BANK OF ENGLAND

Staff Working Paper No. 550
Dynamic term structure models: the best way to enforce the zero lower bound in the United States

Martin M Andreasen and Andrew Meldrum

September 2015

Staff Working Paper No. 550

Dynamic term structure models: the best way to enforce the zero lower bound in the United States

Martin M Andreasen[1] and Andrew Meldrum[2]

Abstract

This paper studies whether dynamic term structure models for US nominal bond yields should enforce the zero lower bound by a quadratic policy rate or a shadow rate specification. We address the question by estimating quadratic term structure models (QTSMs) and shadow rate models (SRMs) with at most four pricing factors. Our findings suggest that QTSMs give a better in-sample fit than SRMs with two and three factors, whereas the SRM marginally dominates with four factors. Loadings from Campbell-Shiller regressions are generally better matched by the SRMs, which also outperform the QTSMs when forecasting bond yields, particularly with four pricing factors.

Key words: Bias-adjustment, forecasting study, quadratic term structure models, sequential regression approach, shadow rate models.

JEL classification: C10, C50, G12.

---

(1) Aarhus University and CREATES. Email: mandreasen@econ.au.dk.
(2) Bank of England. Email: andrew.meldrum@bankofengland.co.uk

Information on the Bank's working paper series can be found at
www.bankofengland.co.uk/research/Pages/workingpapers/default.aspx

1 Introduction

Nominal bond yields have reached historically low levels during the recent financial crisis, with short rates at or close to the zero lower bound (ZLB) in several countries. This development has highlighted the inability of the Gaussian affine term structure model (ATSM) to ensure non-negative bond yields. One way to account for the ZLB is to abandon the affine specification of the policy rate and let this rate be quadratic in the pricing factors with appropriate restrictions. Adopting this extension leads to the class of quadratic term structure models (QTSMs) studied in Ahn, Dittmar & Gallant (2002), Leippold & Wu (2002), and Realdon (2006) among others. Another way to enforce the ZLB is to restrict policy rates to be non-negative by the max function, as done in the class of shadow rate models (SRMs) suggested by Black (1995). The two ways to account for the ZLB imply different dynamics for bond yields but little is currently known about their relative performance in the US. That is, should dynamic term structure models (DTSMs) for US bond yields enforce the ZLB by a quadratic policy rate or a shadow rate specification?

The aim of the present paper is to address this question by comparing the in- and out-of-sample performance of QTSMs and SRMs. We study models with two, three, and four pricing factors to explore how the factor structure affects the relative performance of the two ways to account for the ZLB. Following Dai & Singleton (2002), the performance of DTSMs is commonly evaluated by their ability to match moments from ordinary and risk-adjusted Campbell-Shiller regressions (the so-called LPY tests), as they capture key features of the physical and risk-neutral distributions of bond yields and hence implied term premia. However, none of the ATSMs satisfying the LPY tests in Dai & Singleton (2002) enforce the ZLB, and it is therefore unclear if DTSMs can jointly enforce the ZLB and match term premia to satisfy these tests in the US.[1]

Non-linear DTSMs with latent pricing factors as in QTSMs and SRMs are typically estimated by quasi maximum likelihood (QML) using a non-linear extension of the Kalman filter.[2] However, the asymptotic properties of this QML estimator are generally unknown and it suffers from small-sample biases. We overcome these difficulties by using the sequential regression (SR) approach of Andreasen &

---

[1] Modelling term premia at the ZLB is highly relevant for monetary policy. For example, the recent bond purchases by central banks are likely to affect the economy by reducing term premia according to Gagnon, Raskin, Rernache & Sack (2011) and Joyce, Lasaosa, Stevens & Tong (2011), although Christensen & Rudebusch (2012) argue that the effect on term premia in the US may have been somewhat smaller than found in Gagnon et al. (2011).

[2] Recent applications of this procedure in DTSMs enforcing the ZLB may be found in Ichiue & Ueno (2007), Kim & Singleton (2012), Ichiue & Ueno (2013), Bauer & Rudebusch (2014), and Christensen & Rudebusch (2015).

BANK OF ENGLAND

Christensen (2015), which gives consistent and asymptotically normal estimates under weaker restrictions than typically imposed for likelihood-based inference. For instance, the SR approach allows measurement errors in bond yields to display heteroskedasticity and correlation in both the cross-section and the time series dimension. Building on the work of Andreasen & Christensen (2015), we improve the finite sample properties of the SR approach in two ways. First, a bootstrap is introduced to bias-adjust the estimated physical dynamics of the factors. This extension allows us to explore how small-sample biases affect QTSMs and SRMs, which is an unaddressed issue. Second, a residual-based bootstrap is proposed for the risk-neutral parameters to refine its asymptotic distribution in finite samples. Accordingly, we show how to bootstrap the entire SR approach.

Apart from these robust econometric properties, the SR approach is also attractive from a finance perspective, because the QTSMs and SRMs we consider differ only in their risk-neutral distributions which may be estimated independently of their physical distributions in the first step of the SR approach. Hence, the ability of these models to match in-sample bond yields reported below holds for *any* considered functional form of the market price of risk. Another advantage of the SR approach is its computational simplicity, which allows us to estimate QTSMs and SRMs with four pricing factors, whereas previous studies restrict focus to models with at most three factors.

The performance of DTSMs on US bond yields is typically studied using either a long sample starting in the 1960s or a short sample from around 1990. We find it informative to include both samples because bond yields in the long sample attain very high as well as very low values with frequent changes in conditional volatility, whereas bond yields in the short sample are lower and display relatively stable conditional volatility.[3] Hence, if one believes that the US in the future is likely to experience very high bond yields and frequent changes in volatility, the results from our long sample are likely to be most informative on how to model the ZLB. On the other hand, if one believes that such future bond yields are unlikely, the results from our short sample should probably be preferred.

We highlight the following results from our analysis on monthly US bond yields ending in December 2013. First, the QTSM gives a better in-sample fit than the SRM with two and three pricing factors, whereas the SRM does marginally better than the QTSM with four factors. We also find that both ZLB models clearly outperform the Gaussian ATSMs when measured by in-sample fit. Second, the QTSMs

---

[3]Rudebusch & Wu (2007) argue for the presence of a structural break in US bond yields during the middle or late 1980s. Accounting for this potential break may serve as a second motivation for considering our short sample.

match loadings from ordinary Campbell-Shiller regressions in the sample from 1990 but not in the long sample, whereas the SRMs reproduce these loadings in both samples. We also find that bias-adjusting the estimated physical dynamics of the pricing factors has a sizable effect on term premia in all models, and that the Gaussian ATSMs with three and four factors may have over-estimated the 10-year term premium from 2005 to 2009 by as much as 50 basis points when compared to the corresponding QTSMs and SRMs. Third, the fall in conditional volatility of most bond yields when reaching the ZLB is nicely captured by the QTSMs and the SRMs, although both models struggle to generate the increase in volatility just before reaching the ZLB. Fourth, in an extensive out-of-sample forecasting study from January 2005 to December 2013, we find that the SRM generally performs better than the QTSM, especially in the long sample, and that models accounting for the ZLB outperform the Gaussian ATSM. The SRM is also found to be more robust and less subject to overfitting than the QTSM, as the forecasts in the SRM generally *improve* when adding a fourth factor whereas the opposite generally holds for the QTSM. We finally study two- and three-factor models where the ZLB is enforced by a shadow rate that is an unrestricted quadratic function of the pricing factors. These hybrid models fit bond yields marginally better in-sample than QTSMs but struggle to provide better performance against the LPY tests or conditional volatility in bond yields than the SRMs. We also find that the hybrid models generally deliver less accurate forecasts of bond yields compared to the QTSMs and the SRMs.

Overall, our findings suggest that the best way to enforce the ZLB in the US depends on the sample period and the number of pricing factors considered. In the long sample from 1961, the SRMs with an affine shadow rate in two, three, or four factors are preferred, because they are better than the corresponding QTSMs and the hybrid models at matching term premia, while simultaneously doing well in- and out-of-sample. As for the short sample from 1990, the two- and three-factor QTSMs are preferred because they perform well in LPY-tests and display good properties in- and out-of-sample. When adding a fourth factor in the short sample, we once again recommend the SRM with an affine shadow rate because the corresponding QTSM appears to suffer from overfitting. Our analysis also presents a strong case for a four-factor SRM, as an additional factor improves in-sample fit *and* forecasts of future bond yields compared to the three-factor SRM. Thus, our preferred models clearly differ from the recommended two-factor hybrid model in Kim & Singleton (2012) on Japanese bond yields, showing that one should be cautious of directly extrapolating their results to the US.

The rest of the paper is organized as follows. Section 2 presents the DTSMs considered, and we

describe how these models are estimated by the SR approach in Section 3. In-sample results are reported in Section 4 and the out-of-sample performance is presented in Section 5. We finally study two- and three-factor hybrid models in Section 6. Concluding comments are provided in Section 7.

2 Dynamic term structure models

We start by describing the benchmark Gaussian ATSM before presenting the QTSM and the SRM. The pricing factors in all these models are assumed to be Gaussian under the risk-neutral and physical measures, implying an affine specification for the market price of risk. We do not study the multivariate version of the model by Cox, Ingersoll & Ross (1985) with independent pricing factors or its extension by Dai & Singleton (2000) with correlated factors (the so-called $\mathbb{A}_m(m)$ models), because these models are generally unable to reproduce key moments of term premia (see Dai & Singleton (2002)).

2.1 The Gaussian ATSM

The discrete-time Gaussian ATSM is characterized by three equations. The first specifies the one-period risk-free interest rate r_t to be affine in n_x pricing factors $\mathbf{x}_t$, i.e. $r_t = \alpha + \boldsymbol{\beta}' \mathbf{x}_t$, where α is a scalar and $\boldsymbol{\beta}$ is an $n_x \times 1$ vector. The second equation describes the dynamics of the pricing factors under the risk-neutral measure $\mathbb{Q}$ as a vector autoregressive (VAR) process, i.e.

$$\mathbf{x}_{t+1} = \boldsymbol{\Phi}\boldsymbol{\mu} + (\mathbf{I} - \boldsymbol{\Phi}) \mathbf{x}_t + \boldsymbol{\Sigma} \varepsilon_{t+1}^{\mathbb{Q}} \tag{1}$$

with $\varepsilon_{t+1}^{\mathbb{Q}} \sim \mathcal{NID}(\mathbf{0}, \mathbf{I})$. The no-arbitrage price in time period t of an j-period zero-coupon bond is $P_{t,j} = E_t^{\mathbb{Q}} [\exp\{-r_t\} P_{t+1,j-1}]$, where $E_t^{\mathbb{Q}}$ is the conditional expectation under $\mathbb{Q}$. Letting K denote the longest maturity for the set of zero-coupon bonds considered, we have $P_{t,j}^{ATSM} = \exp\left\{A_j + \mathbf{B}_j' \mathbf{x}_t\right\}$ for $j = 1, 2, ..., K$, where the recursive formulae for A_j and $\mathbf{B}_j$ are easily derived.[4]

The final equation specifies the functional form for the market price of risk $\mathbf{f}(\mathbf{x}_t)$ with dimension $n_x \times 1$. The relationship between the physical measure $\mathbb{P}$ and the $\mathbb{Q}$ measure is given by $\varepsilon_{t+1}^{\mathbb{Q}} = \varepsilon_{t+1}^{\mathbb{P}} + \mathbf{f}(\mathbf{x}_t)$, and the factor dynamics under $\mathbb{P}$ are therefore

$$\mathbf{x}_{t+1} = \boldsymbol{\Phi}\boldsymbol{\mu} + (\mathbf{I} - \boldsymbol{\Phi}) \mathbf{x}_t + \boldsymbol{\Sigma}\mathbf{f}(\mathbf{x}_t) + \boldsymbol{\Sigma} \varepsilon_{t+1}^{\mathbb{P}}$$

---

[4]See the technical appendix accompanying the paper for a derivation of A_j and $\mathbf{B}_j$. The technical appendix is available from the web page of the corresponding author or on request.

with $\varepsilon_{t+1}^{\mathbb{P}} \sim \mathcal{NID}\left(\mathbf{0}, \mathbf{I}\right)$. To obtain an affine process for the pricing factors under $\mathbb{P}$, we let $\mathbf{f}\left(\mathbf{x}_t\right) = \boldsymbol{\Sigma}^{-1}\left(\mathbf{f}_0 + \mathbf{f}_1 \mathbf{x}_t\right)$, where $\mathbf{f}_0$ has dimension $n_x \times 1$ and $\mathbf{f}_1$ is an $n_x \times n_x$ matrix. This implies

$$\mathbf{x}_{t+1} = \mathbf{h}_0 + \mathbf{h_x} \mathbf{x}_t + \boldsymbol{\Sigma} \varepsilon_{t+1}^{\mathbb{P}}, \tag{2}$$

with $\mathbf{h}_0 \equiv \boldsymbol{\Phi}\boldsymbol{\mu} + \mathbf{f}_0$ and $\mathbf{h_x} \equiv \mathbf{I} - \boldsymbol{\Phi} + \mathbf{f}_1$. To obtain stationary bond yields with finite first and second unconditional moments, we require that the process for $\mathbf{x}_t$ under $\mathbb{P}$ is stationary, i.e. all eigenvalues of $\mathbf{I} - \boldsymbol{\Phi} + \mathbf{f}_1$ are inside the unit circle.

The pricing factors are considered to be latent (i.e. unobserved) and a set of normalization restrictions are therefore needed to identify the model. We require i) $\boldsymbol{\beta} = \mathbf{1}$, ii) $\boldsymbol{\mu} = \mathbf{0}$, iii) $\boldsymbol{\Phi}$ to be diagonal, and iv) $\boldsymbol{\Sigma}$ to be triangular.[5] This identification scheme constrains the $\mathbb{Q}$ dynamics for the pricing factors whereas the $\mathbb{P}$ dynamics are unrestricted to simplify estimation within the SR approach.

2.2 The QTSM

The discrete-time QTSM differs from the Gaussian ATSM by letting the policy rate be quadratic in the pricing factors, i.e.

$$r_t = \alpha + \boldsymbol{\beta}' \mathbf{x}_t + \mathbf{x}_t' \boldsymbol{\Psi} \mathbf{x}_t, \tag{3}$$

where $\boldsymbol{\Psi}$ is a symmetric $n_x \times n_x$ matrix. Following Kim & Singleton (2012), we adopt the decomposition $\boldsymbol{\Psi} \equiv \mathbf{A}\mathbf{D}\mathbf{A}'$, where $\mathbf{A}$ is an $n_x \times n_x$ lower triangular matrix with ones on the diagonal and $\mathbf{D}$ is an $n_x \times n_x$ diagonal matrix. Introducing quadratic terms in the policy rate is useful because they allow the model to enforce the ZLB. The non-negativity conditions for bond yields are i) $\alpha \geq \frac{1}{4}\boldsymbol{\beta}'\boldsymbol{\Psi}^{-1}\boldsymbol{\beta}$ and ii) $\boldsymbol{\Psi}$ to be positive semi-definite (see Realdon (2006)). This way of imposing the ZLB may be applied independently of the chosen dynamics for the pricing factors, and a quadratic policy rule therefore serves as a mechanism to enforce the ZLB.

Given the policy rate in (3), we assume the same specification for the factors as in (1), because it gives a closed-form solution for zero-coupon bond prices, i.e. $P_{t,j}^{QTSM} = \exp\left\{\tilde{A}_j + \tilde{\mathbf{B}}_j' \mathbf{x}_t + \mathbf{x}_t' \tilde{\mathbf{C}}_j \mathbf{x}_t\right\}$ for $j = 1, 2, ..., K$, with the recursive formulae for $\tilde{A}_j$, $\tilde{\mathbf{B}}_j$, and $\tilde{\mathbf{C}}_j$ derived in Realdon (2006). Hence, bond yields $y_{t,j}^{QTSM} \equiv -\frac{1}{j} \log P_{t,j}^{QTSM}$ in the QTSM are quadratic in the pricing factors.

For comparability with the Gaussian ATSM, we maintain the affine specification for the market price

---

[5] There exist other normalization schemes, for instance the one recently suggested by Joslin, Singleton & Zhu (2011). We prefer the considered normalization scheme because it is closely related to the one adopted for the QTSM.

of risk, meaning that the $\mathbb{P}$ dynamics for the pricing factors in the QTSM are given by (2). As in the Gaussian ATSM, not all parameters are identified in the QTSM with latent pricing factors. We follow Ahn et al. (2002) and impose the restrictions: i) $\mathbf{\Psi}$ is symmetric with diagonal elements equal to one, ii) $\boldsymbol{\mu} \geq \mathbf{0}$, iii) $\boldsymbol{\beta} = \mathbf{0}$, iv) $\mathbf{\Phi}$ is diagonal, and v) $\mathbf{\Sigma}$ is triangular.[6] This normalization scheme implies unrestricted $\mathbb{P}$ dynamics for the pricing factors and that the ZLB may be enforced by imposing $\alpha = 0$ and $\mathbf{\Psi}$ to be positive semi-definite.

2.3 The SRM

The ZLB may alternatively be enforced in DTSMs by introducing a shadow interest rate $s\left(\mathbf{x}_t\right)$ as suggested by Black (1995). This shadow rate is unconstrained by the ZLB and may attain negative values. Absent transaction and storage costs for money, Black (1995) notes that the nominal interest rate cannot be negative because investors may always hold cash. Hence, we let $r_t = \max\left(0, s\left(\mathbf{x}_t\right)\right)$. As with the quadratic policy rule, the concept of a shadow rate serves as a mechanism to enforce the ZLB and may be applied independently of the functional form for $s\left(\mathbf{x}_t\right)$ and the considered factor dynamics.

For comparability with the Gaussian ATSM, we let the shadow rate be affine in the pricing factors, i.e. $s\left(\mathbf{x}_t\right) = \alpha + \boldsymbol{\beta}'\mathbf{x}_t$, but other specifications may also be considered, as illustrated in Section 6. For the same reason, we also restrict focus to affine processes for the pricing factors under the $\mathbb{Q}$ and the $\mathbb{P}$ measure as given in (1) and (2), respectively. Finally, the identification conditions for the SRM are identical to those for the Gaussian ATSM in Section 2.1.

