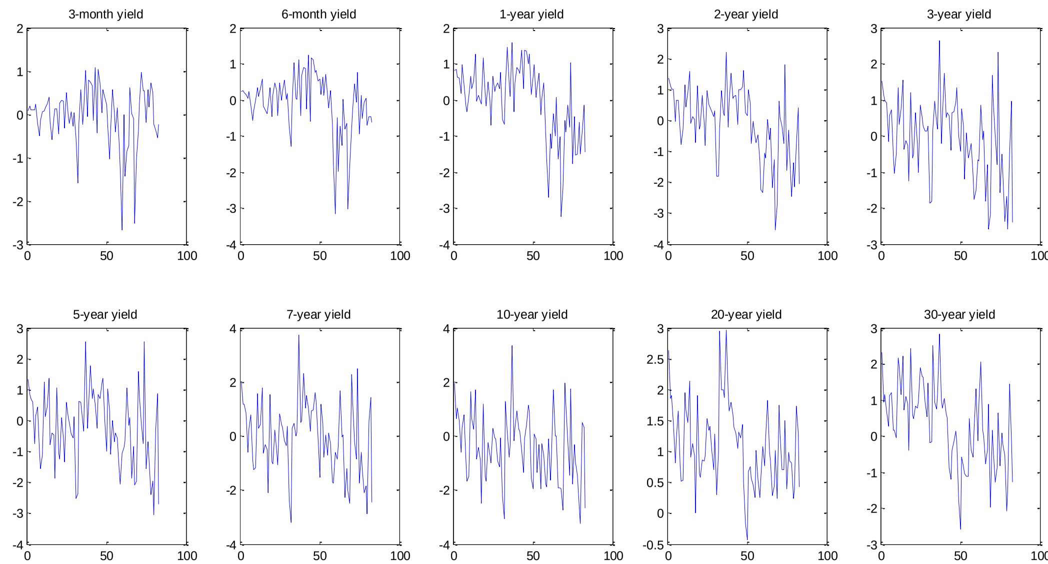


Figure 3.A.9: Normalised forecast errors generated by the $AFNS_0$ model

Note: The normalised forecast errors are defined as $\Phi^{-1}(z_t)$ where z_t denotes the probability integral transforms of 1-month-ahead forecast errors and Φ^{-1} is the inverse of the standard normal cumulative density function. The forecast period is February 1996 to December 2002.

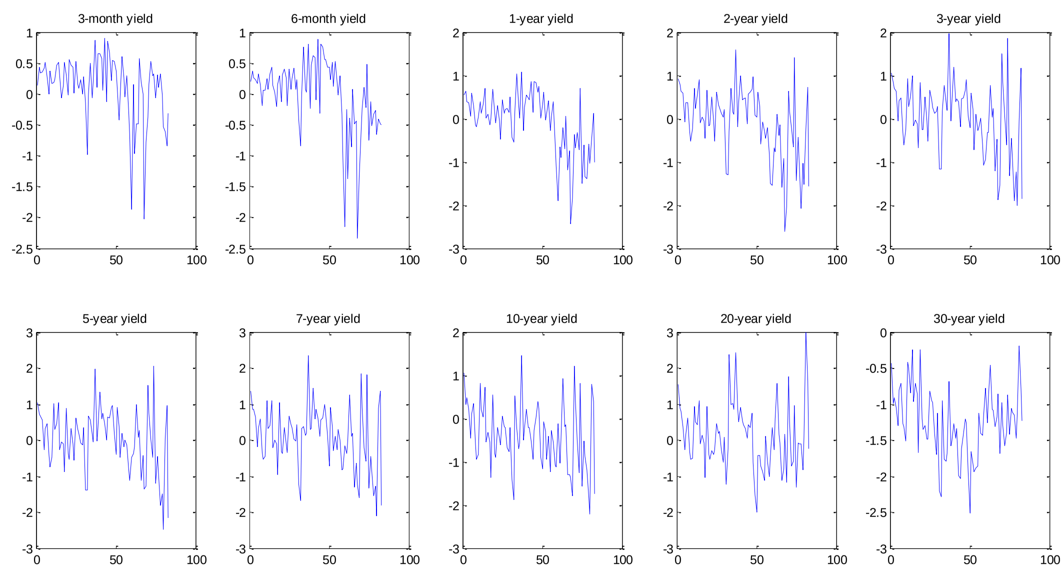


Figure 3.A.10: Normalised forecast errors generated by the $AFNS_1 - L$ model

Note: The normalised forecast errors are defined as $\Phi^{-1}(z_t)$ where z_t denotes the probability integral transforms of 1-month-ahead forecast errors and Φ^{-1} is the inverse of the standard normal cumulative density function. The forecast period is February 1996 to December 2002.

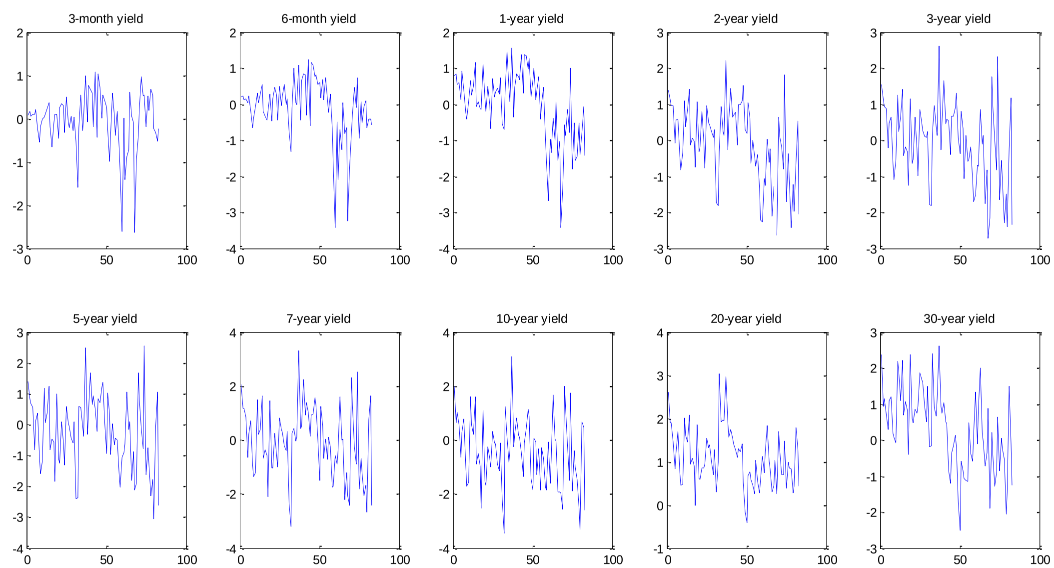


Figure 3.A.11: Normalised forecast errors generated by the $AFNS_1 - C$ model

Note: The normalised forecast errors are defined as $\Phi^{-1}(z_t)$ where z_t denotes the probability integral transforms of 1-month-ahead forecast errors and Φ^{-1} is the inverse of the standard normal cumulative density function. The forecast period is February 1996 to December 2002.

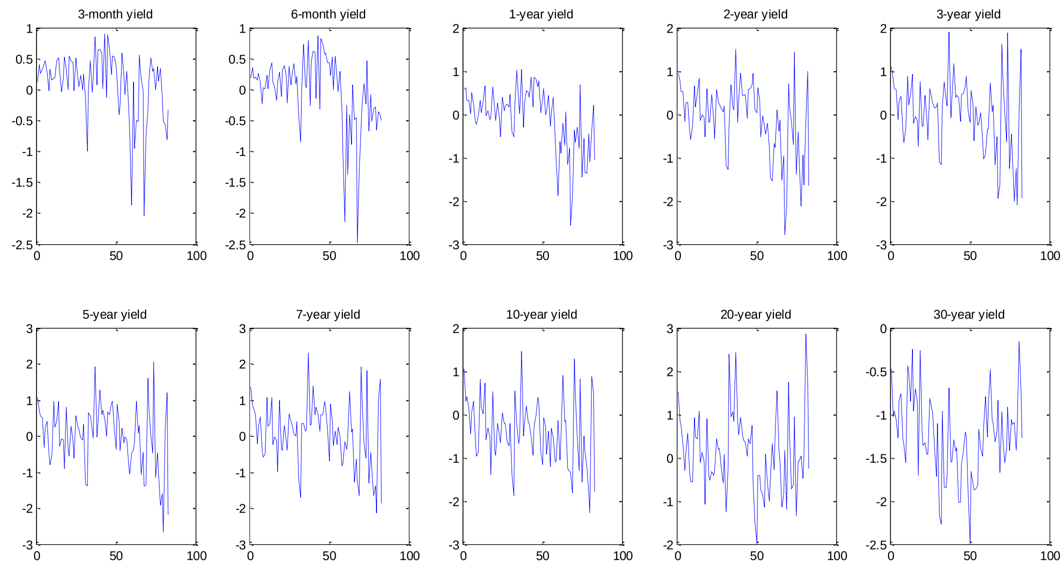


Figure 3.A.12: Normalised forecast errors generated by the $AFNS_2 - LC$ model

Note: The normalised forecast errors are defined as $\Phi^{-1}(z_t)$ where z_t denotes the probability integral transforms of 1-month-ahead forecast errors and Φ^{-1} is the inverse of the standard normal cumulative density function. The forecast period is February 1996 to December 2002.

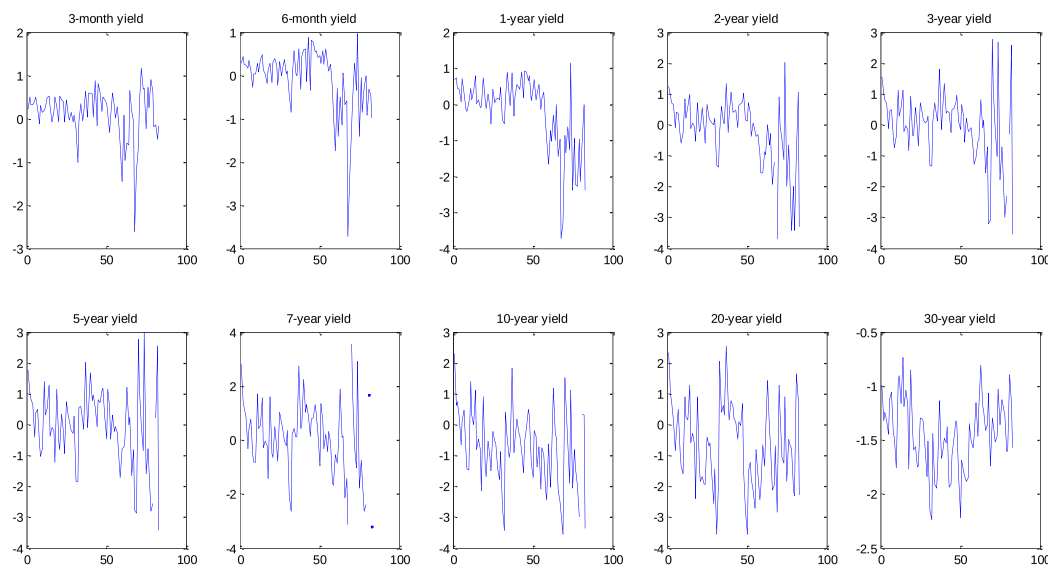


Figure 3.A.13: Normalised forecast errors generated by the $AFNS_3$ model

Note: The normalised forecast errors are defined as $\Phi^{-1}(z_t)$ where z_t denotes the probability integral transforms of 1-month-ahead forecast errors and Φ^{-1} is the inverse of the standard normal cumulative density function. The forecast period is February 1996 to December 2002.

3.B Appendix D: Tables

Table 3.B.1 Descriptive statistics, the US yield curves

Maturity	Mean	Std dev	Skewness	Kurtosis	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(24)$
3	5.085	1.744	-0.060	2.82	0.976	0.508	0.002
6	5.219	1.755	-0.140	2.789	0.975	0.509	0.013
9	5.329	1.763	-0.168	2.747	0.973	0.517	0.036
12	5.481	1.777	-0.196	2.766	0.971	0.520	0.050
18	5.703	1.735	-0.195	2.760	0.968	0.526	0.089
24	5.809	1.660	-0.180	2.742	0.964	0.528	0.127
36	6.063	1.552	-0.116	2.695	0.960	0.538	0.206
48	6.257	1.481	-0.083	2.592	0.959	0.555	0.274
60	6.361	1.440	-0.020	2.442	0.960	0.567	0.326
84	6.604	1.382	0.046	2.207	0.964	0.607	0.410
96	6.699	1.368	0.061	2.129	0.965	0.627	0.451
108	6.741	1.365	0.064	2.062	0.967	0.640	0.472
120	6.735	1.356	0.062	1.984	0.968	0.652	0.493
180	7.164	1.235	0.213	1.887	0.970	0.672	0.542
240	7.248	1.134	0.076	1.776	0.970	0.696	0.582
360	6.765	1.212	0.059	1.743	0.972	0.712	0.589

Notes: The data are monthly Fama-Bliss unsmoothed US zero-coupon bond yields. The sample period is January 1987 to December 2002 and the maturities are reported in months. The last three columns report sample autocorrelations at lags of 1, 12 and 24 months.

Table 3.B.2 Loadings of US yields on first three principal components

Maturity (in months)	First PC	Loadings on Second PC	Third PC
3	-0.279	0.353	-0.541
6	-0.288	0.316	-0.264
9	-0.294	0.277	-0.090
12	-0.298	0.255	0.053
18	-0.296	0.184	0.200
24	-0.286	0.116	0.307
36	-0.269	0.018	0.308
48	-0.256	-0.064	0.268
60	-0.247	-0.116	0.241
84	-0.231	-0.202	0.108
96	-0.226	-0.236	0.040
108	-0.222	-0.262	-0.020
120	-0.217	-0.287	-0.063
180	-0.190	-0.309	-0.121
240	-0.166	-0.326	-0.234
360	-0.176	-0.345	-0.426
% explained	90.03	9.30	0.45

Notes: The table reports eigenvectors of the first three principal components of US zero-coupon bond yields. The final row shows share of variation in all yields explained by each principal component. The sample is from January 1987 to December 2002.

Table 3.B.3 Estimates of parameters of independent-factor DNS model

Mean (μ)		AR(1) coefficient matrix (ϕ)		Error variance matrix (Ω)	
μ_1	0.0699	ϕ_{11}	0.9853	Ω_{11}	6.17x10^{-6}
μ_2	-0.0246	ϕ_{22}	0.9737	Ω_{22}	1.11 x10^{-5}
μ_3	-0.0111	ϕ_{33}	0.9259	Ω_{33}	5.57 x10^{-5}

Table 3.B.4 Estimates of parameters of the Arbitrage-free Nelson-Siegel (AFNS) models with independent-factors specification

Parameters	AFNS models with independent factors				
	$AFNS_0$	$AFNS_1 - L$	$AFNS_1 - C$	$AFNS_2 - LC$	$AFNS_3$
θ_1^P	0.0729	0.0448	0.0520	0.0449	0.0110
θ_2^P	-0.0282	-0.0112	-0.0116	-0.0180	0.0287
θ_3^P	-0.0096	-0.0024	0.1018	0.0772	0.0550
κ_{11}^P	0.0629	0.0439	0.0919	0.0442	0.0993
κ_{22}^P	0.2121	0.0997	0.0992	0.1100	0.1999
κ_{33}^P	1.1092	1.1002	1.0988	0.8999	0.4995
σ_{11}^P	0.0051	0.0481	0.0051	0.0484	0.0465
σ_{22}^P	0.0110	0.0119	0.0111	0.0116	0.0656
σ_{33}^P	0.0263	0.0260	0.0816	0.0961	0.1364
β_{11}^P	-	-	-	-	-
β_{12}^P	-	-	-	-	-
β_{13}^P	-	-	0.0008	-	-
β_{21}^P	-	0.9981	-	1.9755	-
β_{22}^P	-	-	-	-	-
β_{23}^P	-	-	0.0009	-	-
β_{31}^P	-	0.0999	-	0.0006	-
β_{32}^P	-	-	-	-	-
β_{33}^P	-	-	-	-	-
θ_1^Q	-	1965	-	1985	1091
θ_2^Q	-	-	-	-	0.0494
θ_3^Q	-	-	0.1154	0.0800	0.0451
λ	0.5980	0.8190	0.6609	0.8138	0.4942
Max Log L	16280.0	16509.5	16260.1	16496.0	16023.9

Notes: The table reports estimated mean vector under P-dynamics θ^P, the drift matrix K^P, the time-invariant volatility matrix Σ, the volatility sensitivity parameters β, the mean vector under Q-dynamics θ^Q and the λ parameter for different independent-factor specifications of the arbitrage-free Nelson-Siegel (AFNS) models. Estimations are based on monthly yields data from January 1987 to December 2002. The last row reports the maximum log-likelihood values.

Table 3.B.5 RMSE of fitted yields for the Nelson-Siegel Models

Maturity	Nelson-Siegel models with independent factors					
(in months)	DNS	$AFNS_0$	$AFNS_1 - L$	$AFNS_1 - C$	$AFNS_2 - LC$	$AFNS_3$
3	12.26	18.49	11.62	16.17	**11.55**	19.86
6	1.09	7.09	**0.85**	5.40	1.13	7.82
9	7.13	**3.47**	7.28	4.49	7.12	4.99
12	11.19	**9.60**	9.94	9.85	9.77	9.93
18	10.76	10.43	8.40	10.43	**8.23**	10.41
24	5.83	5.93	5.06	5.66	**4.86**	7.21
36	**1.51**	1.98	2.22	1.93	2.44	3.82
48	3.92	**3.72**	4.13	3.80	4.22	3.90
60	7.14	6.82	5.85	7.15	5.84	**4.98**
84	4.25	4.29	4.03	4.37	**4.02**	4.51
96	2.09	2.11	**1.08**	2.15	**1.08**	4.74
108	**2.94**	3.02	4.88	3.03	4.90	6.45
120	8.51	8.23	12.73	**8.06**	12.76	11.37
180	29.45	32.66	16.70	33.4	**16.61**	17.91
240	35.00	42.60	17.36	43.85	17.23	**13.02**
360	37.61	**22.04**	49.81	22.36	50.14	74.74

Notes: The table reports root mean squared errors (RMSEs) of fitted yields with different maturities under independent-factors specification of variants of Nelson-Siegel models. The DNS refers to a standard constant-volatility dynamic Nelson-Siegel model of Diebold and Li (2006), the $AFNS_0$ refers to the constant-volatility arbitrage-free Nelson-Siegel model of Christensen et al. (2011) and $AFNS_i$ refers to the arbitrage-free Nelson-Siegel model with stochastic volatility of Christensen et al. (2010), i denoting the number of factors driving time-varying volatilities and L and C denoting the level and curvature factors, respectively. The bold entries represent minimum RMSEs across maturities. The sample period is January 1987 to December 2002.

Table 3.B.6 Real-time mean forecast errors, Feb 1996 - Dec 2002

	Maturity				
Model	3 month	6 month	1 year	2 year	3 year
1-month-ahead					
$AFNS_0$	-1.778	-3.795	-3.191	-5.840	-4.347
$AFNS_1$-L	3.763	-1.724	-4.393	-6.133	-1.066
$AFNS_1$-C	-2.130	-4.083	-3.387	-5.947	-4.428
$AFNS_2$-LC	3.704	-1.687	-4.281	-6.051	-1.037
$AFNS_3$	3.745	-1.934	-5.140	-7.880**	-3.322
DNS	-5.842**	-7.562**	-6.596*	-8.948**	-7.319**
3-month-ahead					
$AFNS_0$	-10.635	-12.233	-10.911	-12.443	-10.094
$AFNS_1$-L	0.815	-4.634	-7.042	-7.928	-1.959
$AFNS_1$-C	-11.797	-13.305	-11.848	-13.237	-10.831
$AFNS_2$-LC	0.908	-4.641	-7.224	-8.321	-2.409
$AFNS_3$	-3.328	-8.288	-10.571	-12.605	-7.935
DNS	-19.749	-21.422	-20.168	-21.601	-19.001
6-month-ahead					
$AFNS_0$	-21.821	-22.769	-20.392	-20.473	-17.174
$AFNS_1$-L	-0.463	-5.962	-8.269	-8.585	-1.974
$AFNS_1$-C	-24.527	-25.519	-23.200	-23.314	-19.976
$AFNS_2$-LC	-0.454	-6.562	-9.673	-10.602	-4.000
$AFNS_3$	-11.325	-15.644	-17.172	-18.842	-14.371
DNS	-37.881	-39.362	-37.653	-38.116	-34.665
12-month-ahead					
$AFNS_0$	-37.224	-35.959	-30.139	-26.045	-20.664
$AFNS_1$-L	2.827	-0.145	1.204	4.702	12.763
$AFNS_1$-C	-42.327	-41.073	-35.242	-31.031	-25.469
$AFNS_2$-LC	1.847	-2.273	-2.395	0.062	8.169
$AFNS_3$	-18.514	-20.617	-19.221	-18.670	-14.205
DNS	-66.867	-66.100	-60.997	-57.600	-52.433

Notes: The entries are mean forecast errors (actual yields minus forecasts) reported in basis points. '***', '**' and '*' imply significance of a test of zero mean forecast error at 1%, 5% and 10% level of significance, respectively. The test statistic is a t-statistic calculated using Newey-West standard errors.

Table 3.B.7 Real-time mean forecast errors, Feb 1996 - Dec 2002 (Cont.)

	Maturity				
Model	5 year	7 year	10 year	20 year	30 year
1-month-ahead					
$AFNS_0$	-7.461**	-2.905	-10.288***	49.518***	6.393*
$AFNS_1$-L	-0.398	1.640	-16.122***	1.626	-74.702***
$AFNS_1$-C	-7.558**	-3.033	-10.452***	49.305***	6.165*
$AFNS_2$-LC	-0.412	1.634	-16.111***	1.647	-74.690***
$AFNS_3$	-2.723	-0.581	-19.142***	-14.205***	-111.337***
DNS	-10.188***	-5.552	-13.574***	37.154***	-23.696***
3-month-ahead					
$AFNS_0$	-11.931	-6.473	-12.951	48.208***	5.718
$AFNS_1$-L	0.084	2.926	-14.249	3.858	-72.614***
$AFNS_1$-C	-12.658	-7.220	-13.729	47.381***	4.875
$AFNS_2$-LC	-0.317	2.611	-14.477	3.733	-72.710***
$AFNS_3$	-7.362	-5.131	-23.435**	-17.698***	-114.233***
DNS	-20.355	-14.770	-21.986**	29.715***	-30.811***
6-month-ahead					
$AFNS_0$	-17.808	-11.583	-17.309	45.047**	3.251
$AFNS_1$-L	1.020	4.362	-12.514	5.650	-70.937***
$AFNS_1$-C	-20.462	-14.080	-19.631	42.980**	1.271
$AFNS_2$-LC	-0.654	2.996	-13.614	4.861	-71.614***
$AFNS_3$	-14.304	-12.288	-30.561*	-23.952	-119.524***
DNS	-34.830	-28.555	-35.209*	17.157	-43.147***
12-month-ahead					
$AFNS_0$	-19.801	-13.282	-19.000	43.535	2.409
$AFNS_1$-L	16.189	19.096	1.394	17.446	-60.638***
$AFNS_1$-C	-24.235	-17.404	-22.799	40.180	-0.800
$AFNS_2$-LC	12.311	15.811	-1.368	15.381	-62.375***
$AFNS_3$	-15.786	-15.324	-35.047	-29.723	-125.095***
DNS	-51.647	-45.389	-52.290	-0.362	-60.819***

Notes: The entries are mean forecast errors (actual yields minus forecasts) reported in basis points. '***', '**' and '*' imply significance of a test of zero mean forecast error at 1%, 5% and 10% level of significance, respectively. The test statistic is a t-statistic calculated using Newey-West standard errors.

Table 3.B.8 Real-time root mean squared forecast errors, Feb 1996 - Dec 2002

	Maturity				
Model	3 month	6 month	1 year	2 year	3 year
1-month-ahead					
DNS	23.762	26.465	31.006	34.170	31.577
$AFNS_0$	0.964	0.960**	0.990	0.994	0.985
$AFNS_1$-L	1.004	0.961	0.963**	0.964**	0.971
$AFNS_1$-C	0.958	0.956**	0.982	0.982	0.971**
$AFNS_2$-LC	1.004	0.959	0.960**	0.963**	0.972
$AFNS_3$	0.978	0.969	0.994	0.997	0.990
3-month-ahead					
DNS	57.865	63.178	67.123	68.763	62.618
$AFNS_0$	0.946	0.958	0.974	0.976	0.973
$AFNS_1$-L	0.957	0.949	0.957	0.957	0.961
$AFNS_1$-C	0.940*	0.952	0.964	0.963	0.958
$AFNS_2$-LC	0.954	0.944	0.949	0.947	0.952
$AFNS_3$	0.958	0.969	0.987	0.982	0.977
6-month-ahead					
DNS	101.817	106.726	108.57	104.562	94.269
$AFNS_0$	0.963	0.967	0.972	0.966	0.960
$AFNS_1$-L	0.943	0.941	0.951	0.949	0.949
$AFNS_1$-C	0.955	0.960	0.966	0.961	0.957
$AFNS_2$-LC	0.940	0.936	0.943	0.937	0.939
$AFNS_3$	0.949	0.957	0.972	0.967	0.960
12-month-ahead					
DNS	174.863	176.841	171.870	153.302	133.706
$AFNS_0$	0.984	0.988	0.995	0.994	0.990
$AFNS_1$-L	0.921	0.924	0.934	0.928	0.924
$AFNS_1$-C	0.969	0.971	0.975	0.971	0.966
$AFNS_2$-LC	0.919	0.919	0.923	0.912	0.908
$AFNS_3$	0.898	0.906	0.922	0.919	0.910

Notes: The first row of each panel reports root mean squared forecast errors (in basis points) for the benchmark DNS. The rest of the rows report RMSFE relative to DNS (RRMSFE). '**' and '*' imply significance of the Giacomini and White (2006) test of equal MSFE of a model and the benchmark at 1% and 5% level of significance, respectively. The standard errors of the t test statistics are computed with the Newey-West estimator.

Table 3.B.9 Real-time root mean squared forecast errors, Feb 1996 - Dec 2002 (Cont.)

	Maturity				
Model	5 year	7 year	10 year	20 year	30 year
1-month-ahead					
DNS	31.379	28.858	28.217	47.893	37.534
$AFNS_0$	0.970**	0.987	0.959**	1.207**	0.829
$AFNS_1$-L	0.943	0.976	1.037**	0.586**	2.144**
$AFNS_1$-C	0.959**	0.979	0.952**	1.201**	0.819
$AFNS_2$-LC	0.947	0.981	1.037**	0.586**	2.141**
$AFNS_3$	0.961	0.995	1.109**	0.610**	3.048**
3-month-ahead					
DNS	59.053	53.087	48.865	51.135	47.526
$AFNS_0$	0.965	0.979	0.946	1.256**	0.826
$AFNS_1$-L	0.944	0.963	0.934*	0.761*	1.705**
$AFNS_1$-C	0.953	0.967	0.938*	1.235**	0.807
$AFNS_2$-LC	0.940	0.960	0.932*	0.757*	1.701**
$AFNS_3$	0.958	0.971	1.006*	0.786	2.498**
6-month-ahead					
DNS	87.412	77.755	71.252	58.054	64.889
$AFNS_0$	0.952	0.959	0.921	1.255	0.795
$AFNS_1$-L	0.932	0.937	0.881*	0.879	1.296**
$AFNS_1$-C	0.950	0.954	0.919	1.217	0.775
$AFNS_2$-LC	0.926	0.932	0.881*	0.874	1.299**
$AFNS_3$	0.936	0.933	0.943	0.907	1.946**
12-month-ahead					
DNS	117.674	104.695	97.387	70.613	85.278
$AFNS_0$	0.974	0.973	0.921	1.221	0.747
$AFNS_1$-L	0.896	0.892	0.795	0.899	0.914
$AFNS_1$-C	0.953	0.951	0.907	1.181	0.731
$AFNS_2$-LC	0.886	0.885	0.797	0.896	0.926
$AFNS_3$	0.880	0.867*	0.854*	0.904	1.564**

Notes: The first row of each panel reports root mean squared forecast errors (in basis points) for the benchmark DNS. The rest of the rows report RMSFE relative to DNS (RRMSFE). '**' and '*' imply significance of the Giacomini and White (2006) test of equal MSFE of a model and the benchmark at 1% and 5% level of significance, respectively. The standard errors of the t test statistics are computed with the Newey-West estimator.

Table 3.B.10 Real-time forecast coverage Rates, Feb 1996 - Dec 2002

	Maturity									
Model	3 month	6 month	1 year	2 year	3 year	5 year	7 year	10 year	20 year	30 year
1-month-ahead										
$AFNS_0$	0.928	0.867	0.771	0.675	0.675	0.614	0.566	0.542	0.530	0.614
$AFNS_1$-L	0.976	0.940	0.880	0.843	0.795	0.807	0.759	0.747	0.807	0.313
$AFNS_1$-C	0.916	0.867	0.771	0.663	0.663	0.614	0.566	0.542	0.53	0.627
$AFNS_2$-LC	0.976	0.952	0.880	0.843	0.807	0.795	0.771	0.747	0.795	0.313
$AFNS_3$	0.952	0.916	0.855	0.759	0.735	0.663	0.602	0.530	0.506	0.108
DNS	0.94	0.916	0.807	0.735	0.723	0.723	0.675	0.614	0.614	0.747
3-month-ahead										
$AFNS_0$	0.802	0.704	0.593	0.568	0.568	0.556	0.519	0.519	0.543	0.630
$AFNS_1$-L	0.864	0.864	0.802	0.716	0.728	0.716	0.704	0.802	0.790	0.506
$AFNS_1$-C	0.802	0.716	0.605	0.593	0.593	0.543	0.543	0.506	0.556	0.642
$AFNS_2$-LC	0.864	0.864	0.815	0.741	0.716	0.728	0.741	0.790	0.815	0.494
$AFNS_3$	0.864	0.765	0.716	0.667	0.642	0.568	0.568	0.469	0.407	0.148
DNS	0.827	0.778	0.716	0.667	0.679	0.617	0.654	0.667	0.728	0.716
6-month-ahead										
$AFNS_0$	0.654	0.577	0.513	0.449	0.436	0.462	0.410	0.449	0.692	0.577
$AFNS_1$-L	0.769	0.679	0.603	0.564	0.615	0.667	0.705	0.833	0.859	0.564
$AFNS_1$-C	0.679	0.564	0.513	0.474	0.449	0.410	0.410	0.423	0.692	0.564
$AFNS_2$-LC	0.769	0.679	0.603	0.564	0.641	0.705	0.731	0.833	0.859	0.564
$AFNS_3$	0.718	0.628	0.538	0.500	0.513	0.474	0.487	0.385	0.308	0.205
DNS	0.731	0.654	0.603	0.500	0.513	0.474	0.526	0.564	0.782	0.577
12-month-ahead										
$AFNS_0$	0.486	0.458	0.458	0.417	0.458	0.417	0.403	0.431	0.694	0.653
$AFNS_1$-L	0.569	0.528	0.500	0.583	0.667	0.736	0.806	0.847	0.903	0.736
$AFNS_1$-C	0.486	0.472	0.458	0.431	0.458	0.361	0.403	0.431	0.708	0.639
$AFNS_2$-LC	0.569	0.514	0.514	0.611	0.667	0.736	0.806	0.861	0.903	0.736
$AFNS_3$	0.569	0.556	0.528	0.514	0.542	0.486	0.486	0.389	0.389	0.181
DNS	0.528	0.514	0.486	0.472	0.486	0.375	0.431	0.444	0.806	0.556

Notes: The table reports coverage probabilities or proportion of actual yields which fall within 70% intervals. The upper and lower bounds of the interval are the 85th and 15th percentiles of the predictive densities.

Table 3.B.11 Tests of normalised errors of 1-month-ahead forecasts, Feb 1996 - Dec 2002

	Maturity									
Model	3 month		6 month		1 year		2 year		3 year	
(a) Mean										
$AFNS_0$	-0.055	(-0.751)	-0.107	(-1.199)	-0.038	(-0.354)	-0.117	(-0.922)	-0.085	(-0.680)
$AFNS_1$-L	0.096	(1.673)	-0.031	(-0.488)	-0.079	(-1.028)	-0.118	(-1.315)	0.001	(0.014)
$AFNS_1$-C	-0.062	(-0.841)	-0.120	(-1.307)	-0.046	(-0.428)	-0.125	(-1.000)	-0.090	(-0.730)
$AFNS_2$-LC	0.089	(1.560)	-0.041	(-0.630)	-0.091	(-1.185)	-0.133	(-1.468)	-0.011	(-0.122)
$AFNS_3$	0.114	(1.853)	-0.092	(-1.144)	-0.212*	(-1.968)	-0.257*	(-2.022)	-0.099	(-0.737)
DNS	-0.140*	(-2.228)	-0.191*	(-2.571)	-0.125	(-1.409)	-0.192	(-1.864)	-0.168	(-1.654)
(b) Variance										
$AFNS_0$	0.447**	(-3.948)	0.654	(-1.480)	0.967	(-0.175)	1.322	(1.079)	1.267	(0.988)
$AFNS_1$-L	0.269**	(-10.721)	0.340**	(-6.038)	0.486**	(-3.487)	0.664	(-1.830)	0.673	(-1.694)
$AFNS_1$-C	0.442**	(-4.041)	0.686	(-1.206)	0.953	(-0.225)	1.285	(0.962)	1.234	(0.873)
$AFNS_2$-LC	0.267**	(-10.885)	0.341**	(-5.779)	0.482**	(-3.366)	0.670	(-1.624)	0.690	(-1.290)
$AFNS_3$	0.311**	(-6.752)	0.530*	(-1.996)	0.950	(-0.037)	1.324	(0.575)	1.483	(0.525)
DNS	0.326**	(-5.960)	0.453**	(-2.985)	0.643	(-1.824)	0.866	(-0.492)	0.845	(-0.661)
(c) AR(1) Coeff										
$AFNS_0$	0.341**	(4.885)	0.447**	(4.866)	0.553**	(6.425)	0.477**	(4.685)	0.279*	(2.384)
$AFNS_1$-L	0.395**	(5.664)	0.478**	(5.797)	0.536**	(6.194)	0.416**	(3.624)	0.244	(1.813)
$AFNS_1$-C	0.323**	(4.766)	0.426**	(4.777)	0.540**	(6.266)	0.462**	(4.412)	0.255*	(2.163)
$AFNS_2$-LC	0.391**	(5.742)	0.464**	(5.809)	0.523**	(5.915)	0.391**	(3.236)	0.224	(1.585)
$AFNS_3$	0.413**	(6.990)	0.481**	(4.267)	0.567**	(5.441)	0.411**	(2.874)	0.201	(1.131)
DNS	0.357**	(5.700)	0.463**	(5.795)	0.549**	(6.598)	0.444**	(4.191)	0.263*	(2.172)
(d) LR test										
$AFNS_0$	0.000	(0.000)	0.000	(0.001)	0.000	(0.056)	0.000	(0.921)	0.026	(0.516)
$AFNS_1$-L	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.002)	0.012	(0.020)
$AFNS_1$-C	0.000	(0.000)	0.000	(0.003)	0.000	(0.060)	0.000	(0.881)	0.050	(0.535)
$AFNS_2$-LC	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.003)	0.024	(0.033)
$AFNS_3$	0.000	(0.000)	0.000	(0.000)	0.000	(0.072)	0.000	(0.354)	0.013	(0.048)
DNS	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.059)	0.021	(0.129)

Notes: The first panel reports means of normalised errors along with a t-statistic (computed using a Newey-West standard error) for testing the null of zero mean. The second panel reports variance of normalised errors along with a t-statistic (computed by a linear regression of the squared error on a constant, using Newey-West standard error) for testing the null of unit variance. The third panel reports the AR(1) coefficients and t-statistics of its significance, obtained by estimating an AR(1) model with an intercept (with Newey-West standard errors). The fourth column reports the p-values of two Berkowitz (2001) tests: the first for the likelihood ratio test for the joint null of a zero mean, unit variance and no autocorrelation and the second for only zero mean and unit variance. '**' and '*' imply significance at 1% and 5%, respectively. All test statistics are reported in parentheses.

