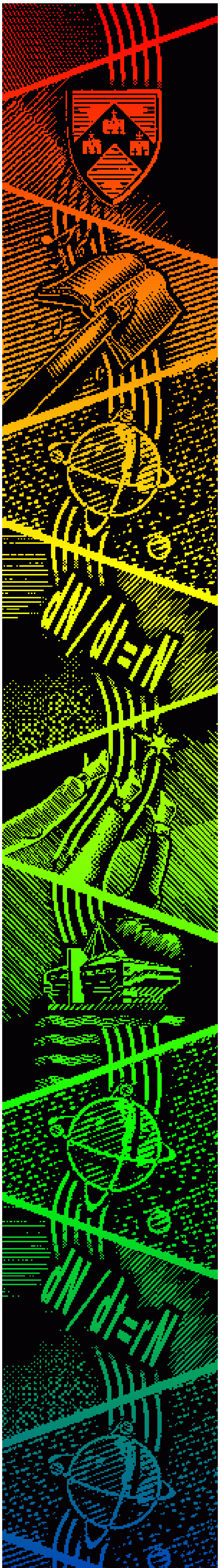

THE UNIVERSITY *of York*

Discussion Papers in Economics

No. 12/27

The Meiselman forward interest rate revision regression as an Affine Term Structure Model

Adam Golinski and Peter Spencer

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

The Meiselman forward interest rate revision regression as an Affine Term Structure Model*†

Adam Golinski and Peter Spencer

Abstract

We adapt the Meiselman (1962) OLS forward rate revision framework to obtain the discrete time analogue of the Heath, Jarrow and Morton (1992) specification and use it for estimating and testing term structure models. Our framework is based upon the Wold representation of the factor dynamics and combines the flexibility of the 'no arbitrage' approach used by practitioners for pricing with the time series domain econometrics used in the 'equilibrium approach' by academic researchers. It allows us to estimate the no-arbitrage term structure under the risk-neutral measure without adopting any specific model of the factor dynamics. Using three different datasets we find that our discrete time Heath et al (1992) no-arbitrage model is not rejected against the unrestricted OLS model of Meiselman (1962). We then develop a dynamic term structure model by specifying a model of a risk premium to link the risk neutral dynamics of the cross section to the real-world factor dynamics. We analyse several different models of the dynamics from the ARFIMA class and

*Corresponding author: Golinski: University of York, Department of Economics and Related Studies, Heslington, YO10 5DD York, UK. Email: adam.golinski@york.ac.uk. Tel.: (0044) 1904 433767.
Spencer: University of York, Department of Economics and Related Studies, Heslington, YO10 5DD York, UK. Email: peter.spencer@york.ac.uk. Tel.: (0044) 1904 433771.

†We are grateful to Karim Abadir, Peter Burridge, Laura Coroneo, Fabrizo Iacone, Peter Phillips, Yongcheol Shin, Peter N Smith, Rob Taylor, Mike Thornton, Mike Wickens, Takashi Yamagata,Tomasz Zastawniak and participants at the York Conference in Honour of P.C.B Phillips.

find that the more flexible models allowing for long memory outperform short memory models and are not rejected against the Heath et al and Meiselman specifications.

Keywords: term structure, Meiselman regression, forward rate revision, Wold representation, long memory.

JEL codes: G12, C58.

1 Introduction

The Affine Term Structure Model ($ATSM$) has greatly enhanced our understanding of the behavior of financial markets and the economy, yielding rich insights into the role of macroeconomic policy and financial factors. Its theoretical structure provides a relatively simple dynamic description of the way that investor expectations and risk aversion affect bond yields, one that generates recursive linear relationships that can readily be taken to the data. The econometric test bench has now given us a multi-factor model that pays careful attention to the effect of risk, reducing the pricing errors implied by the model to very small proportions. The $ATSM$ gives finance teachers a nice way of illustrating important concepts like risk-neutral pricing and change of measure. It gives practitioners a flexible way of fitting arbitrage-free relationships to financial data and pricing new securities and central banks a way of extracting market expectations of inflation from yield data..

The $ATSM$ represents the cross section of yields or forward rates under the risk neutral (or equivalent martingale) probability measure and is very popular with practitioners. Academic studies usually employ the Dynamic Term Structure Model ($DTSM$) which incorporates a dynamic factor model and allows the factors to evolve over time under the real world (or historical) probability measure. Researchers have in the main adopted short memory or autoregressive (AR) structures to represent the dynamics under these measures, which allow the model to adjust exponentially to shocks. Empirically, multi-factor or vector autoregressive (VAR) models usually do a good job in capturing such behavior. These VAR models can be used under both the real world (or historical) probability measure to capture the time series behavior of the explanatory variables and the risk neutral measure to capture the dynamics of investor expectations implicit in the cross-section of yields. However recent studies using the more general Autoregressive Fractionally Integrated Moving Average ($ARFIMA$) specification have suggested that the long run adjustment implicit in long series of interest rate data may not be exponential, but may exhibit long memory (or slower decay than the exponential)[1]. It appears that the very small measurement and misspecification and mispricing errors in the cross-section can allow us to distinguish these from short memory processes.

This paper shows that these AR and $ARFIMA$ representations can be replaced by the more general Wold or moving average (MA) representation[2], in a way that preserves the simple recursive $ATSM$ parameter structure. This representation is normally used in econometrics to describe the behavior of a times series under the real world probability measure in terms of real world shocks. But we also use this to describe its behavior under the risk-neutral measure, as a moving average of risk-neutral shocks. The parameters of the MA representation can be estimated as the slope coefficients of simple OLS forward rate revision equations of the type first estimated by Meiselman (1962). The $ATSM$ can be made arbitrage-free using Jensen restrictions

[1] Shea (1991) found long memory in Treasury bill rates. Connolly and Guner (1999) found evidence of long memory in Treasury bond returns. Tkacz (2001) found a significant long memory parameter for several US and Canadian interest rates. Also, Gil-Alana (2004a and 2004b) found evidence of long memory US rates. Using a semiparametric approach, Iacone (2009) confirmed the presence of long memory in the short maturity nominal interest rates and rejected both short memory and unit root hypotheses.

[2] The moving average (or Wold) *representation* describes any time series as a moving average of lagged stochastic shocks as in an impulse response function (see Hamilton (1994) for example). It can be used to represent any time series and should not be confused with an MA time series *model* which describes a time series in terms of a finite number of lagged shocks.

across the intercept and slope coefficients, analogous to those of the Heath, Jarrow and Morton (1992) (henceforth HJM) continuous time model. We call this the no-arbitrage (NA) model. This procedure offers a simple way of developing an $ATSM$ without restricting the dynamics. Models using AR or $ARFIMA$ restrictions can be tested against this model. Remarkably, David Meiselman's revision regression framework has received scant attention in the finance literature since his original Coles monograph was published fifty years ago and yet it encompasses all of the $ATSMs$ that have been proposed since then.

Because it does not restrict the factor dynamics, the HJM model is quite flexible and is extensively used by practitioners who need such functional forms to fit the cross-section accurately at any moment of time to price derivative securities 'off the curve'. In practice, they use a snapshot of the forward rate structure to back out the volatility function, which is allowed to move in an unconstrained manner over time[3]. However, by recasting the HJM model in discrete time we obtain an NA model that is more flexible. We can extend the arbitrage free framework to handle long memory processes[4]. We can incorporate a dynamic factor model with the flexibility over the change of measure allowed by the discrete time approach[5], exploiting time-series data for forward rates rather than single snapshots of volatility in a time-consistent way[6]. To our knowledge this is the first time that the HJM has been specified in discrete time without restricting the factor dynamics[7].

---

[3] The connection with our Wold-Meiselman model follows from the well-known relationship between the MA representation of any time series and its volatility structure.

[4] Comte and Renault (1996) proposed a term structure model in continuous time based on fractional Brownian motion. However, Rogers (1997) and Cheridito (2003) show that fractional Brownian motion allows arbitrage in continuous time.

[5] Le et al (2010) note that to obtain a continuous time affine yield model, the price of risk must be exponential-affine in the factors, but that this restriction can be relaxed in a discrete time model.

[6] As Dybvig (1997), Lochoff (1993) Backus, Foresi and Zin (1998) and many others have pointed out, these so-called arbitrage free models allow arbitrary shifts in the parameters (and in the case of HJM type models the volatility function) which should in principle be modeled consistently over time.

[7] An early paper by Heath, Jarrow and Morton (1990) uses a binomial pricing model. The LIBOR model (Jamshidian (1997) uses forward neutral measures, which differ with maturity and require

To specify this framework we generalize the univariate Wold-Meiselman approach to allow for three correlated factors[8], which are normally sufficient to characterize the cross-section of yields or forward rates. We modify the standard yield factor approach (Duffie and Kan (1992), Chen and Scott (1993)) by assuming that three forward rate revisions are measured without error, allowing us to back out the three underlying factor revisions. These are then used as risk neutral shocks to model the other forward revisions in the contemporaneous cross-section using the risk-neutral factor dynamics. We start by testing the NA model against the unrestricted OLS model using several cross-section data sets. We find that the NA restrictions are accepted by the data: to our knowledge this is the first time that this test has been passed by a term structure model. We then develop a Dynamic Term Structure Model ($DTSM$) that models the cross section and time series behavior simultaneously. At this stage, we augment the risk-neutral model of the cross section with a Cochrane-Piazzesi (2005) style return forecasting model. This provides estimates of the price of risk which allow us to change from the risk-neutral probability measure used in the cross section to the real world measure used to handle the time series dynamics. We apply this specification to a monthly data set provided by Gurkaynak, Sack and Wright (2007) and test restrictions on the coefficients of the MA representation. The unrestricted (OLS) specification encompasses the NA model as well as models with AR, MA and $ARFIMA$ restrictions on the risk-neutral dynamics. AR models are the workhorse of the term structure literature, but we find that the more general $ARFIMA$ model is superior statistically. Recently, long memory processes have been employed but these are very time consuming when fitted to time series of bond yields rather than forward revisions[9]. The Meiselman regression approach offers a

Monte Carlo or other numerical techniques for their solution.

[8] Multi-factor versions of the HJM model require these factors to be independent.

[9] Backus and Zin (1993) were the first to consider a term structure model with long memory in the

model that is more general theoretically while much simpler to use in practice.

We believe that this new framework, which combines the flexibility of the no-arbitrage approach with the advantages of discrete time series econometrics, will appeal to a wide range of economists, econometricians and finance professionals. It provides a very general framework for testing no-arbitrage, time series and other restrictions using linear-recursive parameter techniques similar to those of the discrete time equilibrium approach, but without the restrictive autoregressive dynamics. It can be used to extracting the expectations and other information needed by central banks as well as for pricing interest rates derivatives.

The paper is set out as follows. The next section sets out the basic theoretical framework of the model, showing how the MA representation of the spot rate can be used to obtain a general $ATSM$. Section 3 sets out the three factor version of the model that is used for testing the no-arbitrage restrictions under the measure Q. It then specifies the model of the price of risk which is used to change to the real world measure used to represent the time series dynamics in the $DSTM$. Section 4 reports the empirical results, starting with the cross-section regression models used to check the validity of the Jensen restrictions on the parameters of the NA model and then the $DSTM$ is used to test the validity of AR and $ARFIMA$ restrictions on the slope parameters. Finally, section 5 offers a brief summary and suggestions for future research in this area.

---

spot rate. Duan and Jacobs (1996) proposed an equilibrium model in which long memory in interest rates arises as a consequence of fractional integration in aggregate consumption growth. They showed that the long memory component can have implications not only for the time series behavior but also for the cross-section, which cannot be compensated by richer short memory dynamics. More recently Golinski and Zaffaroni (2012) considered a 2 factor term structure model with inflation. They found that allowing for long memory in the expected inflation factor significantly improves model performance. Osterrieder and Schotman (2012) proposed a model with long memory process in the short rate and risk premium that can account for predictability of excess bond returns and the strong correlation of long rates with the level factor.

2 The risk neutral dynamics of the forward structure

This section sets out the basic theoretical framework of the model. We start with some basic results from the theory of bond pricing under the risk free measure and then show how the MA representation of the spot rate can be used to obtain a general $ATSM$. This sets out a multi-factor model, which is then specialized in Section 2.4 to a single factor example, to show the relationship with the Meiselman regression. Section 2.5 shows he relationship with the HJM model.

2.1 Discount prices and forward rates under $\mathcal{Q}$

Let the price at time t of a discount bond paying \$1 at period $t+n$ be $P_{n,t}$. Finance theory tells us that asset prices all have the same drift under the risk neutral (or equivalent martingale) measure $\mathcal{Q}$. In the case of a discount bond this implies:

$$P_{n,t} = E_t \left[\exp \left[-\sum\nolimits_{i=0}^{n-1} r_{t+i} \right] \right]. \tag{1}$$

where r_t is the spot rate. We use u and the operator E to denote shocks and expectations under the risk neutral measure $\mathcal{Q}$ and later use $u^{\mathcal{P}}$ and $\mathbb{E}$ to represent them under the real world measure $\mathcal{P}$.

Assuming that this is *conditionally* Gaussian, the value of a discount bond may be expressed using the well known formula for the expected value of a lognormally distributed variable:

$$p_{n,t} = -r_t - E_t \left[\sum\nolimits_{i=1}^{n-1} r_{t+i} \right] + \frac{1}{2} Var \left[\sum\nolimits_{i=1}^{n-1} r_{t+i} \right], \tag{2}$$

where: $p_{n,t} = \log[P_{n,t}]$. We define the forward rate for period $t+n$ as the logarithmic rollover rate $f_{n,t}$ obtained by investing in a bond maturing at $t+n+1$ rather than one maturing at $t+n$:

$$f_{nt} = p_{nt} - p_{n+1,t} \qquad (3)$$
$$= E_t[r_{t+n}] - \frac{1}{2}Var\left[\sum\nolimits_{i=1}^{n} r_{t+i}\right] + \frac{1}{2}Var\left[\sum\nolimits_{i=1}^{n-1} r_{t+i}\right].$$

It is well known that an AR specification of the risk neutral dynamics gives a convenient recursive solution to bond prices under the assumption that bond prices are exponential affine in the factors driving the interest rate. Because an n maturity discount bond becomes an $(n-1)$ maturity bond in the next period, the parameters of this function, known as factor loadings, are recursive in maturity. They also follow an AR restriction (Campbell et al (1997)). We now show that the MA representation preserves the recursive property, but relaxes the AR restriction.

2.2 The Moving Average representation of the spot rate

To estimate the model we use the discrete time MA representation. This can be used to represent the behavior of any mean reverting random variable as an initial value plus an MA of subsequent *i.i.d.* innovations. We assume in this paper that these are Gaussian.

Suppose that the spot rate r_t and the term structure is driven by K homoscedastic factors and that each has a univariate MA representation under the risk-neutral measure:

$$r_t = r_0 + \sum_{k=1}^{K}\sum_{i=0}^{t-1} \beta_{k,i} u_{k,t-i} = r_0 + \sum_{i=0}^{t-1} \boldsymbol{\beta}_i' \mathbf{u}_{t-i}, \qquad (4)$$

with:

$$\mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Omega}) \qquad (5)$$

where $r_0 = E_0[r_t]$ is the initial risk-neutral expectation of the spot rate, $\boldsymbol{\Omega} = \{\sigma_{ij}\}$;

$\sigma_{ij} = \sigma_{ji}$ and $\sigma_{jj} > 0$. $\mathbf{\Omega} = \mathbf{\Sigma}'\mathbf{\Sigma}$ and where $\mathbf{\Sigma}$ is upper (or lower) triangular with positive diagonal entries.