Multivariate SRMs do not attain closed-form expressions for bond prices, and numerical approximations are therefore required.[7] We apply the second-order approximation advocated by Priebsch (2013) formulated in discrete time.[8] That is, bond yields are given by

$$y_{t,j}^{SRM} = \frac{1}{j}E_t^{\mathbb{Q}}\left[\sum_{i=0}^{j-1}r_{t+i}\right] - \frac{1}{2j}Var_t^{\mathbb{Q}}\left[\sum_{i=0}^{j-1}r_{t+i}\right], \tag{4}$$

when preserving all terms up to second order in a Taylor-series expansion of $P_{t,j}^{SRM} = E_t^{\mathbb{Q}}\left[\prod_{i=0}^{j-1}\exp\left(-r_{t+i}\right)\right]$.

---

[6] As an illustration, the normalization restrictions on $\mathbf{\Psi}$ imply that the diagonal elements of $\mathbf{D}$ are $D_{11} = 1$, $D_{22} = 1 - D_{11}A_{21}^2$, and $D_{33} = 1 - A_{31}^2 - \left(1 - A_{21}^2\right)A_{32}^2$ in a model with three pricing factors.

[7] Gorovoi & Linetsky (2004) derive the solution to bond yields in one-factor SRMs with a Gaussian or a square-root process driving $s\left(\mathbf{x}_t\right)$.

[8] Other approximation methods used in the literature include i) lattices (Ichiue & Ueno (2007)), ii) finite-difference methods (Kim & Singleton (2012)), iii) Monte Carlo integration (Bauer & Rudebusch (2014)), iv) an option pricing approximation (Krippner (2012), Christensen & Rudebusch (2015)), and v) ignoring Jensen's inequality term to solve a Gaussian model by a truncated normal distribution (Ichiue & Ueno (2013)).

The required conditional first and second moments for future short rates are obtained using results for the truncated normal and bivariate normal distributions.[9]

3 The estimation procedure

We next present the three steps in the SR approach for estimating the latent pricing factors and model parameters. In doing so, we extend the SR approach with a bias-adjustment when estimating the physical dynamics of the factors and a residual-based bootstrap for the risk-neutral coefficients.

3.1 The SR approach: The model class

The SR approach applies to DTSMs where bond yields are potentially non-linear functions of latent pricing factors and measured with errors v_{t,m_j}, i.e.

$$y_{t,m_j} = g_{m_j}(\mathbf{x}_t; \boldsymbol{\theta}_1) + v_{t,m_j}, \tag{5}$$

where m_j denotes the maturity of the jth observation. For any time period t, we require the measurement errors $\{v_{t,m_j}\}_{j=1}^{n_{y,t}}$ to have zero mean and a finite and positive definite covariance matrix, with $n_{y,t}$ denoting the number of observed bond yields at time t. Apart from the technical regularity conditions in Andreasen & Christensen (2015), no further assumptions are imposed on v_{t,m_j}.

The functional relationship between the pricing factors and bond yields in (5) is parameterized by the risk-neutral coefficients $\boldsymbol{\theta}_1 \equiv \begin{bmatrix} \boldsymbol{\theta}'_{11} & \boldsymbol{\theta}'_{12} \end{bmatrix}'$, with $\boldsymbol{\theta}_{11}$ containng those parameters that only affect the cross-section of bond yields whereas $\boldsymbol{\theta}_{12}$ contains parameters that also enter in the time series dynamics of the pricing factors. For the Gaussian ATSM, the function $\mathbf{g}(\cdot)$ is linear in the factors, i.e. $g_{m_j}^{ATSM}(\mathbf{x}_t; \boldsymbol{\theta}_1^{ATSM}) \equiv -\frac{1}{m_j}\left(A_{m_j} + \mathbf{B}'_{m_j}\mathbf{x}_t\right)$, and we have $\boldsymbol{\theta}_{11}^{ATSM} \equiv \begin{bmatrix} \alpha & diag(\boldsymbol{\Phi})' \end{bmatrix}'$ with $\boldsymbol{\theta}_{12}^{ATSM} \equiv vech(\boldsymbol{\Sigma})$. The QTSM induces a slightly more complicated expression for bond yields because $g_{m_j}^{QTSM}(\mathbf{x}_t; \boldsymbol{\theta}_1^{QTSM}) \equiv -\frac{1}{m_j}\left(\tilde{A}_{m_j} + \tilde{\mathbf{B}}'_{m_j}\mathbf{x}_t + \mathbf{x}'_t\tilde{\mathbf{C}}_{m_j}\mathbf{x}_t\right)$, and the risk-neutral coefficients are in this model given by $\boldsymbol{\theta}_{11}^{QTSM} \equiv \begin{bmatrix} (\boldsymbol{\theta}_1^{ATSM})' & \boldsymbol{\mu}' & \left\{\{A_{i,j}\}_{i=1}^{j-1}\right\}_{j=1}^{n_x} \end{bmatrix}'$ and $\boldsymbol{\theta}_{12}^{QTSM} = \boldsymbol{\theta}_{12}^{ATSM}$. In the SRM, $g_{m_j}^{SRM}(\mathbf{x}_t; \boldsymbol{\theta}_1^{SRM})$ is given by (4) with $\boldsymbol{\theta}_1^{SRM} = \boldsymbol{\theta}_1^{ATSM}$.

All the DTSMs considered share the same linear and unrestricted transition function in (2) for the pricing factors under the $\mathbb{P}$ measure, parameterized by $\boldsymbol{\theta}_2 \equiv \begin{bmatrix} \boldsymbol{\theta}'_{22} & \boldsymbol{\theta}'_{12} \end{bmatrix}'$ and $\boldsymbol{\theta}_{22} \equiv \begin{bmatrix} \mathbf{h}'_0 & vec(\mathbf{h_x})' \end{bmatrix}'$.

---

[9] The details are provided in the technical appendix.

3.2 The SR approach: Step 1

The latent pricing factors are estimated by cross-section regressions, i.e. $\hat{\mathbf{x}}_t(\boldsymbol{\theta}_1) = \arg\min_{\mathbf{x}_t \in \mathbb{R}^{n_x}} Q_t$, where $Q_t \equiv \frac{1}{2n_{y,t}} \sum_{j=1}^{n_{y,t}} \left(y_{t,m_j} - g_{m_j}(\mathbf{x}_t; \boldsymbol{\theta}_1) \right)^2$ for $t = 1, 2, ..., T$. The estimated factors are denoted $\{\hat{\mathbf{x}}_t(\boldsymbol{\theta}_1)\}_{t=1}^T$ because they are computed for a given $\boldsymbol{\theta}_1$. These regressions have a closed-form solution for the Gaussian ATSM as $g_{m_j}^{ATSM}$ is linear in $\mathbf{x}_t$. For the QTSM and the SRM, the cross-section regressions are non-linear in the pricing factors and solved using the Levenberg-Marquardt optimizer with $\hat{\mathbf{x}}_{t-1}(\boldsymbol{\theta}_1)$ serving as good starting values for $t = 2, 3, ..., T$.[10] Although these non-linear regressions converge within a few iterations, some care is needed for the QTSM to find the global optimum. This is illustrated in Figure 1 where we plot the objective functions at four selected dates when filtering out the pricing factors. To facilitate the plotting, we focus on a one-factor QTSM but similar results apply with multiple factors. When only one bond yield is used (the dotted line), the regressions are not identified as there are two solutions. Identification is obtained by including more observations and the negative solution is clearly only a local optimum with 25 bond yields (the widest black line).

< Figure 1 about here >

The model parameters $\boldsymbol{\theta}_1$ are obtained by pooling all squared residuals from the cross-section regressions and minimizing their sum with respect to $\boldsymbol{\theta}_1$, i.e. $\hat{\boldsymbol{\theta}}_1^{step1} = \arg\min_{\boldsymbol{\theta}_1 \in \Theta_1} Q_{1:T}^{step1}$, where $Q_{1:T}^{step1} \equiv \frac{1}{2N} \sum_{t=1}^T \sum_{j=1}^{n_{y,t}} \left(y_{t,m_j} - g_{m_j}(\hat{\mathbf{x}}_t(\boldsymbol{\theta}_1); \boldsymbol{\theta}_1) \right)^2$. Here, $N \equiv \sum_{t=1}^T n_{y,t}$ and Θ_1 denotes the feasible domain of $\boldsymbol{\theta}_1$. Given standard regularity conditions, Andreasen & Christensen (2015) show consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_1^{step1}$ when $n_{y,t} \longrightarrow \infty$ for all t, i.e.

$$\sqrt{N} \left(\hat{\boldsymbol{\theta}}_1^{step1} - \boldsymbol{\theta}_1^o \right) \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \left(\mathbf{A}_o^{\boldsymbol{\theta}_1} \right)^{-1} \mathbf{B}_o^{\boldsymbol{\theta}_1} \left(\mathbf{A}_o^{\boldsymbol{\theta}_1} \right)^{-1} \right), \tag{6}$$

where the "o" refers to the true value. Following Andreasen & Christensen (2015), the expected value of the average Hessian matrix $\mathbf{A}_o^{\boldsymbol{\theta}_1}$ may be estimated consistently by

$$\hat{\mathbf{A}}^{\boldsymbol{\theta}_1} = \frac{1}{N} \sum_{t=1}^T \sum_{j=1}^{n_{y,t}} \left(\hat{\boldsymbol{\Gamma}}_{t,j}^{\boldsymbol{\theta}_1} \right) \left(\hat{\boldsymbol{\Gamma}}_{t,j}^{\boldsymbol{\theta}_1} \right)',$$

---

[10] The main input for the Levenberg-Marquardt optimizer is the Jacobian which is available in closed form for the QTSM. For the SRM, we use a first-order approximation as in Ichiue & Ueno (2013) to compute the Jacobian which equals $\frac{1}{j} \sum_{i=0}^{j-1} F\left(\frac{\alpha + \boldsymbol{\beta}' \mathbb{E}_t^{\mathbb{Q}}[\mathbf{x}_{t+i}]}{\sqrt{\boldsymbol{\beta}' Var_t^{\mathbb{Q}}[\mathbf{x}_{t+i}]\boldsymbol{\beta}}} \right) \boldsymbol{\beta}' (\mathbf{I} - \boldsymbol{\Phi})^i$ for $j = 1, 2, ..., K$ with F denoting the cumulative density of the normal distribution, but the second-order approximation by Priebsch (2013) is otherwise applied in the optimizer. Using the second-order approximation to numerically compute the Jacobian in the optimizer gives identical results but is somewhat slower than the adopted procedure.

BANK OF ENGLAND

where

$$\mathbf{\Gamma}_{t,j}^{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_1) \equiv \frac{\partial \hat{\mathbf{x}}_t'(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} \frac{\partial g_{m_j}(\hat{\mathbf{x}}_t(\boldsymbol{\theta}_1);\boldsymbol{\theta}_1)}{\partial \mathbf{x}_t(\boldsymbol{\theta}_1)} + \frac{\partial g_{m_j}(\hat{\mathbf{x}}_t(\boldsymbol{\theta}_1);\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1}$$

and $\hat{\mathbf{\Gamma}}_{t,j}^{\boldsymbol{\theta}_1} \equiv \mathbf{\Gamma}_{t,j}^{\boldsymbol{\theta}_1}\left(\hat{\boldsymbol{\theta}}_1^{step1}\right)$. The variance of the scaled score function $\mathbf{B}_o^{\boldsymbol{\theta}_1}$ is estimated using an extension of the Newey-West estimator that is robust to heteroskedasticity in the time series dimension, cross-sectional correlation, and autocorrelation in v_{t,m_j}. That is,

$$
\begin{aligned}
\hat{\mathbf{B}}^{\boldsymbol{\theta}_1} =\ & \frac{1}{N} \sum_{t=1}^{T} \sum_{j=1}^{n_{y,t}} \left\{ \hat{\sigma}_t^2 \left(\hat{\mathbf{\Gamma}}_{t,j}^{\boldsymbol{\theta}_1}\right)\left(\hat{\mathbf{\Gamma}}_{t,j}^{\boldsymbol{\theta}_1}\right)' \right. \\
& \left. + \sum_{\substack{k_T=-w_T \\ k_T \neq 0}}^{w_T} \sum_{\substack{k_D=-w_D \\ k_D \neq 0}}^{w_D} \left(1 - \frac{|k_T|}{1+w_T}\right)\left(1 - \frac{|k_D|}{1+w_D}\right)\left(\hat{\mathbf{\Gamma}}_{t,j}^{\boldsymbol{\theta}_1}\right)\left(\hat{\mathbf{\Gamma}}_{t+k_T,j+k_D}^{\boldsymbol{\theta}_1}\right)' \hat{v}_{t,m_j} \hat{v}_{t+k_T,m_{j+k_D}} \right\},
\end{aligned}
$$

where $\hat{\sigma}_t^2 = \frac{1}{n_{y,t}-n_x} \sum_{j=1}^{n_{y,t}} \hat{v}_{t,m_j}^2$ and $\hat{v}_{t,m_j} = y_{t,m_j} - g_{m_j}\left(\hat{\mathbf{x}}_t; \hat{\boldsymbol{\theta}}_1^{step1}\right)$. Here, w_D is the bandwidth for $\hat{v}_{t,m_j}$ in the cross-section dimension when ordered by maturity and w_T is the bandwidth for the time series dimension.

3.3   The SR approach: Step 2

We estimate $\boldsymbol{\theta}_2$ using $\{\hat{\mathbf{x}}_t\}_{t=1}^T$ and moment conditions that correct for uncertainty $\{\mathbf{u}_t\}_{t=1}^T$ in the estimated pricing factors, i.e. $\hat{\mathbf{x}}_t = \mathbf{x}_t^o + \mathbf{u}_t$ where $\mathbf{x}_t^o$ denotes the true factor value. Andreasen & Christensen (2015) show that this corresponds to running the modified regression

$$
\begin{aligned}
\begin{bmatrix} \hat{\mathbf{h}}_{\mathbf{x}}^{step2} & \hat{\mathbf{h}}_0^{step2} \end{bmatrix} =\ & \left(\sum_{t=1}^{T-1}\begin{bmatrix} \hat{\mathbf{x}}_{t+1}\hat{\mathbf{x}}_t' - \widehat{Cov}\left(\mathbf{u}_{t+1},\mathbf{u}_t\right) & \hat{\mathbf{x}}_{t+1} \end{bmatrix}\right) \\
& \times \left(\sum_{t=1}^{T-1}\begin{bmatrix} \hat{\mathbf{x}}_t\hat{\mathbf{x}}_t' - \widehat{Var}\left(\mathbf{u}_t\right) & \hat{\mathbf{x}}_t \\ \hat{\mathbf{x}}_t' & 1 \end{bmatrix}\right)^{-1},
\end{aligned}
\tag{7}
$$

$$
\begin{aligned}
\widehat{Var}\left(\hat{\mathbf{w}}_{t+1}\right)^{step2} =\ & \frac{1}{T-1-n_x-1}\sum_{t=1}^{T-1}\hat{\widehat{\mathbf{w}}}_{t+1}\hat{\widehat{\mathbf{w}}}_{t+1}' - \frac{1}{T-1}\sum_{t=1}^{T-1}\left(\widehat{Var}\left(\mathbf{u}_t\right) + \hat{\mathbf{h}}_{\mathbf{x}}\widehat{Var}\left(\mathbf{u}_t\right)\hat{\mathbf{h}}_{\mathbf{x}}'\right) \\
& + \frac{1}{T-1}\sum_{t=1}^{T-1}\left(\widehat{Cov}\left(\mathbf{u}_{t+1},\mathbf{u}_t\right)\hat{\mathbf{h}}_{\mathbf{x}}' + \hat{\mathbf{h}}_{\mathbf{x}}\widehat{Cov}\left(\mathbf{u}_t,\mathbf{u}_{t+1}\right)\right),
\end{aligned}
\tag{8}
$$

with $\hat{\mathbf{\Sigma}}^{step2}\left(\hat{\mathbf{\Sigma}}^{step2}\right)' = \widehat{Var}\left(\hat{\mathbf{w}}_{t+1}\right)^{step2}$ and $\hat{\widehat{\mathbf{w}}}_{t+1} = \hat{\mathbf{x}}_{t+1} - \hat{\mathbf{h}}_0 - \hat{\mathbf{h}}_{\mathbf{x}}\hat{\mathbf{x}}_t$. The estimators of $Var\left(\mathbf{u}_t\right)$, $Cov\left(\mathbf{u}_{t+1},\mathbf{u}_t\right)$, and $Cov\left(\mathbf{u}_t,\mathbf{u}_{t+1}\right)$ are provided in Andreasen & Christensen (2015). The asymptotic

distribution of $\boldsymbol{\theta}_2$ is given by

$$\sqrt{T}\left(\boldsymbol{\theta}_2^{step2} - \boldsymbol{\theta}_2^o\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \left(\mathbf{R}_o^{\boldsymbol{\theta}_2}\mathbf{S}_o^{-1}\left(\mathbf{R}_o^{\boldsymbol{\theta}_2}\right)'\right)^{-1}\right), \tag{9}$$

when $T \longrightarrow \infty$. Here, $\mathbf{R}_o^{\boldsymbol{\theta}_2}$ denotes the Jacobian of the moment conditions with respect to $\boldsymbol{\theta}_2$ and $\mathbf{S}_o$ is the long-run variance of the moment conditions, obtained by the Newey-West estimator with a bandwidth of one as in Andreasen & Christensen (2015).