Table 3.B.12 Tests of normalised errors of 1-month-ahead forecasts, Feb 1996 - Dec 2002 (Cont.)

	Maturity									
Model	5 year		7 year		10 year		20 year		30 year	
(a) Mean										
$AFNS_0$	-0.253	(-1.920)	-0.073	(-0.488)	-0.430**	(-2.969)	1.097**	(15.068)	0.320*	(2.561)
$AFNS_1$-L	0.004	(0.043)	0.073	(0.758)	-0.405**	(-4.808)	0.105	(1.056)	-1.251**	(-24.109)
$AFNS_1$-C	-0.257*	(-1.976)	-0.080	(-0.546)	-0.440**	(-3.067)	1.096**	(15.115)	0.315*	(2.568)
$AFNS_2$-LC	-0.009	(-0.098)	0.062	(0.638)	-0.410**	(-4.846)	0.099	(1.003)	-1.251**	(-24.408)
$AFNS_3$	-0.084	(-0.592)	0.035	(0.207)	-0.788**	(-5.654)	-0.654**	(-4.391)	-1.443**	(-40.356)
DNS	-0.302**	(-2.830)	-0.160	(-1.435)	-0.439**	(-4.302)	0.848**	(11.201)	-0.570**	(-7.287)
(b) Variance										
$AFNS_0$	1.418	(1.730)	1.846**	(2.900)	1.719**	(3.201)	0.435	(1.792)	1.281*	(2.273)
$AFNS_1$-L	0.734	(-1.301)	0.763	(-1.330)	0.581	(-1.605)	0.804	(-1.240)	0.221*	(2.557)
$AFNS_1$-C	1.383	(1.635)	1.776**	(2.810)	1.685**	(3.000)	0.431	(1.817)	1.235	(2.097)
$AFNS_2$-LC	0.754	(-0.977)	0.780	(-1.047)	0.587	(-1.442)	0.793	(-1.369)	0.216**	(2.606)
$AFNS_3$	1.668	(0.866)	2.295	(1.501)	1.595*	(2.326)	1.817**	(3.293)	0.105**	(5.010)
DNS	0.933	(0.055)	1.025	(0.205)	0.852	(0.215)	0.470	(0.613)	0.501	(-0.719)
(c) AR(1) Coeff										
$AFNS_0$	0.241*	(2.137)	0.240*	(2.304)	0.150	(1.642)	0.445**	(4.451)	0.465**	(5.346)
$AFNS_1$-L	0.226	(1.648)	0.184	(1.509)	0.105	(0.978)	0.343**	(3.702)	0.388**	(4.032)
$AFNS_1$-C	0.224*	(1.994)	0.228*	(2.149)	0.139	(1.493)	0.437**	(4.448)	0.461**	(5.267)
$AFNS_2$-LC	0.224	(1.557)	0.181	(1.434)	0.090	(0.840)	0.339**	(3.602)	0.384**	(3.956)
$AFNS_3$	0.211	(1.340)	0.201	(1.478)	0.217*	(2.034)	0.384**	(4.038)	0.427**	(5.104)
DNS	0.248*	(2.088)	0.226*	(2.039)	0.111	(1.203)	0.394**	(3.924)	0.400**	(4.054)
(d) LR test										
$AFNS_0$	0.001	(0.031)	0.000	(0.001)	0.000	(0.000)	0.000	(0.000)	0.000	(0.045)
$AFNS_1$-L	0.049	(0.080)	0.107	(0.140)	0.000	(0.000)	0.004	(0.079)	0.000	(0.000)
$AFNS_1$-C	0.003	(0.034)	0.000	(0.002)	0.000	(0.000)	0.000	(0.000)	0.000	(0.050)
$AFNS_2$-LC	0.064	(0.110)	0.140	(0.183)	0.000	(0.000)	0.004	(0.068)	0.000	(0.000)
$AFNS_3$	0.001	(0.006)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
DNS	0.004	(0.063)	0.080	(0.407)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)

Notes: The first panel reports means of normalised errors along with a t-statistic (computed using a Newey-West standard error) for testing the null of zero mean. The second panel reports variance of normalised errors along with a t-statistic (computed by a linear regression of the squared error on a constant, using Newey-West standard error) for testing the null of unit variance. The third panel reports the AR(1) coefficients and t-statistics of its significance, obtained by estimating an AR(1) model with an intercept (with Newey-West standard errors). The fourth column reports the p-values of two Berkowitz (2001) tests: the first for the likelihood ratio test for the joint null of a zero mean, unit variance and no autocorrelation and the second for only zero mean and unit variance. '**' and '*' imply significance at 1% and 5%, respectively. All test statistics are reported in parentheses.

Table 3.B.13 Real-time log predictive scores, Feb 1996 - Dec 2002

	Maturity									
Model	3 month	6 month	1 year	2 year	3 year	5 year	7 year	10 year	20 year	30 year
1-month-ahead										
$AFNS_0$	18.484	18.487	18.781	19.047	18.870	18.871	18.887	18.805	20.270	18.891
$AFNS_1$-L	18.918	18.899	19.042	19.144	19.036	18.943	18.852	18.847	18.744	20.769
$AFNS_1$-C	18.479	18.486	18.766	19.03	18.853	18.858	18.881	18.796	20.258	18.866
$AFNS_2$-LC	18.918	18.904	19.057	19.177	19.083	18.983	18.878	18.863	18.740	20.771
$AFNS_3$	18.782	18.782	19.238	19.533	19.471	19.403	19.613	19.386	19.119	21.762
DNS	18.681	18.676	18.853	19.019	18.887	18.871	18.714	18.654	19.698	19.282
3-month-ahead										
$AFNS_0$	20.002	20.259	20.503	20.697	20.456	20.438	20.397	20.164	20.483	19.371
$AFNS_1$-L	20.210	20.289	20.402	20.459	20.310	20.167	20.000	19.812	19.477	20.662
$AFNS_1$-C	19.985	20.239	20.467	20.638	20.388	20.36	20.318	20.107	20.429	19.322
$AFNS_2$-LC	20.204	20.286	20.403	20.465	20.326	20.191	20.012	19.820	19.474	20.679
$AFNS_3$	20.099	20.320	20.722	20.880	20.681	20.642	20.616	20.661	20.141	21.942
DNS	20.092	20.235	20.377	20.494	20.309	20.236	20.003	19.824	19.830	19.758
6-month-ahead										
$AFNS_0$	21.467	21.773	21.940	21.868	21.522	21.502	21.349	21.046	20.703	19.972
$AFNS_1$-L	21.182	21.273	21.354	21.293	21.099	20.923	20.702	20.445	20.056	20.731
$AFNS_1$-C	21.448	21.760	21.930	21.844	21.483	21.442	21.276	20.998	20.625	19.919
$AFNS_2$-LC	21.179	21.269	21.351	21.296	21.109	20.936	20.709	20.456	20.052	20.745
$AFNS_3$	21.156	21.415	21.773	21.811	21.555	21.441	21.290	21.388	20.810	22.291
DNS	21.301	21.479	21.576	21.533	21.271	21.156	20.837	20.625	20.070	20.340
12-month-ahead										
$AFNS_0$	23.752	24.135	24.221	23.662	22.993	22.632	22.312	21.889	21.052	20.305
$AFNS_1$-L	22.385	22.469	22.460	22.133	21.794	21.469	21.230	20.936	20.587	20.726
$AFNS_1$-C	23.689	24.055	24.128	23.556	22.889	22.531	22.215	21.832	20.962	20.264
$AFNS_2$-LC	22.384	22.463	22.445	22.116	21.785	21.469	21.234	20.948	20.593	20.749
$AFNS_3$	22.240	22.457	22.710	22.645	22.363	22.224	21.993	21.926	20.743	22.740
DNS	23.107	23.307	23.279	22.823	22.306	21.950	21.573	21.385	20.650	20.907

Notes: The table reports log predictive density scores calculated with a Gaussian quadratic formula given in Adolfson et al. (2005) and used in Clark (2011). The lower the score the better is the forecast model.

Table 3.B.14 Differences in log predictive scores, Feb 1996 - Dec 2002

	Maturity				
Model	3 month	6 month	1 year	2 year	3 year
1-month-ahead					
$AFNS_0$	-0.197**	-0.189**	-0.072	0.028	-0.017
$AFNS_1$-L	0.237**	0.223**	0.189**	0.125**	0.149**
$AFNS_1$-C	-0.201**	-0.190**	-0.087	0.011	-0.034
$AFNS_2$-LC	0.238**	0.228**	0.203**	0.158**	0.196**
$AFNS_3$	0.101**	0.105	0.384**	0.514**	0.585**
3-month-ahead					
$AFNS_0$	-0.090	0.023	0.126	0.203	0.147
$AFNS_1$-L	0.118	0.054	0.026	-0.035	0.001
$AFNS_1$-C	-0.107	0.003	0.091	0.144	0.079
$AFNS_2$-LC	0.113	0.051	0.026	-0.029	0.017
$AFNS_3$	0.007	0.085	0.345*	0.386	0.372
6-month-ahead					
$AFNS_0$	0.166	0.294	0.364	0.335	0.251
$AFNS_1$-L	-0.119	-0.206	-0.222	-0.240	-0.173
$AFNS_1$-C	0.147	0.281	0.353	0.311	0.211
$AFNS_2$-LC	-0.122	-0.210	-0.226	-0.237	-0.162
$AFNS_3$	-0.145	-0.064	0.196	0.278	0.284
12-month-ahead					
$AFNS_0$	0.645	0.828	0.942	0.839	0.687
$AFNS_1$-L	-0.722	-0.837	-0.819	-0.690	-0.512
$AFNS_1$-C	0.582	0.749	0.849	0.733	0.583
$AFNS_2$-LC	-0.724	-0.844	-0.834	-0.707	-0.521
$AFNS_3$	-0.868	-0.849	-0.569	-0.178	0.057

Notes: The table reports the difference in log scores of a model and that of the DNS benchmark. '**' and '*' imply significance of the Giacomini and White (2006) test of equal log predictive scores at 1% and 5% level of significance, respectively. The standard errors of the t test statistics are computed with the Newey-West estimator.

Table 3.B.15 Differences in log predictive scores, Feb 1996 - Dec 2002 (Cont.)

	Maturity				
Model	5 year	7 year	10 year	20 year	30 year
1-month-ahead					
$AFNS_0$	0.000	0.174	0.150	0.573**	-0.391*
$AFNS_1$-L	0.072	0.138**	0.193**	-0.953**	1.487**
$AFNS_1$-C	-0.013	0.167	0.142	0.560**	-0.415*
$AFNS_2$-LC	0.112	0.164**	0.209**	-0.957**	1.490**
$AFNS_3$	0.532**	0.899**	0.732**	-0.579	2.480**
3-month-ahead					
$AFNS_0$	0.202	0.394*	0.340	0.653**	-0.387
$AFNS_1$-L	-0.068	-0.003	-0.013	-0.353*	0.904**
$AFNS_1$-C	0.125	0.315	0.282	0.599**	-0.436
$AFNS_2$-LC	-0.045	0.009	-0.004	-0.356*	0.921**
$AFNS_3$	0.406	0.613*	0.837*	0.310	2.184**
6-month-ahead					
$AFNS_0$	0.346	0.513	0.421	0.633	-0.369
$AFNS_1$-L	-0.234	-0.135	-0.180	-0.013	0.391*
$AFNS_1$-C	0.285	0.440	0.373	0.556	-0.421
$AFNS_2$-LC	-0.220	-0.127	-0.170	-0.017	0.404**
$AFNS_3$	0.284	0.454	0.763	0.740	1.950**
12-month-ahead					
$AFNS_0$	0.682	0.739	0.504	0.401	-0.602
$AFNS_1$-L	-0.481	-0.342	-0.449	-0.064*	-0.181
$AFNS_1$-C	0.581	0.642	0.447	0.312	-0.643
$AFNS_2$-LC	-0.481	-0.339	-0.437	-0.058*	-0.158
$AFNS_3$	0.274	0.420	0.541	0.092	1.833**

Notes: The table reports the difference in log scores of a model and that of the DNS benchmark. '**' and '*' imply significance of the Giacomini and White (2006) test of equal log predictive scores at 1% and 5% level of significance, respectively. The standard errors of the t test statistics are computed with the Newey-West estimator.

Chapter 4

Forecasting under Structural Change and Long Memory Noise

4.1 Introduction

Dealing with structural change has become one of the most crucial challenges in economic and financial time series modelling and forecasting. In econometrics structural change usually refers to evolution of a parameter of interest of a dynamic model that makes its estimation and/or prediction unstable. The change can be as dramatic as an abrupt shift or permanent break induced, for example, by introduction of a new monetary policy, breakdown of an exchange rate regime or even sudden rise in oil price. In other instances, the change can be slow, smooth and continuous caused for example by gradual progress in technology or production. Empirical evidence of structural change is widespread and well-documented in economic and finance literature. Stock and Watson (1996) investigate many US macroeconomic time series and find instability in both univariate and bivariate relationships by applying standard instability tests and out-of-sample forecast exercises. In finance structural changes are detected in interest rates (e.g., Garcia and Perron (1996), Ang and Bekaert (2002)) and stock prices and returns (e.g., Timmermann (2001), Pesaran and Timmermann (2002)). Such structural change or parameter instability has been identified as one of the main culprits of forecast failures (see Clements and Hendry (1996, 1998), Hendry (2000)) and not surprisingly detection of breaks and forecast strategies in the presence of breaks have attracted a lot of attention from researchers. Nonetheless, real time forecasting of time series which are subject to structural change remains to be a critical challenge to date and is often complicated further by presence of other features of time series such as persistence. Rossi (2012) provides a comprehensive review of strategies that have been developed over the last few decades to tackle the problem of forecasting in face of unforeseen structural changes.

A natural strategy for forecasting in unstable environment would be finding the last change point and using only the post-break data for estimating a model and forecasting. But such strategies may be problematic for various reasons. First, standard tests of structural breaks are not suitable for real time forecasting. Research on break detection tests has gone through significant refinements, e.g., from cases with known single break (Chow (1960)) to unknown multiple breaks (e.g., Andrews (1993), Bai and Perron (1998, 2003)). However, most of these tests require some time to be elapsed after the break for it to be detected and assume that the required post break data is break-free. This makes timely detection of a break almost impossible. Another major criticism of conventional tests is that they are retrospective by nature. This means that they are specifically designed for detecting breaks over a historical sample of given size and are problematic for repetitive use with new arrivals of data. These shortcomings of classical tests prompt emergence of a second class of tests based on sequential testing in statistics literature (see Chu et al.

(1996), Leisch et al. (2000), Zeileis et al. (2005)). These forward looking fluctuation type tests monitor for structural breaks with appearance of new data and are more appropriate for real-time forecasting facing structural changes. Issues with monitoring tests, however, remain. There are still delays, though in smaller margins, in identifying breaks which means that detection of unknown frequent breaks is difficult and estimation of timing of a break is not precise. In addition, small breaks are difficult to track. Second, the amount of post-break data may simply be insufficient for stable estimation of model parameters and consequently, for reliable prediction. Moreover, Pesaran and Timmermann (2007) point out that a trade-off between bias and forecast error variance implies that it is not always optimal to use only post-break data and it is generally beneficial to include some pre-break information.

A second line of strategies involves a more econometric approach that involves formally modelling the break process itself and estimating its characteristics such as timing, size and duration based on historical behaviour of a series. A standard model of this kind is the Markov-switching model of Hamilton (1989) which makes probabilistic inference of whether and when unknown switching of regimes or equivalently, shifts in parameters may occur. While this is a ground-breaking proposition for making stable inference in presence of structural change predictive performance of such models have seriously been questioned. Clements and Krolzig (1998) demonstrate via a Monte Carlo study that despite the true data generating process being Markov-switching regime switching models fail to forecast as accurately as a simple linear AR(1) model in many instances. Research on modelling of structural breaks has continued to evolve rapidly and recent literature records successful forecasting stories of many sophisticated models, mainly founded on Bayesian methods (e.g., Pesaran et al. (2006), Koop and Potter (2007), Giordani and Kohn (2008), Maheu and Gordon (2008)). In one way or another, these models learn about change-points from the past and exploit these information as priors in modelling and forecasting future breaks. A debate that often arises in this context is whether it is convenient to restrict the number of breaks occurring in-sample fixed or to treat it as unknown. A class of models that avoids this argument is the so called time varying parameter (TVP) models which assume that a change occurs each point in time (e.g., Stock and Watson (2007), D'Agostino et al. (2013)). Most of these models document evidence of impressive empirical forecasting ability when evaluated individually. Bauwens et al. (2011) run a horse-race by comparing predictive performance of several of these break models along with simple rolling and recursive regressions. While modelling break processes has a clear advantage in terms of root mean squared forecast error (RMSFE), no single model has been identified to produce superior forecasts consistently. Moreover, when performance is evaluated on the basis of average predictive likelihood criterion a simple break-free method of rolling regression enjoys an upper-hand on most of the break models. Difficulty in finding a single best forecasting model shifts attention to combining forecasts of different models. A well-performing model can forecast badly after a break while a previously poor-performing model can do better. Thus, pooled forecast can result in the least mean squared forecast error (MSFE) even though it was never the best at each point in time. Empirical evidence in the literature strongly speaks for simple combination rules such as averaging forecasts with equal weights across all models. Bayesian Model Averaging (BMA) has also been found to forecast well while forecast combination with time varying weights experiences little success (Rossi (2012)). Clark and McCracken (2010) evaluate ability of different forecast combination strategies in improving forecast accuracy of small-scale macroeconomic VARs in the presence of uncertain forms of model instability. They combine forecasts of many ad hoc strategies designed for tackling structural breaks which include differencing, detrending, intercept corrections, sequential updating of lag orders, estimation with different window lengths, Bayesian shrinkage, among others. They conclude that although BMA provides occasional large forecast improvements, model averaging with equal weights provides good forecast consistency. A recent interesting proposition comes from Castle et al. (2011) who argue that with availability of rich and broad information set even breaks of unknown nature can be predicted. Additional data which are not all directly related to economic phenomena can be potentially used to explain driving forces behind a break. These authors use an automatic model selection approach in order to efficiently exploit a data-rich environment. Simulation study reveals that even though a break is difficult to predict accurately it can be tracked well after its occurrence. Nonetheless, predictive performance is not significantly better than robust mechanisms such as differencing and intercept corrections.

An alternative forecasting approach which earns renewed attention in the literature is adopting methods that do not require any knowledge of structural breaks but are actually robust to them. This class of methods builds on downweighting past information and includes forecasting with rolling window, exponential smoothing or exponentially weighted moving average (EWMA), forecast pooling with window averaging etc. These simple strategies are particularly attractive because they are easy to implement, possibly robust to different types of structural changes and can adjust for breaks without delay which is particularly helpful for real time forecasting. On the downside discarding old data by selecting a fixed discounting rate *a priori* may prove costly when the true data generating process (DGP) is break-free. A significant contribution in this respect is due to Pesaran and Timmermann (2007). These authors advocate two robust strategies: one is selecting a single window by cross-validation based on pseudo-out-of-sample losses and the other is pooling forecasts of the same model constructed over estimation windows of different sizes. They argue that the former may work well in case of well-defined and large breaks while the latter should perform well in situations where the breaks are mild and hence difficult to detect. They provide Monte Carlo results to show that simple forecast averaging performs particularly well when the underlying economic relations are subject to structural breaks. Assenmacher-Wesche and Pesaran (2009) compute forecasts of many economic series of Swiss economy by using vector autoregressive models estimated over windows of different lengths and find that averaging forecasts across windows leads to improvements over largely popular strategy of averaging of forecasts across models. Clark and McCracken (2009) find that averaging forecasts of expanding and rolling windows can be beneficial in presence of structural breaks. Pesaran and Pick (2011) derive theoretical results for random walk and linear regression models proving that averaging over different estimation windows leads to lower bias and smaller RMSFE. They confirm their findings with a successful simulation study and an application to equity returns data. By comparing window-averaging forecast with the exponential smoothing forecast they conclude that the latter is more sensitive to choice of downweighting parameter than the former is to the choice of minimum estimation window.

The issue of structural change occurring in real time and the challenge it poses for time series forecasting are partly but systematically addressed in Eklund et al. (2010). They consider and compare two different approaches to tackle the problem. One requires monitoring for structural breaks and combining forecasts of models estimated using all available data or only post-break data. The other exploits data-downweighting break-robust methods as mentioned previously. On the basis of their Monte Carlo and empirical analysis they establish that the monitoring method appears to be a conservative strategy in the sense that neither its forecast gains nor its losses against a full sample benchmark are substantial. In addition, performance of the rolling window and exponentially weighted moving average (EWMA) methods is sensitive to the choice of window lengths and discount parameters, and averaging forecasts across estimation windows of various sizes performs consistently well in cases where breaks are frequent and small.

One crucial question that Eklund et al. (2010) do not answer is how much to downweight older data. Moreover, their work, like much of the forecasting literature, confines attention solely to structural breaks. The challenge of forecasting under recent and ongoing structural change has been dealt perhaps in the most comprehensive and generic setting in a recent work of Giraitis et al. (2013). Alongside breaks these authors consider various other types of structural changes including deterministic and stochastic trends and smooth cycles. They do not explicitly model the structural change but exploit the typical data-discounting robust-to-break methods such as rolling window, EWMA, forecast averaging over different windows and various extensions of them. Importantly, they make the selection of the tuning parameter which defines the discounting weights data-dependent by minimising the forecast mean squared error. They provide detailed theoretical and simulation analysis of their proposal and convincing evidence of good performance of methods with data-selected discount rate when applied to a number of US macroeconomic and financial time series.

Giraitis et al. (2013) consider persistence through short memory autoregressive dependence in noise process, but they do not explore possibility of long memory which is often considered as a common but crucial property of many economic and financial time series. This chapter extends the work of Giraitis et al. (2013) by offering a more complex, yet realistic forecasting environment where structural change in a dynamic model is accompanied by noises with long range dependence. We consider several simple cases, such as a stationary long memory process and a combination of linear trend and long memory noise and prove theoretically that forecasts generated with a data-tuned downweighting parameter are asymptotically equivalent to optimal fixed value forecasts. Robust methods with data-dependent tuning parameters which have been found to be useful for forecasting time series with short memory noise are then empirically evaluated on their predictive abilities by forecasting time series with various types of structural changes and different levels of long memory persistence.

While this chapter adds a new dimension to the existing challenge of real time forecasting under structural changes it further contributes to an interesting and ongoing argument in econometric literature about possible ‘spurious’ relationship between long range dependence and structural change and potential forecasting difficulties this may create. Many researchers argue that presence of long memory in the data can be easily confused with structural change (e.g., Diebold and Inoue (2001), Gourieroux and Jasiak (2001), Granger and Hyung (2004)). This aggravates the already difficult problem of forecasting under structural change further. Wrongly accounting for one when the other is present or acknowledging only one when both are present may lead to serious forecast failure. Given that it is often difficult to distinguish between the two, it is desirable to establish forecast methods that are robust to structural change and also appropriately account for long memory persistence. Our work is a potential contribution to this end.

The rest of the chapter is structured as follows. Section 4.2 introduces the dynamic model to be forecast and reviews the forecast strategies proposed in Giraitis et al. (2013). We discuss in details how the tuning parameter defining the rate of downweighting is optimally selected from data and how forecasts are constructed based on such data-dependent selection. Section 4.3 contains theoretical justifications of asymptotic optimality of forecasts based on data-tuned discounting strategies. As mentioned before, we discuss only a few specific cases involving long range dependence. In section 4.4 we present Monte Carlo evidence evaluating performance of robust forecast strategies. Section 4.5 justifies practical usefulness of the methods through applications in forecasting a number of UK economic and financial time series. Section 4.6 concludes.

4.2 Econometric Framework

4.2.1 Forecast Strategy

We adopt the forecast settings of Giraitis et al. (2013) who entertain a simple but general location model as given by:

$$y_t = \beta_t + u_t, \qquad t = 1, 2, ..., T \tag{4.2.1}$$

where y_t is the variable to be forecast, β_t is a persistent process of unknown type and u_t is a dependent noise that is independent of β_t. Unlike most of the previous works which focus mainly on structural breaks, this framework offers more flexibility and generality in the sense that it does not impose any structure on β_t and allows it to adopt many other possible structural changes such as deterministic (bounded) and stochastic (unit root) trends.

While Giraitis et al. (2013) specify u_t to be stationary and dependent through a short memory autoregressive process, we contribute by exploring possible long range dependence and non-stationarity in the noise process. Several standard definitions of short and long memory can be drawn from the statistical literature. A time-domain definition says that auto-covariances $\gamma_u(k)$=Cov(u_{t+k}, u_t) of a short memory process u_t are absolutely summable, i.e., $\sum_{k=1}^{\infty} |\gamma_u(k)| < \infty$. The long memory, on the other hand, is defined as the slow decay of $\gamma_u(k) \sim c_\gamma k^{-1+2d}$, as lag length k increases, for some $0 < d < 1/2$ and $c_\gamma > 0$. Unlike short memory, the autocorrelations of long memory processes are non-summable. In frequency domain, long memory would imply explosive low-frequency spectra i.e., the spectral density of u_t, $f_u(\omega) \to \infty$ as frequency $\omega \to 0$. Long- and short-range dependence can also be distinguished in terms of variance of the partial sum process, $S_T = \sum_{t=1}^{T} u_t$. For a short memory process, growth of $Var(S_T)$ is proportional to the number of terms, T. For a long memory process, however, $Var(S_T)$ grows more rapidly and is of order $O(T^{2d+1})$ as $T \to \infty$, $0 < d < 1/2$. One can expect the long-range dependence of noise process u_t to feed into y_t and generate substantial amount of persistence diluting the underlying model structure. Such persistence is a common feature of many economic and financial time series and our aim is to analyse forecasting perspectives of such series which undergo structural change.

The forecast strategy we adopt does not require any particular modelling and estimation of the structure of β_t and relies simply on weighted combination of historical data. Methods based on two types of weighting schemes are particularly popular in practice, namely rolling window method and exponentially weighted moving average (EWMA). Such methods work by choosing a tuning parameter which determines the rate of discounting past information. Previous works that forecast using such data-downweighting methods find performance to be sensitive to the choice of tuning parameters, but they do not provide any guidance on how to select one (see Pesaran and Pick (2011), Eklund et al. (2010)). Clearly, setting the discounting parameter *a priori* to a single fixed value is a risky strategy and unlikely to produce accurate forecasts if a series is subject to structural change. Giraitis et al. (2013) advocate a data-dependent selection of the tuning parameter and provide theoretical justification on how such a selection can be optimal.

The tuning parameter is chosen on the basis of predictive performance evaluated over a part of the sample. The strategy is discussed in details below. Forecast of y_t is based on (local) averaging of past values y_{t-1},..., y_1:

$$\widehat{y}_{t|t-1,H} = \sum_{j=1}^{t-1} w_{tj,H} y_{t-j} = w_{t1,H} y_{t-1} + ... + w_{t,t-1,H} y_1 \qquad (4.2.2)$$

with weights $w_{tj,H} \geq 0$ such that $w_{t1,H} + ... + w_{t,t-1,H} = 1$, parameterised by a single tuning parameter H. The latter defines the rate of downweighting the past observations (e.g., the width of the rolling window). We assume that H takes values in the interval $I_T = [\alpha, H_{\max}]$, where $\alpha > 0$.

Giraitis et al. (2013) propose several interesting extensions built on the flexibility of the location shift model. For example, they show that it is possible to first fit a generic model of conditional mean to the location model and then forecast around the model using robust strategies. This can be helpful if the forecaster has a known preferred model of conditional mean for a series. We include such extensions in our analysis.

4.2.2 Selection of the Tuning Parameter, *H*

Suppose we have a sample of T observations $y_1, ..., y_T$. Then construction of one-step-ahead forecast $\widehat{y}_{T+1|T,H}$ requires selection of the parameter H. This is done by a cross-validation method which holds back the last $T_n = T - T_0 + 1$ observations for a pseudo out-of-sample forecast exercise and chooses the tuning parameter H which yields the smallest mean squared forecast error (MSFE) on this evaluation sample. Thus, the MSFE which is minimised with respect to H forms the objective function and is computed as

$$Q_{T,H} := \frac{1}{T_n} \sum_{t=T_0}^{T} (y_t - \hat{y}_{t|t-1,H})^2, \qquad \hat{H} := \arg \min_{H \in I_T} Q_{T,H} \tag{4.2.3}$$

with starting point of the cross-validation period, $T_0 = o(T)$ and the size, $T_n := T - T_0 + 1$. We define $H_{\max} = T_0 T^{-\delta}, 0 < \delta < 1$. It is assumed that T_0 and $H_{\max}$ are selected such that $T^{2/3} < H_{\max} < T_0 = o(T)$.

Let $H_{opt} = \arg\min_{H \in I_T} \omega_{T,H}$ be the optimal value of the fixed parameter H which minimises MSE $\omega_{T,H} := E(y_{T+1} - \hat{y}_{T+1|T,H})^2$. Giraitis et al. (2013) theoretically prove that the forecast $\hat{y}_{T+1|T,H}$ of y_{T+1}, obtained with data-tuned $\hat{H}$, minimises the asymptotic MSE, $\omega_{T,H}$ in H, hence making the forecast procedure (4.2.2) operational. It is also asymptotically optimal in the following sense:

$$\omega_{T,\hat{H}} = \omega_{T,H_{opt}} + o_p(1), \qquad Q_{T,\hat{H}} = \omega_{T,\hat{H}} + o_p(1),$$

where the quantity $Q_{T,\hat{H}}$ is an estimate of the forecast error $\omega_{T,\hat{H}}$.

Let $\hat{\sigma}^2_{T,u} := T_n^{-1}\sum_{j=T_0}^{T} u_j^2$. Giraitis et al. (2013) show that for a general model $y_t = \beta_t + u_t$ with a number of deterministic and stochastic processes β_t and short memory u_t's,

$$Q_{T,H} = \hat{\sigma}^2_{T,u} + E[Q_{T,H} - \sigma_u^2](1 + o_p(1)), \quad T \to \infty, \quad H \to \infty,$$

uniformly in H. In addition, they verify that the deterministic function $E[Q_{T,H} - \sigma_u^2]$ has a unique minimum. This allows selection of the optimal data-tuned parameter H that asymptotically minimises the objective function $Q_{T,H}$. We shall focus on two cases of $y_t = \beta_t + u_t$ where the noise u_t has long memory and β_t is either a constant or a linear trend.

Assumptions and notation. We construct the weights $w_{tj,H}$ as follows.

Assumption 1 *The function $K(x) \geq 0$, $x \geq 0$ is continuous and differentiable on its support, such that $\int_0^\infty K(u)du = 1$, $K(0) > 0$, and for some $C > 0$, $c > 0$,*

$$K(x) \leq C\exp(-c|x|), \quad |\dot{K}(x)| \leq C/(1 + x^2), \qquad x > 0, \tag{4.2.4}$$

where $\dot{K}$ is the first derivative of K. For $t \geq 1$, $H \in I_T$, set $k_{j,H} = K(j/H)$ and define

$$w_{tj,H} = \frac{k_{j,H}}{\sum_{s=1}^{t-1} k_{s,H}}, \qquad j = 1, \cdots, t-1. \tag{4.2.5}$$

Popular classes of commonly used weights satisfying this assumption include:

(i) *Rolling window* weights, with $K(u) = I(0 \leq u \leq 1)$.

(ii) *Exponential weighted moving average* (EWMA) weights, with $K(u) = e^{-u}$, $u \in [0, \infty)$. Then, with $\rho = \exp(-1/H)$, $k_{j,H} = \rho^j$ and $w_{tj,H} = \rho^j / \sum_{k=1}^{t-1} \rho^k$, $1 \leq j \leq t-1$.

(iii) *Triangular window* weights, with $K(u) = 2(1-u)I(0 \leq u \leq 1)$.

While the rolling window simply averages the H previous observations, the EWMA forecast uses all observations $y_1, \cdots, y_{t-1}$, smoothly downweighting the more distant past.

In addition to $w_{tj,h}$, for technical reasons we will use the weights

$$w_{j,H} = k_{j,H} / \sum_{s=1}^{\infty} k_{s,H}, \quad j \geq 1. \tag{4.2.6}$$

4.3 Theoretical Results and Examples

We illustrate the theoretical properties of the weighted forecast $\hat{y}_{T+1|T,\hat{H}}$ with data selected tuning parameter $\widehat{H}$ by two examples of $y_t = \beta_t + u_t$ where β_t is either a constant or a linear trend and the noise u_t is a stationary long memory process. Our objective is to show that the forecast $y_{T+1|T,\hat{H}}$ of y_{T+1} with optimal turning parameter $\hat{H}$ minimises the forecast MSE in the following sense: $\omega_{T,\hat{H}} = \omega_{T,H_{opt}} + o_p(1)$. Moreover, the property $Q_{T,\hat{H}} = \omega_{T,\hat{H}} + o_p(1)$ allows estimation of the forecast error.

Below, $a \wedge b = \min(a,b)$, $a \vee b = \max(a,b)$ and $I(A)$ is the indicator function; $a_T \sim b_T$ denotes that $a_T/b_T \to 1$, as T increases. We write $o_{p,H}(1)$ or $o_H(1)$ to indicate that $\sup_{H \in I_T} |o_{p,H}(1)| \to_p 0$ or $\sup_{H \in I_T} |o_H(1)| \to 0$, as $T \to \infty$.

The following assumption describes the class of stationary noise processes u_t. It allows u_t to have either short memory (i) or long memory (ii). We denote the k-th order autocovariance function of u_t by $\gamma_u(k) = \mathrm{Cov}(u_k, u_0)$.

Assumption 2 *u_t is a stationary linear process*

$$u_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \; t \in \mathbb{Z}, \qquad \varepsilon_j \sim \mathrm{IID}(0, \sigma_\varepsilon^2), \qquad E\varepsilon_1^4 < \infty. \tag{4.3.1}$$

(i) In the short memory (SM) case we assume that $\sum_{k \in \mathbb{Z}} |\gamma_u(k)| < \infty$, $\sum_{k \geq n} |\gamma_u(k)| = o(\log^{-2} n)$ *and* $s_u^2 := \sum_{k \in \mathbb{Z}} \gamma_u(k) > 0$.