Any future interest rate can be written as the sum of components, known and unknown at t:

$$
\begin{aligned}
r_{t+n} &= \sum\nolimits_{i=0}^{t+n-1} \boldsymbol{\beta}_i' \mathbf{u}_{t+n-i} + r_0 && (6) \\
&= \sum\nolimits_{i=0}^{n-1} \boldsymbol{\beta}_i' \mathbf{u}_{t+n-i} + \sum\nolimits_{i=n}^{t+n-1} \boldsymbol{\beta}_i' \mathbf{u}_{t+n-i} + r_0 && (7) \\
&= \left\{ \boldsymbol{\beta}_0' \mathbf{u}_{t+n} + ... + \boldsymbol{\beta}_{n-1}' \mathbf{u}_{t+1} \right\} + \left\{ \boldsymbol{\beta}_n' \mathbf{u}_t + ... + \boldsymbol{\beta}_{t+n-1}' \mathbf{u}_1 + r_0 \right\} \\
&= \quad \{\text{unknown at } t\} \quad + \quad \{\text{known at } t\}
\end{aligned}
$$

Taking expectations under $\mathcal{Q}$ removes the unknowns and leaves:

$$
E_t[r_{t+n}] = \sum\nolimits_{i=n}^{t+n-1} \boldsymbol{\beta}_i' \mathbf{u}_{t+n-i} + r_0. \tag{8}
$$

This revises the previous expectation for that period by adding the innovation $\boldsymbol{\beta}_n' \mathbf{u}_t$:

$$
E_t[r_{t+n}] = E_{t-1}\left[r_{t+n}\right] + \boldsymbol{\beta}_n' \mathbf{u}_t. \tag{9}
$$

Its variance depends upon the unknown part of (6). Given the i.i.d. assumption, this is given by the well-known formula:

$$
\begin{aligned}
Var[r_{t+n}] &= Var\left[\sum\nolimits_{i=0}^{n-1} \boldsymbol{\beta}_i' \mathbf{u}_{t+n-i}\right] && (10) \\
&= \sum\nolimits_{i=0}^{n-1} \boldsymbol{\beta}_i' \mathbf{\Omega} \boldsymbol{\beta}_i
\end{aligned}
$$

Similarly the variance of the sum in (2) is:

$$Var\left[\sum_{i=1}^{n} r_{t+i}\right] = Var\left[\sum_{i=1}^{n}\sum_{j=0}^{i-1} \boldsymbol{\beta}_j' \mathbf{u}_{t+i-j}\right] \tag{11}$$

$$= \sum_{i=1}^{n}\left(\sum_{j=0}^{i-1} \boldsymbol{\beta}_j'\right) \boldsymbol{\Omega} \left(\sum_{j=0}^{i-1} \boldsymbol{\beta}_j\right) \tag{12}$$

The maturity can be increased using the recursion:

$$Var\left[\sum_{i=1}^{n} r_{t+i}\right] = Var\left[\sum_{i=1}^{n-1} r_{t+i}\right] + \left(\sum_{i=0}^{n-1} \boldsymbol{\beta}_i'\right) \boldsymbol{\Omega} \left(\sum_{i=0}^{n-1} \boldsymbol{\beta}_i\right). \tag{13}$$

Putting this into (3) shows that the forward rate for any future period $t+n$ can be written as the risk neutral expectation minus a Jensen term:

$$\begin{aligned} f_{nt} = \quad & E_t[r_{t+n}] \quad - \quad \frac{1}{2}\left(\sum_{i=0}^{n-1} \boldsymbol{\beta}_i\right)' \boldsymbol{\Omega} \left(\sum_{i=0}^{n-1} \boldsymbol{\beta}_i\right). \\ = & \text{ Expectation } - \quad \text{Jensen term} \end{aligned}$$

and the revision in period t for period $t+n$ can be written as:

$$\begin{aligned} f_{n,t} - f_{n+1,t-1} = & \quad a_n \quad + \quad \boldsymbol{\beta}_n' \mathbf{u}_t; \qquad n \geq 1 \\ = & \text{ Jensen maturity shortening} + \text{Innovation} \end{aligned} \tag{14}$$

subject to the no-arbitrage (NA) condition:

$$\text{where:} \qquad a_n = \frac{1}{2}\left[\left(\sum_{i=0}^{n} \boldsymbol{\beta}_i'\right) \boldsymbol{\Omega} \left(\sum_{i=0}^{n} \boldsymbol{\beta}_i\right) - \left(\sum_{i=0}^{n-1} \boldsymbol{\beta}_i\right)' \boldsymbol{\Omega} \left(\sum_{i=0}^{n-1} \boldsymbol{\beta}_i\right)\right] \tag{15}$$

Under this measure, the forward rate for any specific *period* follows a random walk with a drift due to the Jensen maturity shortening effect. The forward rate for any fixed *maturity* can nevertheless be mean reverting.

2.3 The Hicksian decomposition

The risk neutral innovation can be represented as the sum of a real world innovation plus a risk premium or expected excess return. This reflects the change of measure from $\mathcal{Q}$ to $\mathcal{P}$ which as appendix A explains, respecifies both the drift and the error terms of the revision process, adding the risk factor $\mathbf{\Sigma}'\mathbf{\Lambda}_t$ to the drift and compensating for this by subtracting it from the innovation under $\mathcal{Q}$ to get the innovation under $\mathcal{P}$:

$$\mathbf{u}_t^{\mathcal{P}} = \mathbf{u}_t - \mathbf{\Sigma}'\boldsymbol{\lambda}_{t-1} \quad \text{where: } \mathbb{E}_{t-1}[\mathbf{u}_t^{\mathcal{P}}] = 0; \quad \mathbb{E}_{t-1}[\mathbf{u}_t] = \mathbf{\Sigma}'\boldsymbol{\lambda}_{t-1}. \tag{16}$$

The risk factor is the product of the volatilities ($\mathbf{\Sigma}$) and the vector of prices of risk ($\boldsymbol{\lambda}_t$) . Substituting $\mathbf{u}_t = \mathbf{u}_t^{\mathcal{P}} + \mathbf{\Sigma}'\boldsymbol{\lambda}_{t-1}$ into (14) gives the Hicksian decomposition of the forward rate.

2.4 A single factor Meiselman (1962) regression model

Suppose for a moment that $K = 1$ and that the spot rate r_t and the short forward $f_{1,t-1}$ are measured without error so that we can use the short rate revision as a proxy for the innovation. Recall that if $\beta_0 = 1$, (3) specializes to:

$$f_{1,t-1} = E_{t-1}[r_t] - \frac{1}{2}\sigma^2.$$

Similarly (14) with $f_{0,t} = r_t$ gives the short revision:

$$r_t - f_{1,t-1} = u_t + \frac{1}{2}\sigma^2. \tag{17}$$

allowing us to infer the innovation:

$$u_t = r_t - f_{1,t} - \frac{1}{2}\sigma^2. \tag{18}$$

and substitute back into (14) to get:

$$f_{nt} - f_{n+1,t-1} = \alpha_n + \beta_n \left(r_t - f_{1,t}\right), \quad n = 1, ..., N \tag{19}$$

where:

$$\alpha_n = \frac{1}{2}\sigma^2 \left[\left(\sum_{i=0}^{n} \beta_i \right)^2 - \left(\sum_{i=0}^{n-1} \beta_i \right)^2 - \beta_n \right]$$

and N is the largest maturity in the sample. Adding residuals to allow for measurement error gives a linear model of the cross-section that we recognize as a Meiselman (1962) regression. The single equation model is just-identified for $n = 1$: it yields two reduced form regression coefficients (α_1, β_1) that can be solved for (σ^2, β_1). If additional revisions are added the model is over-identified, imposing no-arbitrage restrictions across the volatility and the Jensen drift term. These take the form of recursion relationships, as in the AR model and section 4.1 gives some empirical examples for a three factor model. This slope-intercept restriction is similar to that implied by the model of HJM.

2.5 Heath Jarrow Morton (1992)

HJM develop a similar arbitrage-free pricing model under $\mathcal{Q}$ in continuous time. They note that absent arbitrage, the drift in all asset prices must be the same. This implies that the instantaneous forward rate $f(t, T, \mathbf{X}(t))$ for period $t + n$ at time t follows a Stochastic Differential Equation with its well known restriction across volatility v

and Jensen-drift α:

$$df(t, t+n, \mathbf{X}(t)) = \alpha(t, t+n, \mathbf{X}(t))dt + v(t, t+n, \mathbf{X}(t))dw(t) \tag{20}$$

where:

$$\alpha(t, t+n, \mathbf{X}(t)) = v(t, t+n, \mathbf{X}(t)) \int_t^{t+n} v(t, s, \mathbf{X}(s))ds$$

allowing v to be 'backed out' from the cross-section forward curve. This model is very popular with practitioners, who fit a different snapshot of the cross-section forward/volatility curve every day. It is one of a range of arbitrage-free models that specify the yield curve in a way that is sufficiently flexible to provide an accurate representation of bonds or option prices at any given instant. Practitioners argue that this type of model is more appropriate than 'equilibrium' models like Vasicek (1979) and Cox, Ingersoll and Ross(1985) which have a small number of parameters. However, these models can be extended using multiple factors and richer specifications of the price of risk, reducing the pricing errors to very small proportions. As noted in the introduction, their use of time series data for the cross-section rather than just a snapshot helps resolve the problem of dynamic inconsistency. Theoretically, there is no difference between time series and volatility modelling since these are uniquely related (by relationships like (10)).

By using a carefully-specified time series model to extract information from both the cross-section and the dynamics of the time series data as we do in section 4.2 of this paper we can potentially get the best of both worlds. Unlike HJM and other arbitrage-free models that only model the cross-section under $\mathcal{Q}$, we can make inferences about the dynamics and the risk premia as well as the cross-section. The so-called arbitrage-free models are silent on these issues. On the other hand, we do not need to use the very long estimation periods used by time series econometricians to extract information about long memory and other low frequency phenomena since we can instead extract this information from the long maturities in the cross-section. Although this only tells us about the risk neutral process, we can make inferences about the real world dynamics given sufficient time series observations to model the price of risk. In practice, the choice of estimation period and cross-section can be flexible, allowing for likely structural breaks through the choice of estimation period, maturity structure and the number and type of model factors and parameters. We return to this subject in section 4.

3 The empirical models

This section sets out the three factor version of the model that is used for testing the no-arbitrage restrictions under the measure $\mathcal{Q}$. We briefly discuss some of the $ARFIMA$ models that can be used to restrict the dynamics under $\mathcal{Q}$. We then specify the model of the price of risk which is used to change to the $\mathcal{P}$ measure used to represent the real world dynamics.

3.1 A three factor model and the risk neutral dynamics

Following the yield factor literature (Duffie and Kan (1996)) we assume that three factors are sufficient to model the yield curve ($K = 3$) and that three revisions, respectively short (s), medium (m) and long term (l), are observed without error. For this reason we call them 'noiseless' or 'noise-free' revisions, distinguishing them from the remaining 'noisy' revisions. The specific maturities depend upon the data set being tested. Collecting these three observations in a vector X_t gives the relationship between these three revisions and the three shocks that drive the system under $\mathcal{Q}$:

$$\mathbf{X}_t = \begin{bmatrix} f_{st} - f_{s+1,t-1} \\ f_{mt} - f_{m+1,t-1} \\ f_{lt} - f_{l+1,t-1} \end{bmatrix} = \begin{bmatrix} a_s \\ a_m \\ a_l \end{bmatrix} + \begin{bmatrix} \beta_{1s} & \beta_{2s} & \beta_{3s} \\ \beta_{1m} & \beta_{2m} & \beta_{3m} \\ \beta_{1l} & \beta_{2l} & \beta_{3l} \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

which we can write as:

$$\mathbf{X}_t = \mathbf{a}_x + \mathbf{B}_x \mathbf{u}_t. \tag{21}$$

We then back out the vector of shocks as:

$$\mathbf{u}_t = \mathbf{B}_x^{-1}\mathbf{X}_t - \mathbf{B}_x^{-1}\mathbf{a}_x, \tag{22}$$

and use this to specify the remaining 'noisy' revisions. These are measured with error and stacked into an $N-3$ column vector Y_t :

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{a}_y + \mathbf{B}_y \mathbf{u}_t + \mathbf{v}_t \qquad (23) \\ &= \left(\mathbf{a}_y - \mathbf{B}_y \mathbf{B}_x^{-1}\mathbf{a}_x\right) + \mathbf{B}_y \mathbf{B}_x^{-1}\mathbf{X}_t + \mathbf{v}_t, \end{aligned}$$

where $\mathbf{v}_t$ is an $N-3-$dimensional vector of measurement errors, $\mathbf{v}_t \sim N(\mathbf{0}, \boldsymbol{\Xi})$. The intercept coefficients $\mathbf{a}_x$ and $\mathbf{a}_y$ in this system follow from the (unrestricted) slope coefficients and the variance structure using (14). Comparing this with an unrestricted OLS regression model of the cross-section then gives a test of the no-arbitrage restrictions reported in section 5.1.

3.2 Restricting the risk neutral dynamics

The dynamics of the factors under Q can be restricted by using models developed by time series econometricians to restrict the βs in (4). For expositional simplicity we present the single factor model here. The traditional assumption about the short rate risk neutral dynamics is that it follows a first order autoregressive ($AR(1)$) process, for which $\beta_0 = 1$ and $\beta_i = \phi^i$ for $i \geq 1$. We consider a broader specification that allows for autoregressive, moving average and fractionally integrated component in the dynamics, i.e. $ARFIMA(p, d, q)$ (see Granger and Joyoux (1980), Hosking (1981)), where p denotes the number of autoregressive lags, q is the number of moving average lags and d is the long memory parameter. In general, the $ARFIMA$ class can be represented as

$$\Phi(L)\left(1-L\right)^{d} x_t = \Theta(L) e_t,$$

where e_t is white noise and $\Phi(L)$ and $\Theta(L)$ are autoregressive and moving average polynomials, respectively. The process is stationary if the autoregressive roots lie outside the unit circle and the long memory parameter is smaller than 1/2. For example, the $ARFIMA(2, d, 2)$ x_t process with zero mean is given by

$$\left(1-\phi_1 L\right)\left(1-\phi_2 L\right)\left(1-L\right)^{d} x_t = \left(1+\theta_1 L\right)\left(1+\theta_2 L\right) e_t,$$

The Wold theorem states that any covariance stationary process has a unique moving average representation, i.e.

$$x_t = \sum_{i=0}^{\infty} \beta_i e_{t-i} = \Phi(L)^{-1}\left(1-L\right)^{-d} \Theta(L) e_{t.}.$$

By noting that

$$\begin{aligned}
\left(1-L\right)^{-d} &= 1 + dL + d\left(d+1\right)L^2/2! + d\left(d+2\right)\left(d+3\right)/3! + \ldots \\
&= \sum_{i=0}^{\infty} \psi_i L, \qquad \text{where } \psi_i = \Gamma\left(i+d\right)/\left[\Gamma\left(d\right)\Gamma\left(i+1\right)\right],
\end{aligned}$$

and where Γ is the gamma function, we can find the β's by matching the coefficients at each lag.

3.3 The price of risk and the real world dynamics

System (23) is the $ATSM$: the model of the cross-section under $\mathcal{Q}$ which relates the noisy revisions $\mathbf{Y}_t$ to the errors backed out from the contemporaneous noise-free revisions X_t. The joint likelihood of the $DTSM$ is a function of the system innovations $\mathbf{u}_t$ and measurement errors $\mathbf{v}_t$, which incorporates a dynamic time series model of X_t. This is informative about the risk premia as well as the dynamics and employs the relationship between the real world and risk neutral innovations (16) where the vector $\boldsymbol{\lambda}_t$ contains the 'prices of risk'. To specify these we follow Duffee (2002):

$$\boldsymbol{\Sigma}'\boldsymbol{\lambda}_t = \boldsymbol{\lambda}_0 + \boldsymbol{\Lambda}_1'\mathbf{Z}_t, \tag{24}$$

where $\mathbf{Z}_t$ is an $M-$vector of variables that drive the risk premia.