3.3.1 The SR approach: Bias-adjusting step 2

It is well-known that the standard moment conditions to estimate VAR models (extended in (7) to account for generated regressors) give biased estimates of $\hat{\mathbf{h}}_{\mathbf{x}}$ in finite samples (see for instance Yamamoto & Kunitomo (1984)). Bauer, Rudebusch & Wu (2012) show that this bias may be substantial for Gaussian ATSMs and have sizeable effects on model-implied term premium. A popular method to reduce the bias is to apply a bootstrap. The bias is then estimated by $\bar{\mathbf{h}}_{\mathbf{x}} - \hat{\mathbf{h}}_{\mathbf{x}}$, where $\bar{\mathbf{h}}_{\mathbf{x}}$ denotes the average estimate of $\mathbf{h}_{\mathbf{x}}$ in the bootstrap, and the bias-adjusted estimate is then given by $\hat{\mathbf{h}}_{\mathbf{x}}^{adj} = \hat{\mathbf{h}}_{\mathbf{x}} - \left(\bar{\mathbf{h}}_{\mathbf{x}} - \hat{\mathbf{h}}_{\mathbf{x}}\right)$. We cannot directly apply the standard bootstrap for VAR models in the SR approach due to the presence of generated regressors, and we therefore generalize it in Appendix A to account for this feature.

Given the persistent nature of the pricing factors in DTSMs, the bias-adjusted estimate $\hat{\mathbf{h}}_{\mathbf{x}}^{adj}$ is unfortunately often pushed into the non-stationary region. To induce stationarity, Kilian (1998) therefore suggests down-scaling the bias-adjustment until all eigenvalues of $\hat{\mathbf{h}}_{\mathbf{x}}^{adj}$ are inside the unit circle. That is, consider $\delta_{i+1} = \delta_i - \varepsilon$, with $\delta_1 = 1$ and $\varepsilon = 0.01$ as in Kilian (1998), and iterate on

$$\hat{\mathbf{h}}_{\mathbf{x}}^{adj,B}\left(\delta\right) = \hat{\mathbf{h}}_{\mathbf{x}} - \delta \times \left(\bar{\mathbf{h}}_{\mathbf{x}} - \hat{\mathbf{h}}_{\mathbf{x}}\right) \tag{10}$$

until all eigenvalues of $\hat{\mathbf{h}}_{\mathbf{x}}^{adj,B}\left(\delta_i\right)$ are inside the unit circle. It should be noted, however, that the value of ε is not derived from any optimality conditions or data-driven selection criteria.

Although Kilian's method to induce stationarity may have minor effects on conditional moments in VAR models, as used for impulse response functions in Kilian (1998), it has substantial effects on any unconditional moments. To realize this, suppose we consider a sequence of grids for δ_i constructed such that the length of the largest eigenvalue of $\hat{\mathbf{h}}_{\mathbf{x}}^{adj,B}\left(\delta_i\right)$ converges to one. This implies that the process for $\mathbf{x}_t$ converges to a non-stationary VAR model with infinite unconditional second moments. In other

words, Kilian's method implies that unconditional moments in the VAR model depend on the grid for δ_i and are therefore not uniquely determined.

As a supplement to Kilian's method, we therefore suggest a data-driven procedure to determine δ. When presenting our method, we first consider the case where $x_{i,t}$ is known, before accounting for measurement errors in $x_{i,t}$ as implied by the SR approach. Our method is based on the observation that the standard estimator of the unconditional variance in $x_{i,t}$, i.e. $\sigma^2_{i,sample} = \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)^2 / (T-1)$ with $\bar{x}_i = \sum_{t=1}^{T} x_{i,t}/T$, is unbiased when $x_{i,t}$ is Gaussian. We therefore suggest to determine δ in (10) by minimizing the distance between $\sigma^2_{i,sample}$ and the variance of $x_{i,t}$ in the VAR model for $i = 1, 2, ..., n_x$. The latter estimate is computed for a given value of δ and is therefore denoted $\sigma^2_{i,VAR}(\delta)$. More formally, we let

$$\hat{\delta} = \arg \min_{\delta \in [\delta_{lower}, 1]} \sum_{i=1}^{n_x} \left(\frac{\sigma^2_{i,VAR}(\delta) - \hat{\sigma}^2_{i,sample}}{\hat{\sigma}^2_{i,sample}} \right)^2 . \tag{11}$$

Monte Carlo evidence in Table 1 suggests that down-scaling the bias *and* the initial estimate of $\mathbf{h_x}$ gives slightly lower bias than only down-scaling the estimated bias when δ is determined using (11). The better performance is related to a larger value of $\hat{\delta}$ when also down-scaling $\hat{\mathbf{h}}_{\mathbf{x}}$, implying that more of the bias-adjustment is preserved. For instance, when using the estimated factor dynamics from the Gaussian ATSM in the long sample and $T = 250$ in our Monte Carlo study, the average of $\hat{\delta}$ across all draws is 0.9921 when down-scaling the bias and $\hat{\mathbf{h}}_{\mathbf{x}}$, whereas the average of $\hat{\delta}$ falls to 0.6950 when only down-scaling the bias. Hence, we prefer the adjustment

$$\hat{\mathbf{h}}_{\mathbf{x}}^{adj,*}(\delta) = \delta \times \left(\hat{\mathbf{h}}_{\mathbf{x}} - \left(\bar{\mathbf{h}}_{\mathbf{x}} - \hat{\mathbf{h}}_{\mathbf{x}} \right) \right) \tag{12}$$

and determine δ using (11). As expected, the Monte Carlo study in Table 1 also shows that the data-driven methods for δ give smaller bias in the unconditional standard deviations of $\mathbf{x}_t$ compared to Kilian's method. Another advantage of (12) is that it always ensures stationarity of VAR models, contrary to the specification in (10). Our method to induce stationarity is summarized in Appendix B, which also describes how to account for measurement errors in $\mathbf{x}_t$ as implied by the SR approach. Unless stated otherwise, we use the bias-adjustment in (12) throughout the paper.

< Table 1 about here >

3.4 The SR approach: Step 3

The elements of $\boldsymbol{\Sigma}$ appear in $\boldsymbol{\theta}_{12}$ which are estimated in both the first and second step of the SR approach. Andreasen & Christensen (2015) suggest letting $\hat{\boldsymbol{\theta}}_{12}^{step3} = \boldsymbol{\Lambda}\hat{\boldsymbol{\theta}}_{12}^{step1} + (\mathbf{I} - \boldsymbol{\Lambda})\hat{\boldsymbol{\theta}}_{12}^{step2}$, where $\boldsymbol{\Lambda}$ is determined by minimizing the variance of $\hat{\boldsymbol{\theta}}_{12}^{step3}$. We generally find that $\hat{\boldsymbol{\Sigma}}^{step1}$ is estimated very inaccurately compared to $\hat{\boldsymbol{\Sigma}}^{step2}$, meaning that the time series estimate $\hat{\boldsymbol{\Sigma}}^{step2}$ cannot be improved by adding cross-section information from $\hat{\boldsymbol{\Sigma}}^{step1}$, i.e. $\boldsymbol{\Lambda} \approx \mathbf{0}$.[11] Hence, the adopted estimate of $\boldsymbol{\Sigma}$ after the first two steps is simply $\hat{\boldsymbol{\Sigma}}^{step2}$.

Based on $\hat{\boldsymbol{\Sigma}}^{step2}$ we then condition on this value and re-estimate $\boldsymbol{\theta}_{11}$, i.e. $\hat{\boldsymbol{\theta}}_{11}^{step3} = \arg\min_{\boldsymbol{\theta}_{11}\in\Theta_{11}} Q_{1:T}^{step3}$, where $Q_{1:T}^{step3} \equiv \frac{1}{2N}\sum_{t=1}^{T}\sum_{j=1}^{n_{y,t}}\left(y_{t,m_j} - g_{m_j}\left(\hat{\mathbf{x}}_t\left(\boldsymbol{\theta}_{11}, \hat{\boldsymbol{\Sigma}}^{step2}\right); \boldsymbol{\theta}_{11}, \hat{\boldsymbol{\Sigma}}^{step2}\right)\right)^2$. Here, Θ_{11} denotes the feasible domain of $\boldsymbol{\theta}_{11}$. Andreasen & Christensen (2015) show consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_{11}^{step3}$ with

$$\widehat{Var}\left(\hat{\boldsymbol{\theta}}_{11}^{step3}\right) = \frac{\hat{\mathbf{V}}_{\boldsymbol{\theta}_{11}}^{step3}\left(\hat{\boldsymbol{\Sigma}}^{step2}\right)}{N} + \hat{\mathbf{K}}\widehat{Var}\left(\hat{\boldsymbol{\Sigma}}^{step2}\right)\hat{\mathbf{K}}'. \tag{13}$$

The term $\hat{\mathbf{V}}_{\boldsymbol{\theta}_{11}}^{step3}\left(\hat{\boldsymbol{\Sigma}}^{step2}\right)/N$ is given by (6) when used on the subset of $\boldsymbol{\theta}_1$ corresponding to $\boldsymbol{\theta}_{11}$, whereas $\mathbf{K} \equiv \partial\hat{\boldsymbol{\theta}}_{11}^{step3}\left(\boldsymbol{\Sigma}\right)/\partial vech\left(\boldsymbol{\Sigma}\right)'$ and estimated as in Andreasen & Christensen (2015). Given the estimated factors $\left\{\hat{\mathbf{x}}_t\left(\hat{\boldsymbol{\theta}}_{11}^{step3}, \hat{\boldsymbol{\Sigma}}^{step2}\right)\right\}_{t=1}^{T}$, we finally update our estimates of $\boldsymbol{\theta}_2$ using (7) and (8).

3.4.1 The SR approach: A residual-based bootstrap for step 3

Although the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{11}^{step3}$ for $n_{y,t} \longrightarrow \infty$ performs well in finite samples with just $n_{y,t} = 25$ according to Andreasen & Christensen (2015), one may nevertheless be hesitant to use it for inference given the relatively small value of $n_{y,t}$. To address this potential concern, we next describe another approximation to the distribution of $\hat{\boldsymbol{\theta}}_{11}^{step3}$ using a bootstrap. We start by re-writing the non-linear regression in (5) in the stacked form

$$\mathbf{Y}_j = \mathbf{G}_j\left(\mathbf{x}_{1:T}; \boldsymbol{\theta}_1\right) + \mathbf{v}_j, \tag{14}$$

---

[11] Similar findings are reported in the Monte Carlo studies for a Gaussian ATSM and a QTSM in Andreasen & Christensen (2015).

where

$$
\mathbf{Y}_j \equiv \begin{bmatrix} y_{1,m_j} \\ y_{2,m_j} \\ ... \\ y_{T,m_j} \end{bmatrix}, \quad \mathbf{G}_j\left(\mathbf{x}_{1:T};\boldsymbol{\theta}_1\right) \equiv \begin{bmatrix} g_{m_j}\left(\mathbf{x}_1;\boldsymbol{\theta}_1\right) \\ g_{m_j}\left(\mathbf{x}_2;\boldsymbol{\theta}_1\right) \\ ... \\ g_{m_j}\left(\mathbf{x}_T;\boldsymbol{\theta}_1\right) \end{bmatrix}, \quad \mathbf{v}_j \equiv \begin{bmatrix} v_{1,m_j} \\ v_{2,m_j} \\ ... \\ v_{T,m_j} \end{bmatrix}
$$

for $j = 1, 2, ..., n_y$. Given that $\boldsymbol{\theta}_1$ *and* $\{\mathbf{x}_t\}_{t=1}^T$ are parameters in the first and third step of the SR approach, we may therefore apply the well-known residual-based bootstrap for a multivariate regression model in a cross-section setting. The restriction $\boldsymbol{\Lambda} = \mathbf{0}$ implies that it is sufficient to bootstrap the third step of the SR approach for inference on $\boldsymbol{\theta}_{11}$.[12] The steps are:

Step A: Run the SR approach on $\{\mathbf{Y}_j\}_{j=1}^{n_y}$ and obtain $\hat{\boldsymbol{\theta}}_1^{step3}$ and $\left\{\hat{\mathbf{x}}_t^{step3}\right\}_{t=1}^T$. The fitted observations are denoted $\hat{\mathbf{Y}}_j \equiv \mathbf{G}_j\left(\hat{\mathbf{x}}_{1:T}^{step3};\hat{\boldsymbol{\theta}}_1^{step3}\right)$ and the estimated residuals are $\hat{\mathbf{v}}_j = \mathbf{Y}_j - \hat{\mathbf{Y}}_j$, where we re-center $\hat{v}_{t,m_j}$ along the cross-section dimension to ensure $\sum_{j=1}^{n_y} \hat{v}_{t,m_j}/n_y = 0$ for all t. Let $b = 1$.

Step B: Fit a pooled stationary autoregressive $AR(p)$ model to $\hat{\mathbf{v}}_j$. The estimated model is denoted $\hat{\mathbf{v}}_j = \sum_{i=1}^p \hat{\phi}_i \hat{\mathbf{v}}_{j-i} + \hat{\boldsymbol{\epsilon}}_j$, where $\hat{\phi}_i$ is a scalar for $i = 1, 2, ..., p$.

Step C: Construct the bootstrap sample $\mathbf{Y}_j^{*,(b)} = \hat{\mathbf{Y}}_j + \hat{\mathbf{v}}_j^{*,(b)}$ using $\hat{\mathbf{v}}_j^{*,(b)} = \sum_{i=1}^p \hat{\phi}_i \hat{\mathbf{v}}_{j-i}^{*,(b)} + \hat{\boldsymbol{\epsilon}}_j^{*,(b)}$ for $j = 1, 2, ..., n_y$, where $\hat{\boldsymbol{\epsilon}}_j^{*,(b)}$ is obtained by resampling with replacement from $\{\hat{\boldsymbol{\epsilon}}_j\}_{j=1}^{n_y}$.

Step D: Condition on $\hat{\boldsymbol{\theta}}_{12}^{step3,(b)} = \hat{\boldsymbol{\theta}}_{12}^{step2,(b)}$, use $\left\{\mathbf{Y}_j^{*,(b)}\right\}_{j=1}^{n_y}$ in the third step of the SR approach to obtain $\hat{\boldsymbol{\theta}}_{11}^{step3,(b)}$ and $\left\{\hat{\mathbf{x}}_t^{step3,(b)}\right\}_{t=1}^T$.

Step E: If $b < \mathcal{B}$, then $b = b + 1$ and go to step C.

Although the residual-based bootstrap is well-known, it is useful to highlight a few details specific to the SR approach. First, in the absence of an intercept in (14), it is necessary to re-center the residuals in Step A to have zero mean and hence that the bootstrap samples in Step C are from a correctly specified model for the conditional mean. Second, we follow Bühlmann (1997) and use an $AR(p)$ model to account for cross-correlation in the residuals when ordered by maturity.[13] Third, by resampling the entire vector $\hat{\boldsymbol{\epsilon}}_j$ in Step C, the variance and covariance structure in the residuals is preserved (see MacKinnon (2009)),

---

[12] See our technical appendix for how to bootstrap the first and third step of the SR approach when $\boldsymbol{\Lambda} \neq \mathbf{0}$.

[13] See Bühlmann (1997) for guidance on how to determine the lag length p in the $AR(p)$ model. The cross-correlation may alternatively be captured using the moving block bootstrap for the residuals. However, unreported simulation results suggest that the $AR(p)$ model is better at capturing the cross-section correlation and hence outperforms the moving block bootstrap in our case.

meaning that the bootstrap accounts for time-variation in the second moments of v_{t,m_j}. Fourth, by drawing from $\hat{\boldsymbol{\theta}}_{12}^{step2,(b)}$ in Step D we condition on the distribution of $\hat{\boldsymbol{\theta}}_{12}^{step2}$ and incorporate this source of uncertainty in the bootstrap. The draws for $\hat{\boldsymbol{\theta}}_{12}^{step2,(b)}$ may be obtained from the bootstrap in Step 2 of the SR approach or from the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{12}^{step2}$ in (9).

Finally, when estimating the AR(p) model in Step B, Appendix C shows that OLS is biased in finite samples due to estimation error in the pricing factors. It is, however, straightforward to correct for this bias by running a preliminary bootstrap with $\mathcal{B}_1 < \mathcal{B}$ draws, where in Step D we use OLS to estimate the AR(p) model $\left\{\hat{\phi}_{j,OLS}^{(b)}\right\}_{j=1}^p$ on the residuals in the bootstrap sample, i.e. on $\hat{\mathbf{v}}_j^{*,(b)} = \mathbf{Y}_j^{*,(b)} -$ $\mathbf{G}_j\left(\hat{\mathbf{x}}_{1:T}^{step3,(b)}; \hat{\boldsymbol{\theta}}_{11}^{step3,(b)}\right)$ for $j = 1, 2, ..., n_y$. The bias-adjusted estimates are then given by $\hat{\phi}_i = 2\hat{\phi}_{i,OLS} - \bar{\phi}_i$ for $i = 1, 2, ..., p$, where $\bar{\phi}_i = \frac{1}{\mathcal{B}_1}\sum_{b=1}^{\mathcal{B}_1} \hat{\phi}_{i,OLS}^{(b)}$ and $\hat{\phi}_{i,OLS}$ denotes the initial OLS estimate in Step B.[14]

3.5 Monte Carlo study: SR approach versus QML

We finally explore the finite sample properties of bootstrapping the entire SR approach in a Monte Carlo study, using a one-factor Gaussian ATSM to reduce the computational burden (see Appendix D for the remaining details of the Monte Carlo study). Hence, we let $\boldsymbol{\theta}_{11} \equiv \begin{bmatrix} \alpha & \Phi_{11} \end{bmatrix}$ and $\boldsymbol{\theta}_2 \equiv \begin{bmatrix} h_0 & h_x & \Sigma_{11} \end{bmatrix}$ in this applicaiton.

Our results in Table 2 show that the asymptotic distributions of $\boldsymbol{\theta}_{11}$ and $\boldsymbol{\theta}_2$ (with bias-adjustment) serve as useful approximations in finite samples with standard errors and rejection probabilities close to their desired values. These results hold even when measurement errors are auto-correlated (Case II), display time-varying heteroskedasticity (Case III), are cross-sectionally correlated (Case IV), or when the three features are combined (Case V). Bootstrapping the SR approach generally provides an refinement to asymptotic inference, as bootstrapped rejection probabilities in most cases are closer to 5% than those from the asymptotic distribution. The largest improvement appears for $\boldsymbol{\theta}_{11}$ with cross-correlation in the measurement errors (i.e. Case IV and V), where the bootstrap corrects the positive bias in the asymptotic standard errors which otherwise generates too low rejection probabilities. The satisfying performance of the bootstrap in these two cases is closely related to the estimated AR(1) model for cross-correlation, where the bias-adjustment returns nearly unbiased estimates of $\phi_1 = 0.40$ with $\hat{\phi}_1 = 0.396$ in both cases

---

[14] Although less likely, if the bias-adjusted estimates in the AR(p) model violate the stationarity requirement, it may be imposed using the same principle as in Section 3.3.1, that is by down-scaling $\left\{\hat{\phi}_i\right\}_{i=1}^p$ by a constant δ which we determine by minimizing the distance between the unconditional variance in the AR(p) model and $\frac{1}{N-T\times n_x}\sum_{t=1}^T\sum_{j=1}^{n_{y,t}} \hat{v}_{t,m_j}^2$.

using $\mathcal{B}_1 = 100$. The corresponding averages of the unadjusted OLS estimates in the Monte Carlo study are $\hat{\phi}_{1,OLS} = 0.336$.