(ii) In the long memory (LM) case we assume that for some $c_\gamma > 0$ and $0 < d < 1/2$,

$$\gamma_u(k) \sim c_\gamma k^{-1+2d}, \qquad k \to \infty. \tag{4.3.2}$$

Under SM Assumption 2 (i), u_t has short memory, while its long-run variance s_u^2 is positive and finite. This case was discussed in Giraitis et al. (2013).

We now proceed to analyse the properties of $Q_{T,H}$, $\hat{H}$ and the forecast error $\omega_{T,\hat{H}}$ under long memory assumption 2(ii).

4.3.1 The Case of a Long Memory Stationary Process

y_t

First, we focus on the forecast in the case when $y_t = \mu + u_t$, $t \geq 1$ is a stationary long memory process. We shall use the following notations:

$$q_{u,H} := E\left(u_0 - \sum_{j=1}^{\infty} w_{j,H} u_{-j}\right)^2 - \sigma_u^2, \tag{4.3.3}$$

$$\lambda_{LM} = c_\gamma \left[\int_0^\infty \int_0^\infty K(x)K(y)|x-y|^{-1+2d} dy dx - 2\int_0^\infty K(x) x^{-1+2d} dx \right]. \tag{4.3.4}$$

Theorem 1 *Suppose that $y_t = \mu + u_t, t \geq 1$, where u_t is a stationary long long memory process with parameter d satisfying Assumption 2 (ii).*

Then, as $T \to \infty$, for $H \in I_T$,

$$Q_{T,H} = \hat{\sigma}_{T,u}^2 + q_{u,H} + o_{p,H}\left(H^{-1+2d}\right), \tag{4.3.5}$$
$$\omega_{T,H} = \sigma_u^2 + q_{u,H} + o_H(H^{-1+2d}),$$

where $q_{u,H} = \lambda_{LM} H^{-1+2d} + o(H^{-1+2d})$, as $H \to \infty$.

Theorem 1 shows that $Q_{T,H}$ is a consistent estimate of $\omega_{T,H}$. The following corollary shows that the forecast $y_{T+1|T,\hat{H}}$ computed with the data-tuned $\hat{H}$ has the same MSE as the forecast $y_{T+1|T,H_{opt}}$ with the tuning parameter H_{opt}.

Corollary 1 *If $q_{u,H}$ reaches its minimum at some finite H_0, then*

$$\begin{aligned} \omega_{T,\hat{H}} &= \omega_{T,H_{opt}} + o_p(1), \\ Q_{T,\hat{H}} &= \omega_{T,\hat{H}} + o_p(1) = \sigma_u^2 + q_{u,H_0} + o_p(1). \end{aligned} \tag{4.3.6}$$

Proof of the Theorem 1 and Corollary 1 can be found in Appendix E.

Remark 1 *Corollary 1 implies that quality of a forecast with tuning parameter $\hat{H}$ is the same as using parameter H_{opt} that minimises the forecast error $\omega_{T,H}$. Observe that for $\lambda_{LM} < 0$, $q_{u,H} = \lambda_{LM} H^{-1+2d} + o(H^{-1+2d})$ implies that $\hat{H}$ will remain finite when T increases. Notice also that $\lambda_{LM} < 0$ for rolling window weights corresponding to kernel function $K(x) = I(0 \leq x \leq 1)$. Indeed,*

$$\lambda_{LM} = c_\gamma \int_0^\infty \int_0^\infty K(x)K(y)|x-y|^{-1+2d} dxdy - 2c_\gamma \int_0^\infty K(x)x^{-1+2d}dx$$

$$= c_\gamma \left(\int_0^1 \int_0^1 |x-y|^{-1+2d} dxdy - 2\int_0^1 x^{-1+2d}dx \right)$$

$$= 2c_\gamma \left(\int_0^1 \int_0^x u^{-1+2d} dudx - \frac{1}{2d} \right) = 2c_\gamma \left(\frac{1}{2d(1+2d)} - \frac{1}{2d} \right) = -\frac{2c_\gamma}{1+2d} < 0.$$

This shows that, using rolling-window weights, the error of the forecast obtained with $\hat{H}$ is smaller than σ_u^2 and $\hat{H}$ remains finite. The fact that under strong dependence the data tuned parameter $\hat{H}$ does not increase with the sample size is in line with the well known fact that a persistent process, e.g. a random walk, can be well forecasted by the last observation, that corresponds to the rolling window with $H = 1$.

4.3.2 The Case of a Deterministic Trend and a Long Memory Noise

Next, we analyse the forecast of $y_t = \beta_t + u_t$, when $\beta_t = at$ is a deterministic trend and u_t is a stationary long memory noise.

Denote

$$q_{\beta,H} := \left(\sum_{j=1}^{\infty} w_{j,H} j \right)^2, \qquad \kappa := \left(\int_0^{\infty} K(x) x dx \right)^2.$$

Notations $q_{u,H}$ and λ_{LM} are as in Theorem 1.

Theorem 2 *Let $y_t = at + u_t, t = 1, \cdots, T$ where u_t is a long memory process satisfying Assumption 2(ii). Then, as $T \to \infty$, for $H \in I_T$,*

$$\begin{aligned} Q_{T,H} &= \hat{\sigma}^2_{T,u} + q_{\beta,H} + q_{u,H} + o_{p,H}(H^2), \\ \omega_{T,H} &= \sigma^2_u + q_{\beta,H} + q_{u,H} + o_H(H^2), \end{aligned} \tag{4.3.7}$$

where $q_{\beta,H} + q_{u,H} = \kappa H^2 + o(H^2)$, as $H \to \infty$.

Theorem 2 allows to establish the following basic properties of the forecast $y_{T+1|T,\hat{H}}$ of a trend stationary process y_t.

Corollary 2 *Under assumptions of Theorem 2, for a linear trend $\beta_t = at$, $\hat{H}$ stays bounded:*

$$\begin{aligned} \omega_{T,\hat{H}} &= \omega_{T,H_{opt}} + o_p(1), \\ Q_{T,\hat{H}} &= \omega_{T,\hat{H}} + o_p(1) = \sigma^2_u + q_{\beta,H_0} + q_{u,H_0} + o_p(1), \end{aligned} \tag{4.3.8}$$

where H_0 is a minimiser of $q_{\beta,H} + q_{u,H}$.

Proof of the Theorem 2 can be found in Appendix E.

Remark 2 *In the presence of a deterministic trend the optimal $\hat{H}$ will take small values and the forecast will be based on averaging over the last few observations.*

4.3.3 Illustrative Examples

We can learn about the adaptability of the proposed robust forecast strategy simply by looking at the evolution of a data-estimated tuning parameter when forecasting in face of structural changes. For interpretational convenience we choose to examine only rolling window forecasts. This also allows us to directly compare the resulting behaviour of the data-tuned windows for long memory noise to that for i.i.d. noise considered in Giraitis et al. (2013). A sample size of $T = 200$ observations is considered and the forecasting starts at $\tau = 100$. We plot a single realisation of data-selected window size $\widehat{H}(t)$ which is computed sequentially at $t = \tau,\ \tau + 1, ...,\ T$. Two different structural change set-ups from our Monte Carlo study in section 4.4 have been considered. **Figure 4.B.1** examines forecasting under break in the mean (Experiment 4) and **Figure 4.B.2** looks at the case of a unit root process (Experiment 11). Depending on noise specifications each figure has two panels: panel (a) referring to i.i.d. u_t and panel (b) referring to u_t with long range dependence generated by $ARFIMA(0, 0.45, 0)$. In each panel the solid line represents the first observation of the data-estimated rolling window when there is a break in the mean (**Figure 4.B.1)** or when there is unit root (**Figure 4.B.2)** and the dotted line represents the starting point when there is no structural change (Experiment 1) based on the same realisations of the noise u_t. The long-dashed line marks the last observation of the estimation window and the vertical distance between it and the solid line (dotted line) measures the size of data-selected rolling window used for constructing forecasts in presence of structural changes (no structural change). In **Figure 4.B.1** the small dashed line marks the first post break observation which is the 110th data point. In case of i.i.d. noise we find similar findings as Giraitis et al. (2013). Immediately after the break the data-dependent forecast method reacts by beginning to use longer data-windows than in no-break situation in order to learn more about the structural change and make subsequent predictions. When more observations become available it soon starts using much smaller samples with the starting point of the data-estimated windows mostly being the first post-break data. Once the switch is made the data-tuned method never uses pre-break information for forecasting and also never uses windows longer than in no structural break case. When u_t has long memory observations are slightly different. Panel (b) of **Figure 4.B.1** shows that similar initial adjustment with longer windows is made immediately after the break, but the switch to post-break information appears to be faster for noises with long range dependence. The window-sizes, when there is structural break, are predominantly smaller than those in no-break experiment, but the margins of difference in window lengths are much smaller than in i.i.d. noise case. There are also periods when data samples coincide for the two experiments. This is probably because persistence through long memory mitigates or conceals the effect of a break in the generated series and a data-based tuning method finds it difficult to distinguish between the break and no-break cases.

Panel (a) of **Figure 4.B.2** shows that for unit root noise processes data-selected window sizes remain much shorter than in no-break case throughout the sample. Panel (b) then confirms that for persistent long memory noises the adaptive method yields even smaller windows but not substantially smaller than those required in absence of a break.

4.4 Monte Carlo Experiments

A next step forward is to conduct a Monte Carlo study to evaluate the performance of adaptive forecast strategies over finite samples of artificially generated data and examine to what extent results comply with theoretical findings. We closely follow Giraitis et al. (2013) in setting up the design of experiments with the main difference being that we adopt long memory dynamics for the noise process in contrast to their short memory specifications. We use the simple location model $y_t = \beta_t + u_t$ to simulate different data series that depend on different structural specifications (e.g., deterministic functions of time or stochastic processes) for β_t component and various long memory dynamics for noises u_t. Alternative forecast methods are compared in terms of $MSFE$ of one-step-ahead forecasts relative to a benchmark of sample mean forecasts. When persistence in u_t is low we expect the benchmark to perform reasonably well. We also include a simple AR(1) and 'last observation' forecasts which are generally considered to capture dependence and unit root dynamics well.

4.4.1 Data Generating Processes

As mentioned earlier we exploit the location shift model (4.2.1) for generating the data:

$$y_t = \beta_t + u_t, \qquad t = 1, 2, ..., T. \tag{4.4.1}$$

While Giraitis et al. (2013) considered short memory i.i.d. and AR(1) noises, we explore several long memory specifications for u_t . We opt to use the widely popular Autoregressive Fractionally Integrated Moving Average (ARFIMA) processes to generate u_t with long range dependence. The $ARFIMA(p, d, q)$ model is defined as:

$$\Phi(L)(1-L)^d u_t = \Theta(L)\epsilon_t, \quad \epsilon_t \sim i.i.d.(0, \sigma_\epsilon^2),$$

where d is the fractional differencing parameter that induces long memory and L is the lag operator. The fractional differencing operator $(1-L)^d$ is defined by the binomial expansion

$$(1-L)^d = \textstyle\sum_{i=0}^{\infty} \binom{d}{i} (-L)^i.$$

The process is stationary and invertible if all the roots of the autoregressive polynomial of order p, $\Phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$ and the moving average part of order q, $\Theta(L) = 1 + \theta_1 L + ... + \theta_q L^q$, lie outside the unit circle, with $|d| < 0.5$. For $|d| > 0.5$ the process becomes non-stationary. Generally, y_t is said to be $I(d)$ if generated by the $ARFIMA(p, d, q)$ model.

We consider stationary and non-stationary $ARFIMA(0, d, 0)$ processes with long memory parameters $d = 0.30$, $d = 0.45$ or $d = 0.75$ indicating different levels of persistence in memory or $ARFIMA(1, d, 0)$ processes with combinations of long memory parameters $d = 0.30$ or 0.75 and short memory AR(1) parameters $\rho = 0.7$ or -0.7. The innovations of the noise processes are i.i.d. standard normal. The component β_t is either a linear or non-linear deterministic trend, a stochastic trend process such as random walk or a process with a break in the mean. We consider eleven data generating processes that are also used in in Giraitis et al. (2013):

$Ex1$. $y_t = u_t$.

$Ex2$. $y_t = 0.05t + 5u_t$.

$Ex3$. $y_t = 0.05t + 3u_t$.

$Ex4$. $y_t = \begin{cases} u_t, \ t \leq t_0 = 0.55T, \\ 1 + u_t, \ t > t_0. \end{cases}$

$Ex5$. $y_t = 2\sin(2\pi t/T) + 3u_t$.

$Ex6$. $y_t = 2\sin(2\pi t/T) + u_t$.

$Ex7$. $y_t = 2T^{-1/2}\sum_{i=1}^{t} v_i + 3u_t$.

$Ex8$. $y_t = 2T^{-1/2}\sum_{i=1}^{t} v_i + u_t$.

$Ex9$. $y_t = 0.5\sum_{i=1}^{t} v_i + 3u_t$.

$Ex10$. $y_t = 0.5\sum_{i=1}^{t} v_i + u_t$.

$Ex11$. $y_t = \sum_{i=1}^{t} u_i$.

In order to get a first-hand idea about the dynamic behaviour of the generated series we plot them for different specifications of β_t and u_t. **Figures 4.B.3 - 4.B.9** present time series plots of y_t for $Ex1$, $Ex3$, $Ex4$, $Ex6$, $Ex7$, $Ex9$ and $Ex11$, respectively. The first two panels of each figure consider cases where u_t are i.i.d. or AR(1) with i.i.d. innovations as studied in Giraitis et al. (2013). The following four panels portray series with long memory noises with u_t following an $ARFIMA(0, d, 0)$ with $d = 0.30$ and $d = 0.75$ and $ARFIMA(1, d, 0)$ with $\rho = 0.7$ and $d = 0.30$ or 0.75.

In $Ex1$, y_t is determined by the noise process alone and there is no structural change. It is not surprising that forecasting an i.i.d. process, as illustrated in panel (a) of **Figure 4.B.3**, would require accounting for long past and the benchmark sample mean should perform the best. Similarly, it is expected that a simple AR(1) benchmark will be difficult to be outperformed when forecasting persistent autoregressive processes such as one plotted in panel (b). Giraitis et al. (2013) confirm these facts through Monte Carlo evidence and also report competitive forecasts for many of the robust methods. It is our interest to investigate how the robust methods fare when the noises have long memory. Panel (c) of **Figure 4.B.3** shows that a weakly stationary long memory process $ARFIMA(0, d, 0)$ with $d = 0.30$ generates enough persistence for it to be visually distinguishable from the i.i.d. process, but it is not as persistent as the short memory AR(1) process with $\rho = 0.7$. However, long term dependence can create false impression of structural change and make prior preference of a forecast model difficult. For example, the $ARFIMA(0, d, 0)$ process with $d = 0.30$ imitates a cyclical trend like movement (see panel (c) of **Figure 4.B.3**) and it is not clear whether a full sample or AR(1) forecast will be accurate. Panel (d) confirms that a higher long memory parameter $d = 0.75$ drives the series to non-stationary territory. However, this clearly induces an impression of an increasing linear trend. Additional persistence through autoregressive dependence make the series even closer to unit root (panel (f)). An AR(1) benchmark should still do well in this case, but ‘last observation’ forecasts should be equally competitive.

Both $Ex2$ and $Ex3$ introduce linear monotonically increasing trends in y_t and differ only in size of variance of noise process. Giraitis et al. (2013)

argue that such linear trends may be unrealistic but they can offer reasonable representations of time series which are detrended through standard techniques such as differencing or filtering. Moreover, **Figure 4.B.4** confirms that the effects of such trends are small enough to be dominated and muted by the noise processes. While linear trends are visually detectable for an i.i.d. and a weakly dependent $ARFIMA(0, d, 0)$ noise process with $d = 0.30$, they become more obscure with increasing persistence. The last two panels of **Figure 4.B.4** confirm that when short and long memory persistence are combined, the trends can vanish completely.

The functional form of y_t in $Ex4$ accommodates structural break in the form of a break in the mean. The break occurs slightly after halfway the sample at time $t_0 = 0.55T$. Giraitis et al. (2013) argue that since the post-break period is greater than $\sqrt{T}$, as required by the theory, the robust forecasting methods should take account of such 'not-too-recent' breaks and yield forecasts that are significantly better than the benchmark sample mean. Their Monte Carlo study confirms their claims. We should note from **Figure 4.B.5** that although the shift in mean can be well identified in i.i.d. or weak long memory series, it becomes more concealed with increasing persistence in the noise process. Dependence in the noise process u_t intensifies the effect of the break and for highly persistent non-stationary long memory series, such as $ARFIMA(1, d, 0)$ with $\rho = 0.7$ and $d = 0.75$ this can even result in a false trend-like movement.

The purpose of $Ex5$ and $Ex6$ is to introduce smooth cyclical bounded trends as observed in standard business cycles. Such trends are less likely to be completely removed from standard detrending and therefore, more realistic than a linear trend. The sample mean benchmark should do poorly, particularly for $Ex6$ where oscillation of the trend is wider compared to the variance of the noise process. It is evident from **Figure 4.B.6** that higher persistence can distort shapes of smooth cycles to substantial extent.

For the remaining data generating processes, β_t are stochastic trends. For $Ex7$ and $Ex8$ the trends are bounded and represent increasingly popular time-varying coefficients type dynamic models. $Ex9$ and $Ex10$ consider unbounded random walk (unit root) process, observed under noise u_t. $Ex11$ analyses a standard random walk model. **Figures 4.B.7 - 4.B.9** show that dynamics of simulated series varies significantly depending on degree of persistence in the noise process, contributed either by short or long memory.

It is evident from the time series plots that long memory can give false impression of structural change. Moreover, persistence in the noise processes induced by long memory or mixture of short and long memory dependence can confound types of structural changes in a time series. It is worth investigating whether typical robust-to-structrual-change methods, such as rolling window and EWMA methods, can perform well in forecasting in presence of long memory. We argue that as long as the choice of tuning parameter is data-dependent such methods can generate forecasts that are comparable to the best possible fixed parameter forecasts.

4.4.2 Forecast Methods

We resort to forecast methods that have been analysed in Giraitis et al. (2013). The range of strategies mainly include forecast methods that discount past data and are known to be robust to historical and ongoing structural changes. Both parametric and nonparametric weights and methods with both fixed and data-dependent discounting parameters are considered. We compare forecasts against a number of simple benchmark models.

Methods with Parametric Weights

These are robust methods where data discounting weights are defined as functions of a tuning parameter. Three methods are discussed which are based on three different types of weights.

Rolling window method. In this method the weights are defined in terms of the parameter H which is essentially the window size and includes the H most recent observations to be used in forecasting. The weight $w_{tj,H}$ attached to y_{t-j} is defined as

$w_{tj,H} = H^{-1}I(1 \leq j \leq H),\ j = 1, 2, \ldots.., t-1$, for $H < t$, and

$w_{tj,H} = (t-1)^{-1}I(1 \leq j \leq t-1)$, for $H \geq t$,

where I is an indicator function.

The weights are flat in the sense that all the observations in the window get equal weights while the older data get zero weights. The one-step-ahead forecast $\widehat{y}_{t|t-1}$ is then simply the average of H previous observations. In the result tables we refer this method as *Rolling* H. Besides selecting H optimally from data we use two fixed window methods with $H = 20$ and 30.

Exponentially weighted moving average (EWMA) method. This method assigns the highest weight to the most recent data point and discounts further past by decreasing weights exponentially fast to zero. The weights used in this methods can be defined as:

$w_{tj,H} = \rho^{t-j} / \left(\sum_{k=1}^{t-1} \rho^k\right),\ 1 \leq j \leq t-1$, with $0 < \rho < 1$.

The closer ρ is to zero the faster is the rate of discounting and the main weights are concentrated on the last few data points. The closer ρ is to one the slower is the rate and significant weights are attached to datum in distant past. In tables this method is denoted as *Exponential* ρ. We consider several fixed value downweighting methods with $\rho = 0.98$, 0.95, 0.80, 0.60, 0.40 and 0.002. The data-tuned parameter is denoted as $\widehat{\rho}$.

Polynomial method. This uses weights

$w_{tj,H} = (t-j)^{-\alpha} / \left(\sum_{k=1}^{t-1} k^{-\alpha} \right), \; 1 \leq j \leq t-1$, with $\alpha > 0$.

The past is downweighted at a slower rate than with exponential weights. This method is referred to as *Polynomial* α. We do not consider any fixed value for α and only report data-dependent downweighting with estimated parameter $\widehat{\alpha}$.

Nonparametric Methods

All the above methods adopt parametric weight functions and in one way or another downweight past data monotonically. While this is sensible in most practical situations, there may be instances when valuing recent data the most may appear unfavourable. For example, if there are a finite number of monetary policy regimes which repeat themselves, then older data from a period when the current regime previously held may be more relevant than more recent data from other regimes (Giraitis et al. (2013)). A nonparametric weighting scheme is used to account for such possibilities. See Giraitis et al. (2013) for a detailed technical explanation about how the method works.

Multiparameter Extensions

Rolling $(\widehat{k}, \widehat{H})$ method. This is an extended two-parameter rolling window method where the downweighting parameter, H is optimally and simultaneously chosen using a 'stable' subsample period $[k, ..., T]$, where k, the starting time of the period, is a second parameter to be estimated. The optimisation procedure requires minimisation of MSE, $Q_{T,kH}$ over both k and H and is given by

$$\begin{aligned} Q_{T,kH} &: = (T-k)^{-1} \textstyle\sum_{t=k}^{T} (\widehat{y}_{t|t-1,H} - y_t)^2, \\ \{\widehat{H}, \widehat{k}\} &: = \arg \min_{H \in I_T, k \in \{k_{\min}, \ldots\ldots, k_{\max}\}} Q_{T,kH} \end{aligned}$$

The one-step-ahead forecast $\widehat{y}_{T+1}$is then constructed based on the optimal subsample $y_{\widehat{k}}, y_{\widehat{k}+1}, ..., y_T$ and the associated tuning parameter $\widehat{H} = \widehat{H}(\widehat{k})$. We should note that $\widehat{H} \leq T - \widehat{k}$. Giraitis et al. (2013) argue and prove theoretically that such two parameter rolling window forecast method is particularly beneficial when there is a break in the mean. Forecasting after the break should require that more of post-break data are used and the irrelevant past are weighted less. Optimally choosing an evaluation subsample rather than using all the available information implies that switching to post-break data is faster than when using the full sample.

Dynamic weighting. Giraitis et al. (2013) propose a more flexible extension of exponential weighting where the weights attached to the first few lags are not determined by parametric functions, but rather freely chosen along with the tuning parameter, H. Thus, like an AR process the first p weights, $w_1, w_2, ..., w_p$ are estimated as additional parameters, while the remaining weights are functions of H. The weight function is defined as

$$\widetilde{w}_{tj,H} = \begin{cases} w_j, & j = 1, ..., p \\ K(j/H), \quad j = p+1, ..., t-1, & H \in I_T \end{cases} \tag{4.4.2}$$

and the final weights are standardised as $\widetilde{w}_{tj,H} = \widetilde{w}_{tj,H} / \left(\sum_{j=1}^{t-1} \widetilde{w}_{tj,H} \right)$ to sum to one. Note that Q_T is jointly minimised over $w_1, w_2, ..., w_p$ and H. We consider a parsimonious representation by specifying $p = 1$ and choose exponential kernel K.

Residual methods. Giraitis et al. (2013) argue that if a time series explicitly allows for modelling the conditional mean of the process and forecaster has a preferred parametric model for it then it might be helpful to first fit the model and use the robust methods to forecast the residuals from the model. The original location model (4.2.1) is restrictive and not suitable for conditional modelling and a more generic forecasting model is therefore, proposed to illustrate the approach:

$$z_t = f(x_t) + y_t, \ t = 1, 2, ...$$

where z_t is the variable of interest, x_t is the vector of predicted variables which may contain lags of z_t, and y_t is the vector of residuals which are unexplained by $f(x_t)$. In the presence of structural change, y_t is expected to contain any remaining persistence in z_t such as trends, breaks or other forms of dependence, and the robust methods should perform well in such scenario. Forecasts of $f(x_t)$ and y_t are then combined to generate improved forecasts of z_t.

Following Giraitis et al. (2013) we adopt the widely popular AR(1) process to model the conditional mean which gives $f(x_t) = \phi z_{t-1}$. The residuals y_t are forecast using either parametric or nonparametric weights discussed above. The forecast of z_{t+1} based on $z_1, z_2, ..., z_t$ is computed as $\widehat{z}_{t+1} = \widehat{\phi} z_t + \widehat{y}_{t+1|t,\widehat{H}}$. Three versions of the residual methods are considered.

Exponential AR method. In this method the tuning parameter H and the autoregressive parameter ϕ are simultaneously estimated by minimising the in-sample mean squared forecast error, $Q_{T,H} = Q_{T,H\phi}$ which is computed by defining $y_t = z_t - \phi z_{t-1}$ and using exponential weights. This method is referred to as *Exponential AR*.

The remaining two methods involve two-step estimation where the autoregressive coefficient ϕ of z_{t-1} is estimated by OLS independently of the parameters associated with forecasting y_t.

Exponential residual method. It forecasts residuals $y_t = z_t - \phi z_{t-1}$ using exponential weights producing $\widehat{H}$ and consequently, the forecast $\widehat{y}_{t+1|t,\widehat{H}}$. In the tables it is denoted as *Exponential Residual*.

Nonparametric residual method. It forecasts residuals $y_t = z_t - \phi z_{t-1}$ using non-parametric methods. We refer to it as *Nonparametric Residual.*

The Benchmark and Other Competitors

Full sample mean. The benchmark forecast for our study is the average of all observations in the sample:

$$\widehat{y}_{benchmark,T+1} = \frac{1}{T}\sum_{t=1}^{T} y_t.$$

AR(1) forecast. We include forecasts based on an AR(1) dynamics which is often considered as a stable and consistent predictor of time series. The one-step-ahead forecast is given by:

$$\widehat{y}_{T+1|T} = \widehat{\theta} y_T.$$

Last observation forecast. For unit root process a simple yet competitive forecast is simply 'no change' forecast:

$$\widehat{y}_{T+1|T} = y_T.$$

Averaging method. Pesaran and Timmermann (2007) advocate simple robust method which is based on the idea of forecast combination with equally weighted forecasts. The one-step-ahead forecast $\overline{y}_{T+1|T}$ is the average of rolling window forecasts $\widehat{y}_{T+1|T,H}$ obtained using all possible window sizes, H that include the last observation:

$$\overline{y}_{T+1|T} = \frac{1}{T}\sum_{H=1}^{T} \widehat{y}_{T+1|T,H}, \quad \widehat{y}_{T+1|T,H} = \frac{1}{H}\sum_{t=T-H+1}^{T} y_t.$$

The method avoids estimation of any discount parameter but usually requires selection of a minimum data-window to be used for forecasting . We ignore such a choice of minimum sample size since forecasts are not significantly sensitive to it. In table this method is referred to as *averaging*.

4.4.3 Monte Carlo Results

The out-of-sample forecast exercise becomes operational by choosing a starting point τ when the first forecast will be made. We subsequently apply all the reported methods to construct one-step-ahead forecasts $\widehat{y}_{t|t-1,H}$, $t = \tau, ..., T$. Forecasts at time t is computed using only information available up to $t-1$

and the forecast evaluation period ends at T. We compare performance of models in terms of their mean squared forecast error ($MSFE$) relative to the benchmark of the sample mean of all available data. $MSFE$ for method j is computed as $MSFE_j = (T - \tau + 1)^{-1} \sum_{t=\tau}^{T} (\widehat{y}_{t|t-1,H}^{(j)} - y_t)^2$ and the relative $MSFE$ is defined as $RMSFE = \frac{MSFE_j}{MSFE_{sm}}$, where $MSFE_{sm}$ corresponds to the benchmark forecast by sample mean.

In what follows we discuss Monte Carlo results of forecasting performance of the adaptive forecasting techniques in predicting time series $y_t = \beta_t + u_t$ with long memory noise u_t and compare them with results for short memory noises reported in Giraitis et al. (2013). Results for different long memory specifications of the noise processes are presented in **Tables 4.C.1 - 4.C.7**. The columns represent data-generating models $Ex1 - Ex11$ which have been discussed in the previous section, and the rows represent different forecasting methods. Entries of the tables are $MSFE$ of different methods relative to sample average, as defined above. Noises in **Tables 4.C.1 - 4.C.3** have been generated by a standard $ARFIMA(0, d, 0)$ model with the long memory parameter $d = 0.30$, $d = 0.45$ and $d = 0.75$, respectively. Note that the first two specifications refer to stationary processes with a moderate and high degree of long memory, while the last refers to a non-stationary integrated process. Following Giraitis et al. (2013) we consider additional forms of persistence in both stationary and non-stationary long memory processes via autoregressions. **Tables 4.C.4 - 4.C.5** report results for $ARFIMA(1, d, 0)$ noise with an AR(1) coefficient $\rho = 0.7$ and long memory parameter $d = 0.30$, 0.75 respectively. **Tables 4.C.6 - 4.C.7** contain results for the $ARFIMA(1, d, 0)$ processes with the same degree of long memory but with a negative AR(1) coefficient of $\rho = -0.7$. The innovations of the noise processes u_t are i.i.d. standard normal.

We begin by discussing the results in **Table 4.C.1** which features $ARFIMA$ (0,0.30,0) noises. $RMSFE$ values below unity suggest that, in general, all the reported forecasting methods, both with fixed and data-driven discounting, are useful in case of noises with moderately strong long memory. Even the simplest case of 'no structural change', $y_t = u_t$ reported in the first column and labelled as $Ex1$ shows that forecasts of the most of the competing methods, including the rolling-window schemes, outperform the benchmark of the full-sample average. The gains are, however, small. This finding is in contrast with the results obtained for stationary i.i.d. process in Giraitis et al. (2013) that record sole dominance of the benchmark over the competitors.[1] Gains over the benchmark are more pronounced when y_t has a persistent component β_t. Then, even naive 'last observation' forecasts are better than the mean forecast in most of the experiments.

Persistence entering y_t through long memory u_t requires stronger discounting and accounting for information contained in the more recent past. This is evident from the performance of fixed parameter exponentially downweighted moving average forecast methods. While exponential downweighting with parameter $\rho = 0.90$ provides the most accurate forecasts for time series with i.i.d. noise processes, discounting with $\rho = 0.80$ gives the best result for time series with *ARFIMA*(0,0.30,0) noise. Extremely strong discounting is penalised, but not as harshly as in the case of short memory i.i.d. series. For example, in the long memory case, the relative $MSFE$ of forecasts with exponential downweighting with parameter $\rho = 0.002$ is 1.253, while in the i.i.d. case, it becomes 1.947. The data-dependent exponential weights do not exactly match the best fixed value forecast method but are reasonably comparable and are never among the worst performing methods. For instance, the exponential weighting method with a fixed $\rho = 0.90$ beats the method with data-based tuned value $\widehat{\rho}$ in a number of experiments such as $Ex1$, $Ex2$, $Ex7$ etc., but is convincingly outperformed by the latter in several occasions such as $Ex6$, $Ex10$, $Ex11$ etc. Results of $Ex11$ where y_t follows a standard random walk

[1] See Table 1 in Giraitis et al. (2013) for Monte Carlo results considering i.i.d. noise.

are particularly different for long memory and i.i.d. noises. For instance, the relative MSFE of the data-tuned exponential weights is only 0.006 for $ARFIMA$ $(0, 0.3, 0)$ noises compared to a much higher value of 0.042 for i.i.d. noises. Forecast methods with optimally chosen exponential weights consistently perform better than the rolling-window methods, but their ability to outperform the polynomial method is more mixed.

A comparison of rolling-window methods reveals that choosing an evaluation period optimally together with the window size helps to improve forecasts of data with long memory noise. The combined methods using both data-dependent window, $\widehat{H}$, and an evaluation period $(\widehat{k}, T)$ consistently outperform methods using $\widehat{H}$ and $k = 1$ in case of $ARFIMA$ $(0, 0.3, 0)$ noises. Performance of these two methods is relatively more comparable in case of i.i.d. noises. Both methods using the data-adjusted rolling-window forecast better than methods with fixed windows of size $H = 20$ and $H = 30$ and also outperform the averaging method of rolling windows advocated by Pesaran and Timmermann (2007). This justifies the use of data-driven choice of the downweighting parameter in rolling window.

Overall, comparison of competing forecasting methods show that the full sample AR(1) forecasts are in general very good compared to the benchmark, but are often outperformed by most of the adaptive data-tuned methods. Forecasts based on the non-parametric method are competitive and those based on the residual methods are impressive. Among the adaptive robust forecasting methods the dynamic weighting method, where the weight attached to the last observation is optimally chosen from data simultaneously with the exponential weighting parameter, consistently provides forecasts that are comparable to the best possible forecasts for all the experiments. The exponential AR method is also equally competitive.

Similar findings as for $ARFIMA$ $(0, 0.3, 0)$ noises hold also for *ARFIMA* (0,0.45,0) and non-stationary $ARFIMA$ $(0, 0.75, 0)$ noise processes presented in **Table 4.C.2** and **Table 4.C.3** respectively. There are, however, a few new important patterns. First, gains of data-tuned methods over the benchmark of sample mean increase with the increase of long memory, particularly for exponentially weighted moving average and dynamic models. For example, the $RMSFE$'s of the data-tuned exponential method in $Ex1$ are 1.085, 0.905, 0.688 and 0.211 for i.i.d. noise and long memory noises with $d = 0.30$, $d = 0.45$ and $d = 0.75$, respectively. Second, exponential weighting with stronger discounting provides better forecasts for processes with stronger long memory and the data-tuned exponential method matches the best fixed parameter method more closely in case of higher persistence. For example, in case of $Ex1$ and long memory process with $d = 0.30$ the smallest $RMSFE$ value of 0.874 is attributable to $\rho = 0.80$ while the $RMSFE$ value of the data-tuned method is 0.905. For $d = 0.75$, however, a much smaller best $RMSFE$ value of 0.208 is generated by $\rho = 0.40$ and the corresponding value for the data-tuned method is 0.211. Third, data-tuned exponentially weighted methods and dynamic methods enjoy larger gains over polynomial methods when long memory increases. Finally, AR(1) forecasts become more and more competitive with increased persistence. For noises following a non-stationary $ARFIMA$ $(0, 0.75, 0)$ process AR(1) is one of the best performing forecast methods across all the experiments.

As mentioned above, **Table 4.C.4** considers performance of forecast methods when noise u_t shows substantial serial dependence through AR(1) coefficient of $\rho = 0.7$ along with long memory persistence, $d = 0.3$. A comparison with results for short memory AR(1) noise reveals that robust adaptive techniques report smaller $RMSFE$'s in many situations for noises with additional time-dependence induced by long memory.[2] The full sample AR(1) forecasts are consistently the best unlike in the short memory scenario where they are

[2] Refer to Table 2 in Giraitis et al. (2013) for Monte Carlo RMSFE results considering short memory noises that have AR(1) dynamics with an autoregressive coefficient of 0.7.

sometimes beaten. Performance of fixed parameter exponentially weighted moving average methods improves with stronger downweighting of past information. Note, for instance, that the forecasts computed using the exponential method with the lowest fixed value of the tuning parameter $\rho = 0.002$ are comparable to the AR(1) forecasts. Advantage of using only the most recent information is further confirmed by equally good 'last observation' forecasts. The adaptive method with optimal data-selected exponential weights almost always matches the best fixed value method.