Substituting these (16) and (24) into (21) and (23) gives the real world dynamics:

$$\mathbf{X}_t = \mathbf{a}_x + \mathbf{B}_x\boldsymbol{\lambda}_0 + \mathbf{B}_x\boldsymbol{\Lambda}_\mathbf{1}'\mathbf{Z}_{t-1} + \mathbf{B}_x\mathbf{u}_t^{\mathcal{P}}, \tag{25}$$

$$\mathbf{Y}_t = \mathbf{a}_y + \mathbf{B}_y\boldsymbol{\lambda}_0 + \mathbf{B}_y\boldsymbol{\Lambda}_\mathbf{1}'\mathbf{Z}_{t-1} + \mathbf{B}_y\mathbf{u}_t^{\mathcal{P}} + \mathbf{v}_t. \tag{26}$$

These resemble Cochrane and Piazzesi (2005) return forecasting equations, although we model the monthly return to a forward rather than the annual return to a discount bond position. The return on an $n-$period bond bought at time t is $r_{n,t+1} = p_{n-1,t+1} - p_{n,t}$, which, from the definition of forward rate (3), is:

$$\begin{aligned} r_{n,t+1} &= (f_{n-1,t+1} + p_{n,t+1}) - (f_{n,t} + p_{n+1,t}) \\ &= (f_{n-1,t+1} - f_{n,t}) + r_{n+1,t+1}. \end{aligned}$$

Thus the forward rate revision is related to the discount bond return by: $f_{n-1,t+1} - f_{n,t} = r_{n,t+1} - r_{n+1,t+1}$. The variables $\mathbf{Z}_t$ driving the price of risk can include spot and forward rates (represented by the 3 dimensional vector X_t) that are spanned as well as macroeconomic and other variables (M_t), that are not spanned by the term structure.

3.4 Estimation procedure

System (23) allows us to estimate the $ATSM$ without restricting the β coefficients. Because we can rotate the factors in such a way that $\mathbf{B}_x$ is an identity matrix, we can use the initial OLS estimates of β's as the starting values in the MLE procedure.

To estimate the DTSM ((25) and (26)), we note that the measurement errors are independent of the state variables. This allows us to write the joint conditional density of the Meiselman revisions under P as:

$$f^{\mathcal{P}}(\mathbf{X}_t, \mathbf{Y}_t | \mathbf{Z}_{t-1}) = |\det(\mathbf{J})|^{-1} f^{\mathcal{P}}(\mathbf{u}_t^{\mathcal{P}}; \mathbf{0}, \mathbf{\Omega}) f^{\mathcal{P}}(\mathbf{v}_t; \mathbf{0}, \mathbf{\Xi}), \tag{27}$$

where $\det(\mathbf{J})$ is the determinant of the Jacobian:

$$\mathbf{J} = \begin{bmatrix} \mathbf{B}_x & \mathbf{0} \\ \mathbf{B}_y & \mathbf{I} \end{bmatrix},$$

and:

$$\mathbf{u}_t^{\mathcal{P}} = \mathbf{B}_x^{-1}\mathbf{X}_t - \mathbf{B}_x^{-1}\mathbf{a}_x - \boldsymbol{\lambda}_0 - \mathbf{\Lambda}_1'\mathbf{Z}_{t-1}, \tag{28}$$

$$\mathbf{v}_t = \mathbf{Y}_t - (\mathbf{a}_y - \mathbf{B}_y\mathbf{B}_x^{-1}\mathbf{a}_x) - \mathbf{B}_y\mathbf{B}_x^{-1}\mathbf{X}_t. \tag{29}$$

Maximizing the function (27) w.r.t. the parameters gives their ML estimates. We first estimate a linear model with unrestricted intercept and slope parameters by OLS and then test (i) the recursive Jensen NA restrictions and (ii) $ARFIMA$ restrictions on the slope parameters against this. In the restricted models the intercepts (and in (ii) the slope coefficients) are determined by recursive relationships involving the basic structural parameters.

The estimation procedure can as usual be speeded up by concentrating the likelihood with respect to the measurement errors $\mathbf{v}_t$which reduces the number of parameters in the non-linear optimization by $N-3$. Moreover, because (25) (and (28)) is just-identified and the price of risk is absent from the equation spanning the noisy revisions (29) we can write the former as an unrestricted form:

$$\mathbf{X}_t = \mathbf{h} + \mathbf{H}\mathbf{Z}_{t-1} + \mathbf{B}_0\mathbf{e}_t^P,$$

which (using the Zellner (1962) unrelated regressions theorem) can be concentrated out of the likelihood using OLS regression formulae. Given the final estimates of the structural parameters $\mathbf{a}_x$ and $\mathbf{B}_x$ the Duffee price of risk parameters can then be backed out as $\mathbf{\Lambda}_1' = \mathbf{B}_x^{-1}\mathbf{H}$ and $\boldsymbol{\lambda}_0 = \mathbf{B}_x^{-1}(\mathbf{h} - \mathbf{a}_x)$. Concentrating the likelihood in this way reduces the number of parameters in the non-linear optimization by another $K \times (M+1)$.

4 Empirical results

This section discusses the various empirical models used to test the restrictions. We start by fitting the $ATSM$ of the cross-section (23) to a variety of different forward rate data sets, comparing them with unrestricted OLS specifications to check the validity of the Jensen NA restrictions on the intercepts. We then develop a set of return forecasting regressions (25) which are combined with (23) to model the cross-section and time series behavior of the forward curve jointly in the $DTSM$. The combined model (25) and (26) is then fitted to a monthly data set for the period 1983-2011 to test the validity of the AR and $ARFIMA$ restrictions on the slope parameters.

4.1 Testing the no-arbitrage restrictions

The NA condition (15) restricts the factor dynamics used to model the cross-section under the $\mathcal{Q}$ measure, allowing us to test these restrictions without specifying the form of the risk premia (or equivalently the $\mathcal{P}$ dynamics)[10]. To test these we optimize this model using MLE and compare the fit with that of an OLS model with unrestricted coefficients. Because these restrictions are recursive, to do this without restricting the slope coefficients we need a complete set of equally spaced maturities in each observation of the cross-section, with a periodicity equal to that of the time series[11]. Equation (14) then shows that the N_y intercept coefficients (N_y being the number of noisy revisions) are determined by recursion formulae involving the slope coefficients and the covariance matrix. In the case of the 3 factor model, the latter has 6 coefficients, so in principle we save $N_y - 6$ degrees of freedom relative to an unrestricted OLS system.

We test the NA restrictions using data from three different sources. The first is the Gurkaynak, Sack and Wright (GSW, 2007) model which can be used to calculate yields and forward rates at any maturity frequency out to 30 years[12] back to 1971.

[10] This is the appropriate framework for such an exercise because a dynamic model such as (25) and (26) is a test of the joint hypothesis of the restricted NA dynamics under $\mathcal{Q}$ and the model of the risk premium.

[11] The time series observations on this cross-section need not be complete however, as in the case of the tests of the Fama-Bliss dataset reported in table 1.

[12] Conveniently, this gives the parameters of the interpolated forward curve, allowing us to obtain forward rates and revisions on a consistent time period/maturity basis. The parameters are available from from http://www.federalreserve.gov/pubs/feds/2006/200628/feds200628.xls.

To avoid measurement problems at the long end where observations are frequently sparse, we restrict the analysis to maturities up to 15 years. To allow for the possibility of a structural break in 1982[13], we run the test on both the full GSW time span and the period since 1983. In this section we use an annual and in the next section we use a monthly GSW data set.

The second source is the McCulloch and Kwon (McC-K, 1993) data set, which has been extensively used in term structure research and provides a time series of the cross-section back to 1946. This data has been updated by the New York Fed and reports annual maturity yield observations out to 20 years. These models respectively generate interpolated discount yield equivalent data using exponential and polynomial spline functions which have the effect of smoothing them, reducing the dimensionality of the resulting estimates. For example, the GSW estimates are generated by the Nelson-Siegel-Svensson model (Svensson (1994)). This is fitted by a six factor function, which means that once six maturity observations are modelled, additional observations are not strictly independent. Consequently our initial strategy was to test the NA restrictions using annual observations and maturities, to keep N relatively low while spanning a wide range of maturities. Finally we tested the NA restrictions using unsmoothed monthly observations for the short end provided by the well-known Fama-Bliss (FB) Treasury Bill file on CRSP[14].

Table 1 reports the results of testing the NA restrictions. The first two columns show the number of parameters for each model and the next two show the number of observations for each data set and the number of 'noisy' revisions used as dependent variables. The next two columns show the maturities of the 'noiseless revisions' used

---

[13] This is found in time series results for short term interest rates under $\mathcal{P}$, see for example Garcia and Perron (1996), Bai (1997), Bai and Perron (1997).

[14] For a detailed description please see the US Treasury Database Guide at: www.crsp.com/documentation.

as explanatory variables and the frequency of the data. The final column reports the value of the standard Likelihood Ratio statistic $\mathcal{LR}$ =2($\ln \mathcal{L}_{OLS} - \ln \mathcal{L}_{NA}$) for the null hypothesis that the data is arbitrage-free (i.e. that OLS and NA models fit the data equally well), where $\ln \mathcal{L}$ is the value of an optimized log-likelihood function). This statistic has a χ^2 distribution with $(k_{OLS} - k_{NA})$ degrees of freedom with the 95% critical value shown in small font. The $\mathcal{LR}$ test does not reject the null hypothesis for any dataset. The $\mathcal{LR}$ test for the McC-K data set gives the $p-$value of 0.16. For GSW data, the $p-$values (not shown) are equal to 1 for all frequencies and time spans. Using the Fama-Bliss unsmoothed unsmoothed observations at the monthly frequency, gives $p-$values ranging from 0.23 to 0.63.

4.2 Testing the risk neutral dynamic restrictions

Having established that the NA restrictions on the intercepts are acceptable in the cross-section, we develop a $DTSM$ that simultaneously extracts information from the time series behavior of the system. We are primarily concerned at this stage to test the validity of the standard AR specification of the factor dynamics and loadings, as well as the more general $ARFIMA$ model. We refer to these as Restricted Risk Neutral Dynamic or '$RRND$' models.

4.2.1 The monthly forward rate data

In this section we use a monthly data set constructed using the GSW parameters of the interpolated forward curve. To allow for the suspected structural break in the time series behavior in 1982 we use the sample January 1983 to December 2011. These models can be estimated using only a small selection of maturities but because the computation cost is minimal we use the complete set of 180 maturities to avoid any possible loss of information. We define the forward rate revisions as $f_{n,t} - f_{n+1,t-1}$, where n denotes the revision horizon. This gives us $62,420$ monthly observations

(180 revisions $\times$ 347 months).

As in the previous section, we assume the 4, 48 and 168 month revisions (the 'noiseless revisions') are measured without error. These maturities have highest correlations with the first three factors from the Principal Component Analysis of the revisions. Figure 1 plots the corresponding 4 month, 4 year and 14 year forward rates, revealing a clear downward trend. Oscillatory behavior at the business cycle frequency is also apparent, which is especially strong for the 4 month rate. The long, random cycles might be suggesting persistent long memory dynamics of the time series that is different from the short memory behavior displayed by $ARMA$ models. Figure 2 plots periodograms of these forward rates. The sharp spike at frequency zero suggest very persistent behavior. Additionally, we can see another spike for the 4 month rates, which could suggest a long memory behavior at the business cycle frequency.

The time series of the forward rate revisions for these maturities are presented in Figure 3. Figures 6 and 7 show the unconditional mean and volatility (as measured by standard deviation) of forward rates and forward revisions, respectively. The term structure of forward rates is upward sloping for maturities up to about 12 years, reaching a maximum of 7.54% at 148 months, and decreasing slowly thereafter. The term structure of the volatility of forward rates is more interesting; it initially increases with maturity up to 2.889% at 11 months, then decreases to 1.997% at 132 months and then appears to increase. Figure 7 shows the unconditional mean and standard deviation of the term structure of forward revisions. The mean forward revisions are negative and their term structure is hump shaped with a maximum value of -0.021% at 167 months. The unconditional standard deviation of the revisions is also humped shaped with a maximum of 0.376% at 22 months, decreasing slightly to 0.325 at the long end. At the short end of the volatility curve there is a sharp kink at 2 months to maturity.

Figures 8 and 9 plot the correlograms of 4 month, 4 year and 14 years of forward rates and forward rate revisions respectively. It is notable especially given the 'irregular' time series behavior shown in Figure 3, that the forward rate revisions display high persistence at long lags. The correlograms of both forward rates and forward revisions also exhibit oscillatory behavior.

4.2.2 Modeling the price of risk

We began this part of the empirical analysis by estimating monthly forward return forecasting regressions for the noise-free forward revisions (25), designed to identify the variables driving the price of risk. We started with the 4 month, 4 year and 14 year forward rates and then added Fama-Bliss annual yields which have been used successfully for annual discount bond return forecasting by Cochrane and Piazzesi (2005). The work of Ludvigson and Ng (2009) and Joslin et al (2012) also suggested the use of macro variables like industrial production growth (IP) and expected inflation (EI)[15]. These data are presented in Figure 4. The variables that are jointly significant in the final model are: the 3, 47 and 167 month forward rates, Fama-Bliss 2 and 5 year yields, IP and EI. To take account of the large sample size, the significance threshold for this test is determined using the Hendry (1995) rule for decreasing the significance level with sample size[16]. The significance level $\alpha = 0.31\%$ for this sample.

[15] Industrial Production is the monthly logarithmic growth in the seasonally adjusted Industrial Production index and Expected Inflation is a monthly series for the median expected price change over the next 12 months compiled by the, University of Michigan Survey Research Center. Both series are available from the website of Federal Reserve Bank of St. Louis.

[16] Hendry (1995) suggests that when the sample size increases, we should be more strict with the hypothesis test and a use smaller significance level for the critical value. As a rule of thumb, Hendry suggested $cT^{-0.9}$, where $c = 1.6$. This gives the critical values for our sample of $p = 0.31$, which is comparable to the 5% level for $T = 50$. If the value of test exceeds this critical value, the null hypothesis (that the two models have the same fit) is rejected. For model selection, Canova (2007) recommends the use of the Schwartz or Bayesian Information Criterion (BIC) which makes a similar adjustment and is asymptotically consistent.

The macroeconomic time series do not exhibit long range persistence (see Figure 5). Although EI exhibits substantial autocorrelation at short lags (the first order autocorrelation coefficient is 0.82), this autocorrelation decays rapidly and becomes insignificant at 26th lag. The first order autocorrelation coefficient for IP is equal to 0.24, which increases to 0.35 at the third lag and then falls, becoming insignificant for lags larger than 10.

Table 2 reports the estimates of the parameters and their significance. There is no obvious pattern of signs and significance across regressors. Industrial production is very significant in the short revision regression, but insignificant in others. Expected inflation is significant in the long revision regression, and marginally significant in the medium term revision. The Fama-Bliss rates are significant in the short rate revision. As can be seen from the lower panel, all variables are significant at the 0.31% significance level.

4.2.3 The risk-neutral dynamics

We first estimate the unrestricted OLS and no-arbitrage models (the results are available on request). The slope parameters are initially unrestricted, as in the exercise of the previous section, but in this exercise we simultaneously estimate the price of risk and hence the real world factor dynamics as in (25) and (26). Given the estimates of the reduced form coefficients $\mathbf{a}_x$ and $\mathbf{B}_x$ and the risk neutral parameters we can then find the price of risk parameters[17]. We then estimate eight models in which the risk-neutral dynamics are restricted using the $AR(1)$, $AR(2)$, $ARMA(1,1)$, $ARFIMA(1,d,0)$, $ARFIMA(2,d,0)$, $ARFIMA(1,d,1)$, $ARFIMA(2,d,1)$ and $ARFIMA(2,d,2)$ specifications discussed in section 3.2. In these $RRND$ models we parametrize all three factors in the same way, giving three short memory and five long memory

[17] As noted in section 3.4, the price of risk parameters $\boldsymbol{\lambda}_0$ and $\boldsymbol{\Lambda}_1$ are concentrated out from the log-likelihood maximization. These results are not reported here but available upon request.

parameterizations of the model. Table 3 reports the estimates of the risk neutral parameters of these models. The asymptotic standard errors are reported in a small font.

Figure 10 plots the moving average representation of the factor dynamics in these models (B_y in (23)). These are similar to the factor loadings of a discount yield model and show the effect of the three risk neutral shocks inferred from the noiseless revisions (using (22)) on the forward rates (and revisions) across the maturity structure. All three factors exhibit 'dual persistence' under $\mathcal{Q}$: long-lasting short and long memory dynamics. The two highest autoregressive roots are typically very close to 0.99 and the smallest is typically about 0.95. When we include the long memory parameter d, for the more parsimonious models $ARFIMA(1,d,0)$ and $ARFIMA(2,d,0)$ one of the factors has $d > 0.5$ and is not covariance-stationary[18]. For models that introduce an MA term, the estimates of all factors are in the stationary region with the estimates of the highest long memory parameter varying from 0.46 ($ARFIMA(1,d,1)$) and 0.40 ($ARFIMA(2,d,2)$). The two other factors also exhibit long memory behavior with the estimates of the long memory parameter of the least persistent factor about 0.09 to 0.12, and 0.24 to 0.30 for the medium persistence factor. The estimates of the MA parameters lie in the invertibility region for the short memory model $ARMA(1,1)$, but for all long memory models at least one factor has non-invertible MA roots. The largest estimate of the MA coefficient is for the $ARFIMA(1,d,1)$ model: 23.21.