For comparability with much of the existing literature, Table 2 also shows maximum likelihood (ML) estimates using the Kalman filter with $\mathcal{IID}$ measurement errors. When the log-likelihood function $\mathcal{L}$ is correctly specified in Case I, we obtain the well-known results that ML is more efficient than the SR approach and provides reliable estimates of standard errors and rejection probabilities, except for the ML estimates of h_0 and h_x due to their well-known biases. The next rows in Table 2 examine the robustness of ML when measurement errors deviate from the $\mathcal{IID}$ assumption used in the Kalman filter. That is, we deliberately run the Kalman filter with misspecified measurement errors and obtain a quasi likelihood function. Table 2 shows that this QML approach is surprisingly robust with nearly no additional bias in the estimates when the measurement errors display auto-correlation, time-varying heteroskedasticity, and cross-correlation. To understand this result, recall that its score function is

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = -\sum_{t=1}^{T} \left(\frac{\partial \mathbf{v}_t}{\partial \theta_i}\right)' \mathbf{F}_t^{-1} \mathbf{v}_t - \sum_{t=1}^{T} tr \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \theta_i} \left(\mathbf{I} - \mathbf{F}_t^{-1} \mathbf{v}_t \mathbf{v}_t'\right)\right) \quad \text{for } i = 1, 2, .., n_\theta \qquad (15)$$

where tr denotes the trace-operator, $\mathbf{v}_t$ is the one-step ahead prediction error for $\mathbf{y}_t$, $\mathbf{F}_t$ is the conditional covariance matrix of $\mathbf{v}_t$, and n_θ refers to the number of estimated parameters. The law of iterated expectations implies $E\left[\sum_{t=1}^{T} \left(\frac{\partial \mathbf{v}_t}{\partial \theta_i}\right)' \mathbf{F}_t^{-1} \mathbf{v}_t\right] = \mathbf{0}$ even when measurement errors deviate from the $\mathcal{IID}$ assumption applied in the Kalman filter, whereas the unconditional expectation of the second sum in (15) is non-zero with the considered misspecifications. Our results therefore suggest that this second sum is nearly zero in our setting where measurement errors are relatively small compared to the variability in bond yields, as also found in empirical studies.[15] However, the usual QML standard errors have a substantially negative bias when the measurement errors are auto-correlated and heteroskedastic, which generates too high rejection probabilities for the QML approach in Case II, III, and V.[16] We emphasize that these shortcomings cannot be expected to disappear when increasing the sample size, as also shown in an unreported Monte Carlo study where the sample size is increased from 250 to 500 observations.[17] It is finally worth noticing from Table 2 that the QML approach in Case II, III, and V no longer dominates

---

[15]Measurement errors are typically within the range of ± 30 basis points in our simulations but may reach levels of ± 60 basis points with time-varying heteroskedasticity. Our technical appendix documents extra biases in the QML estimates when bond yields display even larger measurement errors.

[16]Standard errors and rejection probabilities for the QML estimates are computed by pre- and post-multiplying the variance of the score by the inverse of the Hessian matrix, except in Case I where we use the inverse of the variance of the score. The Hessian matrix is computed as in Harvey (1989).

[17]See the technical appendix for further details.

the SR approach in terms of efficiency.

To summarize, bootstrapping the SR approach delivers nearly unbiased estimates of all parameters and leads to reliable inference even when the measurement errors in bond yields feature auto-correlation, time-varying heteroskedasticity, and cross-correlation, as typically found in empirical applications (see for instance Kim & Singleton (2012) and Adrian, Crump & Moench (2013)). In contrast, likelihood inference based on the Kalman filter and $\mathcal{IID}$ measurement errors generates biased estimates in the $\mathbb{P}$ dynamics for the pricing factors and induces too high rejection probabilities when measurement errors in bond yields are auto-correlated and heteroskedastic.

< Table 2 about here >

4 Empirical results: In-sample performance

This section estimates the considered DTSMs. Section 4.1 presents the data, and we discuss in-sample model fit in Section 4.2 and 4.3. Selected parameter estimates are provided in Section 4.4, while the following subsections explore how well the models match various moments not included in the estimation.

4.1 Data

We use end-of-month nominal US Treasury bond yields from June 1961 to December 2013 as published by the Federal Reserve Board.[18] The SR approach is constructed for a setting with many observables available each time period, and we therefore include more bond yields than typically used when taking DTSMs to the data. Given our interest in the 10-year term structure, we include bond yields in the 0.5- to 3-year maturity range at maturities three months apart, whereas bond yields in the remaining segment of the 10-year term structure are included at maturities six months apart. This gives 25 points on the yield curve, except before September 1971 where bond yields in the 7- to 10-year maturity range are unavailable. These missing observations are accounted for in the SR approach by running the cross-section regressions on the available set of bond yields in a given time period.

As mentioned above, we test the performance of the DTSMs considered on two samples; a long sample from June 1961 to December 2013 ($T = 631$), and a short sample from January 1990 to December 2013 ($T = 288$).

---

[18] All bond yields applied in this paper are computed using the estimated parametric form for the yield curves in Gürkaynak, Sack & Wright (2007).

4.2 Goodness of in-sample fit

A preliminary estimation suggests that all models are badly identified in the SR approach given the standard normalization restrictions. To illustrate this for the Gaussian ATSM, recall that the solution to bond prices with our normalization is

$$A_j = -\alpha + A_{j-1} + \frac{1}{2}\mathbf{B}'_{j-1}\mathbf{\Sigma}\mathbf{\Sigma}'\mathbf{B}_{j-1} \approx -\alpha + A_{j-1}, \tag{16}$$

because $\mathbf{\Sigma}\mathbf{\Sigma}'$ is very small, and

$$\mathbf{B}'_j = -\mathbf{1}' + \mathbf{B}'_{j-1}\left(\mathbf{I} - \mathbf{\Phi}\right). \tag{17}$$

Given that $\mathbf{\Sigma}$ is badly identified from the cross-section dimension of bond yields due to (16), the ordering of the factors is also badly identified.[19] That is, we obtain nearly identical values for the objective functions in the first and third step of the SR approach by changing the order of the eigenvalues in $\mathbf{\Phi}$.[20] To eliminate this identification problem we therefore require that all eigenvalues of $\mathbf{\Phi}$ are increasing in the Gaussian ATSM and in the two ZLB models.[21]

To evaluate the in-sample fit, we start by comparing the objective functions from the first step, which for convenience are reported as $\tilde{Q}^{step1}_{1:T} \equiv 100\sqrt{2 \times Q^{step1}_{1:T}}$, i.e. the standard deviations of all residuals in the sample. The left-hand side of Table 3 shows that with two and three pricing factors, the QTSM outperforms the SRM, which does better than the ATSM. We also find that including a fourth factor improves in-sample fit by more than 50% compared to three-factor models. Adrian et al. (2013) also provide evidence for more than three pricing factors in the Gaussian ATSM, and our results suggest that the same applies for the QTSM and the SRM. Figure 2 shows that the better in-sample performance of four-factor models is explained by a closer fit to short- and long-term bond yields, where the standard deviation of all pricing errors $\hat{\sigma}_{m_j} = 100\sqrt{\frac{1}{T}\sum_{t=1}^{T}\hat{v}^2_{t,m_j}}$ are smaller than 2 basis points. Importantly, the SRM marginally outperforms the QTSM with four pricing factors and hence provides the best in-sample fit. We consider this a somewhat surprising finding, given that the four-factor QTSM has nine additional parameters compared to the corresponding SRM. Note finally that we obtain the same ranking of the models when only focusing on the in-sample fit during the ZLB period from 2009 to 2013, as reported in

---

[19] A similar finding is reported in Ait-Sahalia & Kimmel (2010) using likelihood inference.

[20] Note in relation to (17) that the pricing factors in the regression filter are identified even when eigenvalues under $\mathbb{Q}$ are identical due to Proposition 1 in Joslin et al. (2011).

[21] This empirical observation is related to Hamilton & Wu (2012), showing that eigenvalues of $\mathbf{\Phi}$ must be increasing in Gaussian ATSMs to ensure identification.

BANK OF ENGLAND

the squared brackets in Table 3.

< Figure 2 about here >

The right part of Table 3 reports the scaled objective functions from the third step in the SR approach, i.e. $\tilde{Q}_{1:T}^{step3} \equiv 100\sqrt{2 \times Q_{1:T}^{step3}}$, where $\boldsymbol{\Sigma}$ is estimated from the time series dimension instead of the cross-section dimension as in the first step. For all models and in both samples, $\tilde{Q}_{1:T}^{step3}$ is only marginally higher than $\tilde{Q}_{1:T}^{step1}$, meaning that the in-sample fit of bond yields is nearly unaffected by the alternative estimator of $\boldsymbol{\Sigma}$. The only possible exception is the two-factor SRM in the long sample where $\tilde{Q}_{1:T}^{step1} = 9.414$ and $\tilde{Q}_{1:T}^{step3} = 11.936$. It is therefore reasonable to believe that the dependence on the $\mathbb{P}$ dynamics through $\boldsymbol{\Sigma}$ is minimal in our case, and that results in the third step of the SR approach largely remain robust to the functional form of $\mathbf{f}(\mathbf{x}_t)$. Unreported results show that the in-sample fit is also robust to omitting the bias-adjustment in $\hat{\boldsymbol{\theta}}_2$, partly because $\boldsymbol{\Sigma}$ is badly identified from the cross-section dimension of bond yields and partly because the bias-adjustment in $\hat{\boldsymbol{\Sigma}}^{step2}$ is small.

Based on these findings we conclude that accounting for the ZLB by either QTSMs or SRMs give a better in-sample fit of US bond yields compared to Gaussian ATSMs. We also find that the QTSM outperforms the SRM with two and three pricing factors, whereas the SRM does marginally better than the QTSM with four factors. We therefore conclude that the relative in-sample fit of the two considered mechanisms to enforce the ZLB depends on the richness of the factor structure.

4.3   Model ranking using a QML approach

The in-sample model ranking in the SR approach using squared pricing errors is intuitive and robust to several model misspecifications, but the approach relies on more bond yields in the cross-section dimension than typically considered when estimating DTSMs. For comparability with much of the existing liturature, this section studies the model ranking when using the conventional QML approach on seven bond yields (maturities of 0.5, 1, 2, 3, 5, 7, and 10 years).[22] We further assume that each of these bond yields are contaminated with normally distributed measurement errors with the same standard deviation. The latter implies that all bond yields are weighted equally in the quasi log-likelihood function, as also assumed in our implementation of the SR approach. The quasi log-likelihood function is evaluated

---

[22]We are grateful to an anonymous referee for suggesting this robustness check.

BANK OF ENGLAND

by central difference Kalman filter (CDKF) of Norgaard, Poulsen & Ravn (2000), which is more accuate than the commonly used extended Kalman filter.

Table 4 reports the optimized quasi log-likelihood values $\mathcal{L}_{1:T}^{CDKF}$ from re-estimating the nine models in both samples.[23] Measuring the goodness of in-sample fit by the magnitude of $\mathcal{L}_{1:T}^{CDKF}$, we find that the QTSM is the preferred model with two and three pricing factors, whereas the SRM does best with four factors in the short but not in the long sample. Thus, the in-sample model ranking by the QML estimator coincides broadly with the SR approach, differing only in the long sample where the SR approach marginally prefers the four-factor SRM to the QTSM.

Apart from being computationally more involved than the SR approach, the QML estimator is often biased in small samples and may induce unreliable standard errors with misspecified measurement errors as shown in Section 3.5. The SR approach is robust to both shortcomings when using the suggested bootstrap, and we therefore focus on results from the SR approach in the remainder of the paper.

< Table 4 about here >

4.4 Model estimates

We next discuss the estimated coefficients as implied by the SR approach. In the interest of space, focus is here devoted to models with three pricing factors as typically considered in the literature.[24] Table 5 for the long sample shows that the Gaussian ATSM displays the usual properties with stationary but very persistent factors under both the $\mathbb{Q}$ and the $\mathbb{P}$ measure, as elements in $\hat{\Phi}$ are small but greater than zero and the largest eigenvalue of $\hat{\mathbf{h}}_{\mathbf{x}}$ equals 0.9914. Similar properties hold for the pricing factors in the QTSM, where $\hat{\boldsymbol{\Psi}}$ enforces the ZLB by having eigenvalues of 0.0000, 0.0134, and 2.9866. The requirement of non-negative eigenvalues in $\hat{\boldsymbol{\Psi}}$ is equivalent to imposing $\hat{A}_{21}^2 \leq 1$ and $\hat{A}_{31}^2 + \left(1 - \hat{A}_{21}^2\right) \hat{A}_{32}^2 \leq 1$, implying that the absolute values of $\hat{A}_{12}$ and $\hat{A}_{31}$ cannot exceed one. The presence of a zero eigenvalue means that we are at the boundary (as $\hat{A}_{31}^2 + \left(1 - \hat{A}_{21}^2\right) \hat{A}_{32}^2 \approx 1$), and the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{11}^{step3}$ is therefore unlikely to perform well in this case. The provided bootstrapped confidence intervals should be more reliable, although we acknowledge that the bootstrap may also be inaccurate in this case, because the bootstrap method is inconsistent when parameters are at the boundary of their domain (see Andrews (2000)).[25] Subject to this qualification, we find that the 95% confidence intervals in Table 5 for $\hat{A}_{23}$ and

---

[23] The estimated coefficients from the QML approach are provided in the technical appendix.

[24] The estimated coefficients in models with two and four pricing factors are provided in the technical appendix.

[25] One alternative to the bootstrapped standard errors could be to adopt a Bayesian perspective within the SR approach and use Markov Chain Monte Carlo to draw from the posterior distribution, provided a distributional assumption is imposed

elements in $\hat{\boldsymbol{\mu}}$ are fairly wide and often asymmetric, whereas the intervals for elements in $\hat{\boldsymbol{\Phi}}$ are much tighter and almost symmetric.

The estimates in the SRM are very similar to those in the Gaussian ATSM, and the asymptotic standard errors are for both models close to those from the bootstrap (provided in brackets in Table 5). Despite the strong similarities between the two models we do observe some differences in the estimates, particularly for $\mathbf{h_x}$.

< Table 5 about here >

Table 6 reveals that the pricing factors for all models in the short sample are slightly less persistent than in the long sample when measured by the largest eigenvalue of $\hat{\mathbf{h}}_{\mathbf{x}}$. In the QTSM, the estimates of $\boldsymbol{\Psi}$ imply eigenvalues of 0.0014, 0.0287, and 2.9699, meaning that the short rate is primarily controlled by one pricing factor as in the long sample, given our normalization with $\boldsymbol{\beta} = \mathbf{0}$. We also find that $\hat{\mathbf{h}}_{\mathbf{x}}$ and $\hat{\boldsymbol{\Sigma}}$ for the Gaussian ATSM differ substantially from the corresponding estimates in the SRM.

< Table 6 about here >

We finally examine the accuracy of our approximation to bond yields in the SRM. Using the estimated parameters and state values from the three-factor model in the long sample, Figure 3 shows that the root mean squared errors (RMSEs) of the approximation are about 0.5 basis points and the absolute errors do not exceed 2 basis points. The true value of bond yields at a given state is here computed by the Monte Carlo method. The approximation is even more accurate in the short sample, where the RMSEs and absolute errors are below 0.5 basis points across all maturities. We therefore conclude that the second-order approximation is highly accurate, as also documented in Priebsch (2013).

< Figure 3 about here >

4.5   Matching key moments of bond yields

We next test the models' ability to match moments not included in the estimation. The first set of moments we consider are the unconditional means and standard deviations of bond yields. Following Campbell & Shiller (1991), we also run the ordinary Campbell-Shiller regressions

$$y_{t+1,j-1} - y_{t,j} = \delta_j + \frac{\phi_j}{j-1}\left(y_{t,j} - r_t\right) + u_{t,j}, \tag{18}$$

---

on the measurement errors in bond yields.

BANK OF ENGLAND

where $u_{t,j} \sim \mathcal{IID}(0, Var(u_{t,j}))$.[26] We then explore if the DTSMs can reproduce the empirical pattern in $\{\phi_j\}_{j=2}^K$ and hence capture key moments of the $\mathbb{P}$ dynamics for bond yields, also known as the LPY(i) test. Following Dai & Singleton (2002), a risk-adjusted version of (18) is given by

$$y_{t+1,j-1} - y_{t,j} - (c_{t+1,j-1} - c_{t,j-1}) + \frac{1}{j-1}\theta_{t,j-1} = \delta_j^{\mathbb{Q}} + \frac{\phi_j^{\mathbb{Q}}}{j-1}(y_{t,j} - r_t) + u_{t,j}^{\mathbb{Q}},$$

where $u_{t,j}^{\mathbb{Q}} \sim \mathcal{IID}\left(0, Var\left(u_{t,j}^{\mathbb{Q}}\right)\right)$, $c_{t,j} \equiv y_{t,j} - \frac{1}{j}\sum_{i=0}^{j-1} E_t[r_{t+i}]$ is the spot term premium, and $\theta_{t,j} \equiv f_{t,j} - E_t[r_{t+j}]$ is the forward term premium with $f_{t,j} \equiv -\log(P_{t,j+1}/P_{t,j})$.[27] If the $\mathbb{Q}$ dynamics are correctly specified by the DTSMs (equivalently to well-specified term premia and $\mathbb{P}$ dynamics), then $\phi_j^{\mathbb{Q}} = 1$ for $j = 2, 3, ..., K$. The ability of DTSMs to match these moments is the LPY(ii) test and studies whether the models capture key moments of the $\mathbb{Q}$ dynamics for bond yields.