Among the rolling-window methods, once again methods with data-dependent window outperform the fixed window methods, more convincingly than in the case of long memory noises with no additional short-range dependence. Forecasts based on methods using a data-dependent window, $\widehat{H}$ and evaluation period $(\widehat{k}, T)$ are also more accurate than those based on methods using $\widehat{H}$ and $k = 1$, with gains more pronounced in presence of additional serial dependence than in the long-memory-only situation. Note that while the $RMSFE$'s of fixed window forecast methods are similar for $ARFIMA$ $(0, 0.3, 0)$ noises and $ARFIMA$ $(1, 0.3, 0)$ noises with an AR(1) coefficient of 0.7, $RMSFE$'s of data-tuned rolling-window methods shrink substantially in the latter, confirming adaptability of such methods to higher persistence. Averaging method appears to be one of the worst performing methods. The exponential AR and the residual methods belong to the group of best performers, followed by the dynamic weighting and the polynomial weighting methods.

Results for a non-stationary noise process $ARFIMA$ $(1, 0.75, 0)$ with an AR(1) coefficient of 0.7 retain most of the above findings. Evidential results are presented in **Table 4.C.5**. Performance of data-tuned methods against the full-sample average further improves. Forecasts based on the EWMA methods with data-selected weights and dynamic methods are similar and almost identical to AR(1) and the 'last observation' forecasts. Notably, the exponential AR forecasts can beat the AR(1) forecasts in several experiments, although with marginal gains.

Advantage of data-based adaptive forecasting methods becomes clearly evident when we consider $ARFIMA$ $(1, 0.3, 0)$ with a negative AR coefficient $\rho = -0.7$. **Table 4.C.6** reports corresponding $RMSFE$'s. Although the full sample AR(1) forecast consistently beats the benchmark sample mean, it is outperformed by most of the adaptive forecasting techniques including the rolling window methods. Noteworthy differences between the results of $ARFIMA$ $(1, 0.3, 0)$ noises with positive and negative AR coefficients are that margins of gains over the benchmark are higher in the former and that forecasts using data tuned exponential and rolling-window methods become more comparable in the latter.

Methods adopting data-based selection of the downweighting rate, particularly, the dynamic weighting and the exponential AR methods are the most dominant predictors. The residual methods also generate very good forecasts in most of the experiments. Maximum reduction in relative $MSFE$ of the fixed parameter EWMA methods come from methods with very low discounting rates emphasising necessity of including information of distant past. The optimally chosen exponential weights lead to forecasts that are comparable to the forecasts generated by the best performing fixed parameter methods. There is no significant advantage of optimally choosing the evaluation period, $(\widehat{k}, T)$ along with the window size, $\widehat{H}$ and data-based choice does not always provide better forecasts than the fixed window methods. The 'no-change' forecast is by far the worst candidate reporting $RMSFE$'s which are predominantly much higher than unity.

Forecasting under non-stationary noise generated by *ARFIMA* (1,0.75,0) with an AR coefficient of -0.7 further establishes superiority of the data-tuned forecasting techniques by reporting larger gains over the benchmark. **Table 4.C.7** presents the evidence. It also shows that optimally chosen exponential weights can beat the residual methods with marginal gains.

The Monte Carlo experiments with long memory time series noise u_t generated by $ARFIMA$ models confirm that accuracy of forecasts varies with the degree of persistence in the data and consequently, depends on appropriate down-weighting of past observations. The facts that many of the data-tuned discounting methods always match, if not outperform, the best forecast with fixed downweighting parameter and that the optimal rate of discounting cannot be observed in advance, prove the adequacy of data-tuned adaptive forecasting techniques, particularly when facing structural changes.

4.5 Empirical Application

In this section we examine practical usefulness, if any, of data-tuned discounting methods by applying them to real data. We exploit all the methods previously used in the Monte Carlo experiments to forecast a range of UK time series which are available on quarterly and/or monthly frequencies. Giraitis et al. (2013) investigate predictive performance of same methods by forecasting 97 US quarterly series and find many of them, particularly a EWMA with data-selected downweighting parameter and exponential AR, to be significantly superior to a simple full sample AR(1) benchmark. Our forecast exercise is similar to their design, but it is extended and more detailed in several ways. First, we use latest data that include the recent financial crisis period. Second, we analyse forecasts of both untransformed (raw) and transformed (to stationary) series and data with two different frequencies (quarterly and monthly). Third, we perform robustness check by providing results for two different sub-samples.

The quarterly data consist of 55 series and span a long period of 1971Q1: 2012Q4. The dataset includes economic series related to output, production, employment and inflation and financial series related to interest rates, exchange rates among others. We generate one-quarter-ahead forecasts for the last 22 years of the full sample starting in 1999Q1. We evaluate and compare forecasts over two non-overlapping sub-periods of equal size: the first is [1999Q1: 2001Q4] and the second is [2002Q1: 2012Q4]. The monthly data span a much shorter period ranging from January 1993 to December 2012, but contain a larger information set with 79 series.[3] The full forecast period for the monthly dataset is January 2001 to December 2012 and the sub-sample analysis examines one-month-ahead forecasts over two periods each 6 years long: one ranging from January 2001 to December 2006 and the other ranging from January 2007 to December 2012.

The forecasting methods considered are robust to structural change and include methods with exponential, polynomial and non-parametric weights, rolling-window schemes and residual methods. For each series, we compute $MSFE$ relative to the full sample AR(1) benchmark. Full lists of quarterly and monthly series together with $RMSFE$ results are reported in **Tables 4.C.14 - 4.C.16** of Appendix G. Although we provide a detailed series-by-series comparison of models we emphasise that our goal is not to identify the best forecasting strategy for particular series or datasets, but to examine overall benefit from using data-based discounting.

4.5.1 Results for Quarterly Data

We begin by discussing results for one-quarter-ahead forecasts of untransformed data. **Table 4.C.8** summarises them in terms of a number of descriptive statistics and tests. These include the mean, the median, the minimum and the maximum of the relative $MSFE$'s. The columns DM1 and DM2 report the number of significant Diebold-Mariano tests where the null hypoth-

[3] Note that the quarterly and the monthly datasets share a number of series between them.

esis is that of equal forecast ability of a robust data-downweighting method and the benchmark AR(1). The alternative hypothesis for DM1 is that the AR(1) forecasts are more accurate and for DM2 the downweighting method is better. Most of the series appear to be non-stationary. The full sample unconditional mean and forecast methods using rolling-window weights, non-parametric weights and exponential weights with low discount rates perform poorly compared to the full sample AR(1) benchmark. We, therefore, report results for only those methods, either with fixed or data-tuned downweighting parameters, whose forecasts are reasonably comparable to those of the benchmark.

Majority of the reported data-dependent adaptive methods fare very well against the full sample AR(1). In almost all the cases the median $RMSFE$ is below unity. Results are the most impressive for the EWMA with data-tuned discounting, Exponential AR and the two residual methods. Lower mean than median indicates that the methods can yield substantial gains over the benchmark. The largest gains range between 55%-68%, and the maximum cost is no more than 38%. Moreover, while in 25%-36% of the cases forecasts of these adaptive methods are significantly better than those of the benchmark (indicated by DM2 tests), proportions of significantly worst forecasts do not exceed 7%. The exponential AR is, by all means, the best predictor. It yields the maximum reduction of 17% in the mean $RMSFE$ and 7% in the median. More importantly, it concedes no significant outperformance by the benchmark while significantly beating it for 20 series, the maximum among the competing models. The exponential residual method and EWMA with optimised downweighting parameter also perform very well with average $RMSFE$ gains of 11% and 6%, respectively. Their gains at medians are, however, small. Dynamic weighting and the polynomial weighted moving average method match the benchmark AR(1) in terms of median $RMSFE$ and enjoy healthy proportions of significant DM2 tests, but they lose out on mean $RMSFE$. Although the penalty at mean is trivial for the dynamic method, it is enormous for the polynomial weighting. The cost is reflected on the latter's abnormally high maximum loss and large number of significant DM1 tests which offsets almost all its significant benefits.

Among the fixed parameter methods, the EWMA method with the strongest discounting rate and an associated tuning parameter value $\rho = 0.002$ outperforms the AR(1) benchmark marginally on both mean and median. This implies that most of the series in our sample have unit roots and the last observation often serves as a competitive one-quarter-ahead forecast. Nonetheless, the data-tuned exponential discounting proves to be better than the best fixed discounting by several means. Although the two forecast methods are comparable in terms of median, the former enjoys a 4% average $RMSFE$ gain over the latter. Most importantly, while the best gains over the full-sample AR(1) are comparable for both the methods, the largest cost is much smaller in case of data-dependent downweighting. The largest $RMSFE$ for the best fixed exponential weighting is 2.450 compared to a value of 1.175 for the data-tuned weighting. This suggests that the adaptive method is much safer to use, especially knowing that the optimal rate of downweighting cannot be determined in advance. Lower number of significant DM1 tests and higher number of significant DM2 tests also confirm advantage of data-tuned rate of exponential discounting over the best fixed rate.

The sub-sample results establish more pronounced superiority of adaptive forecasting methods during the first half of the sample. All the methods beat the benchmark AR(1) for the median and mean reductions in $RMSFE$ are often large. For example, the gains at the mean and the median are 25% and 21% for the exponential AR and those for the exponential residual method are 16% and 14%, respectively. The minima indicate that at times benefits relative to the benchmark can be extraordinary with gains as high as 88%. In the second sub-sample domination of adaptive techniques appears to be subdued and there are fewer significant outperformance in favour of them. The fixed parameter EWMA enforcing the strongest discounting performs very well, but not better than its counterpart with optimally chosen downweighting parameter.

In order to evaluate practical importance of forecast methods further, we compare their $MSFE$ relative to the AR(1) benchmark for a selection of 15 economically important series. **Table 4.C.9** reports the results. For each series a bold number indicate the smallest $RMSFE$ and consequently, the best forecast. Supreme forecast performance of exponential AR method, which was documented earlier, is evident. It beats the benchmark for almost all the reported series, sometimes with large gains, e.g., 65% for GDP and 33% for unemployment rate. The largest loss it incurs is merely 8% and arises from forecasting total exports of goods and services. Interestingly, the non-parametric residual method performs very well for certain variables, such as consumption expenditure, CPI and money stock. However, at times there can be costs of considerable amount, e.g., costs of 25% for exchange rate and 14% for Index of production on manufacturing. On average, the EWMA with $\rho = 0.002$ which assigns almost all the weights on the last observation forecasts the best among the fixed discounting methods. But there are occasions, such as forecasting of public sector borrowing, where it performs poorly and a relatively lower discount rate with $\rho = 0.60$ achieves the most accurate prediction. This is where application of data-dependent exponential discounting proves to be particularly useful as it almost always matches the best fixed value method or beat it with small gains. Forecasts of polynomial weighting can often match those of exponential weighting, but there are also costly deviations. For example, the reported $RMSFE$ of the polynomial weighted moving average method for total actual weekly hours worked is 2.642 compared to a much lower value of 1.004 for exponential weighting. Overall, many of the adaptive methods forecast the set of indicator variables well, often with substantial gains over the benchmark and with minor costs at the worst cases.

Following Giraitis et al. (2013) we report the most pronounced gains of two well performing data-dependent adaptive forecast methods - EWMA with optimised discounting parameter and exponential AR. For each method, **Table 4.C.10** lists 20 series with the smallest $MSFE$ relative to the full sample AR(1) benchmark. For many of these series outperformances are large, particularly when forecasting using the exponential AR method. The methods are the most beneficial for forecasting output, production, price and employment related variables.

The quarterly series are predominantly non-stationary. We follow Giraitis et al. (2013) to transform them to be stationary and investigate whether such a transformation affects any of the above findings. The stationarisation advantages the previously discarded poor performing methods and makes all methods more comparable to each other in terms of forecast performance. **Table 4.C.11** summarises the results. Overall, the full sample AR(1) benchmark outperforms almost all the competing forecasting models in terms of mean and median $RMSFE$ and number of significant DM statistics. However, not all is ominous. The EWMA with data-tuned discounting parameter and the exponential AR method match the benchmark at mean and most importantly, yield more significant gains (DM2 tests) than significant losses (DM1 tests). For these two and other adaptive methods such as exponential residual method and methods with polynomial and dynamic weights penalties at mean and median are fairly small. The non-parametric weighting and non-parametric residual methods are not particularly useful.

We should note that the forecast of the EWMA with $\rho = 0.002$ or equivalently the 'last observation', which was a competitive contender in forecasting untransformed quarterly series, loses out to the full sample AR(1) benchmark miserably in the all stationary environment. It experiences more than 50% penalty at both mean and median and there is no significant DM2 statistics in contrast to 39 significant DM1 statistics. Fixed parameter EWMA methods with low discounting rates or high values of the tuning parameter, e.g., $\rho = 0.90, 0.80$, perform rather better. Yet, optimally chosen exponential downweighting method beats almost all the fixed value discounting methods on every evaluation criterion. Most convincing are much lower mean $RMSFE$ and maximum penalty and a much higher number of significant outperformances of the benchmark. Contrasting performance of a fixed discount rate for stationary and non-stationary datasets points out the already identified fact: one discount rate is unlikely to be suitable for every dataset or for every series. Generating reliable forecasts and avoiding severe forecast failures in the face of structural change can be achieved by adaptive forecasting with choosing discounting weights optimally over time.

Predictive performance of rolling-window techniques relative to the AR(1) benchmark is not satisfactory. However, it is worthy of noting that, on average, methods with data-selected window yield better forecasts than the fixed-window and window-averaging methods. For data-based optimally chosen window the maximum forecast gains are much higher and the highest cost is much lower. This once again corroborates importance of optimal selection of downweighting parameter using past information.

Good full-sample forecast performance of adaptive EWMA and exponential AR can also be observed over the two sub-samples. The mean and median $RMSFE$ in the two periods indicate that the adaptive EWMA and the dynamic methods (dynamic weighting and residual methods) enjoy better overall advantage in the first sample. For the rolling-window methods performance is opposite - better forecast in the second half of the sample.

4.5.2 Results for Monthly Data

We forecast the 55 quarterly series over a period of 22 years. Although such a long period can accommodate structural changes with higher probabilities, it leaves us with only 88 one-step-ahead forecasts to evaluate. We, therefore, opt to test the predictive performance of adaptive techniques at monthly frequency. The forecast period span a much shorter period of 12 years, but we have more observations and more importantly, a larger information set to work with.

Similar to untransformed quarterly data we identify most of the monthly series to be non-stationary and to disadvantage many of the competing models including rolling window, estimation window averaging and full sample mean, among others. In order to make reasonable comparisons we, therefore, discard any poor performing methods. We decide to report forecast results of the same set of methods which we present in **Table 4.C.8** for untransformed quarterly data. A summary of $RMSFE$ results in terms of descriptive statistics is presented in **Table 4.C.12**. Unlike quarterly forecasts monthly results are less favourable to data-based adaptive forecasting techniques. The exponential AR and the EWMA method with data-tuned discounting parameter are the only two predictors which beat the benchmark AR(1) at both the mean and the median $RMSFE$. The gains are, however, less pronounced than in quarterly data. For the exponential AR the mean and median reductions are 7% and 5% respectively and for the adaptive EWMA they are merely 5% and 2%, respectively. The superiority of the two methods over the benchmark is further substantiated by large number of significant gains over the benchmark and small number of significant losses. For the other adaptive methods, such as the dynamic weighting and residual methods, outperformances of the benchmark are outnumbered by number of significant DM1 tests (favouring the benchmark). There are, however, positives to take. Importantly enough, penalties at mean or median are not more than 5% for these methods and best outperformances are large compared to much lower worst costs, particularly for residual methods. Results for polynomial weights are far from convincing.

The predictive performance of EWMA methods improve with stronger discounting. Among the fixed parameter EWMA methods the one with a discount rate of $\rho = 0.002$ (which is equivalent to assigning all the weights to the last observation) is rewarded the most. It outperforms the AR(1) benchmark marginally on mean $RMSFE$ and yields more significant DM2 statistics than significant DM1 tests. Nevertheless, this best fixed parameter EWMA is no match for the adaptive EWMA with data-tuned downweighting rate. For the latter the mean $RMSFE$ is 6% lower and the proportion of significant outperformances over the AR(1) is 28% higher. Important of all, it is less prone to forecast failure. In the worst case, the $RMSFE$ of optimised EWMA is 1.103 compared to a much higher value of 1.846 for its best fixed value counterpart. Results of the two sub-samples are similar. For most of the adaptive methods there are small improvements in median $RMSFE$ during the more recent sub-period.

As in quarterly forecast we present the best 20 predictions for optimised EWMA and Exponential AR in **Table 4.C.13**. While the EWMA forecasts sales and production well, the exponential AR enjoys clear advantage on employment and price related variables. Both predict tourism related series much better than the AR(1) benchmark.

4.5.3 Forecast Performance During the Crisis

The recent global financial crisis that set out in late 2007 triggered a recession in the UK economy and adversely affected the dynamics of many of its key indicators. For example, output growth became negative and inflation became volatile. Developing or recongising forecast models and methods which can accurately predict such unusual and abrupt economic changes is of crucial practical interest. Barnett et al. (2012) compare a number of dynamic models with time-varying parameters in forecasting the UK GDP growth, inflation and short term interest rate. These authors conclude that although a single best model is difficult to find, allowing for time-variation of specific types proves to be beneficial, particularly during the crisis. We conduct a similar analysis to assess crisis-period predictive performance of each of the robust methods considered in this study. Figure **4.B.10** plots one-step-ahead forecasts of quarterly GDP growth alongside the actual data over 2008 Q1 - 2010 Q4, a period when the crisis deepened. Figure **4.B.11** reports similar results for CPI inflation. The GDP growth and inflation are computed as log difference of quarterly GDP and CPI index values multiplied by 100.

Figure **4.B.10** clearly shows that the crisis initiated a prolonged period of strong negative growths followed by periods of recovery. In 2008 Q2 the actual growth was about -0.9%, but all the forecast methods predicted positive growth with the benchmark AR(1) being the least biased by forecasting near-zero growth rate. The rolling window methods, using either fixed or estimated window lengths, performed miserably and forecasted near constant positive growth during the entire crisis. Most of the other methods, including the benchmark AR(1), were predicting negative GDP growth by mid 2009. Exponentially weighted moving average methods with stronger discouning of past information performed better. It, however, remained difficult to outperform the last observation, meaning that the growth data were highly persistent during the crisis. Nonetheless, forecasts from a number of methods with data-tuned downweighting rate, such as the adaptive EWMA, Exponential residual, Exponential AR and Dynamic weighting methods, could closely match the last observation on many occasions and importantly, they were, on average, superior to AR(1) forecasts. In 2008 Q4 the GDP growth plummeted to its lowest at -2.1%. The adaptive EWMA and Exponential residual appeared to be the two best predictors by forecasting growths lower than -1.5%. Forecasts of the dynamic weighting method and exponential AR were larger but more accurate than the benchmark AR(1) forecast of -0.7%. These two adaptive methods predicted the recovery of the GDP growth in the year 2009 particularly well. Performance of methods based on nonparametric weights were not satisfactory.

It is evident from **Figure 4.B.11** that inflation became volatile and harder to forecast during the crisis. It increased from 0.6% at 2008 Q1 to 2.2% at the end of next quarter. But subsequent large drops made inflation negative and it reached the trough at -2.5% in 2008 Q4. It recovered and rose to 1% over the following two quarters, and remained fairly stable until 2010 Q2 only to experience some fluctuations at the end of 2010.

The 'no change' forecasts (last observations) were the most accurate during periods of stable inflation, but generated large forecast errors during times of volatility. Interestingly, the exponential AR and adaptive EWMA methods which closely imitated the last observations when forecasting GDP growth avoided such forecast error by behaving very differently based on a much slower rate of discounting. For example, the last observation forecast was about 2.2% in contrast to the actual inflation rate of 0.8% in 2008 Q3. Forecasts of adaptive EWMA and exponential AR were much more reliable being close to 1%. The residual methods were amongst the few which were able to forecast negative inflation in 2009 Q1. They matched the last observation and AR(1) forecasts, but were, in general, less biased, particularly when there were substantial fluctuations. For fixed size rolling window methods, the forecast paths stayed horizontal at around 0.8%. There was no real advantage of adaptively choosing the window size from data except that the method that simultaneously selects the tuning parameter and a stable evaluation period forecasted inflation better after 2009 Q3.

4.6 Conclusion

We look at the problem of forecasting time series which are persistent and also subject to ongoing structural change. Forecast methods that are robust to historical and recent structural changes are of particular interest. These include a class of methods that, in one way or another, downweight past data, such as rolling window regression, forecast averaging across different estimation windows and exponentially weighted moving averages. Our work builds on the contribution of Giraitis et al. (2013) who argue that choosing *a priori* a fixed rate of discounting older data is not optimal provided that the nature of structural change is unknown. They propose a data-based selection of tuning parameter and provide theoretical evidence showing that such a technique minimises mean squared forecast errors asymptotically. They further justify good small-sample performance of their adaptive methods via Monte Carlo simulations and practical usefulness by forecasting many US time series. In their econometric framework Giraitis et al. (2013), however, consider persistence in time series only through short-range dependence in the noise process. We bring long memory into the scenario. Long-range dependence is a common feature of many economic and financial time series and is often confused with structural changes. Presence of both poses a difficult challenge for real time forecasting. We shed light on this aspect by justifying, both theoretically and empirically, efficacy of forecast methods with data-tuned discounting rates in such complex situations.

For the theoretical analysis, we prove asymptotic optimality of forecasts based on data-dependent adaptive methods by considering two specific cases - a stationary long memory process and a linear trend process with long memory noise. We establish that for a persistent process, such as the former, a data-selected tuning parameter is not affected by number of observations and for a deterministic trend, such as the latter, it remains bounded. These reasonably imply that forecasts of a time series with long range dependence will rely on the last observation or averaging of the last few available observations.

Next, for the empirical exercise, we consider different degrees of long memory persistence in the noise process which transmits into the original response series to be forecast. We find that long-range dependence generated by $ARFIMA$ models often creates false impressions of different types of structural changes such as cyclical or monotonic trends and conceals presence of true structural changes. A detailed Monte Carlo study confirms effectiveness of data-tuned robust methods in forecasting in face of ongoing structural change when coupled with long memory noise process. For rolling window methods, a cross-validation based selection of window length almost always results in more accurate forecasts than when fixing the size to a predetermined value. Not surprisingly, forecast performance of EWMA methods appears to be sensitive to choice of the tuning parameter. Different values, meaning different rates of downweighting, achieve the best forecast for different degrees of persistence in noise. Importantly, however, we find that adaptive EWMA methods which update the degree of discounting at each forecast horizon can generate reliable forecasts consistently. In spite of presence of various types of structural changes in a time series and varying level of long memory in the noise process, their forecasts are generally as competitive as the best fixed parameter forecasts. We confirm practical usefulness of data-tuned robust methods by forecasting several economic and financial time series of the UK at both quarterly and monthly frequencies. There are large benefits to gain with rare evidence of adverse penalties.

4.A Appendix: Proofs

4.A.1 Proof of Theorems 1 and 2 and Corollary 1

We follow the same steps of the proof as in Giraitis et al. (2013). By definition

$$Q_{T,H} = T_n^{-1} \sum_{t=T_0}^{T} (y_t - \hat{y}_{t|t-1,\,H})^2 = T_n^{-1} \sum_{t=T_0}^{T} \left(\sum_{j=1}^{t-1} w_{tj,H}(y_t - y_{t-j}) \right)^2,$$

$$\omega_{T,H} = E(y_{T+1} - \hat{y}_{T+1|T,H})^2 = E\left(\sum_{j=1}^{T} w_{T+1,j,H}(y_{T+1} - y_{T+1-j}) \right)^2.$$

We will approximate $Q_{T,H}$ and $\omega_{T,H}$ by the sums

$$Q_{T,H}^{(apr)} = T_n^{-1} \sum_{t=T_0}^{T} \left(\sum_{j=1}^{T_1} w_{j,H}(y_t - y_{t-j}) \right)^2,$$

$$\omega_{T,H}^{(apr)} = E\left(\sum_{j=1}^{T_1} w_{j,H}(y_{T+1} - y_{T+1-j}) \right)^2,$$

replacing $w_{tj,H}$ by $w_{j,H}$ defined by (4.2.6), setting $T_1 = T_0 T^{-\delta/2}$. Since $H_{max} = T_0 T^{-\delta}$, then $T_0/H_{max} = T^{\delta}$, $T_1/H_{max} = T^{\delta/2}$ and $T_1/T_0 \leq T^{-\delta/2}$.

The proof of Theorems 1 and 2 is based on Lemma 1, Lemma 2, and Lemma 3 given in Section 4.A.2. These lemmas divide the proof into 3 steps, establishing required approximations.

Proof of Theorem 1. Write

$$Q_{T,H} = \hat{\sigma}^2_{T,u} + \left[Q_{T,H} - Q_{T,H}^{(apr)}\right] + E\left\{Q_{T,H}^{(apr)} - \hat{\sigma}^2_{T,u}\right\} \quad (4.A.1)$$

$$+ \quad \left[Q_{T,H}^{(apr)} - \hat{\sigma}^2_{T,u} - E\{Q_{T,H}^{(apr)} - \hat{\sigma}^2_{T,u}\}\right].$$

Recall that $H \leq T$ and $0 < d < 1/2$. Uniformly in H, by Lemma 1, $\left[Q_{T,H} - Q_{T,H}^{(apr)}\right] = O(T^{-2})$, by Lemma 2, $E\left\{Q_{T,H}^{(apr)} - \hat{\sigma}^2_{T,u}\right\} = q_{u,H} + O(T^{-2})$

and by Lemma 3,
$\Big[Q^{(apr)}_{T,H} - \hat{\sigma}^2_{T,u} - E\{Q^{(apr)}_{T,H} - \hat{\sigma}^2_{T,u}\}\Big] = o_H(H^{-1+2d})$. Hence,

$$Q_{T,H} = \hat{\sigma}^2_{T,u} + q_{u,H} + o_H(H^{-1+2d}),$$

where by Lemma 2(i), $q_{u,H} = \lambda_{LM}H^{-1+2d} + o_H(H^{-1+2d})$. This completes the proof of the theorem. □

Proof of Theorem 2. Recall equality (4.A.1). Then, uniformly in H, by Lemma 1, $\Big[Q_{T,H} - Q^{(apr)}_{T,H}\Big] = O(T^{-2})$, by Lemma 2, $E\Big\{Q^{(apr)}_{T,H} - \hat{\sigma}^2_{T,u}\Big\} = q_{\beta,H} + q_{u,H} + o_H(H^2)$ and by Lemma 3, $\Big[Q^{(apr)}_{T,H} - \hat{\sigma}^2_{T,u} - E\{Q^{(apr)}_{T,H} - \hat{\sigma}^2_{T,u}\}\Big] = o_H(H^2)$. Hence,

$$Q_{T,H} = \hat{\sigma}^2_{T,u} + q_{\beta,H} + q_{u,H} + o_H(H^2),$$

where by Lemma 2(ii), $q_{\beta,H} + q_{u,H} = \kappa H^2 + o_H(H^2)$. This completes the proof of the theorem. □

Proof of Corollary 1. Proof follows using the same argument as in the case of Corollary 1 in Giraitis et al. (2013). Let $q_{u,H}$ reaches its minimum $c_0 = q_{u,H_0}$ at some finite H_0. Since $\hat{\sigma}^2_{T,u} = \sigma^2_u + o_p(1)$, then (4.3.5) implies that $Q_{T,\hat{H}} = c_0 + o_p(1)$, $\omega_{T,H_{opt}} = c_0 + o(1)$, which in turn implies $\omega_{T,\hat{H}} = Q_{T,\hat{H}} + o(1) = c_0 + o(1)$. Hence, $\omega_{T,\hat{H}} = \omega_{T,H_{opt}} + o(1)$ and $Q_{T,\hat{H}} = \omega_{T,\hat{H}} + o_p(1)$. This proves (4.3.6). □

4.A.2 Main lemmas

Step 1. First, using Lemma 1 we show that

$$Q_{T,H} = Q^{(apr)}_{T,H} + \Big\{Q_{T,H} - Q^{(apr)}_{T,H}\Big\} = Q^{(apr)}_{T,H} + O(T^{-2}),$$

uniformly in H.

Lemma 1 *Under assumptions of Theorems 1 and 2,*

$$E\left[\sup_{H\in I_T}\left|Q_{T,H}-Q_{T,H}^{(apr)}\right|\right]=O(T^{-2}), \tag{4.A.2}$$

$$\sup_{H\in I_T}\left|\omega_{T,H}-\omega_{T,H}^{(apr)}\right|=O(T^{-2}).$$

Proof. The proof is similar to the proof of the Lemma A.1 in Giraitis et al. (2013). We provide here all details.

Notice that by definition of weights, $w_{tk,H}\le 1$ and $w_{j,H}\le 1$. To evaluate $Q_{T,H}-Q_{T,H}^{(apr)}$, for $T_0\le t\le T$ and $j,k\le t-1$ we shall use the following bound: $|w_{tj,H}w_{tk,H}-w_{j,H}w_{k,H}I(j,k\le T_1)|\le |w_{tj,H}w_{tk,H}-w_{j,H}w_{k,H}|+|w_{j,H}w_{k,H}-w_{j,H}w_{k,H}I(j,k\le T_1)|\le |w_{tj,H}-w_{j,H}|w_{tk,H}+w_{j,H}|w_{tk,H}-w_{k,H}|+w_{j,H}w_{k,H}I(j>T_1\ or\ k>T_1)\le |w_{tj,H}-w_{j,H}|+|w_{tk,H}-w_{k,H}|+w_{j,H}I(j>T_1)+w_{k,H}I(k>T_1)\le CT^{-6}$ because $|w_{tj,H}-w_{j,H}|\le CT^{-6}$ by (4.A.24) and $w_{j,H}I(j>T_1)\le CT^{-6}$ by (4.A.23).

Hence,

$$\begin{aligned}
&|Q_{T,H}-Q_{T,H}^{(apr)}|\\
&\quad\le T_n^{-1}\sum_{t=T_0}^{T}\sum_{j,k=1}^{t-1}|w_{tj,H}w_{tk,H}-w_{j,H}w_{k,H}I(j,k\le T_1)|\\
&\qquad\times|(y_t-y_{t-j})(y_t-y_{t-k})|\\
&\quad\le CT_n^{-1}\textstyle\sum_{t=T_0}^{T}\sum_{j,k=1}^{t-1}T^{-6}|(y_t-y_{t-j})(y_t-y_{t-k})|=:j_T.
\end{aligned}$$

Since j_T does not depend on H it remains to show that

$$Ej_T=O(T^{-2}).$$

We have

$$Ej_T\le CT_n^{-1}\sum_{t=T_0}^{T}\sum_{j,k=1}^{t-1}T^{-6}E|(y_t-y_{t-j})(y_t-y_{t-k})|.$$

We use the bound $E|(y_t - y_{t-j})(y_t - y_{t-k})| \leq E(y_t - y_{t-j})^2 + E(y_t - y_{t-k})^2 \leq 2(Ey_t^2 + Ey_{t-j}^2) + 2(Ey_t^2 + Ey_{t-k}^2) \leq 8 \max_{t=1,...,T} Ey_t^2$.

In Theorem 1, $Ey_t^2 = E(\mu + u_t)^2 \leq 2\mu^2 + 2Eu_t^2 = 2\mu^2 + 2Eu_1^2 < \infty$, because u_t is a stationary sequence with finite variance.

In Theorem 2, $Ey_t^2 = E(at + u_t)^2 \leq 2a^2t^2 + 2Eu_t^2 \leq 2a^2T^2 + 2Eu_1^2 \leq CT^2$ for $t = 1, ..., T$.

Therefore, $E|(y_t - y_{t-j})(y_t - y_{t-k})| \leq CT^2$ where C does not depend on t, j, k. So,

$$Ej_T \leq CT_n^{-1}T^2 \sum_{t=T_0}^{T} \sum_{j,k=1}^{t-1} T^{-6} \leq CT_n^{-1}T^{-1}.$$

Since $T_n = T - T_0 + 1 \sim T$, then $Ej_T \leq CT_n^{-1}T^2 \sum_{t=T_0}^{T} \sum_{j,k=1}^{t-1} T^{-6} \leq CT^{-2}$ which proves the first claim of the lemma.

To show the second claim, we use the bound we obtained above: $E|(y_{T+1} - y_{T+1-j})(y_{T+1} - y_{T+1-k})| \leq CT^2$. Then,

$$\begin{aligned}
&|\omega_{T,H} - \omega_{T,H}^{(apr)}| \\
&= E\left|\sum_{j,k=1}^{T}(w_{tj,H}w_{tk,H} - w_{j,H}w_{k,H}I(j,k \leq T_1))\right. \\
&\times(y_{T+1} - y_{T+1-j})(y_{T+1} - y_{T+1-k})| \\
&\leq \sum_{j,k=1}^{T} |w_{tj,H}w_{tk,H} - w_{j,H}w_{k,H}I(j,k \leq T_1)| E|(y_{T+1} - y_{T+1-j})y_{T+1} - y_{T+1-k})| \\
&\leq CT^2 \sum_{j,k=1}^{T} |w_{tj,H}w_{tk,H} - w_{j,H}w_{k,H}I(j,k \leq T_1)| \\
&\leq CT^2 \sum_{j,k=1}^{T} CT^{-6} \leq CT^{-2},
\end{aligned}$$

using the inequalities we obtained above. This proves the second claim. □

Step 2. Next we obtain asymptotics of $E(Q_{T,H}^{(apr)} - \hat{\sigma}_{T,u}^2)$ and $\omega_{T,H}^{(apr)}$.

Lemma 2 *(i) Under assumptions of Theorem 1,*

$$E(Q_{T,H}^{(apr)} - \hat{\sigma}_{T,u}^2) = q_{u,H} + O(T^{-2}), \tag{4.A.3}$$

$$\omega_{T,H}^{(apr)} = \sigma_u^2 + q_{u,H} + O(T^{-2}),$$

where $q_{u,H}$ *is the same as in Theorem 1 and*

$$q_{u,H} = \lambda_{LM} H^{-1+2d} + o_H(H^{-1+2d}). \tag{4.A.4}$$

(ii) Under assumptions of Theorem 2,

$$E(Q^{(apr)}_{T,H} - \hat{\sigma}^2_{T,u}) = q_{\beta,H} + q_{u,H} + o_H(H^2), \tag{4.A.5}$$

$$\omega^{(apr)}_{T,H} = \sigma^2_u + q_{\beta,H} + q_{u,H} + o_H(H^2),$$

where $q_{\beta,H} = \kappa H^2 + o_H(H^2)$*, and*

$$q_{\beta,H} + q_{u,H} = \kappa H^2 + o_H(H^2). \tag{4.A.6}$$

Proof. Since in Theorems 1 and 2, β_t is deterministic, and u_t is a stationary LM sequence with zero mean, $Eu_j = 0$, then

$$E[Q^{(apr)}_{T,H} - \hat{\sigma}^2_{T,u}] = m_{\beta,TH} + m_{u,TH}, \quad \omega^{(apr)}_{T,H} = v_{\beta,TH} + v_{u,TH}, \tag{4.A.7}$$

where $m_{\beta,TH} := T_n^{-1} \sum_{t=T_0}^{T} E\left(\sum_{j=1}^{T_1} w_{j,H}(\beta_t - \beta_{t-j})\right)^2$,

$$\begin{aligned} m_{u,TH} :=& T_n^{-1} \sum_{t=T_0}^{T} E\left(\sum_{j=1}^{T_1} w_{j,H}(u_t - u_{t-j})\right)^2 - \sigma^2_u \\ =& E\left(\sum_{j=1}^{T_1} w_{j,H}(u_0 - u_{-j})\right)^2 - \sigma^2_u, \end{aligned}$$

$$\begin{aligned} v_{\beta,TH} :=& E\left(\sum_{j=1}^{T} w_{j,H}(\beta_{T+1} - \beta_{T+1-j})\right)^2, \\ v_{u,TH} :=& E\left(\sum_{j=1}^{T} w_{j,H}(u_{T+1} - u_{T+1-j})\right)^2. \end{aligned}$$

(i). Suppose that conditions of Theorem 1 are satisfied. Then $\beta_t = \mu$ and hence $m_{\beta,TH} = 0$, $v_{\beta,TH} = 0$.