4.2.4 The real world dynamics

Appendix B shows how the model of the risk premium (24) is used to obtain the model of the real world short rate dynamics ((B-5)) from the model of the risk

[18] However, since all $d < 1$, the factors are still mean-reverting.

neutral short rate dynamics (4). Figure 11 shows the effect of shocks backed out from the three noiseless revisions used as factors (the φ's in (B-5)) and is comparable with figure 10 showing the risk neutral dynamics. These effects are depicted in the form of an impulse response function or moving average representation, showing the response over time of the spot rate to real world shocks that are spanned by the yield curve. While the general pattern is similar to that of 10, these responses tend to be smaller and their 'humps' more muted. The impact of the second and third factors is much weaker than it is under $\mathcal{Q}$, particularly in the case of the $ARFIMA(1, d, 0)$ and $ARFIMA(1, d, 1)$ models. Figure 12 shows the effect of the macro and other unspanned factors (the γ's in (B-5)), which affect the risk premia and the evolution of the yield curve but not the current yield curve. These move away from their initial value of zero to a single peak before mean reverting to zero.

4.2.5 Relative model performance

The standard errors on these cross-sectional parameters are very small. There are two reasons for this. The first is that the likelihood of the cross-section depends upon measurement errors (v_t) that are, as Cochrane and Piazzesi say in their (2008) paper, 'tiny' empirically, so that restrictions that generate tiny perturbations in the yield curve estimates are often rejected. The second is the large sample, which tends to bias classical statistics such as likelihood ratio statistics toward rejection of model simplifications (Hendry (1995), Canova (2007)).

Table 4 reports the number of parameters and log-likelihood values for each model. The table shows the negative of the BIC statistic, making it appropriate to select the model with the highest value of $(-)BIC{=}2\ln\mathcal{L}{-}k\ln T$. The table also reports the likelihood ratio statistic for a test against the $ARFIMA(2, d, 2)$ model[19]. The

[19] For the tests in Panel A the $ARFIMA(2, d, 2)$ model is the restricted model, while for the tests in Panel B it is unrestricted.

number in small font under the $\mathcal{LR}$ test value shows the critical value for $p = 0.0011$, which the analysis of Hendry (1995) suggests will give a null-rejection frequency comparable to that of the conventional 95% value ($p = 0.05$) used in a small sample. Panel A reports these statistics for the OLS and NA models,while Panel B reports the statistics for the $RRND$ models.

Panel A of Table 4 shows that the $ARFIMA(2, d, 2)$ model is preferable to the OLS and NA models. Panel B shows that there is a small increase in the log-likelihood value from $73,814$ for the $AR(1)$ model to $73,821$ for $ARMA(1,1)$, but a much larger increase to $74,258$ for the $AR(2)$ model. However, this model is still inferior to the models with long memory dynamics. The long memory models perform much better than the rest in this comparison. The $ARFIMA(2, d, 2)$ is preferred to all the other models. The MA component seems to play a more significant role in these (compare e.g. $ARFIMA(2, d, 0)$ and $ARFIMA(1, d, 1)$) than it does in the short memory models (compare $AR(2)$ and $ARMA(1,1)$). The long memory parameter plays a double role in the autocorrelation function. First, like the MA component, it can amplify the effect of autocorrelation. The value of loading for the most persistent factors is typically well more than 5 at 180th lag (see Figure 10 for the models with long memory). Second, it makes the rate at which the autocorrelations decay hyperbolic, instead of exponential as in the pure autoregressive models.

4.2.6 Cross sectional errors

Table 5 reports the root mean square error of the cross-sectional errors for each model defined as:

$$RMSE = \sqrt{\frac{1}{TN} \sum_{n=1}^{N} \sum_{t=1}^{T} \left(\widehat{y}_{nt} - y_{nt}\right)^2}, \tag{30}$$

where y_{nt} denotes the revision with maturity n observed at time t and $\widehat{y}_{nt}$ is the fitted value, conditional upon the contemporaneous noiseless revisions. This shows that all models fit the cross section very well, with an average measurement error of just over 10 basis points. The richer models are more flexible, therefore they offer smaller $RMSE$[20]. The improvement is not dramatic at the first sight, but if we assume that the OLS model gives the best possible fit and use this as the baseline then going from the $AR(1)$ model to $ARFIMA(2,d,2)$ we improve the fit[21] by nearly 42%.

4.3 The maturity structure of the unconditional means and coefficients

In this section we look in more detail at the results for four representative specifications: the OLS, NA, $AR(2)$ and $ARFIMA(2,d,2)$ models and the way they capture the behavior of the yield curve at different maturities. First we look at the unconditional means of the noisy forward rate revisions implied by these models and compare these with the sample means. Then we compare the intercept and slope coefficients of the restricted models of the noisy revisions with those of the unrestricted OLS model.

4.3.1 Comparing the model-implied and sample means across maturity

Figure 13 plots the maturity structure of the unconditional mean implied by each of these models for the forward revisions and compares them with the sample mean (shown by the dotted curve). The model means are calculated (using (23)) as:

$$\mathbb{E}\left[\mathbf{Y}_t\right] = \left(\mathbf{a}_y - \mathbf{B}_y\mathbf{B}_x^{-1}\mathbf{a}_x\right) + \mathbf{B}_y\mathbf{B}_x^{-1}\mathbb{E}\left[\mathbf{X}_t\right], \tag{31}$$

[20] Maximising likelihood is not equivalent to minimizing the $RMSE$ of the measurement error. Indeed, we can see that the $ARMA(1,1)$ model has a higher rmse than $AR(1)$, although it has three more parameters .

[21] Calculated as $\frac{total\ rmse_{AR(1)} - total\ rmse_{ARFIMA(2,d,2)}}{total\ rmse_{AR(1)} - total\ rmse_{OLS}} \times 100\%$.

where $\mathbb{E}[\mathbf{X}_t]$ is the vector of sample means. For the noiseless revisions $\mathbf{X}_t$, these coincide with the sample mean, so the curves converge at these maturities (4, 48 and 168 months). The unconditional means implied by these models are very close to those observed in the sample, except for the short maturities. In the case of the $AR(2)$ model the short-end values are much higher than those seen in the data.

4.3.2 Comparing the parameters of the restricted and unrestricted models across maturity

Equation (31) involves both the model intercepts $\left(\mathbf{a}_y - \mathbf{B}_y\mathbf{B}_x^{-1}\mathbf{a}_x\right)$ and the factor loadings or slope coefficients $B_yB_x^{-1}$ in (23). Their separate effects can be seen by comparing the coefficients of the restricted models with those of the unrestricted OLS model, which gives a very good approximation to the data. Figure 14 compares the intercepts $\left(\mathbf{a}_y - \mathbf{B}_y\mathbf{B}_x^{-1}\mathbf{a}_x\right)$ of the restricted models with those of the OLS model (shown by the dotted curve). The NA model replicates the OLS intercept almost perfectly. The intercepts of the $ARFIMA(2,d,2)$ model exhibit a pronounced 'hump' for maturities up to 48 months, but lie within the 95% confidence interval of the OLS model for almost all maturities. However the intercept of the $AR(2)$ model again goes completely astray at the short end of the curve. For example, the OLS intercept is -5.20 bp for the 1 month revision , while the $AR(2)$ intercept is 6.02 bp, a difference of more than 8 OLS standard errors.

Figure 15 presents the loadings $B_yB_x^{-1}$on the three noiseless revisions used to back out the factors. The dots again show the loadings of the unrestricted OLS model. The restricted models have similar loadings. Unsurprisingly the NA model (which has unrestricted slope coefficients) fits better than the $RRND$ models. One difference is clear in the loading of factor 1 shown in the first panel of Figure 15. This shows that the $AR(2)$ and $ARFIMA(2,d,2)$ models overlap each other and have a much shorter hump than the OLS model. The other difference is in the loadings on the third factor for 6 to 42 month maturities, where the loadings for the OLS model are relatively flat but exhibit a 'valley' shape for the $RRND$ models . These long memory models seem to be particularly good at improving the way that the models replicate the behavior of the short maturities.

5 Conclusion

Economists and finance professionals have divided into two camps when modelling the behavior of the term structure of interest rates. On the one hand practitioners tend to follow the so called 'no-arbitrage' approach and use an $ATSM$ to analyze current market prices of fixed income securities and to back-engineer the unrestricted risk-neutral dynamics of the underlying factors and their volatilities. On the other hand economists and finance academics tend to follow the so called 'equilibrium approach' by adopting specific $RRND$ models to describe the dynamic behavior of interest rates and other factors, using a $DSTM$ to model these under both the risk neutral and historical probability measures. These models normally assume the absence of arbitrage, but recently some researchers have compared their performance with unrestricted linear regressions (Hamilton and Wu (2010), Cochrane and Piazzesi (2008) for example).

We show that this divide is somewhat artificial and can be bridged by using the Wold representation to provide a discrete time model-free specification of the risk neutral factor dynamics in the forward rate domain. We start by estimating a Meiselman (1962) OLS model of the revisions to forward rates, without adopting any specific model of the factor dynamics. We modify the standard yield factor approach to develop an $ATSM$ that shows how the innovations implied by (three) revisions that are assumed to be measured without error. We then impose the no-arbitrage restriction on the coefficients, to obtain the analogue of the HJM model in discrete time, but avoiding the restrictions imposed by their continuous time volatility framework. We test the no-arbitrage assumption by comparing this with the unrestricted OLS system and find that this is acceptable across a range of data sets.

We then develop a $DSTM$ by augmenting this model of the cross-section with a Cochrane and Piazzesi (2005) type return forecasting model which defines the price of risk and allows us to model the real world as well as the risk neutral dynamics. This allows the model parameters to be informed by the information contained in the cross section of forward rates and their times series behavior. The change of measure can allow for complete flexibility in the specification of the dynamics under the two measures, but it assumes that the variance structure does not change. The MLE procedure seeks a compromise between the values of the volatility parameters that optimize the fit of the cross section and the values that optimize the dynamic performance of the model, reducing the likelihood compared to an unrestricted OLS model. Nevertheless, the Bayesian Information Criterion, which corrects for the number of parameters and observations, does favour the NA version of the $DTSM$ over the OLS model.

Finally, we use this $DTSM$ to examine different $RRND$ models for the factors under the risk-neutral measure. Since the great majority of term structure models adopt an autoregressive specification, we wanted to check the effect of allowing for richer dynamics, which is very easy to do in this framework. We examined 8 models from the $ARFIMA$ class, allowing for autoregressive, moving average and long memory effects. We find that the model with traditional $AR(1)$ dynamics is outclassed by more flexible models. The model with $AR(2)$ dynamics offers much better performance, but further improvement is achieved using the more flexible $ARFIMA(2,d,1)$ and $ARFIMA(2,d,2)$. models. These long memory models seem to be particularly good at improving the way that the models replicate the behavior of the short maturities.

We are currently extending this research in several directions, shadowing developments in the standard yield curve literature. One is to relax the yield factor assumptions by using latent factors modelled as in the Kalman filter, perhaps distinguishing real and nominal discount factors. Another is to check the spanning assumptions of our $DTSM$, to consider for example whether macro variables affect the forward yield curve directly rather than indirectly through the risk premia (Joslin et al (2012). A related area of research is based on the examination of the form of the risk premium, which is of a restrictive exponential-affine variety in our $DTSM$[22]. Like the continuous time HJM model, this framework can also be used to examine forward and hazard rates in corporate bond prices[23]. Last but not least, it is worth noting that this model framework can be readily adapted to admit square root processes (Cox et al (1985), Sun (1992), Gourieroux and Jasiak (2006)), while preserving the linear-recursive model parameter structure.

References

[1] Backus D., Foresi S. and Zin S., 1998, "Arbitrage Opportunities in Arbitrage-Free Models of Bond Pricing", Journal of Business & Economic Statistics, vol. 16, no 1, 13-26.

[2] Backus D.K. and Zin S.E., 1993, "Long Memory Inflation Uncertainty: Evidence from the Term Structure of Interest Rates", Journal of Money, Credit and Banking, vol. 25, no 3, 681-700.

[3] Bai J., 1997, "Estimation of a Change Point in Multiple Regression Models",

[22] Since Duffee (2002) some contribution in this field has been made (e.g. Duarte (2004), Le, Singleton and Dai (2010)).

[23] (Appendix C of Duffie and Singleton (2003) provides an excellent summary of the corporate HJM literature

Review of Economics and Statistics, vol. 79, no 4, 551-563.

[4] Bai J., Perron P., 2003a, "Computation and analysis of multiple structural change models", Journal of Applied Econometrics, vol. 18, 1-22.

[5] Garcia R. and Perron P., 1996, "An Analysis of the Real Interest Rate under Regime Shifts", The Review of Economics and Statistics, vol. 78, no 1, 111-125.

[6] Campbell J.Y., Lo A.W. and MacKinlay A.C., 1997, "The Econometrics of Financial Markets", Princeton University Press.

[7] Canova F., 2007, "Methods for Applied Macroeconomic Research", Princeton University Press.

[8] Cheridito P., 2002, "Arbitrage in Fractional Brownian Motion Models", Finance and Stochastics, vol. 7, no 4, 533-553.

[9] Chen R. and Scott L., 1993, "Maximum Likelihood Estimation for a Multi-Factor Equilibrium Model of the Term Structure of Interest Rates, Journal of Fixed Income, vol. 4, 14-31.

[10] Cochrane J.H. and Piazzesi M., 2005, "Bond Risk Premia", American Economic Review, vol. 94, no 1, 138-160.

[11] Cochrane J.H. and Piazzesi M., 2008, "Decomposing the Yield Curve", Unpublished working paper.

[12] Comte F. and Renault E., 1996, "Long Memory Continuous Time Models", Journal of Econometrics, vol. 73, no 1, 101-149.

[13] Connolly R. and Guner N., 1999, "Long Memory Characteristics of the Distributions of Treasury Security Yields, Returns and Volatility", Unpublished working paper, Chapel Hill, University of North Carolina.

[14] Cox J.C., Ingersoll J.E. and Ross S.A., 1985, "A Theory of the Term Structure of Interest Rates", Econometrica, vol. 53, 385-467.

[15] Duan J.C. and Jacobs K., 1996, "A Simple Long Memory Equilibrium Interest

Rate Model", Economics Letters, vol. 53, 317-321.

[16] Duarte J., 2004, "Evaluating an Alternative Risk Preference in Affine Term Structure Models", Review of Financial Studies, vol. 17, no 2, 379-404.

[17] Duffee G.R., 2002, "Term Premia and Interest Rate Forecasts in Affine Models", Journal of Finance, vol. 57, no 1, 405-443.

[18] Duffee G.R., 2011, "Forecasting with the Term Structure: The Role of No-Arbitrage Restrictions", Unpublished working paper.

[19] Duffee G.R., 2012, “Bond Pricing and the Macroeconomy,” prepared for the Handbook of the Economics of Finance.

[20] Duffie D. and Singleton K.J., 2003, "Credit Risk: Pricing, Measurement, and Management", Princeton University Press, Princeton, New Jersey.

[21] Dybvig P.H., 1997, "Bond and Bond Option Pricing Based on the Current Term Structure", in Mathematics of Derivative Securities, Michael A. H. Dempster and Stanley Pliska, eds., Cambridge University Press.

[22] Duffie D. and Kan R., 1996, "A Yield-Factor Model of Interest Rates", Mathematical Finance, vol. 6, no 4, 379-406.

[23] Gil-Alana L.A., 2004a, "Long Memory in the U.S. Interest Rate", International Review of Financial Analysis, vol. 13, no 3, 265-276.

[24] Gil-Alana L.A., 2004b, A Joint Test of Fractional Integration and Structural Breaks at a Known Period of Time," Journal of Time Series Analysis, Wiley Blackwell, vol. 25(5), pages 691-700, 09. , "

[25] Golinski A. and Zaffaroni P., 2012, "Long Memory Affine Term Structure Models", Unpublished working paper.