We first study models with three pricing factors as commonly considered in the literature. The ability of these models to match the four types of unconditional moments is illustrated in Figure 4 for the long sample. To illustrate the impact of the bias-adjustment in $\boldsymbol{\theta}_2$, charts to the left report the model-implied moments using the unadjusted estimates of $\boldsymbol{\theta}_2$, whereas the adjustment is imposed in charts to the right. The first row in Figure 4 shows that all models underestimate the average level of bond yields when $\boldsymbol{\theta}_2$ is not bias-adjusted, whereas these moments are matched closely when correcting for the bias in $\boldsymbol{\theta}_2$. The unconditional standard deviations of bond yields are matched by the Gaussian ATSM and the SRM irrespective of whether $\boldsymbol{\theta}_2$ is bias-adjusted or not. The QTSM clearly struggles to match the variability in bond yields, although its performance is improved considerably when $\boldsymbol{\theta}_2$ is bias-adjusted. We further observe that only the Gaussian ATSM and the SRM reproduce the downward sloping pattern in $\{\phi_j\}_{j=2}^K$ and pass the LPY(i) test, whereas the QTSM hardly matches this aspect of bond yields. We acknowledge that the empirical loadings in the ordinary Campbell-Shiller regressions are likely to display some instability across subsamples, as argued by Rudebusch & Wu (2007), and the models are therefore not expected to match these loadings perfectly but only to capture their overall pattern. The LPY(ii) test does not suffer from the same instability issues as the desired regression loadings of one for $\phi_j^{\mathbb{Q}}$ must

---

[26]In practice, we run the regressions $y_{t+m,j-m} - y_{t,j} = \delta_j + \phi_j \frac{m}{j-m}(y_{t,j} - y_{t,m}) + u_{t,j}$ with $m = 6$, i.e. the regressions are done for biannual excess returns. We compute these regressions both on observed bond yields and on simulated data from each of the models to obtain the model-implied regression loadings.

[27]As for (18), in practice we run the regressions $y_{t+m,j-m} - y_{t,j} - (c_{t+m,j-m} - c_{t,j-m}) + \frac{m}{j-m}\theta_{t,j-m} = \delta_j^{\mathbb{Q}} + \phi_j^{\mathbb{Q}}\frac{m}{j-m}(y_{t,j} - y_{t,m}) + u_{t,j}^{\mathbb{Q}}$ with $m = 6$, i.e. the regressions are done for biannual risk-adjusted excess returns. We compute these regressions using observed bond yields and model-implied estimates of term premia obtained at $\{\hat{\mathbf{x}}_t\}_{t=1}^T$.

hold in all subsamples, making the LPY(ii) test potentially more informative. Figure 4 shows that the QTSM with the bias-adjustment is broadly successful at satisfying the LPY(ii) test, whereas the other models imply slightly larger deviations of $\phi_j^{\mathbb{Q}}$ from one.

Turning to the short sample in Figure 5, all models match the average level of bond yields and pass the LPY(i) test due to the bias-adjustment in $\boldsymbol{\theta}_2$. The standard deviations of bond yields are slightly underestimated in the Gaussian ATSM and the SRM, whereas these moments are broadly matched by the QTSM. The last row in Figure 5 suggests that the Gaussian ATSM and the SRM are able to pass the LPY(ii) with $\phi_j^{\mathbb{Q}}$ close to one, whereas the QTSM shows clear deviations from one.

< Figure 4 and 5 about here >

We next study the implications of only using two pricing factors for comparison with the findings in Kim & Singleton (2012) on Japanese bond yields. To conserve space, only results from models estimated with the bias-adjustment in $\boldsymbol{\theta}_2$ are reported. Figure 6 suggests that the SRM with two factors largely reproduces all the moments considered in the long sample, whereas the two-factor QTSM hardly matches any of these moments. The performance of the two-factor QTSM improves substantially in the short sample as seen from Figure 7, where the model gives a satisfying fit to nearly all moments and clearly outperforms the SRM.[28] Hence, the ranking of the two-factor ZLB models based on these moments depends crucially on whether the high and volatile interest rates in the 1970s and 1980s are included in the sample. This finding also illustrates that one should be cautious of extrapolating the results from Japan in Kim & Singleton (2012) to the US, because the historical evolution of bond yields before hitting the ZLB and the number of pricing factors have a substantial impact in these models.

< Figure 6 and 7 about here >

We next examine if the performance of three-factor models may be improved by including a fourth pricing factor. For the long sample in Figure 8, we see marginal improvements for the SRM in matching LPY(i) and LPY(ii), whereas the performance of the QTSM is largely unaffected. A fourth pricing factor has also minor effects in the short sample, as this additional factor only helps the SRM to match the standard deviations of bond yields according to Figure 9.

---

[28] Note that the ability of the two-factor QTSM to match the unconditional mean of bond yields, the standard deviation of bond yields, and pass the LPY(i) test is similar to the results reported in Leippold & Wu (2003) on US data from January 1985 to December 1999.

BANK OF ENGLAND

< Figure 8 and 9 about here >

The ability of DTSMs to pass the LPY tests is largely determined by the models' ability to identify term premia correctly, and we therefore report estimates of the 10-year term premium in Figure 10 for the short sample. To illustrate the impact of the bias-adjustment in $\boldsymbol{\theta}_2$, charts in the first column show the 10-year term premium in the Gaussian ATSM with and without the bias-adjustment. The impact of this correction is sizable, as the low unconditional mean of bond yields without the bias-adjustment implies too low expected policy rates and therefore too high term premia in comparison with the bias-adjusted models. We find the same effect of the bias-adjustment for term premia in QTSMs and SRMs, although not displayed in Figure 10.[29] The remaining charts in Figure 10 show that the Gaussian ATSMs with three and four pricing factors may have over-estimated term premia from 2005 to 2009 by as much as 50 basis points when compared to the corresponding QTSMs and SRMs.

< Figure 10 about here >

We summarize our results for the two ZLB models in Table 7, where we use "Q" and "S" to indicate that a given set of moments are approximately matched by the QTSM and the SRM, respectively. This table reveals that the SRM generally outperforms the QTSM in the long sample, as the two-factor QTSM performs poorly, and all QTSMs struggle to match loadings from the ordinary Campbell-Shiller regressions. The results are more mixed in the short sample with three factors, whereas the QTSM is the preferred model with two pricing factors and the SRM dominates with four factors.

< Table 7 about here >

4.6 Matching conditional volatilities of bond yields

The QTSM allows for heteroskedasticity in bond yields through the quadratic terms in the policy rate, and the model may therefore generate time-variation in the conditional volatility of bond yields both close to the ZLB and when this bound is not binding. The SRM also introduces heteroskedasticity in bond yields, but only when the policy rate is close to zero and its variation is compressed by the ZLB. Hence, the two mechanisms to enforce the ZLB imply different implications for the volatility of bond yields, and this section therefore studies how well the QTSM and the SRM match this feature of the

---

[29]This effect of the bias-adjustment is also present in the long sample, although its magnitude is somewhat smaller.

data. The cyclical variation in volatility for these models is largely unaffected by the number of pricing factors and we therefore only show results for three-factor models.[30]

We use two measures of conditional volatility in the data. The first is the rolling standard deviation of bond yields (denoted $\sigma_{t,j}^{Rolling}$) computed from daily observations with a six-month lookback.[31] As a supplement to these non-parametric estimates is the conditional volatility from a GARCH(1,1) model applied to changes in monthly bond yields (denoted $\sigma_{t,j}^{GARCH}$). Figure 11 shows these estimates at four selected maturities and the model-implied volatilities in the long sample. Overall, the two measures of volatility in the data are fairly similar, although $\sigma_{t,j}^{Rolling}$ is more noisy than $\sigma_{t,j}^{GARCH}$. The QTSM captures most of the gradual increase in volatility during the 1960s and 1970s but does not match the very elevated levels in the early 1980s. A similar finding is reported in Ahn et al. (2002). The gradual fall in volatility from the end of 2008 when the policy rate approaches the ZLB is also largely matched by the QTSM. However, the model is unable to reproduce the increase in volatility for the 0.5-, 2-, and 5-year bond yields just before hitting the ZLB. The SRM predicts constant volatility when the policy rate is far from zero, and the model is therefore unable to reproduce the change in volatility before 2008 but matches the fall in volatility at the ZLB.

< Figure 11 about here >

For the short sample starting in 1990, the QTSM is generally less successful in matching volatility according to Figure 12. To see why, observe that volatility in the QTSM is closely related to the level factor and hence the short rate. As shown in Figure 11, this relationship is able to explain much of the variation in volatility from the 1960s to the 1980s but less successful after 1990. For the SRM, the constant volatility before 2008 performs well given the stable volatility regime, and the model matches the fall in volatility after 2008 when policy rates are constrained by the ZLB.

< Figure 12 about here >

To summarize the relative performance of the two models, we regress volatility in the data on a constant and the model-implied volatility. Table 8 confirms our impression from above that the QTSM provides the best fit in the long sample but not in the short sample, where the SRM dominates. The low R^2

---

[30] The number of pricing factors appears only to affect the level of the conditional volatility, which is somewhat higher with two pricing factors than with three and four factors.

[31] These daily bond yields are obtained from the same source as the monthly bond yields used to estimate the DTSMs.

in these regressions also suggests that both models generally struggle to capture the conditional volatility of bond yields. This may indicate that a more flexible functional form for the policy rate is required in models with Gaussian pricing factors or that the factor dynamics should display heteroskedasticity. We return to this issue in Section 6 where the first extension is considered.

< Table 8 about here >

5   Empirical results: performance out-of-sample

This section studies the models' ability to predict future bond yields from January 2005 to December 2013. This forecasting sample is particularly challenging as it contains bond yields i) far from zero, ii) when hitting the ZLB, and iii) a prolonged period at the lower bound. The forecasting study is carried out by recursively re-estimating all models every month to forecasts bond yields up to 12 months ahead. We do so when starting the sample in 1961 and in 1990.[32]

Figure 13 reports the root mean squared prediction errors (RMSPE) by maturity when the estimation is started in 1961. Columns in Figure 13 refer to the number of pricing factors and rows denote the forecast horizon of 1, 3, 6, and 12 months, respectively. Starting with the two-factor models, the QTSM clearly outperforms the Gaussian ATSM at the 1- and 3-month forecast horizons for all maturities, whereas the two models display similar performance when forecasting 6 and 12 months ahead. The two-factor SRM delivers even better forecasts for short- and medium-term bond yields at the 3-, 6-, and 12-month horizons but struggles when predicting long-term bond yields.

Turning to three-factor models, the QTSM and the SRM have very similar forecasting abilities and dominate the Gaussian ATSM for nearly all maturities and forecast horizons. Importantly, the forecasts from the SRM generally *improve* when including a fourth pricing factor, whereas the opposite applies for the QTSM. This suggests that the parsimonious mechanism to enforce the ZLB in SRMs is more robust and less subject to overfitting than the quadratic specification. A careful inspection of Figure 13 reveals that the three-factor QTSM and the three- and four-factor SRM outperform the random walk for short-term bond yields at all forecast horizons.

< Figure 13 about here >

---

[32] Given that the last 12 months of data are reserved for evaluating the final forecasts, each of the nine models is estimated 96 times on both data sets.

The forecasting results when the estimation is started in 1990 are provided in Figure 14. The overall results are very similar to those obtained for the long sample in Figure 13 and we therefore only highlight the following. First, the two-factor SRM generally benefits from the shorter estimation window as its RMSPEs are either lower than the two other models or very close to the best performing model. Second, the QTSM and the SRM with three pricing factors display similar performance. Third, forecasts again generally *improve* in the SRM when adding a fourth factor whereas the opposite generally holds for the QTSM. Finally, regardless of the considered number of pricing factors, the QTSM and the SRM outperform the Gaussian ATSM at nearly all maturities and forecast horizons.[33]

< Figure 14 about here >

In addition to providing more accurate forecasts than the Gaussian ATSM, the QTSM and the SRM also ensure sensible forecasts as predicted bond yields stay non-negative. The same cannot be guaranteed in the Gaussian ATSM, as we illustrate in Figure 15 for the long sample by showing forecasts for the 0.5-year bond yield on two occasions. The first is the end of December 2008, when the policy rate reached the ZLB. Predicted bond yields in the three-factor Gaussian ATSM barely stay positive at the considered forecast horizons, but do not in the four-factor version, where the 0.5-year bond yield is predicted to turn negative after 5 months. The second row of Figure 15 for the end of May 2010 shows that negative forecasts in the Gaussian ATSM occur with two, three, and four pricing factors and even when the policy rate has been at the ZLB for several years.

< Figure 15 about here >

We summarize the forecasting performance of the three models in Table 9 by reporting the average RMSPEs for all bond yields (i.e. the entire yield curve) at various horizons. To facilitate the reading of this table we adopt two coding schemes. The first uses bold to indicate the model with the lowest RMSPEs when conditioning on the number of pricing factors and the starting point for the estimation. The SRM has 15 bold figures, the QTSM has 8, and the Gaussian ATSM has 1. Based on this finding and the results in Figure 13 and 14 we conclude that the SRM generally performs best out of sample and that both models accounting for the ZLB do better than the Gaussian ATSM.

---

[33]Christensen & Rudebusch (2013) also find that a SRM with three pricing factors outperforms the three-factor Gaussian ATSM when forecasting US bond yields out-of-sample.

Our second coding scheme in Table 9 uses a box to indicate the model with the lowest RMSPEs when comparing its forecasts across the starting point for the estimation, i.e. when comparing individual elements in part $\mathcal{A}$ and $\mathcal{B}$ of Table 9. We surprisingly find that starting the estimation in 1961 generally gives the most accurate forecasts, as part $\mathcal{A}$ has 26 boxed figures whereas part $\mathcal{B}$ only has 10. That is, the best forecasts are in general obtained by using a long sample for the estimation, particularly for the SRM. Any finite sample bias in the estimated $\mathbb{P}$ dynamics is unlikely to explain this finding as we bias-adjust $\hat{\boldsymbol{\theta}}_2$ regardless of the starting point for the sample. Instead, the better forecasting performance from using a long sample is likely to be driven by two features. First, the pricing factors and hence bond yields are more persistent in the long sample compared to the short sample (see Section 4.4) and this is likely to improve forecasts, given the strong performance of the random walk. Second, bond yields in the 1960s were fairly low compared to their average level, meaning that the long sample includes bond yields closer to the levels seen after 2008 than a sample starting in 1990.

< Table 9 about here >

6 A hybrid model

The quadratic terms in the QTSM serve a dual purpose as they enforce the ZLB and generate time-varying conditional volatility in bond yields. Our results in Section 4.4 suggest that the estimates of $\boldsymbol{\Psi}$ are constrained by the non-negativity condition requiring $\boldsymbol{\Psi}$ to be positive semi-definite. Hence, there is a trade-off within the QTSM between enforcing the ZLB and matching other features of the data like the conditional volatility of bond yields. This section explores the potential benefit of eliminating this trade-off by considering models where the shadow rate is an unrestricted quadratic function of the pricing factors. This model was first considered by Kim & Singleton (2012) with two pricing factors and extended below to include a third factor. Given that this model merges the QTSM and the SRM considered above, we refer to it as the hybrid model.

This section is structured as follows. We present the hybrid model in Section 6.1, discuss in-sample fit in Section 6.2, and provide model estimates in Section 6.3. The ability of the hybrid model to match the LPY tests and conditional volatility is examined in Section 6.4. We finally study how well the hybrid model forecasts bond yields out-of-sample in Section 6.5.

6.1 The hybrid model

As in the SRM, we let $r(\mathbf{x}_t) = \max(0, s(\mathbf{x}_t))$ but $s(\mathbf{x}_t)$ is now quadratic in the pricing factors, i.e. $s(\mathbf{x}_t) = \alpha + \boldsymbol{\beta}'\mathbf{x}_t + \mathbf{x}_t'\boldsymbol{\Psi}\mathbf{x}_t$, where $\boldsymbol{\Psi} = \mathbf{ADA}'$ as in the QTSM. The non-negativity of the policy rate is enforced by the shadow rate mechanism and no restrictions are therefore imposed on $\boldsymbol{\Psi}$. This gives the model greater flexibility in matching the level and conditional volatility of bond yields than any of the models considered previously. As in the QTSM and the SRM, we assume affine processes for the pricing factors under the $\mathbb{Q}$ and the $\mathbb{P}$ measure, i.e. (1) and (2) are assumed. The identification conditions for the hybrid model are identical to those for the QTSM in Section 2.2.

In the absence of arbitrage, the price in time period t of an j-period zero-coupon bond is $P_{t,j}^{Hybrid} = E_t^{\mathbb{Q}}\left[\exp\left\{-\sum_{i=0}^{j-1} r(\mathbf{x}_{t+i})\right\}\right]$ for $j = 1, 2, ..., K$. Currently, no closed form solution is available for $P_{t,j}^{Hybrid}$ which must be solved numerically. We use a Monte Carlo (MC) method with anti-thetic sampling, as in Bauer & Rudebusch (2014), to improve efficiency, i.e. we use negatively correlated draws of $\sum_{i=0}^{j-1} r_{t+i}$ when approximating $P_{t,j}^{Hybrid}$. To further increase the efficiency of the MC method, we also introduce anti-control sampling. That is, we first compute a MC estimate of bond prices in the hybrid model using only anti-thetic sampling but also a MC estimate of bond prices in a version of the QTSM with no restrictions on α and $\boldsymbol{\Psi}$, denoted $\hat{P}_{t,j}^{QTSM}$. The latter is useful because bond prices are known in closed form in the QTSM, and the MC error in our first estimate of $P_{t,j}^{Hybrid}$ may then be estimated from $\hat{P}_{t,j}^{QTSM} - P_{t,j}^{QTSM}$ to obtain an even more accurate approximation. The details of our MC procedure are described in Appendix E.