Thus, to prove (4.A.3) we need to show that

$$m_{u,TH} - q_{u,H} = O_H(T^{-2}), \qquad v_{u,TH} - \sigma^2_u - q_{u,H} = O_H(T^{-2}). \tag{4.A.8}$$

By definition, $q_{u,H} = E\left(\sum_{j=1}^{\infty} w_{j,H}(u_0 - u_{-j})\right)^2 - \sigma_u^2$. We showed that $E|(u_0 - u_{-j})(u_0 - u_{-k})| \leq 8Eu_0^2 < \infty$, also we have $\sum_{k=1}^{\infty} w_{k,H} = 1$ by definition of $w_{k,H}$, and $\sum_{j=T_1}^{\infty} w_{j,H} = O(T^{-6})$ by (4.A.23). So,

$|m_{u,TH} - q_{u,H}|$

$\leq \sum_{j,k=1}^{\infty} w_{j,H} w_{k,H} (I(j > T_1) + I(k > T_1)) E|(u_0 - u_{-j})(u_0 - u_{-k})|$

$\leq C \sum_{j,k=T_1}^{\infty} w_{j,H} w_{k,H} (I(j > T_1) + I(k > T_1)) \leq C \sum_{j=T_1}^{\infty} w_{j,H} \sum_{k=1}^{\infty} w_{k,H} \leq CT^{-6}$.

Similarly, by stationarity, $v_{u,TH} := E\left(\sum_{j=1}^{T} w_{j,H}(u_0 - u_{-j})\right)^2$ and $|v_{u,TH} - \sigma_u^2 - q_{u,H}| \leq |E\left(\sum_{j=1}^{T} w_{j,H}(u_0 - u_{-j})\right)^2 - E\left(\sum_{j=1}^{\infty} w_{j,H}(u_0 - u_{-j})\right)^2| \leq CT^{-6}$. This completes the proof of (4.A.8).

It remains to prove (4.A.4). By assumption, $\gamma_u(k) \sim c_\gamma k^{-1+2d}$ as $k \to \infty$. Therefore,

$$
\begin{aligned}
q_{u,H} &= E\left(\sum_{j=1}^{\infty} w_{j,H}(u_0 - u_{-j})\right)^2 - \sigma_u^2 \\
&= \sum_{j,k=1}^{\infty} w_{j,H} w_{k,H} E u_{-j} u_{-k} - 2\sum_{j=1}^{\infty} w_{j,H} E u_0 u_{-j} + \sum_{j=1}^{\infty} w_{j,H} E u_0^2 - \sigma_u^2 \\
&= \sum_{j,k=1}^{\infty} w_{j,H} w_{k,H} \gamma_u(j-k) - 2\sum_{j=1}^{\infty} w_{j,H} \gamma_u(j).
\end{aligned}
$$

By definition, $w_{j,H} = K(j/H)/v_H$ where $v_H = \sum_{j=1}^{\infty} K(j/H) \sim H$ by (4.A.27). Hence, approximating the sum by the integral and change of variables gives

$$
\begin{aligned}
&\sum_{j,k=1}^{\infty} w_{j,H} w_{k,H} \gamma_u(j-k) \sim H^{-2} \sum_{j,k=1}^{\infty} K(j/H) K(k/H) c_\gamma |j-k|^{-1+2d} \\
&\sim H^{-1+2d} H^{-2} \int_0^{\infty} \int_0^{\infty} K(x/H) K(y/H) c_\gamma |x/H - y/H|^{-1+2d} dx dy \\
&\sim H^{-1+2d} \int_0^{\infty} \int_0^{\infty} K(x) K(y) c_\gamma |x-y|^{-1+2d} dx dy.
\end{aligned}
$$

Similarly,

$$2\sum_{j=1}^{\infty} w_{j,H}\gamma_u(j) \sim 2H^{-1}\sum_{j=1}^{\infty} K(j/H)c_\gamma j^{-1+2d}$$
$$\sim 2H^{-1+2d}c_\gamma \int_0^{\infty} K(x)x^{-1+2d}dx$$

which implies $q_{u,H} \sim H^{-1+2d}c_\gamma\big(\int_0^\infty \int_0^\infty K(x)K(y)|x-y|^{-1+2d}dxdy$ $-2\int_0^\infty K(x)x^{-1+2d}dx\big) = H^{-1+2d}\lambda_{LM}$ proving (4.A.4). This completes the proof of part (i).

(ii). Suppose that conditions of Theorem 2 are satisfied. Equalities and (4.A.2) show that to prove (4.A.5) it suffices to show that

$$m_{\beta,TH} - q_{\beta,H} = o_H(T^{-2}), \qquad v_{\beta,TH} - q_{\beta,H} = o_H(T^{-2}). \tag{4.A.9}$$

Since $\beta_t = at$ and $T - T_0 + 1 = T_n$, then $m_{\beta,TH} = T_n^{-1}\sum_{t=T_0}^{T}\Big(\sum_{j=1}^{T_1} w_{j,H}aj\Big)^2$ $= \Big(\sum_{j=1}^{T_1} w_{j,H}aj\Big)^2$, and $v_{\beta,TH} = \Big(\sum_{j=1}^{T} w_{j,H}aj\Big)^2$. On the other hand, $\gamma_{\beta,H} = \Big(\sum_{j=1}^{\infty} w_{j,H}aj\Big)^2$. Then, by equality $a^2 - b^2 = (a-b)(a+b)$,

$$|m_{\beta,TH} - \gamma_{\beta,H}| \le \left|\left(\sum_{j=1}^{T_1} w_{j,H}aj\right)^2 - \left(\sum_{j=1}^{\infty} w_{j,H}aj\right)^2\right|$$
$$\le 2a^2\left(\sum_{j=T_1}^{\infty} w_{j,H}j\right)\left(\sum_{k=1}^{\infty} w_{k,H}k\right).$$

Since $H \le T$, $\sum_{j=T_1}^{\infty} w_{j,H}(j/H) = O(T^{-6})$ by (4.A.23) and $\sum_{j=1}^{\infty} w_{j,H}(j/H) = O(1)$ by (4.A.27), we obtain

$$|m_{\beta,TH} - \gamma_{\beta,H}| \le H^2 O(T^{-6})O(1) = o(T^{-2}),$$

which proves the first claim in (4.A.9). The second claim follows using the same argument.

Property (4.A.6) follows applying to $q_{\beta,H} + q_{u,H}$ property (4.A.4), $q_{u,H} = \lambda_{LM}H^{-1+2d} + o_H(H^{-1+2d})$, and noting that $q_{\beta,H} = (\sum_{j=1}^{\infty} w_{j,H}j)^2 \sim H^2(\int_0^\infty K(x)xdx)^2 = H^2\kappa$ by (4.A.27).

This completes the proof of the lemma. □

Step 3. Here we establish the bound for the stochastic term.

Lemma 3 *(i) Under assumptions of Theorem 1,*

$$E \sup_{H \in I_T} H^{1-2d} \big| Q_{T,H}^{(apr)} - \hat{\sigma}_{T,u}^2 - E\{Q_{T,H}^{(apr)} - \hat{\sigma}_{T,u}^2\} \big| \to 0.$$

(ii) Under assumptions of Theorem 2,

$$E \sup_{H \in I_T} H^{-2} \big| Q_{T,H}^{(apr)} - \hat{\sigma}_{T,u}^2 - E\{Q_{T,H}^{(apr)} - \hat{\sigma}_{T,u}^2\} \big| \to 0. \tag{4.A.10}$$

Proof. Denote $\beta_{tj} = \beta_t - \beta_{t-j}$, $u_{tj} = u_t - u_{t-j}$. Then, $\sum_{j=1}^{T_1} w_{j,H}(y_t - y_{t-j}) = \sum_{j=1}^{T_1} w_{j,H}\beta_{tj} + \sum_{j=1}^{T_1} w_{j,H}u_{tj}$. So,

$$Q_{T,H}^{(apr)} = T_n^{-1} \sum_{t=T_0}^{T} \left(\sum_{j=1}^{T_1} w_{j,H}(y_t - y_{t-j}) \right)^2 \tag{4.A.11}$$

$$= J_{\beta\beta,TH} - 2J_{\beta u,TH} + J_{uu,TH}, \tag{4.A.12}$$

where

$$J_{\beta\beta,TH} = T_n^{-1} \sum_{t=T_0}^{T} \left(\sum_{j=1}^{T_1} w_{j,H}\beta_{tj} \right)^2,$$

$$J_{uu,TH} = T_n^{-1} \sum_{t=T_0}^{T} \left(\sum_{j=1}^{T_1} w_{j,H}u_{tj} \right)^2,$$

$$J_{\beta u,TH} = T_n^{-1} \sum_{t=T_0}^{T} \left(\sum_{j=1}^{T_1} w_{j,H}\beta_{tj} \right) \left(\sum_{k=1}^{T_1} w_{k,H}u_{tk} \right).$$

(i) Since $\beta_{tj} = \mu - \mu = 0$, we have $J_{\beta\beta,TH} = J_{\beta u,TH} = 0$. It remains to show that

$$E \sup_{H \in I_T} H^{1-2d} \big| J_{uu,TH} - \hat{\sigma}_{T,u}^2 - E\{J_{uu,TH} - \hat{\sigma}_{T,u}^2\} \big| \to 0. \tag{4.A.13}$$

We conduct the proof similarly as in the proof of Lemma A.3 of Giraitis et al. (2013), amending it to long memory of u_t's.

Set $w'_{j,H} := w_{j,H} - w_{j+1,H},\ j = 1,\cdots T_1 - 1,\ w'_{T_1,H} := w_{T_1,H},\ \eta_{tj} = \sum_{s=1}^{j} u_{t-s},\ j = 1,\cdots,T_1$ and $h_T := \sum_{j=1}^{T_1} w_{j,H}$. Using summation by parts, write

$$\sum_{j=1}^{T_1} w_{j,H}u_{t-j} = \sum_{j=1}^{T_1-1}(w_{j,H} - w_{j+1,H})\eta_{tj} + w_{T_1,H}\eta_{tT_1} = \sum_{j=1}^{T_1} w'_{j,H}\eta_{tj}.$$

Then, $\sum_{j=1}^{T_1} w_{j,H}u_{tj} = h_T u_t - \sum_{j=1}^{T_1} w_{j,H}u_{t-j} = h_T u_t - \sum_{j=1}^{T_1} w'_{j,H}\eta_{tj}$, and

$$\begin{aligned} J_{uu,TH} &= T_n^{-1}\sum_{t=T_0}^{T}\left(h_T u_t - \sum_{j=1}^{T_1} w'_{j,H}\eta_{tj}\right)^2 \\ &= T_n^{-1}\sum_{t=T_0}^{T}\left\{\left(\sum_{j=1}^{T_1} w'_{j,H}\eta_{tj}\right)^2 - 2h_T\left(\sum_{j=1}^{T_1} w'_{j,H}\eta_{tj}\right)u_t\right\} + h_T^2\hat{\sigma}_{T,u}^2 \end{aligned}$$

Hence,

$$\begin{aligned} J_{uu,TH} - \hat{\sigma}_{T,u}^2 \quad = \quad & \sum_{j,k=1}^{T_1} w'_{j,H}w'_{k,H}\left(T_n^{-1}\sum_{t=T_0}^{T}\eta_{tj}\eta_{tk}\right) - 2h_T\sum_{j=1}^{T_1} w'_{j,H}\left(T_n^{-1}\sum_{t=T_0}^{T}\eta_{tj}u_t\right) \\ & +(h_T^2 - 1)\hat{\sigma}_{T,u}^2. \end{aligned}$$

Denote

$$\begin{aligned} S_{\eta\eta,T,jk} &:= T_n^{-1}\sum_{t=T_0}^{T}\left(\eta_{tj}\eta_{tk} - E\eta_{tj}\eta_{tk}\right), \\ S_{\eta u,T,jj} &:= T_n^{-1}\sum_{t=T_0}^{T}\left(\eta_{tj}u_t - E\eta_{tj}u_t\right). \end{aligned}$$

Then

$$\begin{aligned} R_{H,t} :=& \left|J_{uu,TH} - \hat{\sigma}_{T,u}^2 - E[J_{uu,TH} - \hat{\sigma}_{T,u}^2]\right| \\ &\leq \sum_{j,k=1}^{T_1} |w'_{j,H}w'_{k,H}|\left|S_{\eta\eta,T,jk}\right| + 2\sum_{j=1}^{T_1} w'_{j,H}\left|S_{\eta u,T,jk}\right| + |h_T^2 - 1|(\hat{\sigma}_{T,u}^2 - \sigma_u^2). \end{aligned}$$

Using (4.A.25), we can bound $H^{1-2d}|w'_{j,H}w'_{k,H}| \leq |H^{1/2-d}w'_{j,H}H^{1/2-d}w'_{k,H}| \leq C(jk)^{-3/2-d}$ and $H^{1-2d}|w'_{j,H}| \leq Cj^{-1-d}$. By (4.A.23), $1-h_T^2 \leq 2(1-h_T) \leq 2\sum_{j=T_1}^{\infty} w_{j,H} = O(T^{-6})$, and $h_T \leq \sum_{j=1}^{\infty} w_{j,H} = 1$. Therefore, $H^{1-2d}(1-h_T^2) \leq CHT^{-6} \leq CT^{-5} = o(1)$. Moreover $E\hat{\sigma}^2_{T,u} = T_n^{-1}\sum_{t=T_0}^{T} Eu_t^2 = \sigma_u^2$. Therefore,

$$H^{1-2d}R_{H,t} \leq r_t := C\left(\sum_{j,k=1}^{T_1}(jk)^{-3/2-d}\left|S_{\eta\eta,T,jk}\right| + \sum_{j=1}^{T_1} j^{-1-2d}\left|S_{\eta u,T,jj}\right|\right) + o(1).$$

We will show that for some $\epsilon' > 0$,

$$E\left|S_{\eta\eta,T,jk}\right| \leq C(jk)^{1/2+d}T^{-\epsilon'}, \qquad E\left|S_{\eta u,T,jj}\right| \leq Cj^{1/2+d}T^{-1/2+d}, \quad (4.A.14)$$

which implies (4.A.13):

$$\begin{aligned} Er_t &\leq C\left[\sum_{j,k=1}^{T_1}(jk)^{-3/2-d}E\left|S_{\eta\eta,T,jk}\right| + \sum_{j=1}^{T_1} j^{-1-2d}E\left|S_{\eta u,T,jj}\right|\right] \\ &\leq C\left[T^{-\epsilon}\left(\sum_{j,k=1}^{T_1} j^{-1}\right)^2 + \sum_{j=1}^{T_1} j^{-1/2-d}T^{-1/2+d}\right] \\ &\leq C[T^{-\epsilon}\log^2 T + (T_1/T)^{1/2-d}] \to 0 \text{ as } T \to \infty. \end{aligned}$$

It remains to show (4.A.14). We will use the following general bounds obtained in the proof of Lemma A.4 in Giraitis et al. (2013):

$$ES^2_{\eta\eta,T,jk} \leq CT^{-2}\sum_{t',t=T_0}^{T} E[\eta_{t'j}\eta_{tj}]E[\eta_{t'k}\eta_{tk}], \quad (4.A.15)$$

$$ES^2_{\eta u,T,jj} \leq CT^{-2}\sum_{t',t=T_0}^{T} E[\eta_{t'j}\eta_{tj}]E[u_{t'}u_t].$$

We will show that

(a) for all $T_0 \leq t, t' \leq T$ and $1 \leq j \leq T_1$,

$$|E[\eta_{t'j}\eta_{tj}]E[\eta_{t'k}\eta_{tk}]| \leq C(jk)^{1+2d}, \quad |E[\eta_{t'j}\eta_{tj}]E[u_{t'}u_t]| \leq Cj^{1+2d}. \quad (4.A.16)$$

(b) There exists $0 < \epsilon < 1$ and $\theta > 0$ such that for all $1 \leq j \leq T_1$ and $T_0 \leq t, t' \leq T$ such that $|t - t'| \geq T^{1-\theta}$,

$$|E[\eta_{t'j}\eta_{tj}]E[\eta_{t'k}\eta_{tk}]| \leq C(jk)^{1+2d}T^{-\epsilon} \tag{4.A.17}$$

Now we prove the first claim of (4.A.14). By (4.A.16)-(4.A.17),

$$\begin{aligned} ES^2_{\eta\eta,T,jk} \leq CT^{-2} & \sum_{t',t=T_0:\, |t'-t|\leq T^{1-\theta}}^{T} |E[\eta_{t'j}\eta_{tj}]E[\eta_{t'k}\eta_{tk}]| \\ & + CT^{-2} \sum_{t',t=T_0:\, |t'-t|> T^{1-\theta}}^{T} |E[\eta_{t'j}\eta_{tj}]E[\eta_{t'k}\eta_{tk}]| \\ & \leq C(jk)^{1+2d}T^{-2}[\sum_{t',t=T_0:\, |t'-t|\leq T^{1-\theta}}^{T} 1 + \sum_{t',t=T_0:\, |t'-t|> T^{1-\theta}}^{T} T^{-\epsilon}] \\ & \leq C(jk)^{1+2d}(T^{-\theta} + T^{-\epsilon}). \end{aligned}$$

Hence, $E|S_{\eta\eta,T,jk}| \leq E(S^2_{\eta\eta,T,jk})^{1/2} \leq C(jk)^{1/2+d}T^{-\min(\epsilon,\theta)/2}$ which proves (4.A.14) for $E|S_{\eta\eta,T,jk}|$.

Now we prove the second claim of (4.A.14) for $E|S_{\eta u,T,jk}|$.

$$
\begin{aligned}
ES^2_{\eta u,T,jj} &\leq CT^{-2}\sum_{t',t=T_0}^{T} E[\eta_{t'j}\eta_{tj}]E[u_{t'}u_t] \\
&\leq CT^{-2}j^{1+2d}\sum_{t',t=T_0}^{T} E[u_{t'}u_t] \\
&\leq CT^{-2}j^{1+2d}\sum_{t',t=T_0}^{T} \gamma_u(t^{'}-t) \\
&\leq CT^{-2}j^{1+2d}\sum_{t',t=T_0}^{T} \left(1+\left|t^{'}-t\right|\right)^{-1+2d} \\
&\leq CT^{-2}j^{1+2d}\sum_{t=T_0}^{T}\sum_{k=0}^{T}(1+|k|)^{-1+2d} \\
&\leq CT^{-2}j^{1+2d}TT^{2d} \\
&\leq Cj^{1+2d}T^{-1+2d}
\end{aligned}
$$

Hence, $E|S_{\eta u,T,jj}| \leq Cj^{1/2+d}T^{-1/2+d}$.

Proof of (4.A.16). Recall that long memory assumption $Eu_ju_{j-k} = \gamma_u(k) \sim c_\gamma|k|^{-1+2d}$ as $k \to \infty$. Hence, by stationarity of u_j,

$$
\begin{aligned}
E\eta^2_{tj} &= E\left(\sum_{s=1}^{j} u_{t-s}\right)^2 = E\left(\sum_{s,k=1}^{j} u_{t-s}u_{t-k}\right) \\
&= \sum_{s,k=1}^{j}\gamma_u(s-k) = \sum_{s=1}^{j}\gamma_u(0) + 2\sum_{s=1}^{j}\sum_{k=1}^{s-1}\gamma_u(s-k). \\
&\leq j\gamma_u(0) + 2j\sum_{k=1}^{j}|\gamma_u(k)| \leq C\left(j + j\sum_{k=1}^{j}k^{-1+2d}\right) \\
&\leq C\left(j + j\int_0^j x^{-1+2d}dx\right) \leq Cj^{1+2d}.
\end{aligned}
$$

Thus, using Cauchy-Schwarz inequality we can write

$$|E[\eta_{t'j}\eta_{tj}]| \leq (E\eta_{t'j}^2 E\eta_{tj}^2)^{1/2} \leq Cj^{1+2d}, \tag{4.A.18}$$

$$|E[\eta_{t'j}\eta_{tj}]E[\eta_{t'k}\eta_{tk}]| \leq C(jk)^{1+2d},$$

which proves (4.A.16).

Proof of (4.A.17). Observe the following. Let $t'-t > T^{1-\theta}$. Then for any $s \in [t'-j, t']$, $i \in [t-j, t]$ and $j \leq T_1$, for large T it holds

$$s - i \geq (t'-j) - t = (t'-t) - j = (t'-t)/2$$

because $j \leq T_1 \leq T^{1-\delta/2} = o(T^{1-\theta})$, for our chosen $\theta = \delta/4$. This implies

$$|\gamma_u(s-i)| \leq C|s-i|^{-1+2d} \leq C|t'-t|^{-1+2d} \leq C(T^{1-\theta})^{-1+2d}.$$

Then,

$$\begin{aligned} |E[\eta_{t'j}\eta_{tj}]| &= \left| E\left(\sum_{s=1}^{j} u_{t'-s}\right)\left(\sum_{i=1}^{j} u_{t-i}\right)\right| = \left| E\left(\sum_{s=t'-j}^{t'-1} u_s\right)\left(\sum_{i=t-j}^{t-1} u_i\right)\right| \\ &\leq \sum_{s=t'-j}^{t'-1}\sum_{i=t-j}^{t-1} |\gamma_u(s-i)| \leq C \sum_{s=t'-j}^{t'-1}\sum_{i=t-j}^{t-1} (T^{1-\theta})^{-1+2d} \\ &\leq C(T^{1-\theta})^{-1+2d} j^2. \end{aligned}$$

By definition, $j \leq T_1 \leq T^{1-\delta/2} = T^{-\delta/4}T^{1-\delta/4} = T^{-\theta}T^{1-\theta}$, so we can bound $j^{1-2d} \leq (T^{-\theta}T^{1-\theta})^{1-2d}$. Hence

$$|E[\eta_{t'j}\eta_{tj}]| \leq C(T^{1-\theta})^{-1+2d}(j^{1+2d})(j^{1-2d}) \leq CT^{-\theta(1-2d)}j^{1+2d}. \tag{4.A.19}$$

Therefore,

$$|E[\eta_{t'j}\eta_{tj}]E[\eta_{t'k}\eta_{tk}]| \leq CT^{-\epsilon}(jk)^{1+2d} \tag{4.A.20}$$

with $\epsilon = 2\theta(1-2d)$ which proves (4.A.17).

(ii) In (ii), $\beta_t = at$, $\beta_{tj} = aj$, so $J_{\beta\beta,TH}$ is deterministic and $EJ_{\beta u,TH} = 0$ in (4.A.11). Denote

$$R_H := H^{-2}\big|Q^{(apr)}_{T,H} - \hat\sigma^2_{T,u} - E\{Q^{(apr)}_{T,H} - \hat\sigma^2_{T,u}\}\big|.$$

By (4.A.11),

$$R_H \le 2H^{-2}|J_{\beta u,TH}| + H^{-2}\big|J_{uu,TH} - \hat\sigma^2_{T,u} - E\{J_{uu,TH} - \hat\sigma^2_{T,u}\}\big|$$

$$=: r_{1,H} + r_{2,H}.$$

It remains to show that

$$E \sup_{H\in I_T} r_{l,H} \to 0, \qquad l = 1, 2. \tag{4.A.21}$$

For $l = 2$, (4.A.21) follows from (4.A.13). It remains to show it for $l = 1$. By definition,

$$\begin{aligned} J_{\beta u,TH} &= T_n^{-1}\sum_{t=T_0}^{T}\left(\sum_{j=1}^{T_1} w_{j,H}\beta_{tj}\right)\left(\sum_{k=1}^{T_1} w_{k,H}u_{tk}\right) \\ &= T_n^{-1}\sum_{t=T_0}^{T}\left(\sum_{j=1}^{T_1} w_{j,H}aj\right)\left(\sum_{k=1}^{T_1} w_{k,H}u_{tk}\right) \\ &= \left(\sum_{j=1}^{T_1} w_{j,H}aj\right)\sum_{k=1}^{T_1} w_{k,H}\left(T_n^{-1}\sum_{t=T_0}^{T} u_{tk}\right). \end{aligned}$$

By (4.A.25), for $0 \le \gamma \le 2$, $w_{j,H}(j/H)^\gamma \le Cj^{-1}$, for $j \ge 1$. Hence

$$H^{-2}w_{j,H}jw_{k,H} = w_{j,H}(j/H)^{3/2}w_{k,H}(k/H)^{1/2}(jk)^{-1/2} \le C(jk)^{-3/2}.$$

Hence,

$$\begin{aligned} H^{-2}|J_{\beta u,TH}| &\le C\sum_{j=1}^{T_1} j^{-3/2}\sum_{k=1}^{T_1} k^{-3/2}\left|T_n^{-1}\sum_{t=T_0}^{T} u_{tk}\right|, \\ E\sup_{H\in I_T} r_{1,H} &\le C\sum_{j=1}^{T_1} j^{-3/2}\sum_{k=1}^{T_1} k^{-3/2}E\left|T_n^{-1}\sum_{t=T_0}^{T} u_{tk}\right| \\ &\le \sum_{k=1}^{T_1} k^{-3/2}j_{T,k}, \qquad j^2_{T,k} := E\left(T_n^{-1}\sum_{t=T_0}^{T} u_{tk}\right)^2 \end{aligned}$$

because $\sum_{j=1}^{T_1} j^{-3/2} \le \sum_{j=1}^{\infty} j^{-3/2} < \infty$ and

$$E\left|T_n^{-1}\sum_{t=T_0}^{T} u_{tk}\right| \le \left(E\left(T_n^{-1}\sum_{t=T_0}^{T} u_{tk}\right)^2\right)^{1/2} = j_{T,k}.$$

Note that $j_{T,k}^2 := E(T_n^{-1}\sum_{t=T_0}^{T} u_{tk})^2 \le CT^{-2}\sum_{t,s=T_0}^{T} Eu_{tk}u_{sk}$ where $|Eu_{tk}u_{sk}| \le 2|\gamma_u(t-s)| + |\gamma_u(t-s+k)| + |\gamma_u(t-s-k)|$. Hence, for $1 \le k \le T_1 \le T$,

$$\begin{aligned} j_{T,k}^2 &\le CT^{-2}\sum_{t,s=T_0}^{T}(|\gamma_u(t-s)| + |\gamma_u(t-s+k)| + |\gamma_u(t-s-k)|) \\ &\le CT^{-2}C\left(\sum_{t=-3T}^{3T}|\gamma_u(t)|\right)\left(\sum_{s=T_0}^{T}1\right) \le CT^{-1}\sum_{t=1}^{3T} t^{-1+2d} \\ &\le CT^{-1}\int_0^{3T} x^{-1+2d}dx \le CT^{-1+2d}. \end{aligned}$$

Thus,

$$\begin{aligned} E\sup_{H\in I_T} r_{1,H} &\le C\sum_{k=1}^{T_1} k^{-3/2} j_{T,k} \le CT^{-1/2+d}\sum_{k=1}^{T_1} k^{-3/2} \\ &\le CT^{-1/2+d} \to 0, \end{aligned}$$

because $0 < d < 1/2$ and $\sum_{k=1}^{\infty} k^{-3/2} < \infty$ which proves (4.A.21) and completes the proof of the part (ii) of the lemma. □

4.A.3 Auxiliary Results

Denote $v_{t,H} := \sum_{j=1}^{t-1} k_{j,H}$, $t \ge 1$ and $v_H := \sum_{j=1}^{\infty} k_{j,H}$, $t \ge 1$. Recall definitions $w_{tj,H} = k_{j,H}/v_{t,H}$ and $w_{j,H} = k_{j,H}/v_H$. Below $q_{u,H}$ is an in Theorem 1 and T_1 as in definition of $Q_{T,H}^{(apr)}$.

Lemma 4 *Under Assumption 1, uniformly in $H \in I_T$, $T \ge 1$, the following holds.*

(i) There exists $c > 0$, $C > 0$ *such that for* $0 \le \gamma \le 2$,

$$\nu_H \ge cH, \qquad w_{j,H} \le C(j \vee H)^{-1}, \quad j \ge 1; \tag{4.A.22}$$

$$w_{j,H} \le CT^{-6}, \quad j \ge T_1, \qquad \sum_{j=T_1}^{\infty} w_{j,H}\,(j/H)^{\gamma} = O(T^{-6}); \tag{4.A.23}$$

$$|w_{tj,H} - w_{j,H}| \le CT^{-6},\ T_0 \le t \le T,\ 1 \le j \le t-1; \tag{4.A.24}$$

$$w_{j,H}(j/H)^{\gamma} \le Cj^{-1},\ |w_{j,H} - w_{j+1,H}|\,H^{\gamma} \le Cj^{-2+\gamma},\ j \ge 1. \tag{4.A.25}$$

(ii) As $H \to \infty$,

$$H^{-1}\nu_H \to 1,\ H\sum_{j=1}^{\infty} w_{j,H}^2 \to \int_0^{\infty} K^2(x)dx,\ Hw_{0,H} \to K(0); \tag{4.A.26}$$

$$\sum_{j=1}^{\infty} w_{j,H}(j/H)^{\gamma} \to \int_0^{\infty} K(x)x^{\gamma}dx, \quad 0 \le \gamma \le 2. \tag{4.A.27}$$

Proof *of the first claim of (4.A.22).* With $\epsilon > 0$,

$$\begin{aligned} v_H &= \sum_{j=1}^{\infty} k_{j,H} = \sum_{j=1}^{\infty} K(j/H) \\ &\ge \sum_{j=1}^{[\epsilon H]} K(j/H), \qquad \text{since } \epsilon H \le \infty \\ &= K\left(\frac{1}{H}\right) + K\left(\frac{2}{H}\right) + ... + K\left(\frac{\epsilon H}{H}\right) \\ &\ge \delta\,[\epsilon H] \end{aligned}$$

where $\delta := \inf K(u), 0 \le u \le \epsilon$. Notice that $\delta > 0$ when $\epsilon > 0$ is sufficiently small, because $K(u) \to K(0) > 0$ for $u \to 0$ by Assumption 1. This implies

$$v_H \ge \delta\,[\epsilon H] \ge (\delta\epsilon/2)H = cH,$$

where $c = \delta\epsilon/2$.

Proof *of the second claim of (4.A.22).* Notice that for any positive integer r,

$$\begin{aligned}\exp(-c\,|x|) &= (\exp(c\,|x|))^{-1}\\ &= \left(\sum_{i=0}^{\infty}\frac{(c\,|x|)^i}{i!}\right)^{-1} \leq \left(\frac{c^r\,|x|^r}{r!}\right)^{-1}\\ &= \frac{r!}{c^r}\,|x|^{-r} = c^* x^{-r} \qquad (4.A.28)\end{aligned}$$

where $c^* = \frac{r!}{c^r} > 0$. Then for any $x \geq 0$ we can use (4.2.4) to bound

$$K(x) \leq C\exp(-c\,|x|) \leq Cx^{-r}, r \geq 0. \qquad (4.A.29)$$

Also for any $x \geq 0$,

$$\exp(-\,|x|) \leq 1. \qquad (4.A.30)$$

Then (4.A.29) and (4.A.30) together imply

$$K(x) \leq C\,(\exp -|x|) \leq C(\frac{1}{x} \wedge 1).$$

Since $j/H \geq 1$ is positive we can then write

$$K\,(j/H) \leq C\,((j/H)^{-1} \wedge 1).$$

This together with the first claim of (4.A.22) implies,

$$\begin{aligned}w_{j,H} = \frac{K(j/H)}{v_H} &\leq \frac{C((j/H)^{-1} \wedge 1)}{C^*H}\\ &= C'\left(j^{-1} \wedge H^{-1}\right) = C'(j \vee H)^{-1},\end{aligned}$$

where $C' = \frac{C}{C^*} > 0$.

Properties: For $a > 0, b > 0$ and $c > 0$, $(a \wedge b)\frac{1}{c} = \left(\frac{a}{c} \wedge \frac{b}{c}\right) = \left(\frac{c}{a} \vee \frac{c}{b}\right)^{-1}$.

Proof *of the first claim of (4.A.23).* Note that $T_1 = T_0 T^{-\delta/2}$ and $H_{\max} = T_0 T^{-\delta}$. Hence, $\frac{H_{\max}}{T_1} \leq T^{-\delta/2}$. The inequality (4.A.29) implies that we can bound $K(x) \leq C\,|x|^{-(m+4)}$ by choosing m such that $\frac{m\delta}{2} \geq 6$. We can then write

$$\begin{aligned} k_{j,H} &= K(j/H) \leq C(j/H)^{-(m+4)} \\ &= C(H/j)^{m+4} = C(H/j)^{m}(H/j)^{4} \\ &\leq C\left(H_{\max}/T_1\right)^{m}(H/j)^{4}, \qquad \text{since } H_{\max} \geq H \text{ and } T_1 \leq j \\ &\leq C\left(T^{-\delta/2}\right)^{m}(H/j)^{4} = CT^{-\delta m/2}(H/j)^{4} \\ &\leq CT^{-6}(H/j)^{4}, \end{aligned} \tag{4.A.31}$$

since $T^{\frac{-m\delta}{2}} \leq T^{-6}$.

Since $H^{-1}(H/j)^4 \leq 1$ we can use (4.A.31) and the first claim of (4.A.22) to write

$$\begin{aligned} w_{j,H} &= \frac{K_{j,H}}{v_H} \leq \frac{CT^{-6}(H/j)^4}{C'H} \\ &= C^{*}T^{-6}H^{-1}(H/j)^4 \leq C^{*}T^{-6}, \end{aligned} \tag{4.A.32}$$

where $C^* = \frac{C}{C'} > 0$.

Proof *second claim of (4.A.23).* Since $\frac{j}{H} \geq 1$ and $0 \leq \gamma \leq 2$, we can use (4.A.32) to bound

$$\begin{aligned} \sum_{j=T_1}^{\infty} w_{j,H}(j/H)^{\gamma} &\leq \sum_{j=T_1}^{\infty} w_{j,H}(j/H)^{2} \\ &\leq \sum_{j=T_1}^{\infty} CT^{-6}H^{-1}\left(H/j\right)^{4}\left(H/j\right)^{-2} \\ &= CT^{-6}H^{-1}\sum_{j=T_1}^{\infty}\left(H/j\right)^{2} = CT^{-6}H\sum_{j=T_1}^{\infty} j^{-2} \end{aligned}$$

For any $c \geq 1$ and $n \geq 1$, we can approximate sum by integral to write

$$\sum_{j=T_1}^{\infty} j^{-n} \sim \int_{T_1}^{\infty} x^{-n} dx \leq c \left[-\frac{x^{-n+1}}{-n+1} \right]_{T_1}^{\infty} \leq \frac{c}{T_1^{n-1}} \tag{4.A.33}$$

Hence, $\sum_{j=T_1}^{\infty} w_{j,H}(j/H)^{\gamma} \leq CT^{-6}HT_1^{-1} \leq CT^{-6} = O\left(T^{-6}\right)$, since $\frac{H}{T_1} \leq 1$.