[26] Gourieroux C. and Jasiak J., 2006, "Autoregressive gamma processes," Journal of Forecasting, vol. 25, no 2, 129-152.

[27] Granger C.W.J. and Joyeux R., 1980, "An Introduction to Long Memory Time

Series Models and Fractional Differencing", Journal of Time Series Analysis, vol. 1, no 1, 15-29.

[28] Gurkaynak R.S., Sack B. and Wright J.H., 2007, "The U.S. Treasury Yield Curve: 1961 to the Present", Journal of Monetary Economics, vol. 54, no 8, 2291-2304.

[29] Hamilton J.D., 1994, "Time Series Analysis", Princeton University Press.

[30] Heath D., Jarrow R. and Morton A., 1992, "Bond Pricing and the Term Structure of Interest Rates", Journal of Financial and Quantitative Analysis, vol. 25, 419-440.

[31] Heath D., Jarrow R. and Morton A., 1990, "Bond Pricing and the Term Structure of Interest Rates: A Discrete Time Approximation", Journal of Financial and Quantitative Analysis, vol. 25, no 4, 419-440.

[32] Hendry D.F., 1995, "Dynamic Econometrics", Oxford University Press, Oxford.

[33] Hosking J.R.M., 1981, "Fractional Differencing", Biometrica, vol. 68, 165-176.

[34] Iacone F., 2009, "A Semiparametric Analysis of the Term Structure of the U.S. Interest Rates", 2009, Oxford Bulletin of Economics and Statistics, vol. 71, no 4, 475-490.

[35] Jamshidian F., 1997, "LIBOR and Swap Market Models and Measures", Finance and Stochastics, vol. 1, no 4, 293-330.

[36] Joslin S., Priebsch M. and Singleton K.J., 2010, "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks", Unpublished working paper.

[37] Le A., Singleton K.J. and Dai Q., 2010, "Discrete-Time Affine$\mathbb{Q}$ Term Structure Models with Generalized Market Prices of Risk", Review of Financial Studies, vol. 23, no 5, 2184-2227.

[38] Lochoff R. ,1993, “The Contingent Claims Arms Race”, Journal of Portfolio

Management, vol. 20, no 1, 88-92.

[39] Ludvigson S.C. and Ng S., 2009, "Macro Factors in Bond Risk Premia", Review of Financial Studies, vol. 22, no 12, 5027-5067.

[40] McCulloch, H., and Kwon H., 1993, "U. S. Term Structure Data, 1947-1991", Unpublished working paper 93-6, Ohio State University.

[41] Meiselman D., 1962, "The Term Structure of Interest Rates", Prentice Hall, New Jersey.

[42] Osterrieder D. and Schotman P., 2012, "The Volatility of Long-Term Bond Returns: Persistent Interest Shocks and Time-Varying Risk Premiums", Unpublished working paper.

[43] Rogers, L.C.G., 1997, "Arbitrage with fractional Brownian motion", Mathematical Finance, vol. 7, 95-105.

[44] Shea G.S., 1991, "Uncertainty and Implied Variance Bounds in Long Memory Models of the Interest Rate Term Structure", Empirical Economics, vol. 16, no 3, 287-312.

[45] Sun, T., 1992, "Real and Nominal Interest Rates: A Discrete-Time Model and Its Continuous Time Limit," Review of Financial Studies, vol. 5, 581–611.

[46] Svensson L., 1994, "Estimating and Interpreting Forward Interest Rates: Sweden 1992–1994", IMF working paper 94/114.

[47] Tkacz G., 2001, "Estimating the Fractional Order of Integration of Interest Rates Using a Wavelet OLS Estimator", Studies in Nonlinear Dynamics & Econometrics, vol. 5, no 1, 1068-1068.

[48] Vasicek, O.A., 1977, "An Equilibrium Characterization of the Term Structure", Journal of Financial Economics, vol. 5, 177-188.

[49] Zellner A., 1962, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", Journal of the American Statistical

Association, vol. 57, no 298, 348-368

Appendices

Appendix A The change of measure

The change of measure from $\mathcal{Q}$ to $\mathcal{P}$ depends on the Radon-Nikodym derivative or $\frac{d\mathcal{P}}{d\mathcal{Q}}(\mathbf{X}_t|\mathbf{X}_{t-1})$ which transforms the conditional densities multiplicatively[24]:

$$f^{\mathcal{P}}(\mathbf{X}_t|\mathbf{X}_{t-1}) = \frac{d\mathcal{P}}{d\mathcal{Q}}(\mathbf{X}_t|\mathbf{X}_{t-1}) \times f(\mathbf{X}_t|\mathbf{X}_{t-1}). \tag{A-1}$$

where f and $f^{\mathcal{P}}$ are densities under $\mathcal{Q}$ and $\mathcal{P}$ respectively conditional upon the information set $\mathbf{X}_{t-1}$, where $\mathbf{X}_{t-1}$ is the relevant information set. We assume that this multiplier (and hence the distribution of $\mathbf{e}_t^{\mathcal{P}}$ under $\mathcal{P}$) is exponential affine:

$$\frac{d\mathcal{P}}{d\mathcal{Q}}(\mathbf{X}_t|\mathbf{X}_{t-1}) = \exp[-\frac{1}{2}\boldsymbol{\lambda}'_{t-1}\boldsymbol{\lambda}_{t-1} - \boldsymbol{\lambda}'_{t-1}\boldsymbol{\Sigma}'^{-1}\mathbf{e}_t^{\mathcal{P}}] \tag{A-2}$$

where $\boldsymbol{\lambda}_{t-1}$ is the price of risk vector[25]. Given (5), the conditional density of the factor shocks under the risk neutral measure $\mathcal{Q}$ is:

$$f(\mathbf{e}_t) = (2\pi)^{-\frac{3}{2}} |\boldsymbol{\Omega}|^{-\frac{1}{2}} \exp[-\frac{1}{2}\mathbf{e}_t'\boldsymbol{\Omega}^{-1}\mathbf{e}_t].$$

Substituting these two relationships and (16) into (A-1) and consolidating the exponents gives the density of the factor shocks under P conditional upon the information

[24] Although there are two kinds of innovations in our model, shocks to the factors and measurement errors, these are independent and serially uncorrelated. Consequently the latter are not priced and do not affect this transform.

[25] In other words, the Stochastic Discount Function (SDF) $\frac{dP}{dQ}(\mathbf{X}_{t+1}|\mathbf{X}_t)e^{-r_t}$is specified as.$\exp[-r_t - \frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t - \boldsymbol{\lambda}_t'\mathbf{P}'^{-1}\mathbf{e}_{t+t}^{P}]$.

set $\mathbf{X}_{t-1}$:

$$f^{\mathcal{P}}(\mathbf{e}_t^{\mathcal{P}}) = (2\pi)^{-\frac{3}{2}} |\boldsymbol{\Omega}|^{-\frac{1}{2}} \exp[-\frac{1}{2}\left(\mathbf{e}_t^{\mathcal{P}} - \boldsymbol{\Sigma}'\boldsymbol{\lambda}_{t-1}\right)' \boldsymbol{\Omega}^{-1}\left(\mathbf{e}_t^{\mathcal{P}} - \boldsymbol{\Sigma}'\boldsymbol{\lambda}_{t-1}\right) + \boldsymbol{\lambda}'_{t-1}\boldsymbol{\Sigma}'^{-1}\mathbf{e}_t^{\mathcal{P}} + \frac{1}{2}\boldsymbol{\lambda}'_{t-1}\boldsymbol{\lambda}_{t-1}]$$
$$= (2\pi)^{-\frac{3}{2}} |\boldsymbol{\Omega}|^{-\frac{1}{2}} \exp[-\frac{1}{2}\mathbf{e}_t^{\mathcal{P}\prime}\boldsymbol{\Omega}^{-1}\mathbf{e}_t^{\mathcal{P}}].$$

which satisfies $\mathcal{E}_{t-1}[\mathbf{e}_t^{\mathcal{P}}] = \mathbf{0}$ as reported in (16).

Appendix B Derivation of the dynamics under the $\mathcal{P}$ measure

This appendix shows how the model of the risk premium (24) is used to obtain the model of the real world short rate dynamics ((B-5), below) from the model of the risk neutral short rate dynamics (4). We focus on the relationship betwen the Wold coefficients under the two measures and ignore intercept terms by assuming for expositional simplicity that r_0 and $\boldsymbol{\lambda}_0$ are zero. The risk neutral dynamics of the vector of forward rates ($\mathbf{f}_t$) with maturities $\{a, b, c\}$ that span the term structure follow directly from (4):

$$\mathbf{f_t} = \mathbf{f}_0 + \sum_{j=0}^{t-1} \mathbf{B}'_j \mathbf{u}_{t-j}, \tag{B-1}$$

where $\mathbf{B}_0 = \mathbf{B}_x$ is defined in (21) and

$$\mathbf{B}'_j = \begin{bmatrix} \beta_{1,a+j} & \beta_{2,a+j} & \beta_{3,a+j} \\ \beta_{1b+j} & \beta_{2,b+j} & \beta_{3,b+j} \\ \beta_{1,c+j} & \beta_{2,c+j} & \beta_{3,c+j} \end{bmatrix}; \quad j \geq 1.$$

For expositional simplicity it is assumed that $\mathbf{f}_0 = \mathbf{0}$.

Joslin et al (2012) note that the dynamics under P can be richer than under Q in the sense that they can be driven by macroeconomic and other unspanned factors ($\mathbf{m}_t$) as well as the factors that span the term structure ($\mathbf{f}_t$). In this paper we model this using an *AR OLS* regression specification for these additional variables[26]

$$\mathbf{m}_t = \boldsymbol{\Theta}\mathbf{f}_{t-1} + \boldsymbol{\Phi}\mathbf{m}_{t-1} + \mathbf{w}_t \quad t \geq 1. \tag{B-2}$$

(with zero intercepts for mean-adjusted data). Substituting recursively for $\mathbf{m}_t$ yields:

$$\begin{aligned}\mathbf{m}_t &= \sum_{j=0}^{t-1} \Phi^j (\Theta \mathbf{f}_{t-1-j} + \mathbf{w}_{t-j}) \\ &= \sum_{j=0}^{t-2} \Phi^j \Theta \mathbf{f}_{t-1-j} + \sum_{j=0}^{t-1} \Phi^j \mathbf{w}_{t-j}. \quad t \geq 2.\end{aligned}$$

(where the first summation is truncated because $\mathbf{f}_0 = 0$). Substituting (B-1) we get:

$$\begin{aligned}\mathbf{m}_t &= \sum_{j=0}^{t-1} \Phi^j \mathbf{w}_{t-j} + \sum_{j=0}^{t-2} \Phi^j \Theta \sum_{i=1}^{t-1-j} \mathbf{B}'_{i-1} \mathbf{u}_{t-j-i} \\ &= \sum_{j=0}^{t-1} \Phi^j \mathbf{w}_{t-j} + \sum_{j=1}^{t-1} \left(\sum_{i=0}^{j-1} \Phi^j \Theta \mathbf{B}'_{j-i-1} \right) \mathbf{u}_{t-j} \\ &= \sum_{j=0}^{t-1} \Phi^j \mathbf{w}_{t-j} + \sum_{j=1}^{t-1} \mathbf{A}'_j \mathbf{u}_{t-j} \quad \text{where: } \mathbf{A}'_j = \sum_{i=0}^{j-1} \Phi^i \Theta \mathbf{B}'_{j-i-1}\end{aligned} \tag{B-3}$$

Writing the vector $\mathbf{Z}'_t = (\mathbf{f}'_t, \mathbf{m}'_t)$, and assuming $\boldsymbol{\lambda}_0 = \mathbf{0}$ the price of risk (24) can be written as:

$$\mathbf{P}'\boldsymbol{\lambda}_t = \boldsymbol{\Lambda}'_{1f}\mathbf{f}_t + \boldsymbol{\Lambda}'_{1m}\mathbf{m}_t, \tag{B-4}$$

which together with (B-3) implies that the dynamics of the factors under the P measure are of the form:

$$r_t = \sum_{i=0}^{t-1} \boldsymbol{\varphi}'_i \mathbf{u}^P_{t-i} + \sum_{i=1}^{t-1} \boldsymbol{\gamma}_i \mathbf{w}_{t-i}. \tag{B-5}$$

[26] This is not reported but available upon request.

where

$$\mathbf{u}_t^P = \mathbf{u}_t - \mathbf{P}'\boldsymbol{\lambda}_{t-1}. \qquad \text{(B-6)}$$

Sucessively substituting (B-6), (B-4), (B-3) and (B-1) back into (B-5):

$$\begin{aligned} r_t &= \sum_{i=0}^{t-1} \boldsymbol{\varphi}_i'(\mathbf{u}_{t-i} - (\boldsymbol{\Lambda}_{11}'\mathbf{f}_{t-1-i} + \boldsymbol{\Lambda}_{12}'\mathbf{m}_{t-1-i})) + \sum_{i=1}^{t-1} \boldsymbol{\gamma}_i \mathbf{w}_{t-i} \\ &= \sum_{i=0}^{t-1} \boldsymbol{\varphi}_i'\mathbf{u}_{t-i} + \sum_{i=1}^{t-1} \boldsymbol{\gamma}_i \mathbf{w}_{t-i} - \sum_{i=0}^{t-2} \boldsymbol{\varphi}_i'\{\boldsymbol{\Lambda}_{11}' \sum_{j=1}^{t-1-i} \mathbf{B}_{j-1}' e_{t-i-j} \\ &\quad + \boldsymbol{\Lambda}_{12}' \sum_{j=2}^{t-1-i} \mathbf{A}_{j-1}'\mathbf{u}_{t-i-j} + \sum_{j=1}^{t-1-i} \Phi^{j-1}\mathbf{w}_{t-i-j}])\} \end{aligned}$$

Equating this with (4) gives a set of recursive restrictions across the coefficients of the forward rates spanning the term structure, which determines the φ_is in terms of their shorter maturity values and the P-coefficients:

$$\begin{aligned} \sum_{i=0}^{t-1} \boldsymbol{\varphi}_i'\mathbf{u}_{t-i} &= \sum_{i=1}^{t-1} \boldsymbol{\beta}_i \mathbf{u}_{t-i} + \sum_{i=0}^{t-2} \boldsymbol{\varphi}_i'\boldsymbol{\Lambda}_{11}'\mathbf{B}_0' e_{t-i-1} + \sum_{i=0}^{t-2} \boldsymbol{\varphi}_i' \sum_{j=2}^{t-1-i} [\boldsymbol{\Lambda}_{11}'\mathbf{B}_{j-1}' + \boldsymbol{\Lambda}_{12}'\mathbf{A}_{j-1}']\mathbf{u}_{t-i-j} \\ &= \boldsymbol{\beta}_0\mathbf{u}_t + (\boldsymbol{\beta}_1 - \boldsymbol{\varphi}_0\boldsymbol{\Lambda}_{11}'\mathbf{B}_0')\mathbf{u}_{t-1} \\ &\quad - \sum_{i=2}^{t-1} (\boldsymbol{\beta}_i + \sum_{j=0}^{i-2} \boldsymbol{\varphi}_j'[\boldsymbol{\Lambda}_{11}'\mathbf{B}_{i-j-1}' + \boldsymbol{\Lambda}_{12}'\mathbf{A}_{i-j-1}'] + \boldsymbol{\varphi}_i'\boldsymbol{\Lambda}_{11}'\mathbf{B}_0')\mathbf{u}_{t-i} \end{aligned}$$

Hence:

$$\boldsymbol{\varphi}_0 = \boldsymbol{\beta}_0;$$

$$\boldsymbol{\varphi}_1 = \boldsymbol{\beta}_1 + \boldsymbol{\varphi}_0\boldsymbol{\Lambda}_{11}'\mathbf{B}_0';$$

$$\boldsymbol{\varphi}_i' = (\boldsymbol{\beta}_i + \sum_{j=0}^{i-2} \boldsymbol{\varphi}_j'[\boldsymbol{\Lambda}_{11}'\mathbf{B}_{i-j-1}' + \boldsymbol{\Lambda}_{12}'\mathbf{A}_{i-j-1}'] + \boldsymbol{\varphi}_{i-1}'\boldsymbol{\Lambda}_{11}'\mathbf{B}_0') \text{ where: } \mathbf{A}_i' = \sum_{j=0}^{i-1} \Phi^j \Theta \mathbf{B}_{i-j-1}' \quad i \geq 2,$$