6.2 Goodness of in-sample fit

The hybrid models nest the QTSMs, which give the best in-sample fit with two and three pricing factors as shown in Section 4.2. Hence, the hybrid models should fit bond yields at least as well as the QTSMs in the first step of the SR approach. Considering the case with two pricing factors, the hybrid model has $\tilde{Q}_{1:T}^{step1} = 7.826$ and $\tilde{Q}_{1:T}^{step3} = 7.901$, implying a better fit of bond yields than the QTSM, where $\tilde{Q}_{1:T}^{step1} = 8.327$ and $\tilde{Q}_{1:T}^{step3} = 8.780$. The advantage of the two-factor hybrid model is somewhat smaller in the short sample with $\tilde{Q}_{1:T}^{step1} = 6.299$ and $\tilde{Q}_{1:T}^{step3} = 6.655$, given that the corresponding QTSM has $\tilde{Q}_{1:T}^{step1} = 6.554$ and $\tilde{Q}_{1:T}^{step3} = 6.650$.

The difference in fit between the hybrid and the quadratic models becomes even smaller when including an extra pricing factor. The three-factor hybrid model has $\tilde{Q}_{1:T}^{step1} = 2.677$ and $\tilde{Q}_{1:T}^{step3} = 2.709$, which is

a marginal improvement compared to the QTSM where we have $\tilde{Q}_{1:T}^{step1} = 2.704$ and $\tilde{Q}_{1:T}^{step3} = 2.719$. For the short sample, the hybrid model provides a very tight fit with $\tilde{Q}_{1:T}^{step1} = 1.600$ and $\tilde{Q}_{1:T}^{step3} = 1.618$, but so does the QTSM with $\tilde{Q}_{1:T}^{step1} = 1.614$ and $\tilde{Q}_{1:T}^{step3} = 1.632$.

Thus, relaxing the constrains on α and $\boldsymbol{\Psi}$ by enforcing the ZLB using a shadow rate specification provide only a small improvement in the in-sample fit compared to the QTSM with two and three pricing factors. Although somewhat disappointing from the perspective of the hybrid model, this finding is encouraging for the QTSM, because it means that the model does not lose much in-sample fit by enforcing the ZLB through restrictions on $\boldsymbol{\Psi}$.

6.3   Model estimates

The estimated paramters for the hybrid models are provided in Table 10.[34] With two pricing factors, we first note that $\boldsymbol{\Psi}$ in the long sample is positive definite with eigenvalues 0.639 and 1.362, whereas $\boldsymbol{\Psi}$ is indefinite in the short sample with eigenvalues -0.024 and 2.024. Hence, the shadow rate in the long sample only turns negative in the hybrid model due to the sizable negative intercept for the policy rate, i.e. $\alpha = -0.0076$.

As for the three-factor hybrid model, all elements in $\hat{\mathbf{A}}$ exceed one when omitting the ZLB restriction on $\boldsymbol{\Psi}$, although most predominantly in the long sample. This implies that $\boldsymbol{\Psi}$ is indefinite, with eigenvalues of $\{-2.403, -0.028, 5.431\}$ in the long sample, meaning that the short rate is controlled by two pricing factors instead of one in the QTSM. We also find that $\boldsymbol{\Psi}$ is indefinite in the short sample with eigenvalues of $\{-0.153, -0.010, 3.163\}$. The positive eigenvalue is here substantially larger than the two negative eigenvalues, implying that the policy rate is mainly controlled by one pricing factor as in the QTSM within the short sample. We also note that many of the standard errors are fairly wide for the three-factor hybrid model in the short sample, indicating that it is hard to estimate this flexible model accurately when starting the estimation in 1990.[35]

< Table 10 about here >

---

[34]The estimation of the hybrid models is computationally demanding with the MC approximation to bond yields. Furthermore, the objective function in the first step of the SR approach has several local optima. We address these challenges by using the CMA-ES optimizer of Hansen, Müller & Koumoutsakos (2003), capable of optimizing multi-model objective functions and implemented with multiprocessing in FORTRAN on a computer cluster to make the estimation feasible. The estimation is carried out with multiple starting values and with 60 CPUs per optimization.

[35]This finding also explains why we have not attempted to estimate a hybrid model with four pricing factors.

Bond yields in the hybrid models are approximated using just 500 draws in the MC method. Figure 16 considers the three-factor model and shows that this approximation is very accurate, with the largest RMSE being just 0.77 basis points when evaluating bond yields at $\{\hat{\mathbf{x}}_t\}_{t=1}^T$ in the long sample. Without anti-control sampling, the largest RMSE increases to 4.63 basis points as shown in the top right chart of Figure 16, documenting the benefit of anti-control sampling. The bottom row of Figure 16 shows that our MC approximation is even more accurate in the short sample, with the largest RMSE being only 0.07 basis points.

< Figure 16 about here >

6.4 The LPY tests and conditional volatility

Figure 17 shows that the hybrid model with three pricing factors preserves the ability of the corresponding QTSM to match the mean level and the unconditional volatility of bond yields in the two samples. The three-factor hybrid model is, however, less successful in passing the LPY(i) test in both samples and struggles to match LPY(ii) in the short sample. We also note that the two-factor hybrid model in the short sample actually does better on the LPY tests than with three factors, which is similar to our findings in Section 4.5 for QTSMs.

< Figure 17 about here >

The ability of the three-factor hybrid model to match conditional volatility in bond yields is summarized in Table 11, where we run the volatility regressions from Section 4.6. We see small improvements in the long sample for the 0.5- and 2-year bond yields as the R^2 increases from 0.33 and 0.38 in the QTSM to 0.40 and 0.43 in the hybrid model, respectively, when using the GARCH measure of volatility in the data. This is highlighted by the boxed figures in Table 11. For bond yields at the 5- and 10-year maturity, we do not find any improvements compared to the three-factor QTSM. In the short sample, we somewhat surprisingly do not find that the three-factor hybrid model provides a better fit of volatility than the QTSM and the SRM. Unreported results show that volatility in the hybrid model and the QTSM are very similar and closely linked to the short rate, which moves less closely with volatility after 1990 as argued in Section 4.6.

< Table 11 about here >

6.5 Performance out-of-sample

Figures 18 and 19 explore the forecasting ability of the hybrid models. We generally find that the performance of the hybrid model with two pricing factors is dominated by the two-factor QTSM in the long sample and by the two-factor SRM in the short sample.

Forecasts from the hybrid model with three pricing factors in the long sample are similar to those in the QTSM and the SRM for bond yields within the 0.5- to 7-year maturity range. Beyond the 7-year maturity, the performance of the three-factor hybrid model gradually deteriorates, particularly at the 1- and 3-month forecast horizons. We see the same pattern in the three-factor hybrid model when starting the estimation in 1990, except that the deteriorating performance of the hybrid model starts from the 3-year maturity. Hence, the hybrid model with three pricing factors delivers less accurate forecasts of medium- and particularly long-term bond yields compared to the QTSM and the SRM. To understand this finding, recall that the hybrid model only differs from the QTSM by having a more flexible $\mathbb{Q}$ dynamics. This makes overfitting of the policy rate more likely in the three-factor hybrid model and its effects are gradually propagated through the yield curve by the no-arbitrage pricing, thereby generating less accurate forecasts of medium- and long-term bond yields.

< Figure 18 and 19 about here >

7 Conclusion

This paper studies the performance of QTSMs and SRMs on post-war US bond yields. Accounting for the ZLB, the QTSM gives a better in-sample fit than the SRM with two and three pricing factors, whereas the SRM does marginally better than the QTSM with four pricing factors. QTSMs generally struggle to match loadings from ordinary Campbell-Shiller regressions in the long sample from 1961, whereas these moments are better matched by the SRMs. In an out-of-sample forecasting study from January 2005 to December 2013, we find that the SRM generally outperforms the QTSM, and that models accounting for the ZLB do better than the Gaussian ATSM. The SRM is also found to be more robust and less subject to overfitting than the QTSM, as forecasts in the SRM generally improve when including a fourth pricing factor whereas the opposite generally holds in the QTSM. Importantly, the QTSM and the SRM ensure sensible forecasts as predicted bond yields stay non-negative, whereas they easily turn negative in the Gaussian ATSM.

In an attempt to improve the performance of the QTSM, we also study models where the shadow rate is an unrestricted quadratic function of the pricing factors. These hybrid models with two and three pricing factors fit bond yields marginally better in-sample than the other ZLB models but are generally outperformed by the QTSMs and the SRMs when forecasting bond yields out-of-sample. The three-factor hybrid model also struggles to provide a better fit of conditional volatility compared to the QTSM, at least when the model is estimated solely on bond yields. It is likely that the ability of these hybrid models to match conditional volatility of bond yields could be improved by also including this time series in the estimation as in Monfort, Pegoraro, Renne & Roussellet (2014) or by matching option prices as in Kim (2008). Another way to improve the fit of conditional volatility in bond yields would be to maintain the affine specification for the shadow rate and instead introduce heteroskedasticity in the dynamics of the pricing factors. Given the strong performance of the SRM, this extension seems particularly promising and deserves attention in future research.

A Step 2 in the SR approach: A bootstrap bias-adjustment

The standard bootstrap for a VAR model without measurement errors in the pricing factors generates the sampling distribution for moments involving $\hat{\mathbf{x}}_t$ and $\widehat{\mathbf{w}}_{t+1}$ in (7) and (8). The variability in the remaining moments in (7) and (8) related to the measurement errors is accounted for by resampling with replacement from $\left\{\widehat{Var}\left(\mathbf{u}_t\right)\right\}_{t=1}^{T}, \left\{\widehat{Cov}\left(\mathbf{u}_{t+1}, \mathbf{u}_t\right)\right\}_{t=1}^{T-1}$, and $\left\{\widehat{Cov}\left(\mathbf{u}_t, \mathbf{u}_{t+1}\right)\right\}_{t=1}^{T-1}$. The suggested bootstrap for a VAR model with measurement errors in the pricing factors is therefore:

Step A: Use (7) and (8) to obtain $\hat{\boldsymbol{\theta}}_2$. Compute the residuals $\widehat{\mathbf{w}}_{t+1} = \hat{\mathbf{x}}_{t+1} - \hat{\mathbf{h}}_0 - \hat{\mathbf{h}}_\mathbf{x}\hat{\mathbf{x}}_t$ for $t = 1, 2, ..., T-1$. Let $b = 1$.

Step B: Resample with replacement from $\left\{\widehat{\mathbf{w}}_{t+1}\right\}_{t=1}^{T-1}$ to generate a bootstrap sample of length $T-1$ using

$$\mathbf{x}_{t+1}^* = \hat{\mathbf{h}}_0 + \hat{\mathbf{h}}_\mathbf{x}\mathbf{x}_t^* + \widehat{\mathbf{w}}_{t+1}^* \quad \text{for } t = 1, 2, ..., T-1.$$

where $\widehat{\mathbf{w}}_{t+1}^*$ denote independent draws from $\left\{\widehat{\mathbf{w}}_{t+1}\right\}_{t=1}^{T-1}$.

Step C: Generate $\left\{\widehat{Var}\left(\mathbf{u}_t\right)^*\right\}_{t=1}^{T}, \left\{\widehat{Cov}\left(\mathbf{u}_{t+1}, \mathbf{u}_t\right)^*\right\}_{t=1}^{T-1}$, and $\left\{\widehat{Cov}\left(\mathbf{u}_t, \mathbf{u}_{t+1}\right)^*\right\}_{t=1}^{T-1}$ by resampling with replacement from $\left\{\widehat{Var}\left(\mathbf{u}_t\right)\right\}_{t=1}^{T}, \left\{\widehat{Cov}\left(\mathbf{u}_{t+1}, \mathbf{u}_t\right)\right\}_{t=1}^{T-1}$, and $\left\{\widehat{Cov}\left(\mathbf{u}_t, \mathbf{u}_{t+1}\right)\right\}_{t=1}^{T-1}$.

Step D: Use the draws from Step B and C in (7) and (8) to obtain $\hat{\mathbf{h}}_0^{(b)}$, $\hat{\mathbf{h}}_\mathbf{x}^{(b)}$, and $\hat{\boldsymbol{\Sigma}}^{(b)}$.

Step E: If $b < \mathcal{B}$, then $b = b + 1$ and go to step B.

The bootstrap bias-adjusted estimate of $\mathbf{h}_\mathbf{x}$ is then given by

$$\hat{\mathbf{h}}_\mathbf{x}^{adj} = \hat{\mathbf{h}}_\mathbf{x} - \left(\bar{\mathbf{h}}_\mathbf{x} - \hat{\mathbf{h}}_\mathbf{x}\right) = 2\hat{\mathbf{h}}_\mathbf{x} - \bar{\mathbf{h}}_\mathbf{x},$$

where $\bar{\mathbf{h}}_\mathbf{x} \equiv \frac{1}{\mathcal{B}}\sum_{b=1}^{\mathcal{B}}\hat{\mathbf{h}}_\mathbf{x}^{(b)}$. The bias-adjusted estimates of $\mathbf{h}_0$ and $\boldsymbol{\Sigma}$ are obtained as in Engsted & Pedersen (2012). That is, we obtain an unbiased estimate of $\mathbf{h}_0$ by letting

$$\hat{\mathbf{h}}_\mathbf{0}^{adj} = \left(\mathbf{I} - \hat{\mathbf{h}}_\mathbf{x}^{adj}\right)\hat{E}\left[\hat{\mathbf{x}}_t\right],$$

where $\hat{E}\left[\hat{\mathbf{x}}_t\right] \equiv 1/T \sum_{t=1}^{T}\hat{\mathbf{x}}_t$ remains an unbiased estimator of the sample mean even with measurement errors in $\mathbf{x}_t$. This is because $E\left[\mathbf{u}_t\right] = \mathbf{0}$, given a sufficiently large cross-section panel of bond yields as required in the SR approach, i.e. this property follows from consistency of the regression-filter when the cross-section dimension tends to infinity. Finally, the bias-adjusted estimate of $\hat{\boldsymbol{\Sigma}}^{adj}$ is computed using

$$\widehat{\mathbf{w}}_{t+1}^{adj} = \hat{\mathbf{x}}_{t+1} - \hat{\mathbf{h}}_\mathbf{0}^{adj} - \hat{\mathbf{h}}_\mathbf{x}^{adj}\hat{\mathbf{x}}_t \quad \text{for } t = 1, 2, ..., T-1$$

and a direct modification of (8), i.e.

$$
\begin{aligned}
\widehat{Var}\left(\mathbf{w}_{t+1}\right)^{adj} &= \frac{1}{T - 1 - n_x - 1}\sum_{t=1}^{T-1}\widehat{\mathbf{w}}_{t+1}^{adj}\left(\widehat{\mathbf{w}}_{t+1}^{adj}\right)' \\
&- \frac{1}{T-1}\sum_{t=1}^{T-1}\left(\widehat{Var}\left(\mathbf{u}_t\right) + \hat{\mathbf{h}}_\mathbf{x}^{adj}\widehat{Var}\left(\mathbf{u}_t\right)\left(\hat{\mathbf{h}}_\mathbf{x}^{adj}\right)'\right) \\
&+ \frac{1}{T-1}\sum_{t=1}^{T-1}\left(\widehat{Cov}\left(\mathbf{u}_{t+1}, \mathbf{u}_t\right)\left(\hat{\mathbf{h}}_\mathbf{x}^{adj}\right)' + \hat{\mathbf{h}}_\mathbf{x}^{adj}\widehat{Cov}\left(\mathbf{u}_t, \mathbf{u}_{t+1}\right)\right),
\end{aligned}
$$

where we have imposed the standard degrees of freedom adjustment. Hence, $\hat{\boldsymbol{\Sigma}}^{adj}$ is obtained from a Cholesky decomposition of $\widehat{Var}\left(\mathbf{w}_{t+1}\right)^{adj}$.

B Inducing stationarity in VAR models: A data-driven method

This section presents a data-driven method to determine δ by minimizing the distance between the unconditional variances of the factors in the sample and the unconditional variances implied by the VAR model. To compute the variances in the bias-adjusted VAR model, we consider

$$\hat{\mathbf{h}}_{\mathbf{x}}^{adj}\left(\delta\right) = \delta \times \left(\hat{\mathbf{h}}_{\mathbf{x}} - \left(\bar{\mathbf{h}}_{\mathbf{x}} - \hat{\mathbf{h}}_{\mathbf{x}}\right)\right)$$

and

$$\hat{\mathbf{h}}_{\mathbf{0}}^{adj}\left(\delta\right) = \left(\mathbf{I} - \hat{\mathbf{h}}_{\mathbf{x}}^{adj}\left(\delta\right)\right)\hat{E}\left[\hat{\mathbf{x}}_t\right].$$

For given values of $\hat{\mathbf{h}}_{\mathbf{x}}^{adj}\left(\delta\right)$ and $\hat{\mathbf{h}}_{\mathbf{0}}^{adj}\left(\delta\right)$, we may then compute the residuals as

$$\widehat{\mathbf{w}}_{t+1}^{adj}\left(\delta\right) = \hat{\mathbf{x}}_{t+1} - \hat{\mathbf{h}}_{\mathbf{0}}^{adj}\left(\delta\right) - \hat{\mathbf{h}}_{\mathbf{x}}^{adj}\left(\delta\right)\hat{\mathbf{x}}_t \quad \text{for } t = 1, 2, ..., T-1,$$

and estimate the variance of the innovations by

$$
\begin{aligned}
\widehat{Var}\left(\mathbf{w}_{t+1}\left(\delta\right)\right)^{adj} \;=\; & \frac{1}{T-1-n_x-1}\sum_{t=1}^{T-1}\widehat{\mathbf{w}}_{t+1}^{adj}\left(\delta\right)\left(\widehat{\mathbf{w}}_{t+1}^{adj}\left(\delta\right)\right)' \\
& -\frac{1}{T-1}\sum_{t=1}^{T-1}\left(\widehat{Var}\left(\mathbf{u}_t\right) + \hat{\mathbf{h}}_{\mathbf{x}}^{adj}\left(\delta\right)\widehat{Var}\left(\mathbf{u}_t\right)\left(\hat{\mathbf{h}}_{\mathbf{x}}^{adj}\left(\delta\right)\right)'\right) \\
& +\frac{1}{T-1}\sum_{t=1}^{T-1}\left(\widehat{Cov}\left(\mathbf{u}_{t+1}, \mathbf{u}_t\right)\left(\hat{\mathbf{h}}_{\mathbf{x}}^{adj}\left(\delta\right)\right)' + \hat{\mathbf{h}}_{\mathbf{x}}^{adj}\left(\delta\right)\widehat{Cov}\left(\mathbf{u}_t, \mathbf{u}_{t+1}\right)\right).
\end{aligned}
$$

Hence, the unconditional variance in the VAR model is given by

$$vec\left(\mathbf{V}_{\mathbf{x}_t}\left(\delta\right)\right) = \left(\mathbf{I}_{m^2} - \hat{\mathbf{h}}_{\mathbf{x}}^{adj}\left(\delta\right) \otimes \hat{\mathbf{h}}_{\mathbf{x}}^{adj}\left(\delta\right)\right)^{-1} vec\left(\widehat{Var}\left(\mathbf{w}_{t+1}\left(\delta\right)\right)^{adj}\right),$$

where the diagonal of $\mathbf{V}_{\mathbf{x}_t}\left(\delta\right)$ gives the factor variance in the VAR model, denoted $\sigma_{i,VAR}^2\left(\delta\right)$ for $i = 1, 2, ..., n_x$.