Proof *of (4.A.24).* To show (4.A.24) we first verify that

$$|v_H - v_{t,H}| \leq CT^{-6}H, t \geq T_0$$

for some $C > 0$. Note that $H_{\max} = T_0 T^{-\delta} \Rightarrow \frac{H_{\max}}{T_0} \leq T^{-\delta}$. Using (4.A.29) we can bound $K(x) \leq c\,|x|^{-(m+1)}$ for any $m > 0$. Then, $k_{j,H} = K\left(j/H\right) \leq c\left(H/j\right)^{m+1}$ and

$$\begin{aligned} v_H - v_{t,H} &= \sum_{j=1}^{\infty} k_{j,H} - \sum_{j=1}^{t-1} k_{j,H} \\ &= \sum_{j=t}^{\infty} k_{j,H} \leq \sum_{j=t}^{\infty} c\left(H/j\right)^{m+1} \\ &= cH^{m+1} \sum_{j=t}^{\infty} j^{-(m+1)} \end{aligned}$$

Using (4.A.33) we can write $\sum_{j=t}^{\infty} j^{-(m+1)} \leq \frac{c^*}{t^m}$. Choosing m such that $\delta m \geq 6$ we can then write

$$\begin{aligned} v_H - v_{t,H} &\leq cH^{m+1}\frac{1}{t^m} = c\left(H/t\right)^m H \\ &\leq c\left(\frac{H_{\max}}{T_0}\right)^m H, \qquad \text{since } H_{\max} \geq H \text{ and } T_0 \leq t \\ &\leq cT^{-\delta m}H, \qquad \text{since } \frac{H_{\max}}{T_0} \leq T^{-\delta} \\ &\leq cT^{-6}H, \end{aligned} \tag{4.A.34}$$

since $T^{-\delta m} \leq T^{-6}$. Finally,

$$\begin{aligned}
|w_{tj,H} - w_{j,H}| &= k_{j,H} \left| v_{t,H}^{-1} - v_H^{-1} \right| \\
&= \frac{k_{j,H}}{v_{t,H} v_H} |v_H - v_{t,H}| = w_{tj,H} |v_H - v_{t,H}| v_H^{-1} \\
&\leq |v_H - v_{t,H}| v_H^{-1}, \text{ since } w_{tj,H} \leq 1 \\
&\leq \frac{CT^{-6} H}{C' H} = C^* T^{-6}
\end{aligned}$$

using (4.A.34) and first claim of (4.A.22).

Proof *of first claim of (4.A.25).* Using (4.2.4) and (4.A.29) we can bound $K(x) \leq C |x|^{-(m+4)}$ for any $m > 0$. This implies for any $x \geq 0$ and $0 \leq \gamma \leq 2$,

$$K(x) \leq C x^{-(\gamma+1)} \Rightarrow K(x) x^{\gamma} \leq \ C x^{-1} \tag{4.A.35}$$

Using (4.A.35) and first part of (4.A.22),we can then write

$$\begin{aligned}
w_{j,H} (j/H)^{\gamma} &= \frac{K(j/H)}{v_H} (j/H)^{\gamma} \\
&\leq \frac{C (j/H)^{-1}}{c' H} = C^* j^{-1},
\end{aligned}$$

where $C^* = \frac{C}{C'} > 0$.

Proof *of second claim of (4.A.25).* Using first claim of (4.A.22) we can write

$$\begin{aligned}
H^{\gamma} |w_{j+1,H} - w_{j,H}| &= H^{\gamma} v_H^{-1} |K((j+1)/H) - K(j/H)| \\
&\leq H^{\gamma} \frac{1}{cH} \frac{1}{H} \frac{|K((j+1)/H) - K(j/H)|}{\frac{1}{H}} \\
&\leq c^* H^{-2+\gamma} \left| \dot{K}(\xi) \right|
\end{aligned}$$

for some $\xi \in [jH^{-1}, (j+1)H^{-1}]$ and using the central limit theorem. By (4.2.4), $\left|\dot{K}(x)\right| \le \frac{c}{1+x^2} \le \frac{c}{x^2}$. This implies

$$\begin{aligned}
\left|\dot{K}(\xi)\right| &\le c\xi^{-2} \\
&\le c\xi^{-2+\gamma}, \qquad \text{since } \ 0 \le \gamma \le 2 \\
&\le c\,(j/H)^{-2+\gamma}, \quad \text{since } \ \xi \ge \frac{j}{H}, \xi^{-2+\gamma} \le \left(\frac{j}{H}\right)^{-2+\gamma} \\
&= c\,(H/j)^{2-\gamma}
\end{aligned}$$

Hence,

$$H^{\gamma}\left|w_{j+1,H} - w_{j,H}\right| \le c^{*}H^{-2+\gamma}c\,(H/j)^{2-\gamma} = c^{**}j^{-2+\gamma}$$

where $c^{**} = c^{*}c > 0$.

Proof *of first claim of (4.A.26).* Approximating sum by integral we can write

$$\begin{aligned}
\frac{1}{H}v_H &= \frac{1}{H}\sum_{j=1}^{\infty} K\left(\frac{j}{H}\right) \sim \frac{1}{H}\int_{1}^{\infty} K\left(\frac{u}{H}\right)du = \frac{1}{H}\int_{\frac{1}{H}}^{\infty} K(x)\,H dx \\
&= \int_{\frac{1}{H}}^{\infty} K(x)\,dx.
\end{aligned}$$

As $H \to \infty$, $\frac{1}{H} \to 0$ and hence, $\frac{1}{H}v_H \to \int_0^{\infty} K(x)dx = 1$.

Proof *of second claim of (4.A.26).* We can write

$$\begin{aligned}
H\sum_{j=1}^{\infty} w_{j,H}^2 &= H\sum_{j=1}^{\infty}\left(\frac{K\left(\frac{j}{H}\right)}{v_H}\right)^2 \\
&\sim H\int_0^{\infty}\left(\frac{K\left(\frac{u}{H}\right)}{v_H}\right)^2 du \\
&= \frac{1}{H}\int_{\frac{1}{H}}^{\infty}\left(\frac{K(x)}{\frac{1}{H}v_H}\right)^2 H dx
\end{aligned}$$

Using first claim of (4.A.26) and the fact that $\frac{1}{H} \to 0$ as $H \to \infty$, we get

$$H \sum_{j=1}^{\infty} w_{j,H}^2 \to \int_0^{\infty} K^2(x)dx, \qquad \text{as } H \to \infty.$$

Proof *of third claim of (4.A.26).* Using first claim of (4.A.26) we can write

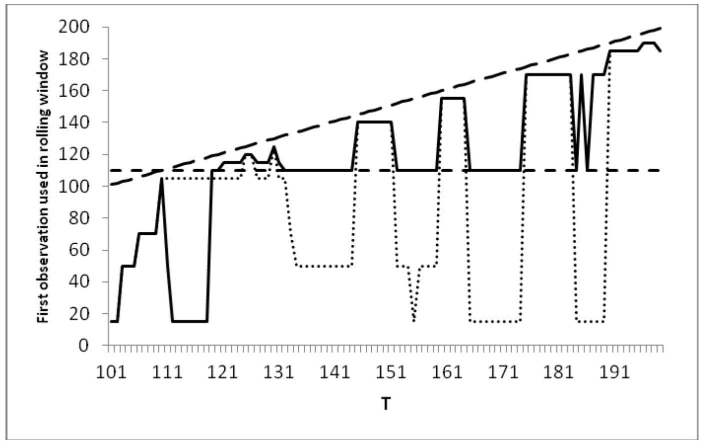
$$Hw_{0,H} = \frac{K(0)}{\frac{1}{H}v_H} \to K(0).$$

Proof *of (4.A.27).* Approximating sum by integral we can write

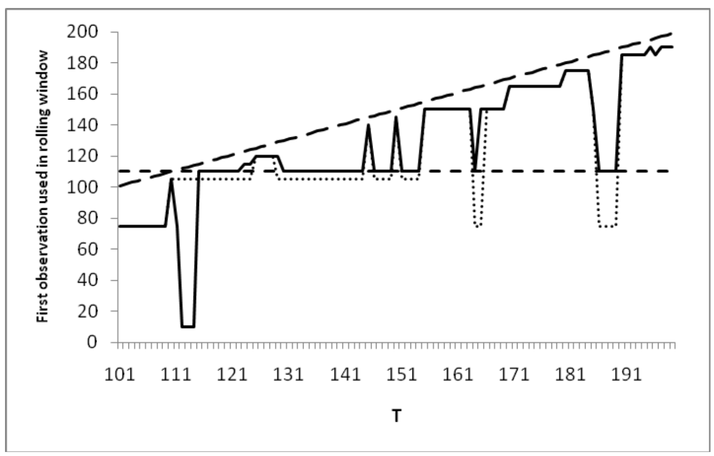
$$\begin{aligned}
\sum_{j=1}^{\infty} w_{j,H} \left(\frac{j}{H}\right)^{\gamma} &\sim \int_1^{\infty} \frac{K\left(\frac{u}{H}\right)}{v_H} \left(\frac{u}{H}\right)^{\gamma} du \\
&= \frac{1}{H} \int_0^{\infty} \frac{K\left(\frac{u}{H}\right)}{\frac{1}{H}v_H} \left(\frac{u}{H}\right)^{\gamma} du \\
&= \frac{1}{H} \int_{\frac{1}{H}}^{\infty} \frac{K(x)}{\frac{1}{H}v_H} x^{\gamma} H dx \\
&\sim \int_0^{\infty} K(x) x^{\gamma} dx, \qquad 0 \le \gamma \le 2
\end{aligned}$$

using the first part of (4.A.26) and the fact that $\frac{1}{H} \to 0$ as $H \to \infty$. □

4.B Appendix F: Figures



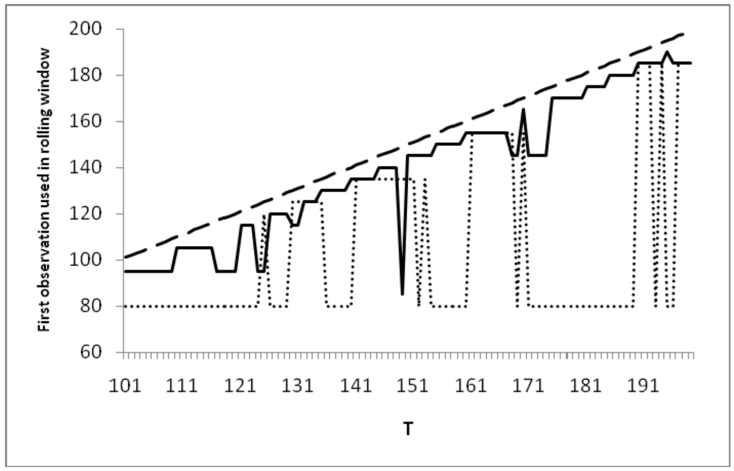
(a) u_t ~ i.i.d.



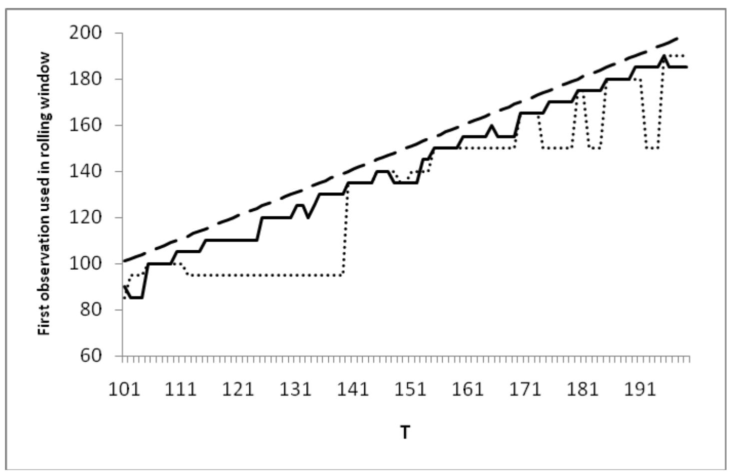
(b) u_t ~ ARFIMA(0,0.45,0)

Figure 4.B.1: Realisation of the data-selected rolling window for a structural break.

Note: The solid line represents the starting point of the window for a structural break model with a break at observation 110 (Experiment 4 of Monte Carlo Study), and the dashed line (long dashes) shows the last observation in the window. The dashed line (short dashes) indicates the first post break observation, and the dotted line the beginning of the window when there is no structural change.



(a) u_t ~ i.i.d.



(b) u_t ~ ARFIMA(0,0.45,0)

Figure 4.B.2: Realisation of the data-selected rolling window for a random walk.

Note: The solid line represents the starting point of the window for a random walk model with a break at observation 110 (Experiment 11 of Monte Carlo Study), and the dashed line (long dashes) shows the last observation in the window. The dashed line (short dashes) indicates the first post break observation, and the dotted line the beginning of the window when there is no structural change.

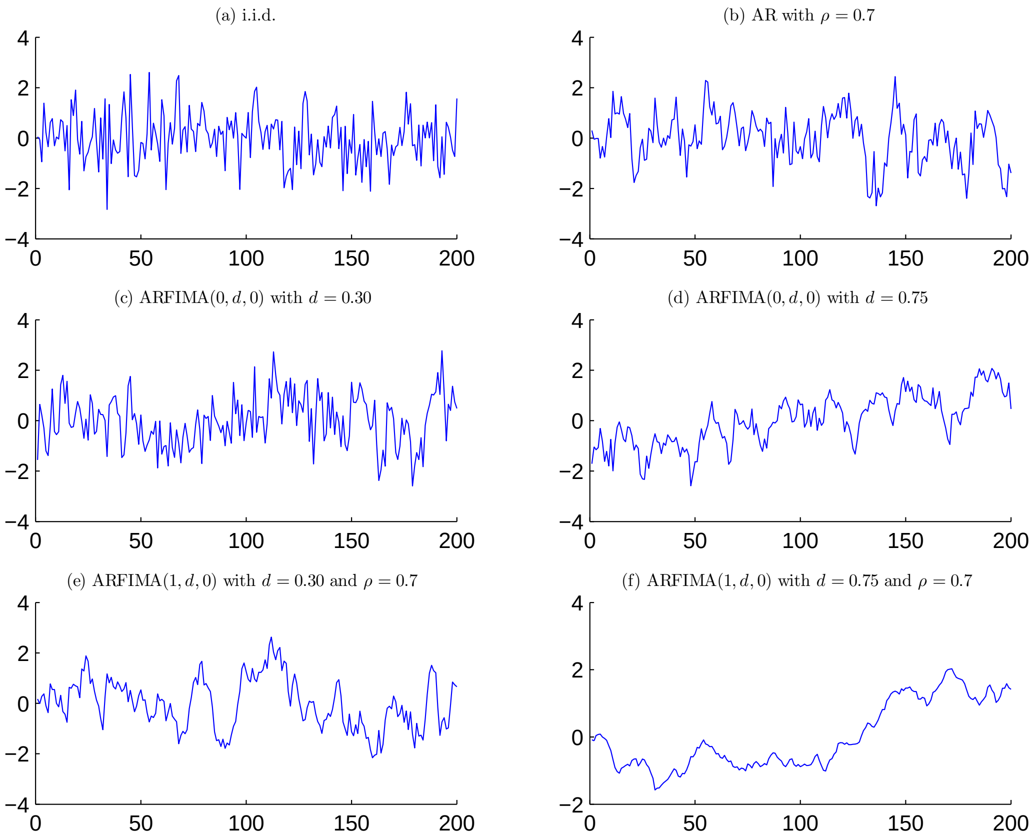


Figure 4.B.3: Plots of simulated y_t with different dynamics for noise process u_t (Experiment 1)

Note: Data are generated using the model: $y_t = u_t$. This is the case of no structural change. The panels specify alternative dynamics for noise u_t. Innovations for AR and ARFIMA noise processes are i.i.d. standard normal.

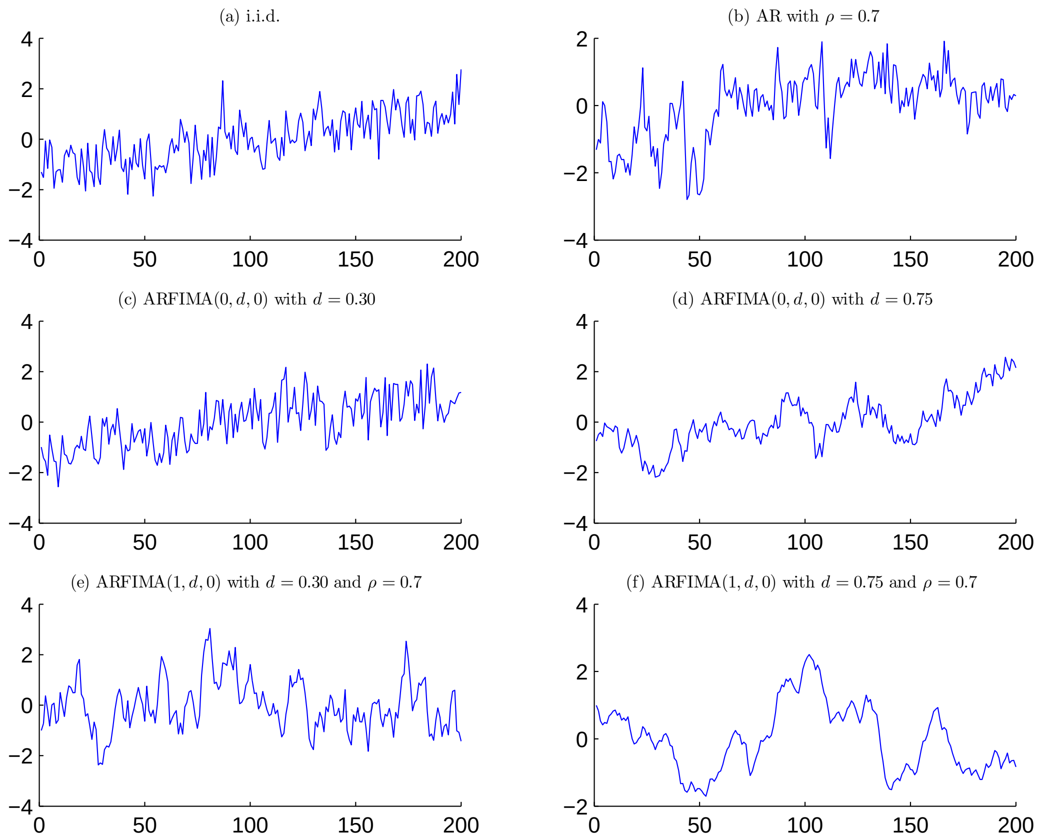


Figure 4.B.4: Plots of simulated y_t with different dynamics for noise process u_t (Experiment 3)

Note: Data are generated using the model: $y_t = 0.05t + 3u_t$. This introduces a monotonic linear trend. The panels specify alternative dynamics for noise u_t. Innovations for AR and ARFIMA noise processes are i.i.d. standard normal.

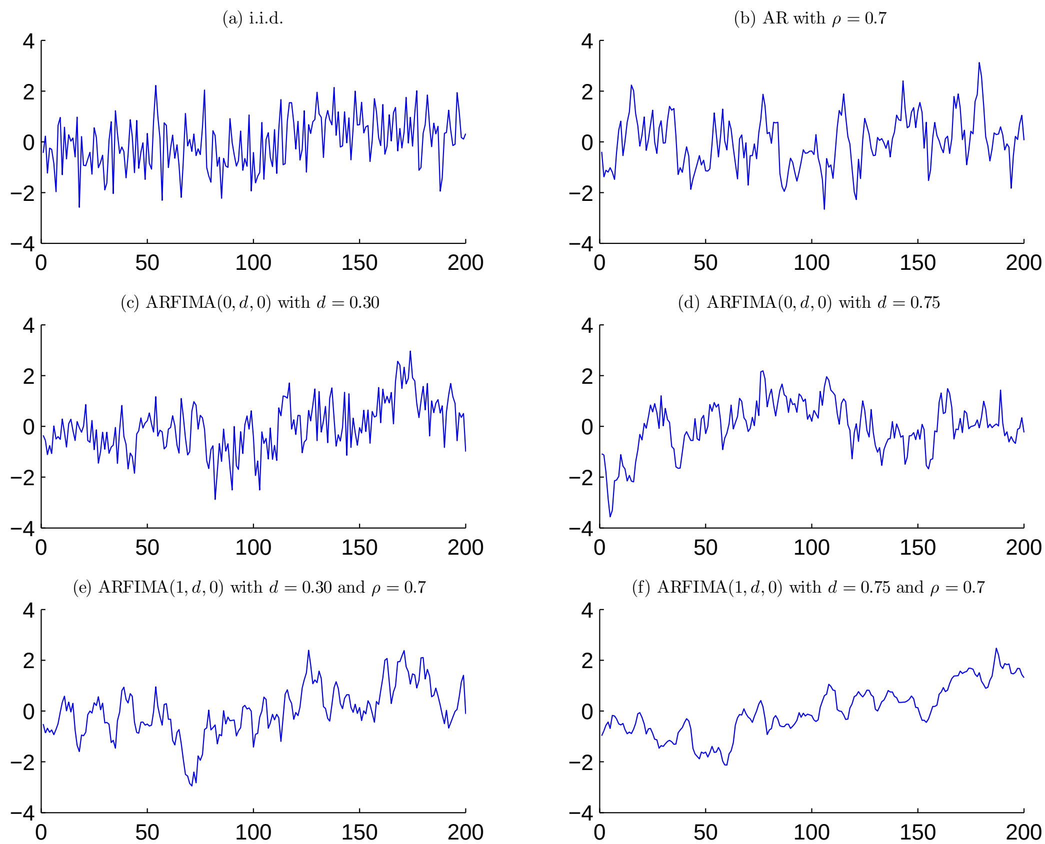


Figure 4.B.5: Plots of simulated y_t with different dynamics for noise process u_t (Experiment 4)

Note: Data are generated using the model: $y_t = u_t,\ t \leq t_0 = 0.55T$ and $y_t = 1 + u_t,\ t > t_0$. This introduces a break in the mean. The panels specify alternative dynamics for noise u_t. Innovations for AR and ARFIMA noise processes are i.i.d. standard normal.

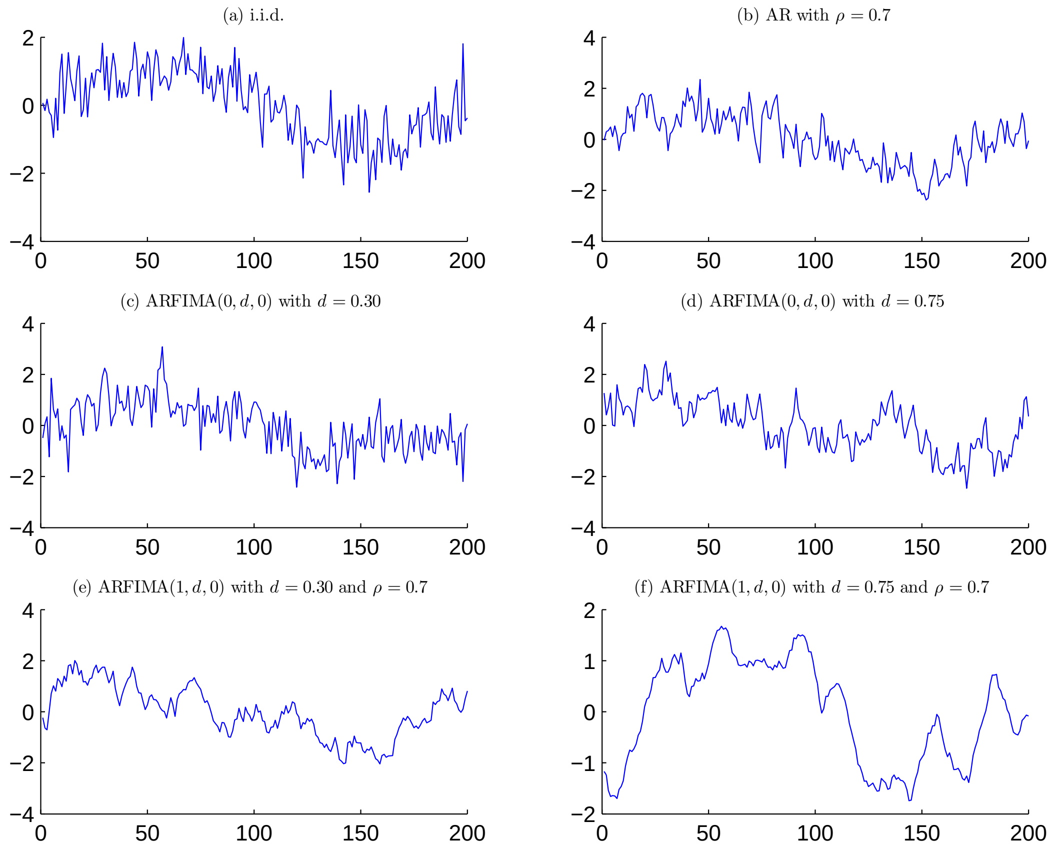


Figure 4.B.6: Plots of simulated y_t with different dynamics for noise process u_t (Experiment 6)

Note: Data are generated using the model: $y_t = 2\sin(2\pi t/T) + u_t$. This introduces a smooth cyclical trend. The panels specify alternative dynamics for noise u_t. Innovations for AR and ARFIMA noise processes are i.i.d. standard normal.

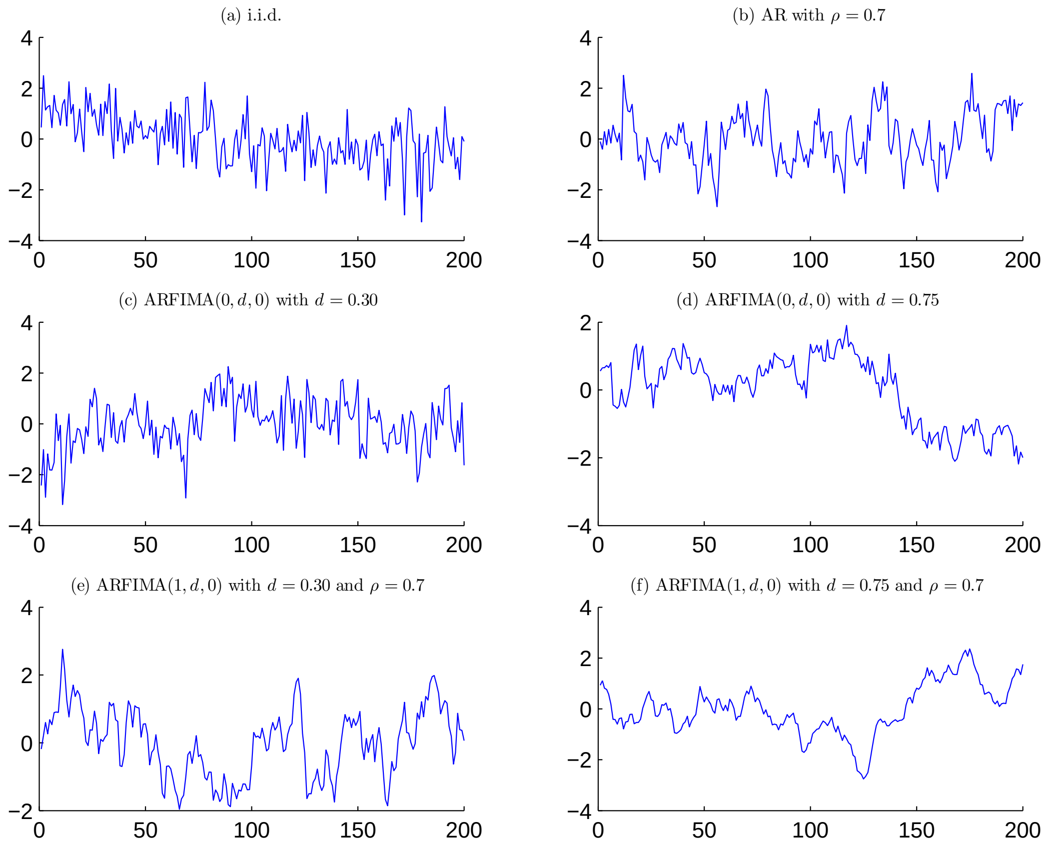


Figure 4.B.7: Plots of simulated y_t with different dynamics for noise process u_t (Experiment 8)

Note: Data are generated using the model: $y_t = 2T^{-1/2}\sum_{i=1}^{t} v_i + u_t$. This introduces a bounded stochastic trend. The panels specify alternative dynamics for noise u_t. Innovations for AR and ARFIMA noise processes are i.i.d. standard normal.

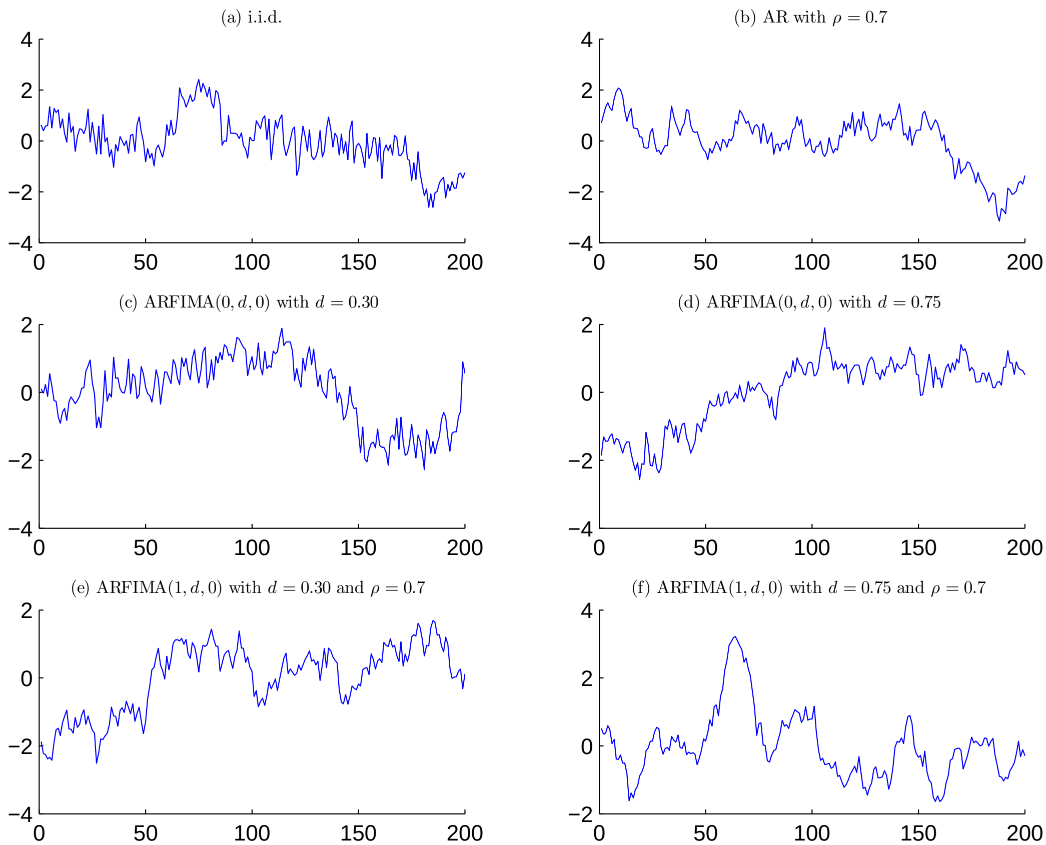


Figure 4.B.8: Plots of simulated y_t with different dynamics for noise process u_t (Experiment 10)

Note: Data are generated using the model: $y_t = 0.5\sum_{i=1}^{t} v_i + u_t$. This introduces an unbounded stochastic trend process, such as random walk with noise. The panels specify alternative dynamics for noise u_t. Innovations for AR and ARFIMA noise processes are i.i.d. standard normal.

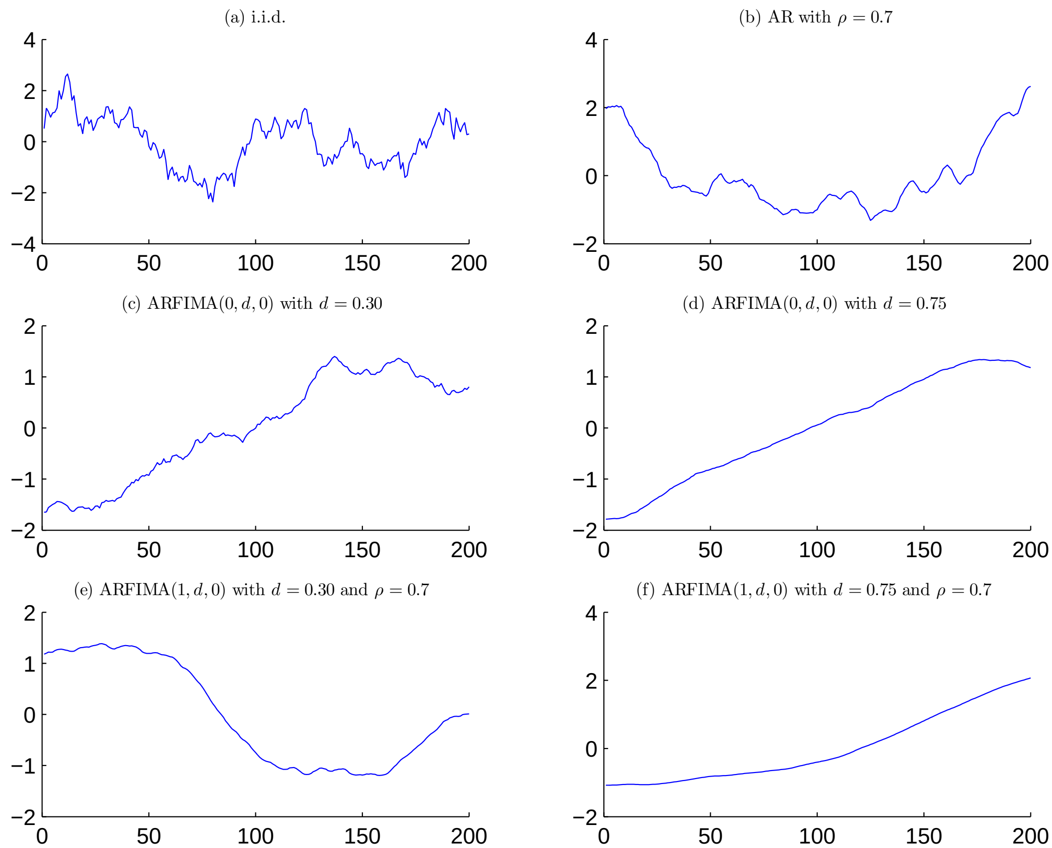


Figure 4.B.9: Plots of simulated y_t with different dynamics for noise process u_t (Experiment 11)

Note: Data are generated using the model: $y_t = \sum_{i=1}^{t} u_i$. This is a standard driftless random walk process. The panels specify alternative dynamics for noise u_t. Innovations for AR and ARFIMA noise processes are i.i.d. standard normal.