This leaves a set of Duffee (2012) style restrictions across the effects of macro & other unspanned variables on the term structure:

$$\sum_{i=0}^{t-1} \boldsymbol{\gamma}_i \mathbf{w}_{t-i} = \sum_{i=0}^{t-2} \boldsymbol{\varphi}_i' \sum_{j=1}^{t-1-i} \boldsymbol{\Lambda}_{12}' \Phi^{j-1} \mathbf{w}_{t-i-j}$$
$$= \sum_{i=1}^{t-1} \left(\sum_{j=0}^{i-1} \boldsymbol{\varphi}_j' \boldsymbol{\Lambda}_{12}' \Phi^{i-j-1} \right) \mathbf{w}_{t-i}$$

Hence:

$$\boldsymbol{\gamma}_0 = 0$$
$$\boldsymbol{\gamma}_i = \left(\sum_{j=0}^{i-1} \boldsymbol{\varphi}_j' \boldsymbol{\Lambda}_{12}' \Phi^{i-j-1} \right), \quad i \geq 1.$$

Tables

Number of parameters (k)		Number of observations		Maturities of regressors	Frequency of data and maturities	$\mathcal{LR}$ $c_{5\%}$
OLS	NA	T	N_y			
Panel A: Annual GSW, 0-14y, 1971-2011						
55	50	40	11	4,48,168	monthly	0.043 11.071
Panel B: Annual GSW, 0-14y, 1983-2011						
55	50	28	11	4,48,168	monthly	0.098 11.071
Panel C: Annual McC-K, 0-19y, 1946-2010						
80	70	64	16	3,10,18	annual	14.183 18.307
Panel D: Monthly Fama, 0-11m, 01/1964-03/2000						
40	38	434	8	3,6,12	monthly	0.720 5.992
Panel E: Monthly Fama, 0-11m, 01/1964-03/2000 & 10/2008-12/2011						
40	38	472	8	3,6,12	monthly	0.843 5.992
Panel F: Monthly Fama, 0-11m, 01/1983-03/2000						
40	38	206	8	3,6,12	monthly	4.606 5.992
Panel G: Monthly Fama, 0-11m, 01/1983-03/2000 & 10/2008-12/2011						
40	38	244	8	3,6,12	monthly	2.962 5.992

Table 1: Comparison of the unrestricted (OLS) and No Arbitrage (NA) ATSMsThe columns show: the number of parameters for the OLS and NA models; the number of observations and dependent variables; the maturities of the 'noiseless revisions' used as regressors; the frequency of the data and the Likelihood Ratio statistic and (in small typeface) the p-value for the NA model against the OLS model. Panels A to G present results for different datasets: Gurkaynak-Sack-Wright, McCulloch-Kwan, and Fama-Bliss, reporting frequency, forward rate span and time span. The NA restriction are accepted in all these data sets.

Regressand	$\mathbf{b}_{0,0}$	IP_t	EI_t	$y^{FB}_{24,t}$	$y^{FB}_{60,t}$	$f_{3,t}$	$f_{47,t}$	$f_{167,t}$	R^2
					Panel A				
$f_{3,t+1} - f_{4,t}$	-0.122	0.070	0.044	0.870	-0.903	-0.284	0.177	0.115	0.141
	0.115	0.025	0.028	0.185	0.278	0.069	0.147	0.037	
$f_{47,t+1} - f_{48,t}$	0.086	-0.046	-0.066	0.436	-0.371	-0.127	-0.039	0.112	0.052
	0.140	0.031	0.034	0.225	0.339	0.084	0.179	0.045	
$f_{167,t+1} - f_{168,t}$	0.485	-0.001	-0.078	0.206	-0.604	0.043	0.476	-0.146	0.068
	0.128	0.028	0.031	0.205	0.309	0.077	0.163	0.041	
Joint significance test					Panel B				
$F(3, 336)$	7.357	8.124	4.834	7.545	5.183	6.688	7.244	18.564	
crit.value	4.715	4.715	4.715	4.715	4.715	4.715	4.715	4.715	

Table 2: Forward return forecasting regressions (24). These regress the noisy revisions on the noiseless revisions and determine the price of risk. The standard errors of the parameters are reported in small font. In Panel B the joint significant test is reported with the critical value at 0.31% significance level. The sample is January 1983 to January 2012.

	ϕ_1	ϕ_2	d	θ_1	θ_2	σ_{ii}	Correlations f_1	f_2	f_3
Panel A: $AR(1)$									
factor 1	0.995					4.121	1		
	0.000					0.029			
factor 2	0.993					3.417	−0.950	1	
	0.000					0.031	0.001		
factor 3	0.955					0.342	0.214	−0.337	1
	0.001					0.002	0.016	0.022	
Panel B: $AR(2)$									
factor 1	1.982	−0.983				0.006	1		
	0.000	0.000				0.000			
factor 2	1.465	−0.469				0.484	0.254	1	
	0.000	0.000				0.000	0.000		
factor 3	1.444	−0.469				0.209	0.190	−0.607	1
	0.000	0.000				0.002	0.022	0.003	
Panel C: $ARMA(1,1)$									
factor 1	0.994			0.178		4.875	1		
	0.000			0.018		0.076			
factor 2	0.993			−0.139		5.898	−0.977	1	
	0.000			0.018		0.128	0.000		
factor 3	0.955			0.479		0.228	0.233	−0.317	1
	0.000			0.256		0.041	0.008	0.009	
Panel D: $ARFIMA(1,d,0)$									
factor 1	0.989		0.518			0.105	1		
	0.000		0.001			0.000			
factor 2	0.991		0.249			0.430	−0.809	1	
	0.000		0.001			0.001	0.002		
factor 3	0.945		0.119			0.464	0.563	−0.827	1
	0.000		0.002			0.004	0.015	0.018	
Panel E: $ARFIMA(2,d,0)$									
factor 1	0.999	−0.010	0.508			0.113	1		
	0.000	0.000	0.001			0.000			
factor 2	0.994	−0.003	0.243			0.454	−0.734	1	
	0.000	0.000	0.001			0.000	0.002		
factor 3	1.030	−0.077	0.102			0.461	0.478	−0.840	1
	0.000	0.000	0.000			0.004	0.014	0.008	
Panel F: $ARFIMA(1,d,1)$									
factor 1	0.989		0.458	0.855		0.124	1		
	0.000		0.001	0.031		0.000			
factor 2	0.999		0.288	0.697		0.324	−0.877	1	
	0.000		0.001	0.008		0.001	0.003		
factor 3	0.951		0.091	23.210		0.021	0.575	−0.815	1
	0.000		0.001	0.317		0.000	0.026	0.051	
Panel G: $ARFIMA(2,d,1)$									
factor 1	0.989	−0.001	0.411	2.382		0.130	1		
	0.000	0.000	0.000	0.012		0.000			
factor 2	0.990	0.001	0.298	2.025		0.257	−0.953	1	
	0.000	0.001	0.000	0.006		0.000	0.000		
factor 3	1.161	−0.201	0.107	0.016		0.431	0.607	−0.770	1
	0.000	0.000	0.001	0.003		0.003	0.008	0.007	
Panel H: $ARFIMA(2,d,2)$									
factor 1	0.996	−0.007	0.396	1.097	−0.357	0.216	1		
	0.000	0.000	0.000	0.008	0.006	0.000			
factor 2	0.994	−0.003	0.252	1.844	−0.179	0.288	−0.929	1	
	0.000	0.000	0.000	0.006	0.004	0.000	0.001		
factor 3	1.180	−0.220	0.118	3.218	−0.415	0.119	0.533	−0.755	1
	0.000	0.000	0.001	0.048	0.010	0.001	0.011	0.008	

Table 3: The risk neutral dynamic parameters of the RRND models. The standard errors are reported in small font. The sample period is January 1983 to January 2011.

Model	k	$ln\mathcal{L}$	$(-)BIC$	$\mathcal{LR}$ c_α
Panel A				
OLS model	918	$76,031$	$141,926$	$2,646.3$ 2874.3
NA model	741	$75,544$	$142,906$	$1,671.66$ 1853.6
Panel B				
$AR(1)$	213	$73,814$	$145,277$	$1,788.0$ 39.811
$AR(2)$	216	$74,258$	$146,131$	900.2 34.358
$ARMA(1,1)$	216	$73,821$	$145,258$	$1,774.2$ 34.358
$ARFIMA(1,d,0)$	216	$74,539$	$146,694$	337.4 34.358
$ARFIMA(2,d,0)$	219	$74,590$	$146,762$	236.7 28.450
$ARFIMA(1,d,1)$	219	$74,639$	$146,860$	138.3 28.450
$ARFIMA(2,d,1)$	222	$74,676$	$146,902$	63.3 21.646
$ARFIMA(2,d,2)$	225	$74,708$	$146,932$	−

Table 4: Likelihood statistics for the DTSMs (OLS, NA and RRND models). The columns show: the number of parameters; the loglikelihood, Baysian Information Criterion and the Likelihood Ratio statistic and (in small typeface) the p=0.0011 value for the test of the ARFIMA(2,d,2)) model against the others. In panel A the likelihood ratio is below this critical value, favoring ARFIMA(2,d,2)) as the restricted model. In panel B the likelihood ratio is above this critical value, favoring ARFIMA(2,d,2)) as the unrestricted model.

Model	Total RMSE
OLS	10.029
NA	10.049
$AR(1)$	10.320
$AR(2)$	10.275
$ARMA(1,1)$	10.321
$ARFIMA(1,d,0)$	10.263
$ARFIMA(2,d,0)$	10.238
$ARFIMA(1,d,1)$	10.223
$ARFIMA(2,d,1)$	10.212
$ARFIMA(2,d,2)$	10.198

Table 5: RMSE of cross-sectional errors. These are calculated using (29) and reported in basis points.

Figures

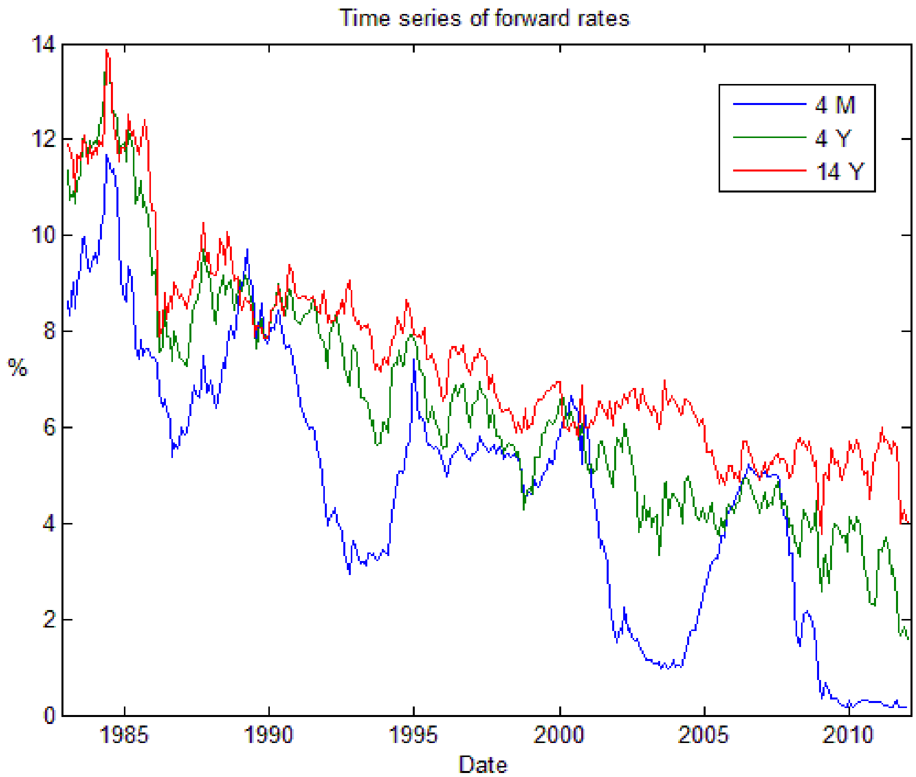

Figure 1: Time series of 4 month, 4 year and 14 year forward rates.

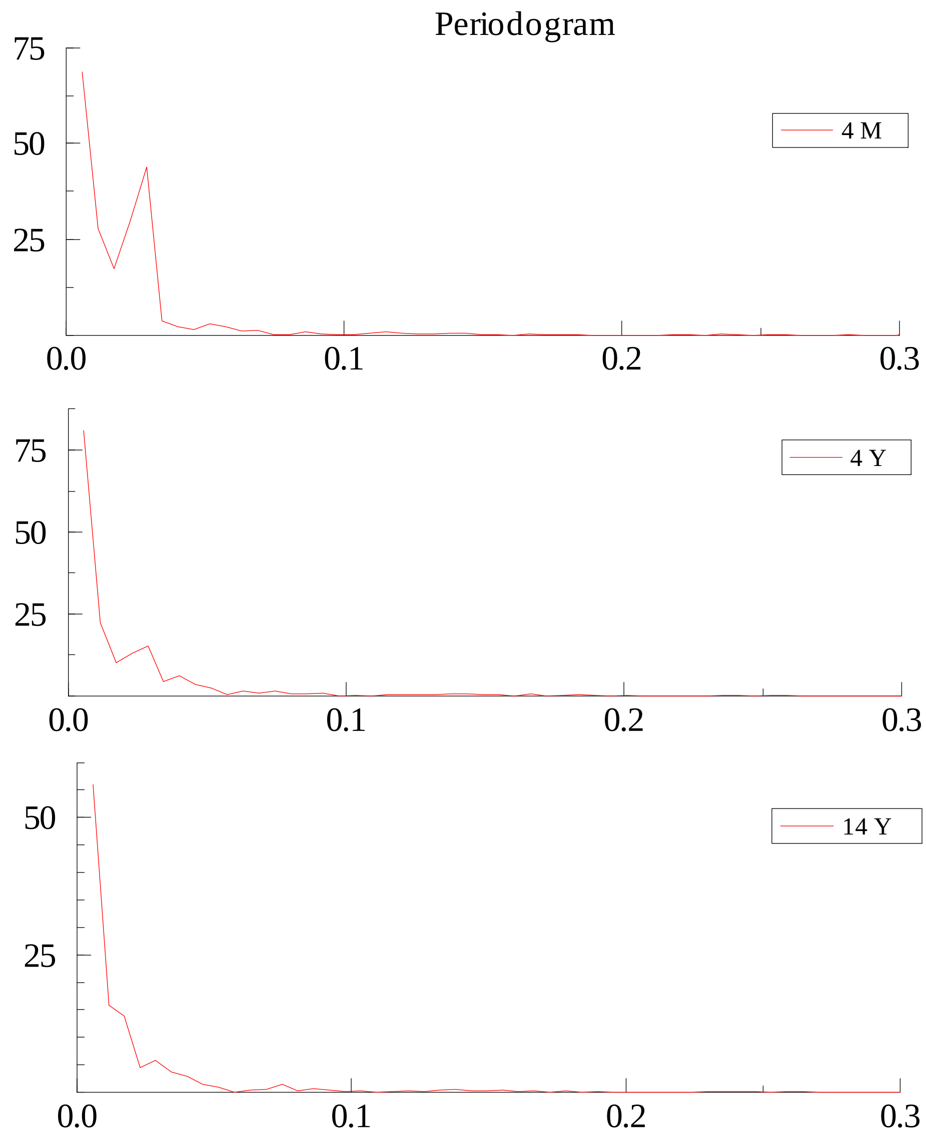

Figure 2: Periodogram of the 4 month, 4 and 14 year forward rates.