To compute the model-independent unconditional variances of the factors as implied by $\{\hat{\mathbf{x}}_t\}_{t=1}^{T}$, the unconditional mean of the ith pricing factor is estimated by $\hat{E}\left[\hat{x}_{i,t}\right] = 1/T\sum_{t=1}^{T}\hat{x}_{i,t}$. We also have

$$
\begin{aligned}
& \frac{1}{T-1}\sum_{t=1}^{T}\left(\hat{x}_{i,t} - \hat{E}\left[\hat{x}_i\right]\right)^2 \\
= \; & \frac{1}{T-1}\sum_{t=1}^{T}\left(x_{i,t}^o + u_{i,t} - \hat{E}\left[\hat{x}_i\right]\right)^2 \\
= \; & \frac{1}{T-1}\sum_{t=1}^{T}\left(x_{i,t}^o - \hat{E}\left[\hat{x}_i\right]\right)^2 + \frac{1}{T-1}\sum_{t=1}^{T}u_{i,t}^2 + 2\frac{1}{T-1}\sum_{t=1}^{T}\left(x_{i,t}^o - \hat{E}\left[\hat{x}_i\right]\right)u_{i,t} \\
= \; & \frac{1}{T-1}\sum_{t=1}^{T}\left(x_{i,t}^o - \hat{E}\left[\hat{x}_i\right]\right)^2 + \frac{1}{T-1}\sum_{t=1}^{T}Var\left(u_{i,t}\right) + 2\frac{1}{T-1}\sum_{t=1}^{T}\left(x_{i,t}^o - \hat{E}\left[\hat{x}_i\right]\right)u_{i,t}
\end{aligned}
$$

for $i = 1, 2, ..., n_x$, where the last line follows by considering $u_{i,t}^2$ as a point estimate of $Var\left(u_{i,t}\right)$. A similar argument is used when computing standard errors robust to heteroskedasticity. Clearly,

$$\frac{1}{T-1}\sum_{t=1}^{T}\left(x_{i,t}^{o}-\hat{E}\left[\hat{x}_{i}\right]\right)^{2}\xrightarrow{p}Var\left(x_{i,t}^{o}\right) \text{ as } T\longrightarrow\infty. \text{ We also have for } T\longrightarrow\infty, \text{ that}$$

$$\frac{1}{T-1}\sum_{t=1}^{T}\left(x_{i,t}^{o}-\hat{E}\left[\hat{x}_{i}\right]\right)u_{i,t}\xrightarrow{p}E\left[\left(x_{i,t}^{o}-E\left[x_{i,t}^{o}\right]\right)u_{i,t}\right]=E\left[x_{i,t}^{o}u_{i,t}\right],$$

as $E\left[u_{i,t}\right]=0$ for $i=1,2,...,n_x$. We next recall that the measurement errors in the factors $u_{i,t}$ are a function of the measurement errors in the yields, denoted $\mathbf{v}_t$. Moreover, $\mathbf{v}_t$ is by assumption uncorrelated with the innovations to the factors $\boldsymbol{\varepsilon}_t$ at all leads and lags, which drives the evolution of the factors. Hence, $E\left[x_{i,t}^{o}u_{i,t}\right]=0$, at least up to a first-order approximation. Thus,

$$\frac{1}{T-1}\sum_{t=1}^{T}\left(\hat{x}_{i,t}-E\left[x_{i,t}^{o}\right]\right)^{2}\xrightarrow{p}Var\left(x_{i,t}^{o}\right)+E\left[Var\left(u_{i,t}\right)\right].$$

This implies that the unconditional variance of the ith pricing factor from the sample may be estimated by

$$\hat{\sigma}_{i,sample}^{2}=\frac{1}{T-1}\sum_{t=1}^{T}\left(\hat{x}_{i,t}-\hat{E}\left[\hat{x}_{i}\right]\right)^{2}-\frac{1}{T}\sum_{t=1}^{T}Var\left(u_{i,t}\right).$$

We then suggest letting the scaling parameter δ be given by

$$\hat{\delta}=\arg\min_{\delta\in[\delta_{lower},1]}\sum_{i=1}^{n_x}\left(\frac{\sigma_{i,VAR}^{2}\left(\delta\right)-\hat{\sigma}_{i,sample}^{2}}{\hat{\sigma}_{i,sample}^{2}}\right)^{2} \tag{19}$$

where $\delta_{lower}>0$. The constraint on the domain of δ is imposed because at $\delta=0$, we have $\hat{\mathbf{h}}_{\mathbf{x}}^{adj}\left(\delta=0\right)=\mathbf{0}$ and $\hat{\mathbf{h}}_{\mathbf{0}}^{adj}\left(\delta=0\right)=\hat{E}\left[\hat{\mathbf{x}}_t\right]$, meaning that the two estimators of the unconditional variances in (19) nearly coincides as they only differ by $\frac{1}{T}\sum_{t=1}^{T}Var\left(u_{i,t}\right)$.

C Bias in OLS for the cross-sectional AR model in the bootstrap

We consider an AR(1) model for simplicity, but similar arguments extend directly to the AR(p) model. That is, we consider $v_{t,m_j}=\phi_1 v_{t,m_{j-1}}+\epsilon_{t,m_j}$ for $j=2,3,...,n_y$ and $t=1,2,...,T$. The variable v_{t,m_j} is unobserved and replaced by the fitted residuals, i.e. $\hat{v}_{t,m_j}=\phi_1\hat{v}_{t,m_{j-1}}+\hat{\epsilon}_{t,m_j}$. Here, $\hat{v}_{t,m_j}\equiv y_{t,m_j}-g_{m_j}\left(\hat{\mathbf{x}}_t;\boldsymbol{\theta}_1^{o}\right)-c_t$, where c_t is the re-centering constant ensuring $\sum_{j=1}^{n_y}\hat{v}_{t,m_j}/n_y=0$ for all t. Further, let $\hat{v}_{t,m_j}\equiv v_{t,m_j}+u_{t,m_j}^{v}$ where u_{t,m_j}^{v} denotes the estimation error in v_{t,m_j}. Hence,

$$\begin{aligned}u_{t,m_j}^{v}&\equiv\hat{v}_{t,m_j}-v_{t,m_j}\\&=y_{t,m_j}-g_{m_j}\left(\hat{\mathbf{x}}_t;\boldsymbol{\theta}_1^{o}\right)-c_t-\left(y_{t,m_j}-g_{m_j}\left(\mathbf{x}_t^{o};\boldsymbol{\theta}_1^{o}\right)\right)\\&=g_{m_j}\left(\mathbf{x}_t^{o};\boldsymbol{\theta}_1^{o}\right)-g_{m_j}\left(\hat{\mathbf{x}}_t;\boldsymbol{\theta}_1^{o}\right)-c_t.\end{aligned}$$

Moreover, we have also have

$$\begin{aligned}\hat{\epsilon}_{t,m_j}&=\hat{v}_{t,m_j}-\phi_1\hat{v}_{t,m_{j-1}}\\&=v_{t,m_j}+u_{t,m_j}^{v}-\phi_1\left(v_{t,m_{j-1}}+u_{t,m_{j-1}}^{v}\right)\\&=\epsilon_{t,m_j}+u_{t,m_j}^{v}-\phi_1 u_{t,m_{j-1}}^{v}.\end{aligned}$$

Analyzing the moment condition for the OLS estimator, we get

$$
\begin{aligned}
E\left[\sum_{t=1}^{T}\sum_{j=2}^{n_y}\hat{v}_{t,m_{j-1}}\hat{v}_{t,m_j}\right] &= \phi_1 E\left[\sum_{t=1}^{T}\sum_{j=2}^{n_y}\hat{v}_{t,m_{j-1}}\hat{v}_{t,m_{j-1}}\right] + E\left[\sum_{t=1}^{T}\sum_{j=2}^{n_y}\hat{v}_{t,m_{j-1}}\hat{\epsilon}_{t,m_j}\right] \\
&= \phi_1 E\left[\sum_{t=1}^{T}\sum_{j=2}^{n_y}\hat{v}_{t,m_{j-1}}\hat{v}_{t,m_{j-1}}\right] + E\left[\sum_{t=1}^{T}\sum_{j=2}^{n_y}\hat{v}_{t,m_{j-1}}\left(\epsilon_{t,m_j}+u^v_{t,m_j}-\phi_1 u^v_{t,m_{j-1}}\right)\right] \\
&= \phi_1 E\left[\sum_{t=1}^{T}\sum_{j=2}^{n_y}\hat{v}_{t,m_{j-1}}\hat{v}_{t,m_{j-1}}\right] + E\left[\sum_{t=1}^{T}\sum_{j=2}^{n_y}\hat{v}_{t,m_{j-1}}\left(u^v_{t,m_j}-\phi_1 u^v_{t,m_{j-1}}\right)\right]
\end{aligned}
$$

because $E\left[\hat{v}_{t,m_{j-1}}\epsilon_{t,m_j}\right]=0$. But $E\left[\hat{v}_{t,m_{j-1}}\left(u^v_{t,m_j}-\phi_1 u^v_{t,m_{j-1}}\right)\right]\neq 0$ and this generates a bias in the OLS estimator of ϕ_1.

D   Details for Monte Carlo study in Section 3.5

The data generating process for the Monte Carlo study is a one-factor Gaussian ATSM where we let $\alpha=0.008$, $\Phi_{11}=0.01$, $\Sigma_{11}=5.5\times10^{-4}$, $h_0=-0.0002$, and $h_x=0.96$. This calibration ensures that the one-factor model roughly matches the level and variability of the US yield curve from 1990 to 2013.

We consider a general specification for measurement errors in bond yields to accommodate various deviations from the standard assumption of independent and identical errors. More precisely, we assume that bond yields are generated as follows

$$
\mathbf{y}_t = \mathbf{A} + \mathbf{B}x_t + \sqrt{R_v(t)}\mathbf{u}_t
$$

$$
\mathbf{u}_t = \boldsymbol{\lambda}\mathbf{u}_{t-1} + \boldsymbol{\Omega}\mathbf{z}_{\mathbf{u},t},
$$

$$
x_{t+1} = h_0 + h_x x_t + \Sigma_{11}\varepsilon^{\mathbb{P}}_{t+1}
$$

where $\boldsymbol{\lambda}=diag\left(\rho_{Time}\right)$ is an $n_y\times n_y$ diagonal matrix with ρ_{Time} along the diagonal and $\boldsymbol{\Omega}$ is a lower triangular $n_y\times n_y$ matrix where $\boldsymbol{\Omega}(i,j)=\phi_1^{|i-j|}$ for $j\leq i$. That is, ρ_{Time} controls the degree of auto-correlation in the measurement errors and ϕ_1 determines the degree of cross-sectional dependence using two AR(1) models. Here, $\mathbf{z}_{\mathbf{u},t}\sim\mathcal{NID}\left(\mathbf{0},\mathbf{I}\right)$. We also allow for heteroskedasticity in the measurement errors along the time series dimension by letting the conditional variance evolve according to

$$
n_t = (1-\rho_{Rv})R_v + \rho_{Rv}R_v(t-1) + \sigma_{Rv}z_t
$$

$$
R_v(t) = \begin{cases} 0.05^2 & \text{for } n_t < 0.05^2 \\ n_t & \text{else} \end{cases}
$$

where $z_t\sim\mathcal{NID}\left(0,1\right)$ and independent of $\mathbf{z}_{\mathbf{u},t}$. This specification allows for persistence in the conditional variance through ρ_{Rv} and ensures that the conditional standard deviation is at least 5 basis points. The chosen specification has measurement errors with a standard deviation of 10 basis points, i.e. $R_v=0.10^2$, and we let $\sigma_{Rv}=R_v/2$.

To apply the SR approach for this one-factor model, we let $\boldsymbol{\theta}_{11}\equiv\left[\begin{array}{cc}\alpha & \Phi_{11}\end{array}\right]'$, $\boldsymbol{\theta}_{12}\equiv\left[\begin{array}{c}\Sigma_{11}\end{array}\right]$, and $\boldsymbol{\theta}_{22}\equiv\left[\begin{array}{cc}h_0 & h_x\end{array}\right]$. All the risk-neutral parameters are estimated in step 1 as described in Section 3.2, and $\boldsymbol{\theta}_2$ is obtained in step 2 using the bias-adjustment described in Section 3.3.1. In step 3, we let $\boldsymbol{\Lambda}=\mathbf{0}$ and re-estimate $\boldsymbol{\theta}_{11}$ as described in Section 3.4.

BANK OF ENGLAND

E   The hybrid model: Monte Carlo approximation to bond yields

For a given state vector $\mathbf{x}_t$, the Monte Carlo (MC) approximation to bond prices is $\hat{P}_{t,j} = \frac{1}{M} \sum_{s=1}^{M} P_{t,j}^s$ where

$$P_{t,j}^s \equiv \frac{1}{2} \left(\exp\left\{ -\sum_{i=0}^{j-1} r\left(\mathbf{x}_{t+i}^s\right) \right\} + \exp\left\{ -\sum_{i=0}^{j-1} r\left(\tilde{\mathbf{x}}_{t+i}^s\right) \right\} \right).$$

Here, $\left\{ \{\mathbf{x}_{t+i}^s\}_{i=1}^{j-1} \right\}_{s=1}^{M}$ are generated using the $\mathcal{IID}$ draws $\left\{ \{\boldsymbol{\varepsilon}_{t+i}^s\}_{i=1}^{j-1} \right\}_{s=1}^{M}$ under the $\mathbb{Q}$ measure, while $\left\{ \{\tilde{\mathbf{x}}_{t+i}^s\}_{i=1}^{j-1} \right\}_{s=1}^{M}$ are constructed using $\left\{ \{-\boldsymbol{\varepsilon}_{t+i}^s\}_{i=1}^{j-1} \right\}_{s=1}^{M}$ to induce negative correlation across the draws, i.e. anti-thetic sampling. Hence, we let $M = S/2$ to obtain S draws. To implement anti-control sampling, we also compute the MC approximation to bond yields in a version of the QTSM with no restrictions on α and $\boldsymbol{\Psi}$, denoted $\hat{P}_{t,j}^{QTSM}$. That is, $\hat{P}_{t,j}^{QTSM} = \frac{1}{M} \sum_{s=1}^{M} P_{t,j}^{QTSM,s}$ where

$$\begin{aligned} P_{t,j}^{QTSM,s} &\equiv \frac{1}{2} \exp\left\{ -\sum_{i=0}^{j-1} \left(\alpha + \boldsymbol{\beta}' \mathbf{x}_{t+i}^s + \left(\mathbf{x}_{t+i}^s\right)' \boldsymbol{\Psi} \mathbf{x}_{t+i}^s \right) \right\} \\ &\quad + \frac{1}{2} \exp\left\{ -\sum_{i=0}^{j-1} \left(\alpha + \boldsymbol{\beta}' \tilde{\mathbf{x}}_{t+i}^s + \left(\tilde{\mathbf{x}}_{t+i}^s\right)' \boldsymbol{\Psi} \tilde{\mathbf{x}}_{t+i}^s \right) \right\}. \end{aligned}$$

The MC error in this version of the QTSM is $e_{t,j}^{QTSM} = P_{t,j}^{QTSM} - \hat{P}_{t,j}^{QTSM}$, where $P_{t,j}^{QTSM}$ denotes the exact solution. The adjusted MC estimate of bond prices in the hybrid model is then

$$\hat{P}_{t,j}^{Hybrid}\left(b_{t,j}\right) = \hat{P}_{t,j} + b_{t,j} \left(P_{t,j}^{QTSM} - \hat{P}_{t,j}^{QTSM} \right),$$

where the scaling parameter $b_{t,j}$ is state and maturity dependent. As shown in Chapter 16 of Munk (2011), we may alternatively adjust each draw of $P_{t,j}^{Hybrid,s}$, i.e.

$$P_{t,j}^{Hybrid,s}\left(b_{t,j}\right) = P_{t,j}^s + b_{t,j} \left(P_{t,j}^{QTSM} - P_{t,j}^{QTSM,s} \right).$$

The scaling parameter $b_{t,j}$ is set to minimize the variance of $P_{t,j}^{Hybrid,s}$. That is

$$\min_{b_{t,j}} Var\left(P_{t,j}^{Hybrid,s}\left(b_{t,j}\right) \right) = Var\left(P_{t,j}^s \right) + b_{t,j}^2 Var\left(P_{t,j}^{QTSM,s} \right) - 2 b_{t,j} Cov\left(P_{t,j}^s, P_{t,j}^{QTSM,s} \right),$$

implying

$$b_{t,j}^* = \frac{Cov\left(P_{t,j}^s, P_{t,j}^{QTSM,s} \right)}{Var\left(P_{t,j}^{QTSM,s} \right)} = \rho\left(P_{t,j}^s, P_{t,j}^{QTSM,s} \right) \sqrt{\frac{Var\left(P_{t,j}^s \right)}{Var\left(P_{t,j}^{QTSM,s} \right)}},$$

where $\rho\left(P_{t,j}^s, P_{t,j}^{QTSM,s} \right)$ is the correlation coefficient. Evaluating $Var\left(P_{t,j}^{Hybrid,s} \right)$ at $b_{t,j}^*$ gives $Var\left(P_{t,j}^s \right) \left(1 - \rho\left(P_{t,j}^s, P_{t,j}^{QTSM,s} \right)^2 \right)$, meaning that the variance of $Var\left(P_{t,j}^{Hybrid,s} \right)$ is reduced if $\rho\left(P_{t,j}^s, P_{t,j}^{QTSM,s} \right) \neq 0$. Ideally, $\rho\left(P_{t,j}^s, P_{t,j}^{QTSM,s} \right) \approx \pm 1$, which implies $Var\left(P_{t,j}^{Hybrid,s} \right) \approx 0$. The values of $Var\left(P_{t,j}^{QTSM,s} \right)$, $Var\left(P_{t,j}^s \right)$, and $\rho\left(P_{t,j}^s, P_{t,j}^{QTSM,s} \right)$ are unknown but easily estimated by simple averages from the simulated paths.