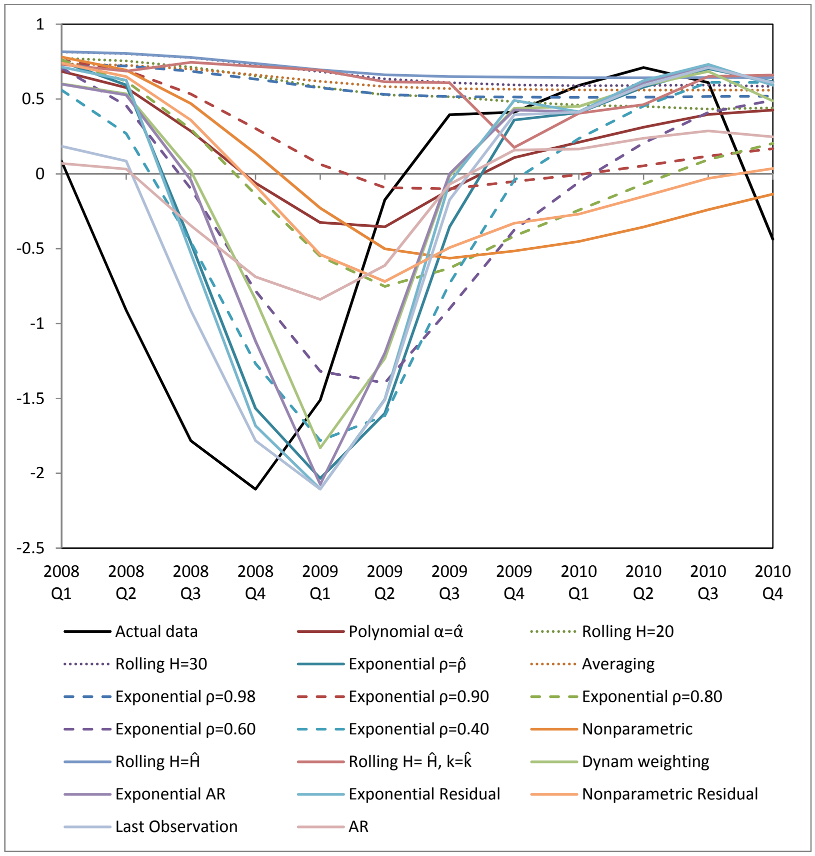


Figure 4.B.10: Actual and forecasted GDP growth, 2008 Q1 - 2010 Q4.

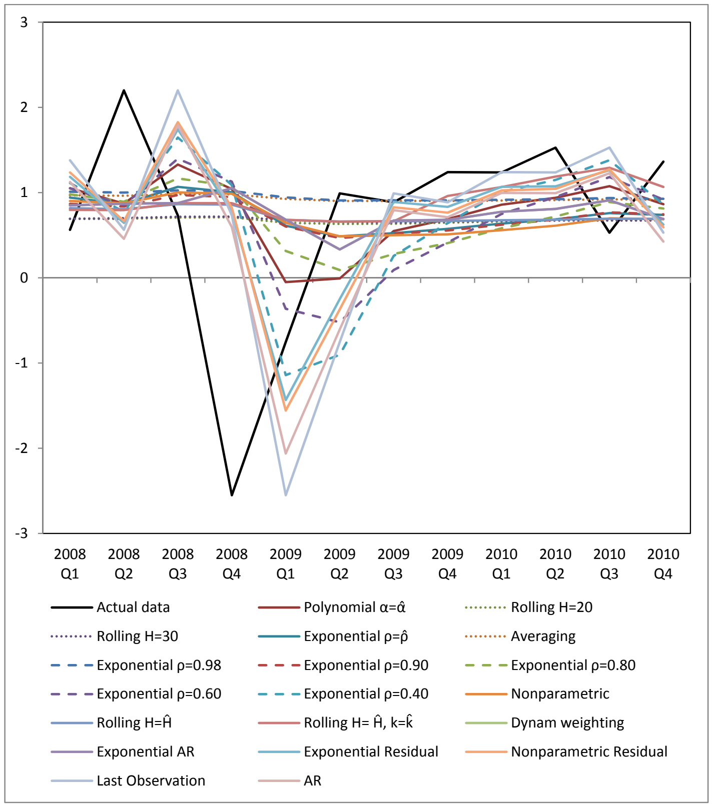


Figure 4.B.11: Actual and forecasted inflation, 2008 Q1 - 2010 Q4.

4.C Appendix G: Tables

Table 4.C.1: Monte Carlo Results. $T = 200$. $u_t \sim$ ARFIMA$(0, 0.3, 0)$. Relative MSFE's of one-step-ahead forecasts with respect to the full sample mean benchmark.

		Experiments										
Method		$Ex1$	$Ex2$	$Ex3$	$Ex4$	$Ex5$	$Ex6$	$Ex7$	$Ex8$	$Ex9$	$Ex10$	$Ex11$
Exponential	$\rho = \hat{\rho}$	0.905	0.664	0.419	0.680	0.673	0.211	0.847	0.621	0.568	0.194	0.006
Rolling	$H = \hat{H}$	0.959	0.716	0.461	0.738	0.730	0.252	0.901	0.693	0.644	0.284	0.055
Rolling	$H = 20$	0.972	0.725	0.481	0.790	0.774	0.377	0.921	0.763	0.733	0.513	0.304
	$H = 30$	0.977	0.756	0.531	0.831	0.836	0.534	0.941	0.819	0.795	0.627	0.453
Exponential	$\rho = 0.98$	0.950	0.773	0.593	0.820	0.819	0.553	0.917	0.797	0.773	0.609	0.471
	$\rho = 0.90$	0.890	0.654	0.419	0.682	0.671	0.240	0.837	0.636	0.591	0.295	0.099
	$\rho = 0.80$	0.874	0.639	0.405	0.656	0.649	0.208	0.820	0.603	0.551	0.223	0.038
	$\rho = 0.60$	0.902	0.656	0.416	0.670	0.669	0.207	0.845	0.612	0.552	0.193	0.016
	$\rho = 0.40$	0.977	0.707	0.448	0.720	0.724	0.222	0.914	0.657	0.589	0.192	0.010
	$\rho = 0.002$	1.253	0.903	0.573	0.913	0.927	0.282	1.166	0.831	0.745	0.228	0.006
Averaging		0.965	0.790	0.615	0.845	0.839	0.583	0.933	0.822	0.800	0.655	0.527
Nonparametric		0.951	0.681	0.424	0.707	0.693	0.220	0.884	0.654	0.602	0.272	0.062
Polynomial	$\alpha = \hat{\alpha}$	0.865	0.646	0.461	0.666	0.660	0.279	0.816	0.628	0.586	0.297	0.024
Rolling	$H = \hat{H}$, $k = \hat{k}$	0.908	0.681	0.447	0.701	0.698	0.238	0.852	0.653	0.601	0.243	0.038
Dynamic Weighting		0.890	0.650	0.415	0.667	0.663	0.212	0.833	0.615	0.560	0.196	0.007
Exponential AR		0.893	0.655	0.419	0.676	0.670	0.214	0.837	0.619	0.567	0.197	0.005
Exponential Residual		0.877	0.657	0.433	0.673	0.675	0.247	0.829	0.632	0.582	0.218	0.006
Nonparametric Residual		0.900	0.662	0.436	0.675	0.678	0.245	0.844	0.636	0.585	0.215	0.006
Last Observation		1.255	0.904	0.575	0.915	0.929	0.282	1.168	0.833	0.747	0.228	0.006
AR		0.849	0.665	0.466	0.668	0.676	0.252	0.809	0.632	0.581	0.208	0.006

Table 4.C.2: Monte Carlo Results. $T = 200$. $u_t \sim$ ARFIMA$(0, 0.45, 0)$. Relative MSFE's of one-step-ahead forecasts with respect to the full sample mean benchmark.

		Experiments										
Method		$Ex1$	$Ex2$	$Ex3$	$Ex4$	$Ex5$	$Ex6$	$Ex7$	$Ex8$	$Ex9$	$Ex10$	$Ex11$
Exponential	$\rho = \hat{\rho}$	0.688	0.576	0.402	0.563	0.568	0.173	0.639	0.452	0.415	0.147	0.003
Rolling	$H = \hat{H}$	0.811	0.688	0.483	0.677	0.669	0.230	0.756	0.568	0.523	0.270	0.032
Rolling	$H = 20$	0.909	0.776	0.566	0.783	0.779	0.376	0.865	0.727	0.690	0.527	0.243
	$H = 30$	0.944	0.812	0.618	0.834	0.850	0.536	0.902	0.801	0.765	0.636	0.397
Exponential	$\rho = 0.98$	0.897	0.795	0.644	0.807	0.809	0.548	0.861	0.766	0.735	0.606	0.440
	$\rho = 0.90$	0.745	0.632	0.445	0.621	0.618	0.222	0.702	0.535	0.498	0.279	0.073
	$\rho = 0.80$	0.683	0.579	0.403	0.563	0.565	0.180	0.641	0.468	0.431	0.193	0.025
	$\rho = 0.60$	0.661	0.557	0.389	0.544	0.548	0.169	0.619	0.437	0.403	0.153	0.009
	$\rho = 0.40$	0.685	0.576	0.403	0.566	0.570	0.173	0.642	0.445	0.411	0.144	0.005
	$\rho = 0.002$	0.823	0.695	0.490	0.693	0.691	0.208	0.780	0.530	0.491	0.159	0.003
Averaging		0.926	0.822	0.670	0.840	0.838	0.581	0.889	0.804	0.774	0.655	0.497
Nonparametric		0.785	0.658	0.450	0.642	0.637	0.202	0.735	0.541	0.498	0.251	0.041
Polynomial	$\alpha = \hat{\alpha}$	0.680	0.584	0.450	0.570	0.576	0.255	0.648	0.490	0.466	0.259	0.017
Rolling	$H = \hat{H}$, $k = \hat{k}$	0.735	0.628	0.447	0.613	0.611	0.205	0.684	0.510	0.467	0.212	0.022
Dynamic Weighting		0.662	0.559	0.394	0.552	0.554	0.172	0.624	0.442	0.405	0.149	0.003
Exponential AR		0.672	0.565	0.398	0.560	0.560	0.174	0.631	0.450	0.411	0.150	0.002
Exponential Residual		0.664	0.564	0.409	0.564	0.565	0.197	0.632	0.453	0.425	0.162	0.002
Nonparametric Residual		0.676	0.574	0.413	0.569	0.570	0.193	0.642	0.455	0.429	0.159	0.002
Last Observation		0.825	0.696	0.491	0.694	0.693	0.209	0.782	0.531	0.492	0.159	0.003
AR		0.643	0.557	0.413	0.550	0.551	0.192	0.614	0.445	0.417	0.150	0.003

Table 4.C.3: Monte Carlo Results. $T = 200$. $u_t \sim$ ARFIMA$(0, 0.75, 0)$. Relative MSFE's of one-step-ahead forecasts with respect to the full sample mean benchmark.

		Experiments										
Method		*Ex*1	*Ex*2	*Ex*3	*Ex*4	*Ex*5	*Ex*6	*Ex*7	*Ex*8	*Ex*9	*Ex*10	*Ex*11
Exponential	$\rho = \hat{\rho}$	0.211	0.187	0.182	0.197	0.220	0.174	0.230	0.188	0.194	0.135	0.001
Rolling	$H = \hat{H}$	0.391	0.363	0.347	0.407	0.403	0.330	0.431	0.370	0.390	0.302	0.014
Rolling	$H = 20$	0.658	0.626	0.616	0.715	0.663	0.577	0.694	0.646	0.679	0.606	0.190
	$H = 30$	0.761	0.731	0.723	0.830	0.777	0.692	0.798	0.740	0.772	0.714	0.341
Exponential	$\rho = 0.98$	0.706	0.676	0.672	0.745	0.711	0.650	0.727	0.680	0.711	0.653	0.406
	$\rho = 0.90$	0.390	0.359	0.350	0.408	0.398	0.333	0.419	0.371	0.389	0.317	0.051
	$\rho = 0.80$	0.288	0.260	0.252	0.289	0.296	0.242	0.312	0.267	0.278	0.212	0.014
	$\rho = 0.60$	0.228	0.204	0.198	0.220	0.236	0.190	0.248	0.207	0.214	0.155	0.004
	$\rho = 0.40$	0.208	0.186	0.180	0.197	0.217	0.173	0.227	0.187	0.193	0.136	0.002
	$\rho = 0.002$	0.216	0.193	0.188	0.201	0.227	0.181	0.238	0.193	0.199	0.137	0.001
Averaging		0.755	0.724	0.720	0.812	0.760	0.694	0.781	0.725	0.767	0.701	0.460
Nonparametric		0.377	0.341	0.329	0.387	0.384	0.316	0.407	0.354	0.369	0.288	0.026
Polynomial	$\alpha = \hat{\alpha}$	0.325	0.292	0.297	0.299	0.327	0.284	0.338	0.289	0.302	0.237	0.014
Rolling	$H = \hat{H}$, $k = \hat{k}$	0.316	0.284	0.281	0.319	0.322	0.264	0.340	0.287	0.310	0.231	0.011
Dynamic Weighting		0.211	0.189	0.182	0.200	0.220	0.175	0.229	0.190	0.195	0.138	0.001
Exponential AR		0.211	0.188	0.184	0.199	0.220	0.175	0.230	0.189	0.195	0.136	0.000
Exponential Residual		0.215	0.196	0.189	0.203	0.226	0.183	0.235	0.196	0.199	0.143	0.000
Nonparametric Residual		0.215	0.195	0.189	0.203	0.227	0.182	0.237	0.195	0.200	0.142	0.000
Last Observation		0.217	0.193	0.188	0.201	0.227	0.181	0.238	0.193	0.199	0.137	0.001
AR		0.202	0.183	0.178	0.189	0.212	0.170	0.222	0.182	0.188	0.132	0.001

Table 4.C.4: Monte Carlo Results. $T = 200$. $u_t \sim$ ARFIMA$(1, 0.30, 0)$ with $\rho = 0.7$. Relative MSFE's of one-step-ahead forecasts with respect to the full sample mean benchmark.

		Experiments										
Method		$Ex1$	$Ex2$	$Ex3$	$Ex4$	$Ex5$	$Ex6$	$Ex7$	$Ex8$	$Ex9$	$Ex10$	$Ex11$
Exponential	$\rho = \hat{\rho}$	0.241	0.212	0.204	0.194	0.196	0.062	0.226	0.165	0.152	0.080	0.002
Rolling	$H = \hat{H}$	0.682	0.604	0.568	0.536	0.566	0.197	0.642	0.489	0.449	0.258	0.041
Rolling	$H = 20$	0.947	0.854	0.791	0.778	0.803	0.392	0.891	0.735	0.698	0.548	0.269
	$H = 30$	0.977	0.897	0.820	0.838	0.868	0.550	0.930	0.804	0.771	0.674	0.426
Exponential	$\rho = 0.98$	0.886	0.826	0.777	0.781	0.804	0.552	0.858	0.752	0.726	0.622	0.446
	$\rho = 0.90$	0.638	0.571	0.536	0.509	0.535	0.203	0.605	0.467	0.432	0.274	0.084
	$\rho = 0.80$	0.480	0.426	0.401	0.379	0.397	0.133	0.453	0.340	0.309	0.173	0.030
	$\rho = 0.60$	0.348	0.307	0.290	0.276	0.284	0.092	0.326	0.240	0.219	0.115	0.011
	$\rho = 0.40$	0.282	0.249	0.236	0.225	0.229	0.073	0.264	0.193	0.177	0.091	0.005
	$\rho = 0.002$	0.238	0.210	0.202	0.192	0.194	0.061	0.224	0.164	0.151	0.079	0.002
Averaging		0.930	0.865	0.817	0.825	0.846	0.584	0.903	0.801	0.771	0.674	0.493
Nonparametric		0.699	0.616	0.577	0.538	0.578	0.180	0.658	0.490	0.441	0.249	0.050
Polynomial	$\alpha = \hat{\alpha}$	0.311	0.269	0.270	0.251	0.255	0.093	0.289	0.235	0.226	0.137	0.023
Rolling	$H = \hat{H}$, $k = \hat{k}$	0.512	0.449	0.420	0.405	0.420	0.140	0.481	0.363	0.335	0.185	0.025
Dynamic Weighting		0.280	0.244	0.234	0.222	0.226	0.069	0.262	0.191	0.172	0.089	0.003
Exponential AR		0.244	0.217	0.206	0.199	0.199	0.064	0.230	0.170	0.155	0.082	0.001
Exponential Residual		0.241	0.216	0.206	0.198	0.198	0.068	0.228	0.170	0.157	0.083	0.001
Nonparametric Residual		0.253	0.223	0.215	0.204	0.206	0.066	0.238	0.175	0.161	0.085	0.002
Last Observation		0.238	0.210	0.202	0.192	0.194	0.061	0.224	0.164	0.151	0.079	0.002
AR		0.225	0.199	0.191	0.183	0.184	0.060	0.211	0.157	0.145	0.077	0.002

Table 4.C.5: Monte Carlo Results. $T = 200$. $u_t \sim$ ARFIMA$(1, 0.75, 0)$ with $\rho = 0.7$. Relative MSFE's of one-step-ahead forecasts with respect to the full sample mean benchmark.

		Experiments										
Method		$Ex1$	$Ex2$	$Ex3$	$Ex4$	$Ex5$	$Ex6$	$Ex7$	$Ex8$	$Ex9$	$Ex10$	$Ex11$
Exponential	$\rho = \hat{\rho}$	0.030	0.033	0.028	0.044	0.030	0.010	0.034	0.039	0.037	0.044	0.001
Rolling	$H = \hat{H}$	0.231	0.251	0.224	0.279	0.235	0.093	0.256	0.235	0.224	0.183	0.014
Rolling	$H = 20$	0.601	0.634	0.582	0.654	0.610	0.378	0.628	0.577	0.565	0.490	0.189
	$H = 30$	0.722	0.744	0.714	0.768	0.756	0.574	0.764	0.715	0.697	0.634	0.338
Exponential	$\rho = 0.98$	0.649	0.663	0.641	0.688	0.670	0.544	0.677	0.639	0.634	0.584	0.410
	$\rho = 0.90$	0.272	0.292	0.265	0.309	0.274	0.135	0.288	0.269	0.259	0.216	0.051
	$\rho = 0.80$	0.148	0.162	0.143	0.174	0.149	0.060	0.159	0.150	0.143	0.120	0.014
	$\rho = 0.60$	0.077	0.085	0.074	0.095	0.078	0.028	0.084	0.083	0.078	0.070	0.004
	$\rho = 0.40$	0.049	0.054	0.047	0.064	0.050	0.017	0.055	0.056	0.053	0.053	0.002
	$\rho = 0.002$	0.030	0.033	0.028	0.044	0.030	0.010	0.034	0.039	0.037	0.043	0.001
Averaging		0.705	0.724	0.695	0.755	0.735	0.591	0.738	0.693	0.688	0.640	0.467
Nonparametric		0.238	0.259	0.232	0.272	0.239	0.095	0.254	0.236	0.225	0.181	0.025
Polynomial	$\alpha = \hat{\alpha}$	0.059	0.070	0.058	0.058	0.065	0.015	0.064	0.070	0.068	0.100	0.019
Rolling	$H = \hat{H}$, $k = \hat{k}$	0.143	0.159	0.140	0.190	0.148	0.058	0.166	0.156	0.148	0.125	0.011
Dynamic Weighting		0.042	0.047	0.040	0.061	0.043	0.013	0.047	0.051	0.048	0.050	0.001
Exponential AR		0.024	0.027	0.023	0.047	0.025	0.008	0.029	0.037	0.036	0.045	0.000
Exponential Residual		0.029	0.032	0.027	0.052	0.030	0.009	0.034	0.041	0.040	0.047	0.000
Nonparametric Residual		0.034	0.037	0.032	0.050	0.034	0.011	0.038	0.043	0.041	0.047	0.000
Last Observation		0.030	0.033	0.028	0.044	0.030	0.010	0.034	0.039	0.037	0.043	0.001
AR		0.030	0.032	0.028	0.044	0.030	0.010	0.033	0.038	0.037	0.043	0.001

Table 4.C.6: Monte Carlo Results. $T = 200$. $u_t \sim$ ARFIMA$(1, 0.30, 0)$ with $\rho = -0.7$. Relative MSFE's of one-step-ahead forecasts with respect to the full sample mean benchmark.

		Experiments										
Method		$Ex1$	$Ex2$	$Ex3$	$Ex4$	$Ex5$	$Ex6$	$Ex7$	$Ex8$	$Ex9$	$Ex10$	$Ex11$
Exponential	$\rho = \hat{\rho}$	1.006	0.764	0.519	0.755	0.778	0.259	0.946	0.697	0.629	0.239	0.022
Rolling	$H = \hat{H}$	1.031	0.778	0.531	0.763	0.789	0.276	0.963	0.709	0.647	0.296	0.071
Rolling	$H = 20$	1.004	0.754	0.526	0.772	0.783	0.381	0.942	0.763	0.706	0.500	0.334
	$H = 30$	1.000	0.775	0.570	0.811	0.836	0.538	0.950	0.815	0.767	0.629	0.487
Exponential	$\rho = 0.98$	0.995	0.807	0.637	0.821	0.841	0.561	0.948	0.809	0.768	0.606	0.491
	$\rho = 0.90$	1.034	0.763	0.514	0.737	0.766	0.267	0.953	0.688	0.621	0.299	0.114
	$\rho = 0.80$	1.126	0.827	0.553	0.781	0.824	0.259	1.031	0.715	0.638	0.242	0.050
	$\rho = 0.60$	1.353	0.992	0.662	0.928	0.987	0.303	1.235	0.840	0.745	0.241	0.027
	$\rho = 0.40$	1.702	1.245	0.831	1.162	1.241	0.380	1.551	1.047	0.927	0.278	0.022
	$\rho = 0.002$	2.914	2.132	1.420	1.982	2.126	0.652	2.655	1.779	1.578	0.445	0.026
Averaging		0.995	0.814	0.652	0.837	0.850	0.588	0.952	0.831	0.794	0.652	0.546
Nonparametric		1.057	0.771	0.515	0.738	0.770	0.245	0.969	0.687	0.617	0.272	0.074
Polynomial	$\alpha = \hat{\alpha}$	1.001	0.842	0.617	0.828	0.851	0.343	0.967	0.757	0.696	0.331	0.115
Rolling	$H = \hat{H}$, $k = \hat{k}$	1.035	0.797	0.554	0.776	0.807	0.273	0.974	0.711	0.644	0.256	0.050
Dynamic Weighting		0.759	0.553	0.373	0.537	0.549	0.171	0.699	0.513	0.452	0.197	0.023
Exponential AR		0.737	0.557	0.375	0.545	0.551	0.173	0.692	0.520	0.459	0.192	0.019
Exponential Residual		0.736	0.594	0.474	0.606	0.613	0.380	0.712	0.651	0.628	0.355	0.027
Nonparametric Residual		0.755	0.583	0.463	0.598	0.599	0.370	0.717	0.645	0.621	0.357	0.027
Last Observation		2.926	2.141	1.426	1.990	2.134	0.654	2.666	1.787	1.584	0.447	0.026
AR		0.749	0.833	0.770	0.827	0.819	0.492	0.810	0.844	0.817	0.377	0.026

Table 4.C.7: Monte Carlo Results. $T = 200$. $u_t \sim \text{ARFIMA}(1, 0.75, 0)$ with $\rho = -0.7$. Relative MSFE's of one-step-ahead forecasts with respect to the full sample mean benchmark.

		Experiments										
Method		$Ex1$	$Ex2$	$Ex3$	$Ex4$	$Ex5$	$Ex6$	$Ex7$	$Ex8$	$Ex9$	$Ex10$	$Ex11$
Exponential	$\rho = \hat{\rho}$	0.521	0.495	0.388	0.460	0.472	0.156	0.521	0.371	0.357	0.146	0.001
Rolling	$H = \hat{H}$	0.584	0.558	0.439	0.530	0.527	0.186	0.584	0.435	0.428	0.236	0.016
Rolling	$H = 20$	0.730	0.705	0.575	0.720	0.708	0.381	0.739	0.610	0.628	0.505	0.197
	$H = 30$	0.800	0.775	0.645	0.804	0.807	0.562	0.817	0.707	0.728	0.637	0.342
Exponential	$\rho = 0.98$	0.764	0.747	0.651	0.758	0.758	0.556	0.778	0.678	0.695	0.596	0.405
	$\rho = 0.90$	0.554	0.529	0.422	0.513	0.510	0.202	0.556	0.421	0.419	0.257	0.054
	$\rho = 0.80$	0.512	0.487	0.383	0.455	0.464	0.158	0.512	0.370	0.359	0.177	0.016
	$\rho = 0.60$	0.544	0.517	0.403	0.467	0.492	0.160	0.542	0.379	0.362	0.146	0.005
	$\rho = 0.40$	0.640	0.610	0.473	0.542	0.581	0.186	0.639	0.439	0.417	0.150	0.003
	$\rho = 0.002$	1.025	0.981	0.758	0.858	0.936	0.298	1.029	0.694	0.657	0.210	0.001
Averaging		0.796	0.782	0.682	0.804	0.795	0.592	0.818	0.717	0.737	0.643	0.460
Nonparametric		0.549	0.522	0.409	0.497	0.496	0.171	0.550	0.405	0.399	0.225	0.029
Polynomial	$\alpha = \hat{\alpha}$	0.583	0.565	0.460	0.523	0.534	0.243	0.582	0.441	0.434	0.247	0.020
Rolling	$H = \hat{H}$, $k = \hat{k}$	0.535	0.505	0.399	0.480	0.489	0.168	0.536	0.387	0.377	0.186	0.012
Dynamic Weighting		0.439	0.415	0.326	0.393	0.393	0.129	0.432	0.319	0.308	0.144	0.001
Exponential AR		0.442	0.415	0.325	0.388	0.394	0.126	0.434	0.315	0.303	0.132	0.001
Exponential Residual		0.634	0.617	0.498	0.579	0.591	0.246	0.643	0.495	0.481	0.196	0.001
Nonparametric Residual		0.636	0.614	0.494	0.575	0.586	0.242	0.638	0.493	0.477	0.199	0.001
Last Observation		1.029	0.985	0.760	0.861	0.940	0.299	1.033	0.697	0.659	0.211	0.001
AR		0.715	0.687	0.557	0.618	0.663	0.260	0.707	0.526	0.516	0.193	0.001

Table 4.C.8: Summary statistics of empirical relative MSFEs for 55 quarterly UK series (untransformed) with the AR(1) forecast as benchmark.

		Full sample 1991Q1:2012Q4						First sub-sample 1991Q1:2001Q4						Second sub-sample 2002Q1:2012Q4					
Method		Mean	Median	Min	Max	DM1	DM2	Mean	Median	Min	Max	DM1	DM2	Mean	Median	Min	Max	DM1	DM2
Exponential	$\rho = \hat{\rho}$	0.939	0.972	0.445	1.175	0	16	0.931	0.963	0.363	1.140	1	23	0.939	0.974	0.497	1.250	1	16
Exponential	$\rho = 0.60$	2.286	2.073	0.674	4.469	44	1	2.539	2.487	0.692	4.675	43	2	2.100	1.883	0.643	4.472	38	1
	$\rho = 0.40$	1.434	1.388	0.550	2.292	42	3	1.498	1.524	0.468	2.374	40	2	1.373	1.353	0.666	2.211	32	5
	$\rho = 0.002$	0.982	0.978	0.427	2.450	2	15	0.979	0.980	0.341	2.757	3	21	0.975	0.977	0.498	2.153	3	18
Polynomial	$\alpha = \hat{\alpha}$	1.961	1.002	0.651	16.982	12	13	2.588	0.977	0.658	28.652	10	21	1.519	1.005	0.646	14.366	7	12
Dynamic Weighting		1.031	0.991	0.452	1.692	6	13	1.013	0.965	0.379	2.018	7	20	1.029	0.986	0.553	2.035	4	14
Exponential AR		0.827	0.929	0.320	1.282	0	20	0.746	0.836	0.122	1.241	2	23	0.872	0.941	0.423	1.307	1	13
Exponential Residual		0.891	0.957	0.315	1.374	3	16	0.785	0.863	0.118	1.630	3	21	0.942	0.981	0.399	1.379	6	13
Nonparametric Residual		0.902	0.992	0.319	1.251	4	14	0.840	0.935	0.136	1.248	1	13	0.930	1.013	0.351	1.253	4	10

Table 4.C.9: Relative MSFE for 15 key economic series (quarterly) for the UK with the AR(1) forecast as benchmark.

Series	Polynomial $\alpha = \hat{\alpha}$	Exponential $\rho = \hat{\rho}$	Exponential $\rho = 0.60$	Exponential $\rho = 0.40$	Exponential $\rho = 0.002$	Dynamic weighting	Exponential AR	Exponential residual	Nonparametric residual
Gross Domestic Product (chained volume measures)	0.879	0.878	3.829	1.957	0.881	0.884	**0.347**	0.379	0.561
Household final consumption expenditure: National concept	0.939	0.937	4.469	2.136	0.941	0.947	0.350	0.383	**0.347**
Real households disposable income	0.914	0.916	1.743	1.097	0.922	0.923	0.859	0.866	**0.843**
Total exports: Trade in Goods & Services	**0.977**	**0.977**	1.888	1.333	0.978	0.980	1.084	1.054	1.032
Total imports: Trade in Goods & Services	0.982	0.982	2.307	1.528	0.983	0.984	**0.951**	1.033	0.995
In employment: Aged 16+	0.992	0.992	3.626	1.963	0.994	0.993	**0.635**	0.682	0.892
Unemployment rate: Aged 16+	1.001	1.001	3.640	2.005	1.004	1.086	**0.665**	0.732	0.978
Total actual weekly hours worked	2.642	1.004	2.491	1.513	0.990	1.054	**0.958**	1.096	1.072
Exchange rate : US dollar per Pound Sterling	1.029	1.028	1.717	1.337	1.030	1.213	**1.003**	1.210	1.251
IOP: Production	0.973	0.972	2.589	1.613	0.974	1.156	**0.912**	1.036	1.009
IOP: Manufacturing	0.995	0.995	2.493	1.624	0.996	1.692	**0.988**	1.096	1.135
Consumer Price Index: Goods and services	0.758	0.757	3.077	1.535	0.759	0.766	0.428	0.421	**0.411**
M4	1.038	1.037	3.927	2.001	1.040	1.042	0.734	1.048	**0.692**
Public sector finances: Net Borrowing	0.822	0.684	0.674	0.760	1.145	0.614	**0.608**	0.946	0.916
LIBOR: 3 month rate	0.965	0.953	2.855	1.734	0.956	1.141	**0.849**	1.059	1.035

Table 4.C.10: 20 quarterly UK series (untransformed) with the smallest relative MSFE for optimised *Exponential* and *Exponential AR* forecast using the AR(1) forecast as benchmark.

Optimised *Exponential*		*Exponential AR*	
IOP: Mining and Quarrying	0.445	Claimant count rate	0.320
Exchange rate: Swiss franc per Pound Sterling	0.642	Total compensation of employees	0.324
GDP deflator	0.650	Gross Value Added at basic prices	0.337
Public sector finances: Net Borrowing	0.684	Gross Domestic Product (chained volume measures)	0.347
Consumer Price Index: Goods and services	0.757	Household final consumption expenditure: National concept	0.350
Exchange rate: Japanese yen per Pound Sterling	0.825	GDP deflator	0.422
IOP: Manufacture of coke and refined petroleum product	0.841	GDP at market prices: Current price	0.424
Output per Worker: Whole Economy	0.859	Consumer Price Index: Goods and services	0.428
Gross Domestic Product (chained volume measures)	0.878	IOP: Mining and Quarrying	0.461
Gross Value Added at basic prices	0.879	Output per Worker: Whole Economy	0.516
Total compensation of employees	0.901	Economically active	0.550
Notes and coins	0.903	Public sector finances: Net Borrowing	0.608
Balance of Payments: Trade in Goods & Services	0.903	Exchange rate: Swiss franc per Pound Sterling	0.629
Economically active	0.905	In employment: UK: Aged 16+	0.635
GDP at market prices	0.913	Unemployment rate: Aged 16+	0.665
IOP: Manufacture of textiles wearing apparel and leather products	0.914	General Government: Final consumption expenditure	0.701
Real households disposable income	0.916	UK Employee Jobs	0.724
Household final consumption expenditure: National concept	0.937	M4	0.734
BOP: IM: Fuels other than oil	0.939	RPI: Percentage change over 12 months	0.766
IOP: Manufacture of Food products beverages and tobacco	0.946	Exchange rate:Japanese yen per Pound Sterling	0.829

Table 4.C.11: Summary statistics of empirical relative MSFEs for 55 quarterly UK series (transformed to stationary) with the AR(1) forecast as benchmark.

		Full Sample 1991Q1:2012Q4						First Sub-sample 1991Q1:2001Q4						Second Sub-sample 2002Q1:2012Q4					
Method		Mean	Median	Min	Max	DM1	DM2	Mean	Median	Min	Max	DM1	DM2	Mean	Median	Min	Max	DM1	DM2
Exponential	$\rho = \hat{\rho}$	0.999	1.040	0.638	1.273	3	12	0.995	0.999	0.680	1.348	3	5	1.016	1.034	0.540	1.464	5	7
Rolling	$H = \hat{H}$	1.183	1.057	0.658	3.385	16	5	1.275	1.106	0.674	4.232	15	2	1.144	1.055	0.522	2.871	14	5
Rolling	$H = 20$	1.243	1.026	0.758	5.132	13	4	1.415	1.049	0.751	8.370	10	1	1.161	0.999	0.504	5.152	7	6
	$H = 30$	1.407	1.066	0.811	10.385	14	2	1.685	1.072	0.653	21.508	16	1	1.246	1.002	0.490	7.386	10	4
Exponential	$\rho = 0.98$	1.507	1.056	0.825	14.876	18	1	1.921	1.084	0.695	31.366	17	1	1.261	1.023	0.667	7.690	10	6
	$\rho = 0.90$	1.114	1.018	0.667	3.941	8	5	1.229	1.040	0.727	5.659	8	2	1.064	1.017	0.536	2.897	10	8
	$\rho = 0.80$	1.079	1.069	0.658	3.127	9	7	1.113	1.021	0.687	3.845	11	4	1.068	1.079	0.591	2.691	9	8
	$\rho = 0.60$	1.102	1.129	0.674	2.260	12	6	1.089	1.046	0.687	2.434	9	5	1.112	1.132	0.643	2.155	9	6
	$\rho = 0.40$	1.188	1.196	0.679	1.912	26	4	1.174	1.091	0.664	2.050	12	2	1.196	1.262	0.669	2.094	16	4
	$\rho = 0.002$	1.532	1.527	0.688	3.097	39	0	1.542	1.387	0.810	3.177	31	0	1.523	1.504	0.611	3.438	34	0
Averaging		1.448	1.056	0.837	11.615	16	2	1.752	1.065	0.696	23.649	16	1	1.268	1.026	0.624	8.340	10	5
Nonparametric		1.147	1.073	0.636	3.602	14	7	1.229	1.088	0.673	4.325	14	4	1.115	1.072	0.545	3.162	11	4
Polynomial	$\alpha = \hat{\alpha}$	1.028	1.025	0.757	2.404	8	10	1.055	1.017	0.697	2.848	7	4	1.024	1.027	0.663	2.204	9	8
Rolling	$H = \hat{H}$, $k = \hat{k}$	1.138	1.067	0.697	2.639	11	3	1.185	1.089	0.777	3.273	9	1	1.117	1.046	0.619	2.723	10	6
Dynamic Weighting		1.017	1.035	0.614	1.436	9	6	0.993	0.999	0.705	1.615	6	4	1.039	1.070	0.553	1.608	6	5
Exponential AR		0.998	1.026	0.608	1.302	4	7	0.981	1.012	0.697	1.278	4	6	1.018	1.041	0.545	1.395	4	6
Exponential Residual		1.033	1.039	0.754	1.290	4	3	1.025	1.019	0.785	1.337	5	4	1.041	1.025	0.702	1.312	3	1
Nonparametric Residual		1.058	1.083	0.743	1.211	16	2	1.067	1.081	0.767	1.247	10	1	1.057	1.088	0.709	1.209	9	2
Last Observation		1.535	1.529	0.689	3.108	39	0	1.546	1.391	0.812	3.187	31	0	1.526	1.505	0.611	3.451	34	0

Table 4.C.12: Summary statistics of empirical relative MSFEs for 79 monthly UK series (untransformed) with the AR(1) forecast as benchmark.