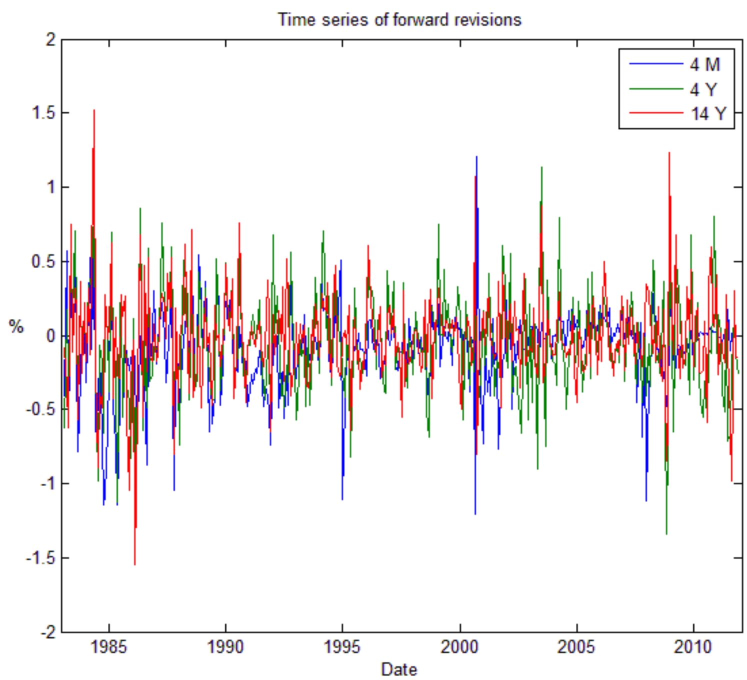

Figure 3: Time series of 4 month, 4 year and 14 year forward rates revisions. The forward rate revisions are defined as $f_{n-1,t+1} - f_{n,t}$

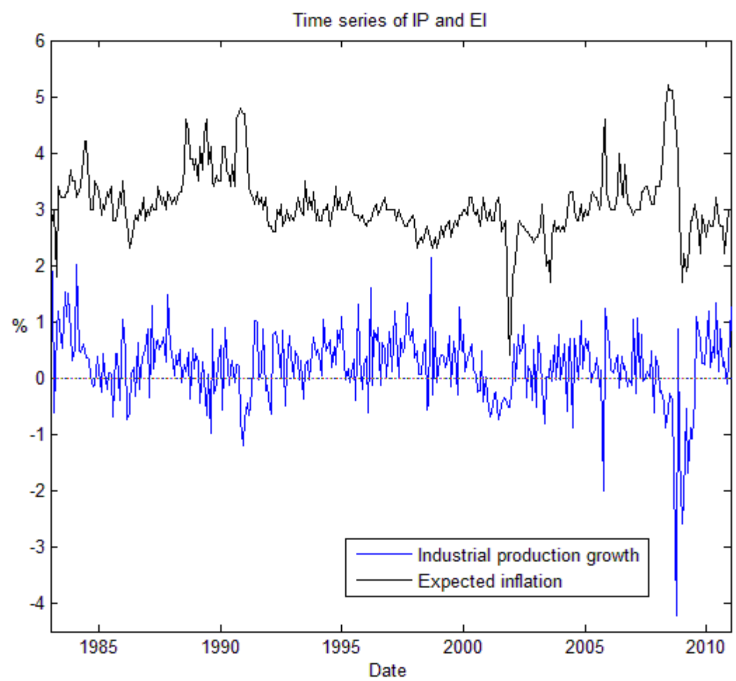

Figure 4: Time series of monthly changes in industrial production and 12 month expected inflation.

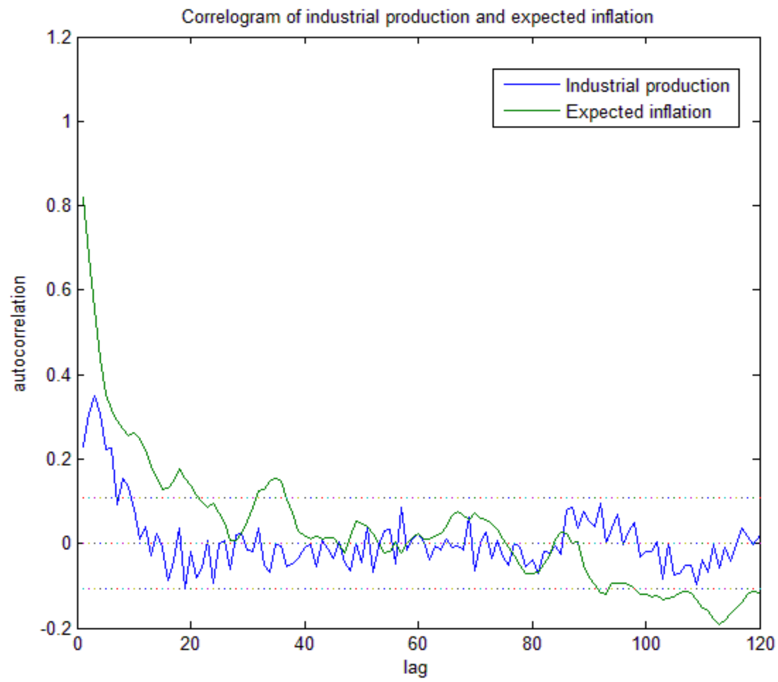

Figure 5: Correlogram of expected inflation and industrial production growth.

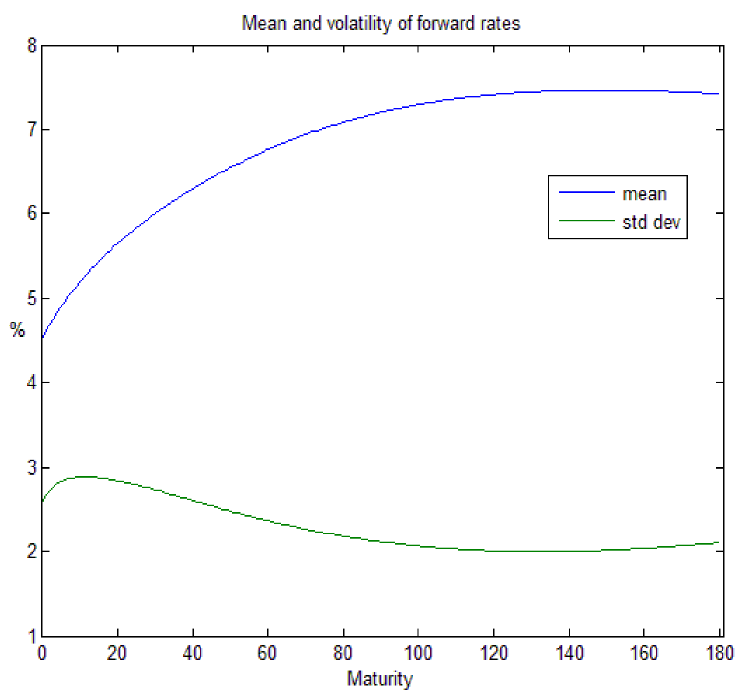

Figure 6: Unconditional mean and standard deviation of forward rates by maturity. The sample period is January 1983 to January 2011.

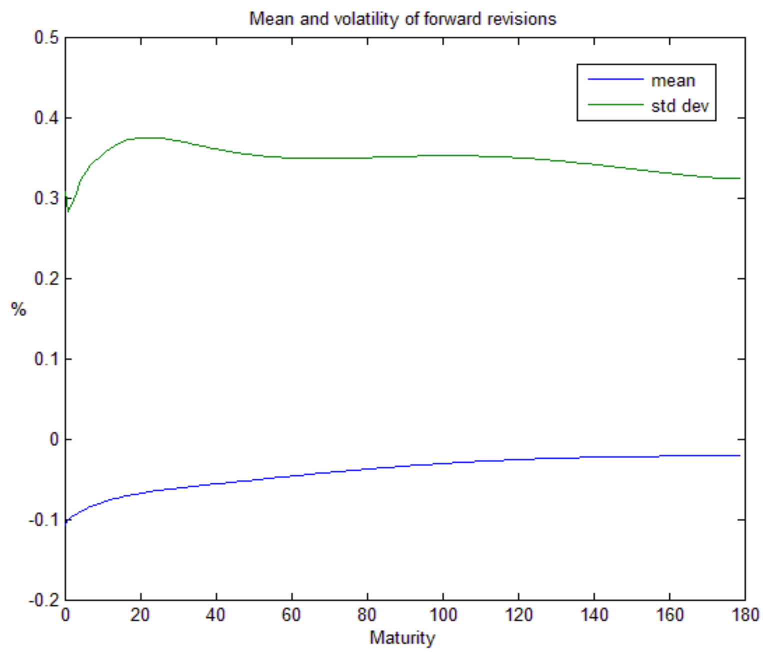

Figure 7: Unconditional mean and standard deviation of forward rate revisions by maturity. The sample period is January 1983 to January 2011.

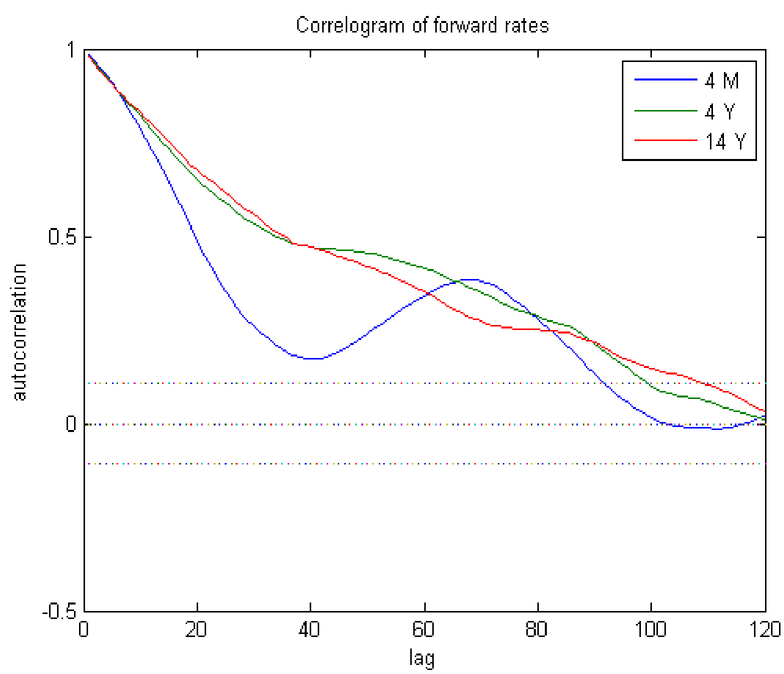

Figure 8: Correlogram of 4 month, 4 year and 14 year forward rates. The area between the dotted lines denotes the 95% confidence interval around zero. The sample period is January 1983 to January 2011.

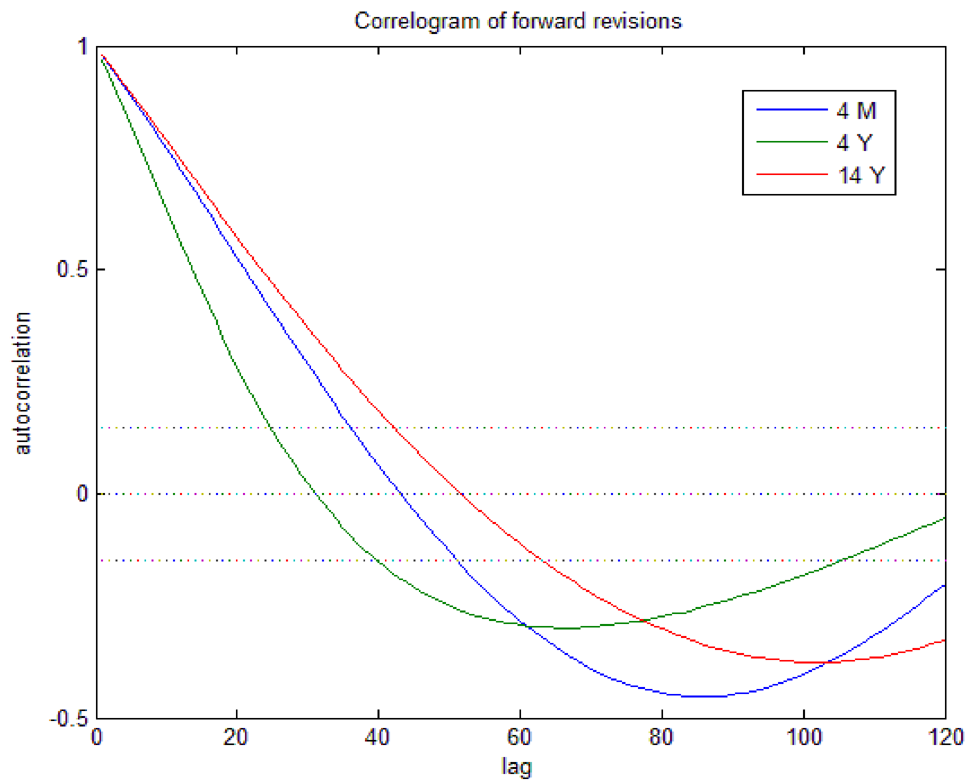

Figure 9: Correlogram of 4 month, 4 year and 14 year forward rate revisions. The area between the dotted lines denotes the 95% confidence interval around zero. The sample period is January 1983 to January 2011.

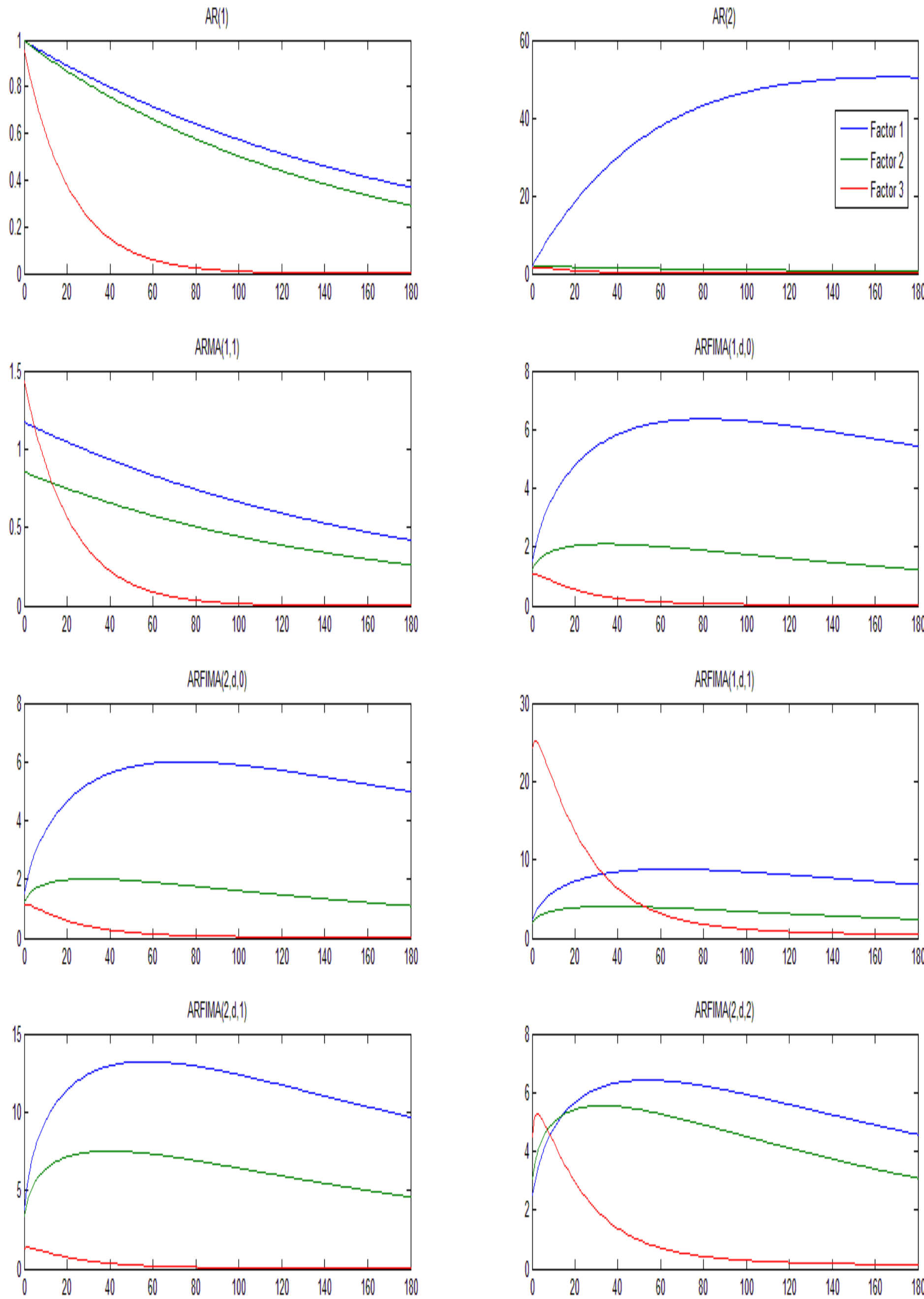

Figure 10: Moving average representations of the factor dynamics under the risk neutral measure (4.2.3). These are similar to the factor loadings of a discount yield model and show the effect of risk neutral shocks inferred from the noiseless revisions on the forward rates and revisions at different maturities.