		Full sample Jan 2001:Dec 2012						First sub-sample Jan 2001:Dec 2006						Second sub-sample Jan 2007:Dec 2012					
Method		Mean	Median	Min	Max	DM1	DM2	Mean	Median	Min	Max	DM1	DM2	Mean	Median	Min	Max	DM1	DM2
Exponential	$\rho = \hat{\rho}$	0.954	0.975	0.652	1.103	2	22	0.947	0.991	0.584	1.082	4	19	0.962	0.977	0.623	1.225	3	20
Exponential	$\rho = 0.60$	1.815	1.681	0.659	4.415	50	6	1.685	1.384	0.637	5.132	47	9	1.896	1.687	0.624	4.325	53	6
	$\rho = 0.40$	1.255	1.271	0.721	2.084	47	11	1.207	1.164	0.656	2.320	38	13	1.284	1.251	0.706	2.192	42	10
	$\rho = 0.002$	1.010	0.994	0.767	1.846	5	13	1.013	1.003	0.633	1.894	8	8	1.008	0.993	0.862	1.829	8	15
Polynomial	$\alpha = \hat{\alpha}$	4.347	1.791	0.845	62.434	50	8	4.486	1.445	0.670	79.439	45	7	4.308	1.028	0.777	82.884	28	16
Dynamic Weighting		1.038	1.000	0.639	1.901	21	16	1.005	0.995	0.584	1.554	18	11	1.058	1.008	0.661	2.399	18	11
Exponential AR		0.925	0.947	0.427	1.288	6	21	0.921	0.981	0.239	1.398	7	17	0.920	0.940	0.570	1.162	2	14
Exponential Residual		1.035	1.054	0.397	1.421	23	10	1.032	1.057	0.141	1.525	18	12	1.029	1.031	0.540	1.490	11	7
Nonparametric Residual		0.999	1.049	0.355	1.179	17	12	0.977	1.061	0.145	1.194	21	14	1.002	1.041	0.525	1.208	12	6

Table 4.C.13: 20 monthly UK series with the smallest relative MSFE for optimised *Exponential* and *Exponential AR* forecast using the AR(1) forecast as benchmark.

Optimised *Exponential*		*Exponential AR*	
RPI: Household goods	0.652	M4 retail	0.427
Overseas visits to UK : Earnings	0.666	Claimant count rate	0.607
RSI: Textiles, clothing & footwear	0.682	RPI: All Services	0.622
CPI: Health	0.750	RPI: Household goods	0.637
RSI: Predominantly non-food stores	0.792	RSI: Textiles, clothing & footwear	0.704
UK visits abroad: Expenditure abroad	0.795	Nominal zero coupon bond yield (1-year maturity)	0.708
UK visits abroad: All visits	0.802	Overseas visits to UK : Earnings	0.708
RPI: Consumer durables	0.820	In employment: Aged 16+	0.711
IOP: Manufacture of transport equipment	0.828	Unemployed: Aged 16+	0.713
RSI: Household goods stores	0.841	RPI: Housing	0.722
Public debt: net borrowing	0.852	RPI: All items excluding mortgage interest	0.751
Employmentl: Aged 16+	0.890	RPI: Travel & leisure	0.789
RSI: Non-specialised stores	0.893	CPI: Health	0.792
RSI: Other non-food stores	0.895	RPI: All items retail prices index	0.805
IOP: Manuf Of Basic Pharmaceutical Prods	0.904	Consumer Price Index: Goods and services	0.806
RPI: Personal expenditure	0.909	RSI: Predominantly non-food stores	0.808
IOP: Manuf of rubberplastics prods & other non-metallic mineral products	0.914	RPI: Consumer durables	0.811
IOP: Manufacture of textiles wearing apparel and leather products	0.917	UK visits abroad: All visits	0.816
RSI: Predominantly food stores	0.931	Public debt net borrowing	0.834

Table 4.C.14: Relative MSFE results for UK quarterly series (untransformed) with full sample AR(1) as the benchmark forecast: Full sample (1991Q1: 2012Q4)

Series	Polynomial $\alpha = \hat{\alpha}$	Exp $\rho = \hat{\rho}$	Exp $\rho = 0.60$	Exp $\rho = 0.40$	Exp $\rho = 0.002$	Dynamic weighting	Exp AR	Exp residual	Nonparametric residual
Gross Domestic Product at market prices	0.914	0.913	4.072	1.988	0.916	0.922	0.424	0.483	0.439
GDP deflator	0.651	0.650	2.547	1.242	0.651	0.658	0.422	0.497	0.357
Gross Value Added at basic prices	0.880	0.879	3.826	1.961	0.883	0.885	0.337	0.370	0.568
Total adjustment to basic prices (General Government + Rest of the World)	0.987	0.987	2.540	1.562	0.988	0.991	0.932	0.893	0.922
Household final consumption expenditure: National concept	0.939	0.937	4.469	2.136	0.941	0.947	0.350	0.383	0.347
General Government: Final consumption expenditure	0.964	0.963	3.000	1.555	0.964	0.969	0.701	0.666	0.641
Total Gross Fixed Capital Formation	0.972	0.972	2.341	1.495	0.973	0.970	0.929	1.156	1.038
Changes in inventories including alignment adjustment	1.051	1.175	1.132	1.219	1.421	1.077	1.061	1.025	1.106
Net income from abroad (Current price)	1.226	1.079	1.074	1.016	1.069	0.990	1.064	1.048	1.069
Real households disposable income	0.914	0.916	1.743	1.097	0.922	0.923	0.859	0.866	0.843
Total compensation of employees (Uses)	0.902	0.901	4.364	2.035	0.904	0.910	0.324	0.337	0.319
Exports: Total Trade in Goods & Services (Current Price)	0.977	0.977	1.888	1.333	0.978	0.980	1.084	1.054	1.032
Imports: Total Trade in Goods & Services (Current Price)	0.982	0.982	2.307	1.528	0.983	0.984	0.951	1.033	0.995
Balance of Payments: Trade in Goods & Services: Total balance	2.318	0.903	0.928	0.864	0.997	0.883	0.939	1.026	1.065
BOP: EX: Food, beverages and tobacco	2.697	0.963	1.448	1.027	0.971	1.012	0.984	1.142	0.992
BOP: EX: Semi-manufactures	0.973	0.976	1.733	1.251	0.973	0.976	1.041	0.813	1.018
BOP: EX: Finished manufactures	2.667	1.022	1.294	1.099	0.966	1.013	1.282	1.304	1.115
BOP: IM: Fuels other than oil	1.002	0.939	2.078	1.490	0.935	1.151	0.857	1.061	1.105
BOP: IM: Food, beverages and tobacco	10.397	1.007	1.954	1.223	0.994	1.016	1.064	0.957	0.915
BOP: IM: Semi-manufactures	0.988	0.986	2.119	1.377	0.987	0.989	1.025	0.925	0.997
BOP: IM: Finished manufactures	3.679	0.966	1.385	1.136	0.966	0.998	1.229	1.253	1.094
BOP: IM: Oil	1.201	0.962	1.847	1.341	0.958	0.986	1.062	0.998	1.065
UK Employee Jobs	5.263	0.984	3.546	1.901	0.987	0.985	0.724	0.789	0.889
Output per Worker: Whole Economy	0.859	0.859	3.137	1.692	0.861	0.864	0.516	0.597	0.633
In employment, aged 16+	0.992	0.992	3.626	1.963	0.994	0.993	0.635	0.682	0.892
Economically active	0.906	0.905	3.286	1.728	0.908	0.909	0.550	0.578	0.570
Unemployment rate, aged 16+	1.001	1.001	3.640	2.005	1.004	1.086	0.665	0.732	0.978
Economic activity rate, aged 16+	1.560	1.026	2.073	1.376	1.024	1.533	0.964	0.976	1.086
Economic inactivity rate, aged 16+	1.560	1.026	2.073	1.376	1.024	1.533	0.964	0.976	1.086
Claimant count rate	0.996	0.996	4.325	2.292	1.001	1.270	0.320	0.315	0.956
Total actual weekly hours worked	2.642	1.004	2.491	1.513	0.990	1.054	0.958	1.096	1.072
Exchange rate : Swiss franc per Pound stirling	2.806	0.642	1.417	0.944	0.643	0.710	0.629	0.735	0.701
Exchange rate : Swedish kroner per Pound stirling	1.296	1.017	2.044	1.435	1.004	1.467	0.957	1.026	1.067
Exchange rate : Japanese yen per Pound stirling	0.826	0.825	1.735	1.171	0.827	1.143	0.829	0.811	0.939
Exchange rate : US dollar per Pound stirling	1.029	1.028	1.717	1.337	1.030	1.213	1.003	1.210	1.251
IOP: Production	0.973	0.972	2.589	1.613	0.974	1.156	0.912	1.036	1.009
IOP: Mining and quarrying	1.212	0.445	0.846	0.550	0.427	0.452	0.461	0.464	0.438
IOP: Manufacturing	0.995	0.995	2.493	1.624	0.996	1.692	0.988	1.096	1.135
IOP: Manufacture of Food products beverages and tobacco	5.154	0.946	1.310	1.070	0.981	1.075	0.990	1.077	1.073
IOP: Manufacture of textiles wearing apparel and leather products	0.914	0.914	2.529	1.519	0.915	1.124	0.836	0.815	0.871
IOP: Manufacture of coke and refined petroleum product	1.275	0.841	0.836	0.846	1.018	0.916	0.916	0.952	0.969
IOP: Manufacture of chemicals and chemical products	0.987	0.986	1.936	1.388	0.987	1.074	0.986	1.142	1.036
IOP: Manufacturing of computer electronic & optical products	1.034	0.985	2.229	1.465	0.986	0.991	1.018	1.067	1.112
IOP: Manufacture of electrical equipment	1.010	1.003	1.729	1.283	1.003	1.243	1.051	1.051	1.115
Consumer Price Index: Goods and services	0.758	0.757	3.077	1.535	0.759	0.766	0.428	0.421	0.411
Retail Price Index: Percentage change over 12 months	1.008	1.008	2.260	1.653	1.011	1.410	0.766	0.936	1.153
Notes and coins	16.982	0.903	1.183	0.914	0.995	0.949	0.883	0.937	0.982
M4	1.038	1.037	3.927	2.001	1.040	1.042	0.734	1.048	0.692
LIBOR : 3 month rate	0.965	0.953	2.855	1.734	0.956	1.141	0.849	1.059	1.035
Nominal Zero coupon bond yield (10-year maturity)	4.202	1.013	1.549	1.187	0.946	1.095	1.040	1.214	0.986
Nominal Zero coupon bond yield (2-year maturity)	2.609	0.964	1.846	1.319	0.948	1.078	1.024	1.374	1.023
Nominal Zero coupon bond yield (5-year maturity)	3.952	1.000	1.577	1.202	0.944	1.092	1.090	1.257	1.003
Private non-financial corporations one quarter growth rate	1.069	1.011	1.251	1.531	2.450	0.940	0.937	0.921	0.930
Public sector finances: Net Borrowing	0.822	0.684	0.674	0.760	1.145	0.614	0.608	0.946	0.916

Table 4.C.15: Relative MSFE results for UK monthly series (untransformed) with full sample AR(1) as the benchmark forecast: Full sample (Jan 2001: Dec 2012)

Series	Polynomial $\alpha = \hat{\alpha}$	Exp $\rho = \hat{\rho}$	Exp $\rho = 0.60$	Exp $\rho = 0.40$	Exp $\rho = 0.002$	Dynamic weighting	Exp AR	Exp residual	Nonparametric residual
Nominal zero coupon bond yield (1-year maturity)	0.992	0.992	2.966	1.801	0.995	1.857	0.708	0.799	1.114
Nominal zero coupon bond yield (3-year maturity)	1.059	0.988	2.218	1.497	0.990	1.229	0.912	1.012	1.135
Nominal zero coupon bond yield (5-year maturity)	0.983	0.983	1.955	1.374	0.984	1.254	0.964	1.112	1.114
Nominal zero coupon bond yield (10-year maturity)	1.021	0.988	1.454	1.133	0.981	1.198	1.081	1.009	1.089
Exchange rate: Canadian dollar per Pound Sterling	2.440	0.989	1.818	1.278	0.988	1.199	1.064	1.091	1.093
Exchange rate: Swiss franc per Pound Sterling	1.165	0.987	2.389	1.531	0.989	1.162	0.945	1.067	1.027
Exchange rate: Swedish kroner per Pound Sterling	8.936	1.007	1.556	1.167	0.994	1.051	1.068	1.042	1.090
Exchange rate: Norwegian kroner per Pound Sterling	3.946	1.009	1.681	1.252	0.999	1.063	1.103	1.036	1.094
Exchange rate: Danish kroner per Pound Sterling	1.311	0.997	2.143	1.448	1.002	1.062	1.011	1.156	1.077
Exchange rate: Japanese yen per Pound Sterling	1.175	0.986	2.533	1.607	0.987	1.056	0.864	1.062	1.003
Exchange rate: Hong Kong Dollar per Pound Sterling	1.972	1.007	2.356	1.522	1.004	1.154	1.010	1.142	1.084
Exchange rate: New Zealand Dollar per Pound Sterling	4.119	0.993	1.722	1.271	0.994	1.161	1.023	1.253	1.074
Exchange rate: South African Rand Dollar per Pound Sterling	1.158	0.998	1.883	1.327	0.999	1.085	1.288	1.399	1.058
Exchange rate: US dollar Dollar per Pound Sterling	1.619	1.008	2.290	1.495	1.005	1.167	1.018	1.141	1.084
Exchange rate: Euro per Pound Sterling	1.293	1.001	2.120	1.438	1.002	1.092	1.023	1.157	1.079
Effective exchange rate index	1.376	1.006	2.080	1.406	0.997	1.169	1.024	1.245	1.088
RSI: Predominantly food stores	24.114	0.931	1.160	0.905	0.962	0.972	0.936	0.904	0.958
RSI: Non-specialised stores	2.341	0.893	1.012	0.874	0.977	0.969	0.963	1.051	1.046
RSI: Predominantly non-food stores	8.870	0.792	0.840	0.803	0.987	0.800	0.808	1.060	1.069
RSI: Other non-food stores	4.323	0.895	0.958	0.877	0.998	0.908	0.919	1.119	1.081
RSI: Textiles, clothing & footwear	5.772	0.682	0.675	0.734	0.992	0.700	0.704	1.232	1.059
RSI: Household goods stores	0.966	0.841	0.850	0.826	1.007	0.902	0.907	1.069	1.093
UK visits abroad	0.845	0.802	0.794	0.800	1.015	0.819	0.816	1.273	1.082
OS visits to UK	1.273	0.666	0.687	0.771	1.033	0.705	0.708	1.031	1.019
UK visits abroad	2.024	0.795	0.787	0.805	1.004	0.819	0.857	1.421	1.072
FTSE adj closing price	2.629	1.035	1.735	1.250	1.015	1.071	1.112	1.022	1.052
FTSE return	1.039	1.087	1.182	1.353	1.812	1.088	1.092	1.064	1.049
Public debt net borrowing	0.918	0.852	1.069	1.295	1.846	0.834	0.834	0.920	0.903
CPI inflation: All items	1.791	1.018	2.100	1.440	1.006	1.540	1.011	1.102	1.140
CPI inflation: Food and non-alcoholic beverages	1.022	1.008	2.052	1.429	1.010	1.431	1.013	1.106	1.179
CPI inflation: Health	2.766	0.750	0.849	0.743	0.767	0.789	0.792	0.822	0.792
CPI inflation: Transport	1.329	1.021	2.174	1.464	1.011	1.901	1.006	1.012	1.160
CPI inflation: Communication	2.223	0.987	1.607	1.189	1.003	1.272	1.017	1.100	1.071
CPI inflation: Miscellanious goods and services	1.199	1.044	1.375	1.112	1.059	1.151	1.080	1.050	1.146
CPI inflation: Goods	4.005	1.021	1.961	1.407	1.004	1.213	0.967	1.136	1.122
CPI inflation: Services	1.375	1.058	1.514	1.157	1.058	1.194	1.090	1.049	1.176
M4	0.971	0.972	2.377	1.393	0.973	0.976	0.945	1.265	0.828
M4 retail	0.964	0.961	4.415	2.084	0.964	0.982	0.427	0.397	0.355
Notes & Coins	17.797	0.950	1.065	0.944	0.981	0.958	0.946	0.963	1.013
M1	0.972	0.970	2.106	1.332	0.970	0.975	0.890	1.143	0.902
IOP: Production	0.930	0.953	1.320	1.026	0.991	0.937	0.983	1.006	1.045
IOP: Mining and quarrying	27.550	1.018	1.048	0.976	0.990	0.963	0.984	1.063	1.034
IOP: Manufacturing	1.829	0.961	1.363	1.050	1.019	0.914	0.928	1.049	1.087
IOP: Manufacture of Food products beverages and tobacco	2.608	0.991	1.079	0.962	1.035	1.055	1.054	1.126	1.057
IOP: Manufacture of textiles wearing apparel and leather products	6.126	0.917	1.135	0.942	0.987	0.953	0.982	1.078	1.084
IOP: Manufacture of wood and paper products and printing	2.643	0.965	1.096	0.952	0.980	1.007	0.987	1.145	0.983
IOP: Manufacture of coke and refined petroleum product	1.843	0.934	0.963	0.923	1.018	0.942	0.974	1.263	1.010
IOP: Manufacture of chemicals and chemical products	3.920	1.045	1.507	1.144	1.007	1.286	1.070	1.232	1.081
IOP: Manuf Of Basic Pharmaceutical Prods & Pharmaceutical Preparations	4.258	0.904	0.909	0.874	1.006	0.903	0.914	1.276	1.090

Table 4.C.16: Relative MSFE results for UK monthly series (untransformed) with full sample AR(1) as the benchmark forecast: Full sample (Jan 2001: Dec 2012) (Cont.)

Series	Polynomial $\alpha = \hat{\alpha}$	Exp $\rho = \hat{\rho}$	Exp $\rho = 0.60$	Exp $\rho = 0.40$	Exp $\rho = 0.002$	Dynamic weighting	Exp AR	Exp residual	Nonparametric residual
IOP: Manufacture of basic metals and metal products	1.962	1.023	1.372	1.055	0.996	1.000	1.020	1.144	1.032
IOP: Manufacturing of computer electronic & optical products	4.132	0.954	1.116	0.962	1.012	0.962	0.973	1.278	1.094
IOP: Manufacture of electrical equipment	2.938	1.006	1.225	1.019	1.020	1.036	1.035	1.034	1.090
IOP: Manufacture of machinery and equipment n.e.c.	1.665	0.964	1.481	1.098	0.976	0.926	0.946	1.020	1.010
IOP: Manufacture of transport equipment	2.423	0.828	0.999	0.876	0.990	0.838	0.844	1.243	1.023
Long term indicator of consumer goods and services	0.935	0.934	2.877	1.614	0.936	0.943	0.806	0.800	0.734
In employment: Aged 16+	0.891	0.890	2.800	1.601	0.892	0.912	0.711	0.795	0.809
Unemployed: Aged 16+	62.434	0.964	3.265	1.768	0.966	0.973	0.713	0.736	0.852
Economically active	1.003	1.001	2.754	1.605	1.003	1.011	0.867	0.944	0.833
Unemployment rate: Aged 16+	35.214	0.972	2.960	1.652	0.971	0.983	0.856	0.816	0.895
Economic activity rate: Aged 16+	4.954	1.019	1.297	1.082	1.001	1.101	1.078	1.208	1.086
Claimant count rate	2.271	0.956	3.552	1.844	0.956	0.967	0.607	0.608	0.878
Total actual weekly hours worked	0.986	1.004	1.489	1.148	0.980	1.017	1.001	1.066	1.049
Financial ins lending to public sector	1.412	1.103	3.131	1.708	1.085	1.081	0.883	1.006	0.960
Financial ins lending to private sector	5.753	0.950	1.053	0.934	1.008	0.946	0.949	1.054	1.069
RPI: All items exc mortgage int payments and indirect taxes	0.973	0.971	2.232	1.340	0.969	0.979	0.927	0.809	0.784
RPI: All items	0.935	0.934	2.876	1.613	0.936	0.943	0.805	0.800	0.735
RPI: Total food	1.024	1.023	2.610	1.569	1.021	1.030	0.958	0.956	0.895
RPI: Housing	6.654	0.955	3.351	1.819	0.958	0.981	0.722	0.623	0.837
RPI: Fuel & light	1.846	1.010	2.516	1.588	1.012	1.010	1.071	1.193	1.033
RPI: Household goods	5.112	0.652	0.659	0.721	0.983	0.639	0.637	1.130	1.007
RPI: Household services	1.019	1.018	2.145	1.404	1.019	1.024	0.985	0.900	0.917
RPI: Personal expenditure	1.860	0.909	1.078	0.988	0.906	0.850	0.901	1.009	0.878
RPI: Travel & leisure	0.946	0.944	2.632	1.606	0.946	1.125	0.789	0.817	0.923
RPI: Consumer durables	4.401	0.820	0.810	0.852	1.010	0.811	0.811	1.233	1.051
RPI: All items excluding mortgage interest	0.947	0.945	2.885	1.608	0.947	0.955	0.751	0.715	0.680
RPI: All goods	7.257	0.982	1.849	1.289	0.962	1.026	0.947	0.978	0.937
RPI: All services	0.977	0.975	3.503	1.784	0.977	0.989	0.622	0.600	0.547

Bibliography

Adolfson, M., Lindé, J., & Villani, M. (2005). Forecasting performance of an open economy dynamic stochastic general equilibrium model. *Sveriges Riksbank Working Paper, 190.*

Anderson, N., & Sleath, J. (2001). New estimates of the UK real and nominal yield curves. *Bank of England Working Paper, 126.*

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica, 61*(4), 821-856.

Ang, A., & Bekaert, G. (2002). Regime switches in interest rates. *Journal of Business & Economic Statistics, 20*(2), 163-182.

Ang, A., & Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary economics, 50*(4), 745-787.

Assenmacher-Wesche, K., & Pesaran, M. H. (2009). A VECX model of the Swiss economy. *Swiss National Bank Economic Studies*, *6.*

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, *66*(1), 47-78.

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics, 18*(1), 1-22.

Barnett, A, Mumtaz, H. & Theodoridis, K. (2012) Forecasting UK GDP Growth, Inflation and Interest Rates under Structural Change: A Comparison of Models with Time-Varying Parameters, *Bank of England Working Paper, 450.*

Bauwens, L., Koop, G., Korobilis, D., & Rombouts, J. (2011). A comparison of forecasting procedures for macroeconomic series: the contribution of structural break models. *SIRE Discussion Paper, 2011-33.*

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics, 19*(4), 465-474.

Bernanke, B. S., & Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics, 50*(3), 525-546.

Bianchi, F., Mumtaz, H., & Surico, P. (2009). The great moderation of the term structure of UK interest rates. *Journal of Monetary Economics, 56*(6), 856-871.

Björk, T., & Christensen, B. J. (1999). Interest rate dynamics and consistent forward rate curves. *Mathematical Finance, 9*(4), 323-348.

Bliss, R. R. (1996). Testing term structure estimation methods. *Working Paper 96-12, Federal Reserve Bank of Atlanta.*

Bliss, R. R. (1997). Movements in the term structure of interest rates. *Economic Review,* (Q4), 16-33.

Castle, J. L., Fawcett, N. W., & Hendry, D. F. (2011). Forecasting breaks and forecasting during breaks. *Oxford Handbook of Economic forecasting*, Oxford University Press, 315-354.

Cheridito, P., Filipović, D., & Kimmel, R. L. (2007). Market price of risk specifications for affine models: Theory and evidence. *Journal of Financial Economics, 83*(1), 123-170.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica, 28*(3), 591-605.

Christensen, J. H., Diebold, F. X., & Rudebusch, G. D. (2009). An arbitrage-free generalized Nelson–Siegel term structure model. *The Econometrics Journal, 12*(3), C33-C64.

Christensen, J. H., Diebold, F. X., & Rudebusch, G. D. (2011). The affine arbitrage-free class of Nelson–Siegel term structure models. *Journal of Econometrics, 164*(1), 4-20.

Christensen, J. H., & Lopez, J. A. (2008). Common risk factors in the US treasury and corporate bond markets: an arbitrage-free dynamic Nelson-Siegel modelling approach. *Manuscript, Federal Reserve Bank of San Francisco.*

Christensen, J. H., Lopez, J. A., & Rudebusch, G. D. (2010). Can spanned term structure factors drive stochastic volatility? *Manuscript, Federal Reserve Bank of San Francisco.*

Chu, C. S. J., Stinchcombe, M., & White, H. (1996). Monitoring structural change. *Econometrica, 65*(4), 1045-1065.

Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics, 29*(3).

Clark, T. E., & McCracken, M. W. (2009). Improving forecast accuracy by combining recursive and rolling forecasts. *International Economic Review, 50*(2), 363-395.

Clark, T. E., & McCracken, M. W. (2010). Averaging forecasts from VARs with uncertain instabilities. *Journal of Applied Econometrics, 25*(1), 5-29.

Clements, M. P., & Hendry, D. F. (1996). Intercept corrections and structural change. *Journal of Applied Econometrics, 11*(5), 475-494.

Clements, M., & Hendry, D. (1998). *Forecasting economic time series.* Cambridge University Press.

Clements, M. P., & Hendry, D. F. (2001). *Forecasting non-stationary economic time series.* The MIT Press.

Clements, M. P., & Krolzig, H. M. (1998). A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP. *The Econometrics Journal, 1*(1), 47-75.

Collin-Dufresne, P., Goldstein, R. S., & Jones, C. S. (2009). Can interest rate volatility be extracted from the cross section of bond yields?. *Journal of Financial Economics, 94*(1), 47-66.

Cox, J. C., Ingersoll Jr, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica, 53*(2), 385-407.

D'Agostino, A., Gambetti, L., & Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics, 28,* 82-101.

Dai, Q., & Singleton, K. J. (2000). Specification analysis of affine term structure models. *The Journal of Finance, 55*(5), 1943-1978.

De Jong, F. (2000). Time series and cross-section information in affine term-structure models. *Journal of Business & Economic Statistics, 18*(3), 300-314.

De Pooter, M. (2007). Examining the Nelson-Siegel class of term structure models. *Tinbergen Institute Discussion Paper, TI 2007-043/4.*

De Pooter, M., Ravazzolo, F., & Van Dijk, D. (2010). Term structure forecasting using macro factors and forecast combination. *International Finance Discussion Papers, 993.*

Diebold, F. X., Gunther, T., and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review, 39,* 863-883.

Diebold, F. X., & Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics, 105*(1), 131-159.

Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics, 130*(2), 337-364.

Diebold, F. X. & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 13*(3), 253-63.

Diebold, F. X., and Rudebusch, G. D. (2013). *Yield Curve Modelling and Forecasting: The Dynamic Nelson-Siegel Approach.* Princeton University Press.

Diebold, F. X., Rudebusch, G. D., & Aruoba, S. B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of econometrics, 131*(1), 309-338.

Duffee, G. R. (2011). Forecasting with the term structure: The role of no-arbitrage restrictions. *Working papers, the Johns Hopkins University, Department of Economics.*

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance, 57*(1), 405-443.

Duffee, G. R. (1999). Estimating the price of default risk. *Review of Financial Studies, 12,* 197-226.

Duffie, D. R., & Kan, R. (1996). A yield-factor model of interest rates. *Mathematical finance, 6*(4), 379-406.

Egorov, A. V., Hong, Y., & Li, H. (2006). Validating forecasts of the joint probability density of bond yields: Can affine models beat random walk? *Journal of Econometrics, 135*(1), 255-284.

Eklund, J., Kapetanios, G., & Price, S. (2010). Forecasting in the presence of recent structural change. *Bank of England Working Paper, 406.*

Estrella, A., & Mishkin, F. S. (1998). Predicting US recessions: financial variables as leading indicators. *Review of Economics and Statistics, 80*(1), 45-61.

Fackler, P. (2000). Moments of affine diffusions. *Unpublished working paper, North Carolina State University.*

Favero, C. A., Niu, L., & Sala, L. (2012). Term structure forecasting: No-arbitrage restrictions versus large information set. *Journal of Forecasting, 31*(2), 124-156.

Fisher, M., & Gilles, C. (1996). Estimating exponential-affine models of the term structure. *Unpublished working paper, Federal Reserve Board, Washington, DC.*

Garcia, R., & Perron, P. (1996). An analysis of the real interest rate under regime shifts. *The Review of Economics and Statistics, 78*(1), 111-125.

Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics, 25*(4), 595-620.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica, 74*(6), 1545-1578.

Giordani, P., & Kohn, R. (2008). Efficient Bayesian inference for multiple change-point and mixture innovation models. *Journal of Business & Economic Statistics, 26*(1), 66-77.

Giordani, P., & Villani, M. (2010). Forecasting macroeconomic time series with locally adaptive signal extraction. International *Journal of Forecasting, 26*(2), 312-325.

Giraitis, L., Kapetanios, G., & Price, S. (2013). Adaptive forecasting in the presence of recent and ongoing structural change. *Journal of Econometrics,* in press.

Gourieroux, C., & Jasiak, J. (2001). Memory and infrequent breaks. *Economics Letters, 70*(1), 29-41.

Granger, C. W., & Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance, 11*(3), 399-421.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica,* 57(2), 357-384.

Hautsch, N., & Yang, F. (2012). Bayesian inference in a stochastic volatility Nelson-Siegel model. *Computational Statistics & Data Analysis*, *56*(11), 3774-3792.

Hendry, D. F. (2000). On detectable and non-detectable structural change. *Structural Change and Economic Dynamics, 11*(1), 45-65.

Hong, Y., & Li, H. (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies, 18*(1), 37-84.

Hördahl, P., Tristani, O., & Vestin, D. (2006). A joint econometric model of macroeconomic and term-structure dynamics. *Journal of Econometrics, 131*(1), 405-444.

Jacobs, K., & Karoui, L. (2009). Conditional volatility in affine term-structure models: Evidence from Treasury and swap markets. *Journal of Financial Economics*, *91*(3), 288-318.

Joslin, S., Singleton, K. J., & Zhu, H. (2011). A new perspective on Gaussian dynamic term structure models. *Review of Financial Studies, 24*(3), 926-970.

Joyce, M., Kaminska, I., & Lildholdt, P. (2008). Understanding the real rate conundrum: An application of no-arbitrage finance models to the UK real yield curve. *Bank of England working paper, 358.*

Kaminska, I. (2012). A no-arbitrage structural vector autoregressive model of the UK yield curve. *Oxford Bulletin of Economics and Statistics, 75*(5), 680-704.

Koop, G., & Potter, S. M. (2007). Estimation and forecasting in models with multiple breaks. *The Review of Economic Studies, 74*(3), 763-789.

Koopman, S. J., Mallee, M. I., & Van der Wel, M. (2010). Analyzing the term structure of interest rates using the dynamic Nelson–Siegel model with time-varying parameters. *Journal of Business & Economic Statistics*, *28*(3), 329-343.

Leisch, F., Hornik, K., & Kuan, C. M. (2000). Monitoring structural changes with the generalized fluctuation test. *Econometric Theory, 16*(6), 835-854.

Lildholdt, P., Panigirtzoglou, N., & Peacock, C. (2007). An affine macro-factor model of the UK yield curve. *Bank of England working paper, 322.*

Litterman, R. B., & Scheinkman, J. (1991). Common factors affecting bond returns. *The Journal of Fixed Income, 1*(1), 54-61.

Litterman, R. B., Scheinkman, J., & Weiss, L. (1991). Volatility and the yield curve. *The Journal of Fixed Income, 1*(1), 49-53.

Maheu, J. M., & Gordon, S. (2008). Learning, forecasting and structural breaks. *Journal of Applied Econometrics, 23*(5), 553-583.

Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics, 135*(1), 499-526.

Moench, E. (2008). Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach. *Journal of Econometrics, 146*(1), 26-43.

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modelling of yield curves. *Journal of Business, 60*(4), 473-489.

Pesaran, M. H., Pettenuzzo, D., & Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks. *The Review of Economic Studies, 73*(4), 1057-1084.

Pesaran, M. H., & Pick, A. (2011). Forecast combination across estimation windows. *Journal of Business & Economic Statistics, 29*(2), 307-308.

Pesaran, M. H., & Timmermann, A. (2002). Market timing and return prediction under model instability. *Journal of Empirical Finance, 9*(5), 495-510.

Pesaran, M. H., & Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics, 137*(1), 134-161.

Piazzesi, M. (2010). Affine term structure models. *Handbook of Financial Econometrics, 1,* 691-766, Elsevier, North-Holland.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics, 23*(3), 470-472.

Rossi, B. (2012). Advances in forecasting under instability. In: Elliott. G. & Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, Elsevier, North-Holland.

Rudebusch, G. D., & Wu, T. (2008). A macro-finance model of the term structure, monetary policy and the economy. *The Economic Journal, 118*(530), 906-926.

Stock, J. H., & Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics, 14*(1), 11-30.

Stock, J. H., & Watson, M. W. (2003). Forecasting Output and Inflation: The Role of Asset Prices. *Journal of Economic Literature, 16,* 788-829.

Stock, J. H., & Watson, M. W. (2007). Why has US inflation become harder to forecast?. *Journal of Money, Credit and Banking, 39*(s1), 3-33.

Svensson, L. E. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994. *National Bureau of Economic Research working paper, 4871.*

Taylor, J. B. (1993). Discretion versus policy rules in practice. *In Carnegie-Rochester conference series on public policy, 39,* 195-214.

Timmermann, A. (2001). Structural breaks, incomplete information, and stock prices. *Journal of Business & Economic Statistics, 19*(3), 299-314.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics, 5*(2), 177-188.

Wu, T. (2006). Macro factors and the affine term structure of interest rates. *Journal of Money, Credit and Banking, 38*, 1847-1875.

Zeileis, A., Leisch, F., Kleiber, C., & Hornik, K. (2005). Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics, 20*(1), 99-121.