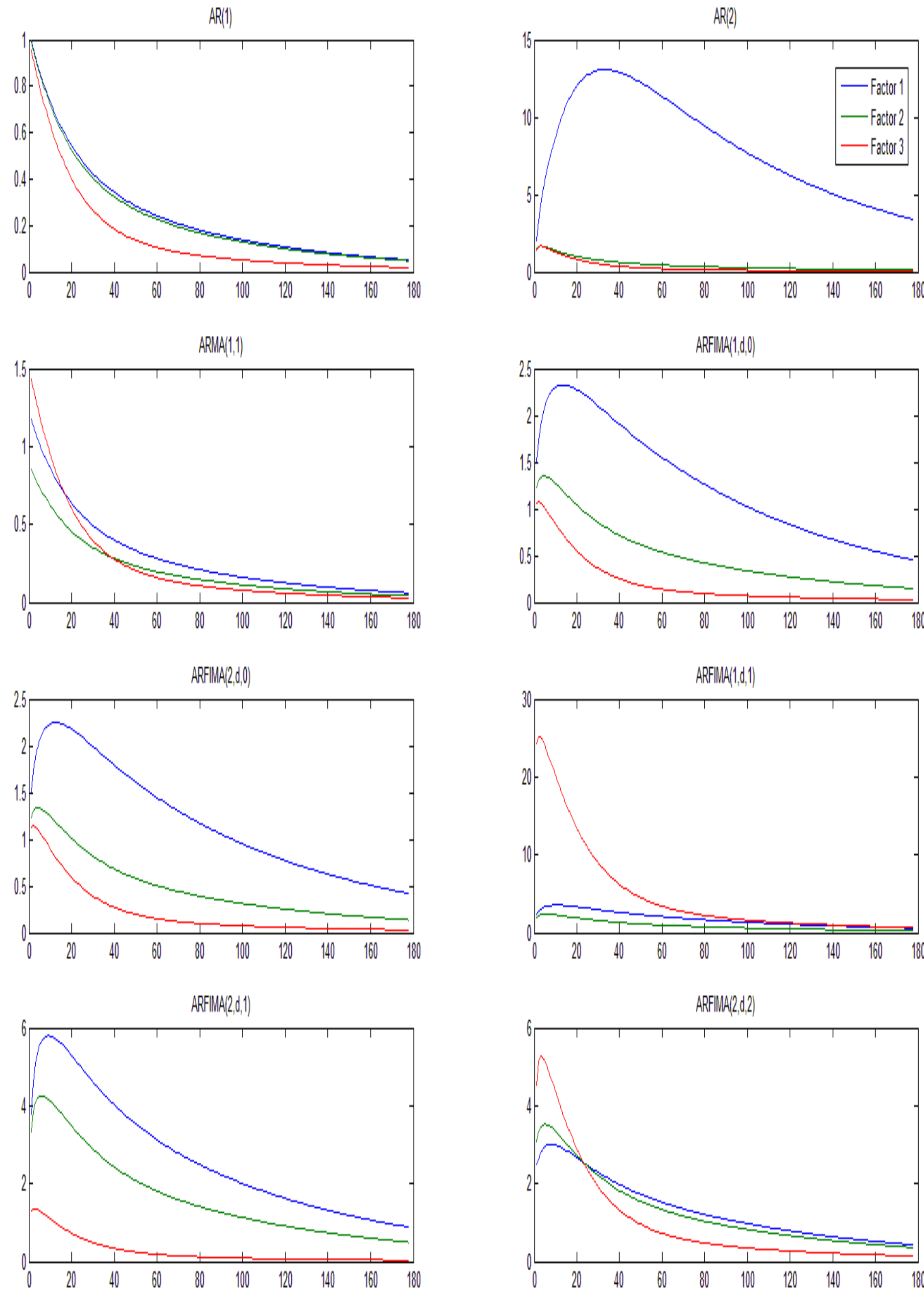

Figure 11: Impulse responses for the spanned factors. These show the response over time of the short rate to the real world shocks spanned by the yield curve (backed out from the noiseless revisions) under the risk-neutral measure. They are given by the parameters φ defined in Appendix B. This figure is comparable to the previous figure showing the risk neutral dynamics.

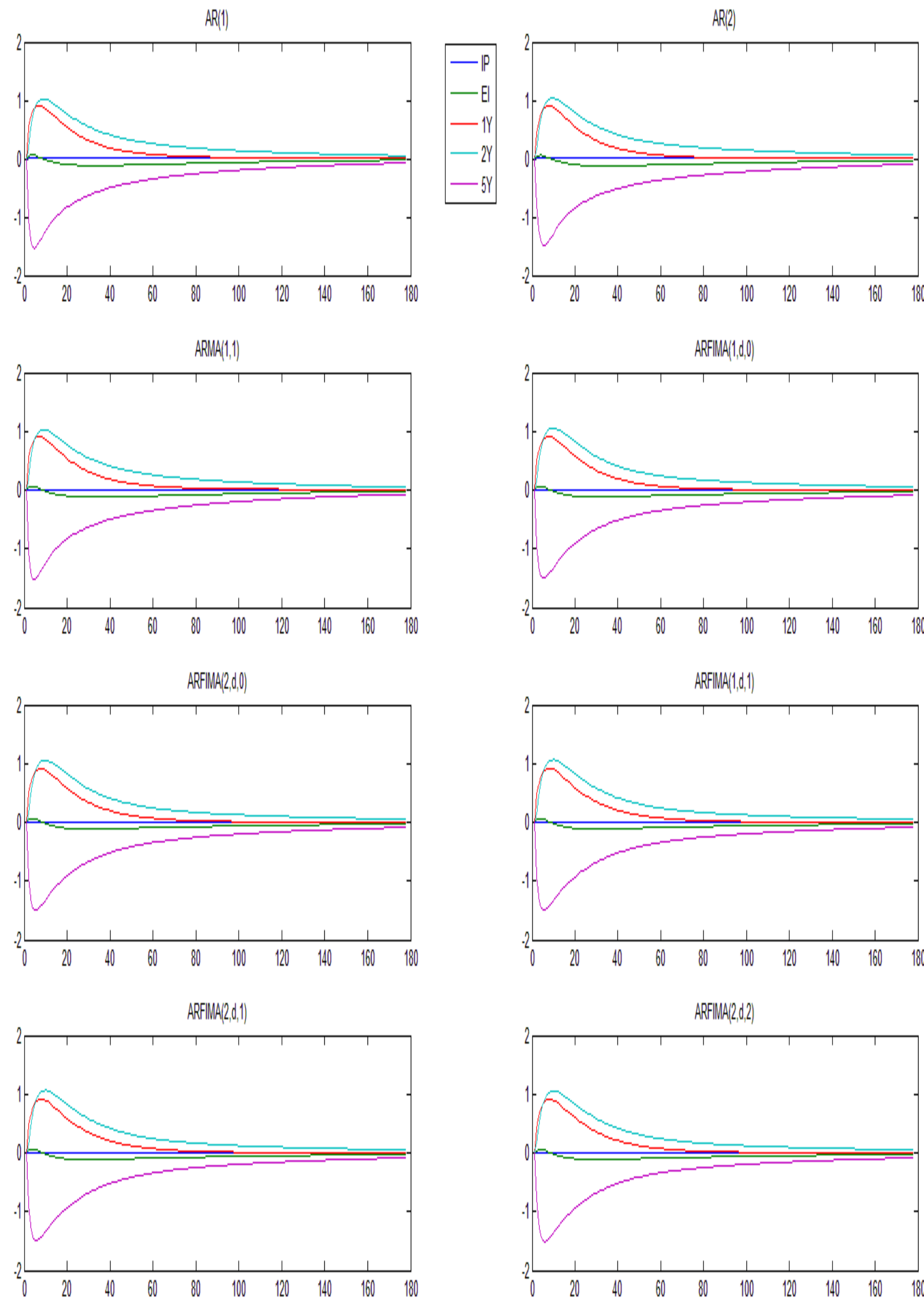

Figure 12: Impulse responses for the unspanned factors ($\mathbf{m}_t$ in (37)). These are defined under the real-world measure and show the response over time of the short rate to macro and other unspanned real world shocks. They move away from their initial value of zero to a single peak before mean reverting to zero. They are given by the parameters γ defined in Appendix B. These unspanned factors are 'latent' in the sense that they do not affect the yield curve under the risk-neutral measure.

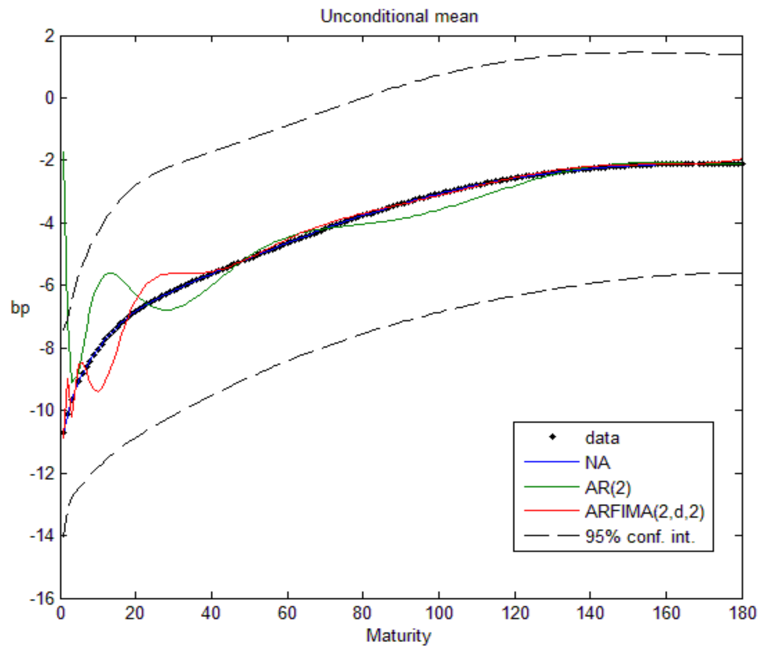

Figure 13: Sample and model-implied unconditional means of the forward rate revisions by maturity. The dashed line shows the 95% confidence interval for the data sample.

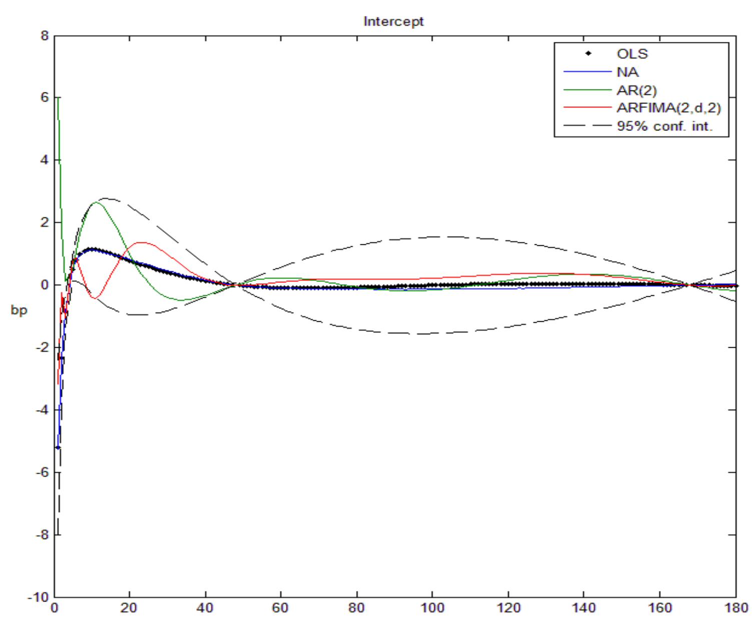

Figure 14: Intercepts of restricted and unrestricted (OLS) models. The dashed line shows the 95% confidence band for the latter.

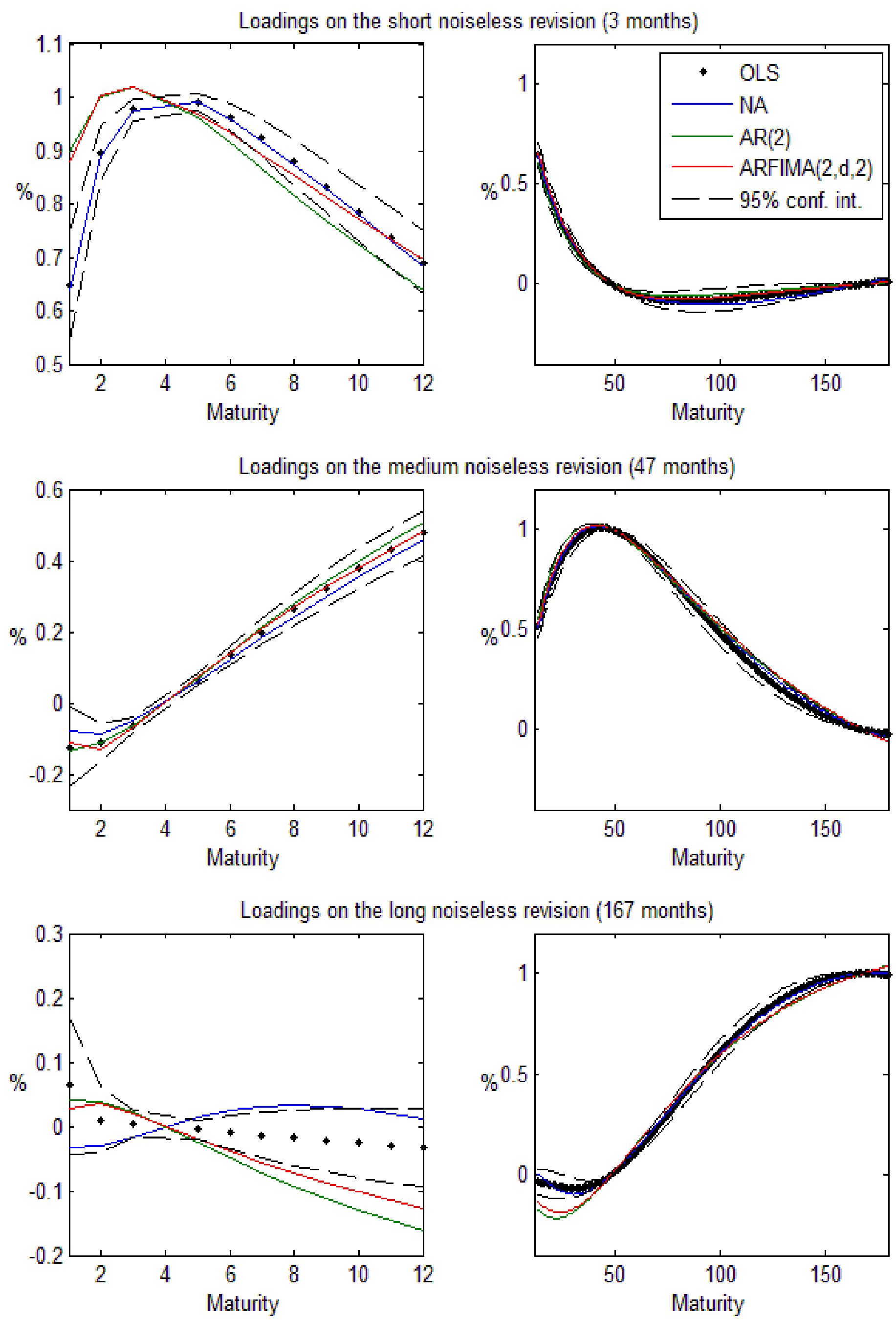

Figure 15: Loadings on the three noiseless revisions (4, 48 year and 14 year maturities) in the restricted and unrestricted (OLS) models ($B_y B_x^{-1}$). The dots denote the loadings of the unrestricted OLS model