References

Adrian, T., Crump, R. K. & Moench, E. (2013), 'Pricing the term structure with linear regressions', *Journal of Financial Economics* **110**, 110–138.

Ahn, D.-H., Dittmar, R. F. & Gallant, A. R. (2002), 'Quadratic term structure models: Theory and evidence', *The Review of Financial Studies* **15**(1), 243–288.

Ait-Sahalia, Y. & Kimmel, R. L. (2010), 'Estimating affine multifactor term structure models using closed-form likelihood expansions', *Journal of Financial Economics* **98**, 113–144.

Andreasen, M. M. & Christensen, B. J. (2015), 'The SR approach: A new estimation procedure for non-linear and non-Gaussian dynamic term structure models', *Journal of Econometrics* **184**, 420–451.

Andrews, D. (2000), 'Inconsistency of the bootstrap when a parameter is on the boubound of the parameter space', *Econometrica* **68**(2), 399–405.

Bauer, M. D. & Rudebusch, G. D. (2014), 'Monetary policy expectations at the zero lower bound', *Federal Reserve Bank of San Francisco Working Paper 2013-18* .

Bauer, M. D., Rudebusch, G. D. & Wu, J. C. (2012), 'Correcting estimation bias in dynamic term structure models', *Journal of Business and Economic Statistics* **3**, 454–467.

Black, F. (1995), 'Interest rates as options', *The Journal of Finance* **50**(5), 1371–1376.

Bühlmann, P. (1997), 'Sieve bootstrap for time series', *Bernoulli* **3**(2), 123–148.

Campbell, J. Y. & Shiller, R. J. (1991), 'Yield spread and interest rate movements: A bird's eye view', *The Review of Economic Studies* **58**(3), 495–514.

Christensen, J. H. E. & Rudebusch, G. D. (2012), 'The response of interest rates to US and UK quantitative easing', *The Economic Journal* **122**, 385–414.

Christensen, J. H. E. & Rudebusch, G. D. (2013), 'Modeling Yields at the Zero Lower Bound: Are Shadow Rates the Solution?', *Federal Reserve Bank of San Francisco Working Paper 2013-39* .

Christensen, J. H. E. & Rudebusch, G. D. (2015), 'Estimating shadow-rate term structure models with near-zero yields', *Journal of Financial Econometrics* **13**(2), 226–259.

Cox, J. C., Ingersoll, J. E. & Ross, S. A. (1985), 'A theory of the term structure of interest rates', *Econometrica* **53**(2), 385–407.

Dai, Q. & Singleton, K. J. (2000), 'Specification analysis of affine term structure models', *Journal of Finance* **55**, 1946–1978.

Dai, Q. & Singleton, K. J. (2002), 'Expectation puzzles, time-varying risk premia and affine models of the term structure', *Journal of Financial Economics* **63**, 415–441.

Engsted, T. & Pedersen, T. Q. (2012), 'Return predictability and intertemporal asset allocation: Evidence from a bias-adjusted VAR model', *Journal of Empirical Finance* **19**, 241–253.

Gagnon, J., Raskin, M., Rernache, J. & Sack, B. (2011), 'Large-Scale Asset Purchases by the Federal Reserve: Did They Work?', *Federal Reserve Bank of New York, Research Paper Series - Economic Policy Review* **17**(1), 41–59.

Gorovoi, V. & Linetsky, V. (2004), 'Black's model of interest rates as options, eigenfunction expansions and japanese interest rates', *Mathematical Finance* **14**(1), 49–78.

Gürkaynak, R., Sack, B. & Wright, J. (2007), 'The U.S. Treasury yield curve: 1961 to the present', *Journal of Monetary Economics* **54**, 2291–2304.

Hamilton, J. D. & Wu, J. C. (2012), 'Identification and estimation of Gaussian affine term structure models', *Journal of Econometrics* **168**, 315–331.

Hansen, N., Müller, S. D. & Koumoutsakos, P. (2003), 'Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES)', *Evolutionary Computation* **11**(1), 1–18.

Harvey, A. C. (1989), 'Forecasting, structural time series models and the kalman filter', *Cambridge University Press* .

Ichiue, H. & Ueno, Y. (2007), 'Equilibrium interest rate and the yield curve in a low interest rate enviroment', *Bank of Japan Working Paper* **07-E-18**.

Ichiue, H. & Ueno, Y. (2013), 'Estimating term premia at the zero bound: An analysis of Japanese, US, and UK yields', *Bank of Japan Working Paper Series* **13-E-8**.

Joslin, S., Singleton, K. J. & Zhu, H. (2011), 'A new perspective on Gaussian dynamic term structure models', *The Review of Financial Studies* **24**, 926–970.

Joyce, M. A. S., Lasaosa, A., Stevens, I. & Tong, M. (2011), 'The Financial Market Impact of Quantitative Easing in the United Kingdom', *International Journal of Central Banking* **7**(3), 113–161.

Kilian, L. (1998), 'Small-sample confidence intervals for impulse response functions', *The Review of Economics and Statistics* **80**, 218–230.

Kim, D. H. (2008), 'Zero Bound, Option-Implied PDF's, and Term Structure Models', *Federal Reserve Board Finance and Economics Discussion Series 2008-31* .

Kim, D. H. & Singleton, K. J. (2012), 'Term structure models and the zero bound: An empirical investigation of Japanese yields', *Journal of Econometrics* **170**, 32–49.

Krippner, L. (2012), 'Modifying Gaussian term structure models when interest rates are near the zero lower bound', *Reserve Bank of New Zealand, Discussion Paper Series* .

Leippold, M. & Wu, L. (2002), 'Asset pricing under the quadratic class', *The Journal of Financial and Quantitative Analysis* **37**(2), 271–295.

Leippold, M. & Wu, L. (2003), 'Design and estimation of quadratic term structure models', *European Finance Review* **7**, 47–73.

MacKinnon, J. G. (2009), *'Bootstrap hypothesis testing' in "Handbook of Computational Econometrics"*, John Wiley and Sons, Ltd.

Monfort, A., Pegoraro, F., Renne, J.-P. & Roussellet, G. (2014), 'Staying at Zero with Affine Processes: A New Dynamic Term Structure Model', *Working Paper* .

Munk, C. (2011), *Fixed Income Modelling*, Oxford University Press.

Norgaard, M., Poulsen, N. K. & Ravn, O. (2000), 'Advances in derivative-free state estimation for nonlinear systems', *Automatica* **36:11**, 1627–1638.

Priebsch, M. A. (2013), 'Computing arbitrage-free yields in multi-factor Gaussian shadow-rate term structure models', *Federal Reserve Board Finance and Economics Discussion Series 2013-63* .

Realdon, M. (2006), 'Quadratic term structure models in discrete time', *Finance Research Letters* **3**, 277–289.

Rudebusch, G. D. & Wu, T. (2007), 'Accounting for a shift in term structure behavior with no-arbitrage and macro-finance models', *Journal of Money, Credit and Banking* **39**(2-3), 395–422.

Yamamoto, T. & Kunitomo, N. (1984), 'Asymptotic bias of the least square estimator for mulitivariate autoregressive models', *Annals of the Institute of Statistical Mathematics* **36**(4), 419–430.

Table 1: Monte Carlo study: Bias-adjustment in VAR models

The Monte Carlo study is implemented without measurement errors in the pricing factors and with $M = 5,000$ draws, where each bootstrap adjustment is computed with $\mathcal{B} = 5,000$ bootstrap replications. The data generating processes (DGP) are the estimated VAR models for the pricing factors under the physical measure in the Gaussian ATSM reported in Tables 5 and 6. The notation $\text{Bias}(\mathbf{h}_0)$ indicates the total absolute bias for $\mathbf{h}_0$ and similarly for the other rows. When computing the total absolute bias of the unconditional standard deviation in the pricing factors, denoted $\text{Bias}(\{\sigma_{x_i}\}_{i=1}^{n_x})$, only the stationary draws are used. Bold figures indicate the lowest bias among the two data-driven methods.

		OLS	Standard bootstrap	Kilian's method	Data-driven methods: $\hat{\mathbf{h}}_{\mathbf{x}}^{adj,B}(\delta)$	$\hat{\mathbf{h}}_{\mathbf{x}}^{adj,*}(\delta)$
DGP: ATSM from 1961-2013						
$T = 250$	$\text{Bias}(\mathbf{h}_0)$	0.0004	0.0002	0.0002	0.0003	0.0003
	$\text{Bias}(\mathbf{h}_{\mathbf{x}})$	0.1563	0.0547	0.0642	0.0850	**0.0747**
	$\text{Bias}(\mathbf{\Sigma} \times 100)$	0.0012	0.0006	0.0006	0.0007	0.0007
	$\text{Bias}(\{\sigma_{x_i}\}_{i=1}^{n_x})$	0.0015	0.0017	0.0278	0.0010	**0.0008**
	Pct of nonstationary draws	0.48	30.98	0.48	0.48	0.00
$T = 500$	$\text{Bias}(\mathbf{h}_0)$	0.0002	0.0001	0.0001	0.0001	0.0001
	$\text{Bias}(\mathbf{h}_{\mathbf{x}})$	0.0676	0.0115	0.0152	0.0234	**0.0190**
	$\text{Bias}(\mathbf{\Sigma} \times 100)$	0.0005	0.0002	0.0003	0.0003	0.0003
	$\text{Bias}(\{\sigma_{x_i}\}_{i=1}^{n_x})$	0.0008	0.0021	0.0249	0.0023	**0.0017**
	Pct of nonstationary draws	0.14	20.16	0.14	0.14	0.00
DGP: ATSM from 1990-2013						
$T = 250$	$\text{Bias}(\mathbf{h}_0)$	0.0086	0.0027	0.0032	0.0043	**0.0031**
	$\text{Bias}(\mathbf{h}_{\mathbf{x}})$	3.7685	1.2012	1.4260	1.8965	**1.2938**
	$\text{Bias}(\mathbf{\Sigma} \times 100)$	0.0129	0.0045	0.0062	0.0077	**0.0068**
	$\text{Bias}(\{\sigma_{x_i}\}_{i=1}^{n_x})$	0.0092	0.0280	0.2913	0.0211	**0.0196**
	Pct of nonstationary draws	0.22	25.78	0.22	0.22	0.00
$T = 500$	$\text{Bias}(\mathbf{h}_0)$	0.0035	0.0006	0.0006	0.0007	**0.0006**
	$\text{Bias}(\mathbf{h}_{\mathbf{x}})$	1.5484	0.2233	0.2394	0.2949	**0.2334**
	$\text{Bias}(\mathbf{\Sigma} \times 100)$	0.0037	0.0017	0.0018	0.0019	0.0019
	$\text{Bias}(\{\sigma_{x_i}\}_{i=1}^{n_x})$	0.0050	0.0193	0.0853	0.0192	**0.0187**
	Pct of nonstationary draws	0.00	5.28	0.00	0.00	0.00

Table 2: Monte Carlo study: Bootstrapping the SR approach

The Monte Carlo study is carried out for $T = 250$ and $n_y = 25$ using the maturities selected in Section 4.1. We use $M = 5,000$ replications and $B = 1,000$ in the bootstrap. The asymptotic standard errors are computed using (13) with $w_T = 10$ when $\rho_{Time} = 0.6$ and $w_D = 5$ when $\rho_{Cross} = 0.4$, otherwise $w_T = 0$ and $w_D = 0$. Standard resampling is used, unless $\phi_1 = 0.4$, where we let $p = 1$, and estimate ϕ_1 using a preliminary bootstrap with 100 draws. Standard errors for $\hat{\theta}_2^{step3}$ are computed using (9). Bold figures indicate the best method for computing standard errors and rejection probabilities in the SR approach. The (Q)ML estimates are obtained using the Kalman filter with $\mathcal{IID}$ measurement errors with a diagonal covariance matrix. Standard errors and rejection probabilities for the (Q)ML estimates are computed by pre- and post-multiplying the variance of the score with the inverse of the Hessian matrix, except in Case I where we use the inverse of the variance of the score.

		Estimation bias		Standard errors ($\times 10^{-3}$)					Reject probabilities: 5% level			
		(Q)ML	SR	(Q)ML True	(Q)ML Bias$_{Asymp}$	SR True	SR Bias$_{Asymp}$	SR Bias$_{Boot}$	(Q)ML Asymp	Asymp	SR Boot*	SR Boot**
Case I: $\mathcal{IID}$ errors	α	0.0000	0.0000	0.0446	0.0007	0.0779	**-0.0001**	0.0006	5.22	5.94	5.54	**5.40**
	Φ_{11}	0.0000	0.0000	0.0609	0.0022	0.0631	**-0.0008**	-0.0036	**4.04**	**4.78**	6.50	6.38
	h_0	-0.0001	0.0000	0.1192	-0.0043	0.1230	**-0.0127**	0.0157	8.24	6.74	1.68	**6.02**
	h_x	-0.0211	-0.0017	23.9026	-2.0449	24.6930	-3.7802	**1.2100**	9.68	8.74	2.14	**6.80**
	Σ_{11}	0.0000	0.0000	0.0123	0.0000	0.0241	**0.0001**	0.0004	5.10	5.52	**5.12**	4.32
Case II: autocorrelated errors ($\rho_{Time} = 0.6$)	α	0.0000	0.0000	0.0940	-0.0428	0.0925	0.0083	**0.0008**	27.94	3.98	**5.10**	4.72
	Φ_{11}	0.0000	0.0000	0.1422	-0.0700	0.1425	0.0255	**-0.0078**	30.96	2.32	6.82	**6.26**
	h_0	-0.0001	0.0000	0.1191	-0.0144	0.1228	**-0.0126**	0.0156	10.86	6.76	1.80	**5.82**
	h_x	-0.0211	-0.0017	23.8851	-3.8651	24.6804	-3.7660	**1.2044**	12.62	8.64	2.06	**6.52**
	Σ_{11}	-0.0000	-0.0000	0.0249	-0.0110	0.0241	**0.0001**	0.0003	27.30	5.62	**5.36**	4.32
Case III: autocorrelated and heteroskedastic errors ($\rho_{Time} = 0.6$, $\rho_{Rv} = 0.98$)	α	0.0000	0.0000	0.1170	-0.0496	0.1105	0.0148	**0.0008**	25.36	3.10	5.04	**4.98**
	Φ_{11}	0.0000	0.0000	0.2114	-0.1046	0.2117	0.0340	**-0.0134**	30.80	2.24	6.96	**6.30**
	h_0	-0.0001	0.0000	0.1185	-0.0141	0.1227	**-0.0126**	0.0148	10.68	6.78	1.70	**6.06**
	h_x	-0.0208	-0.0018	23.7893	-3.8145	24.6864	-3.7665	**1.0424**	12.34	8.58	2.14	**6.78**
	Σ_{11}	0.0000	0.0000	0.0279	-0.0105	0.0242	**0.0000**	0.0003	22.78	6.06	5.74	**4.76**
Case IV: cross-correlated errors ($\phi_1 = 0.4$)	α	0.0000	0.0000	0.0640	-0.0020	0.0827	0.0692	**0.0013**	6.32	0.12	5.26	**5.10**
	Φ_{11}	0.0000	0.0000	0.0950	-0.0031	0.0958	0.0431	**-0.0030**	5.46	0.84	**5.98**	6.68
	h_0	-0.0001	0.0000	0.1191	-0.0144	0.1228	**-0.0126**	0.0155	10.92	6.76	1.72	**6.06**
	h_x	-0.0211	-0.0017	23.9087	-3.8776	24.6892	-3.7739	**1.1670**	12.58	8.68	2.12	**6.92**
	Σ_{11}	0.0000	0.0000	0.0169	-0.0005	0.0241	**0.0001**	0.0003	5.96	5.62	**5.22**	4.26
Case V: autocorrelated, heteroskedastic, and cross-correlated errors ($\rho_{Time} = 0.6$, $\rho_{Rv} = 0.98$, $\phi_1 = 0.4$)	α	0.0000	0.0000	0.1720	-0.0796	0.1465	0.0593	**0.0050**	28.04	1.18	**4.98**	5.22
	Φ_{11}	0.0000	0.0000	0.3297	-0.1638	0.3296	0.0250	**-0.0116**	31.00	3.68	**6.10**	6.92
	h_0	-0.0001	0.0000	0.1188	-0.0141	0.1224	**-0.0124**	0.0139	10.80	6.76	1.84	**5.86**
	h_x	-0.0208	-0.0018	23.7808	-3.8131	24.6797	-3.7511	**0.8452**	12.28	8.74	2.24	**6.66**
	Σ_{11}	0.0000	0.0000	0.0378	-0.0163	0.0244	**-0.0001**	0.0002	25.84	6.52	6.26	**5.06**

For the rejection probabilities, i) Asymp denotes the asymptotical t-test $\left| \frac{\hat{\theta}_{11} - \theta_{11}^o}{SE(\hat{\theta}_{11})^{asymp}} \right| \geq 1.96$, ii) Boot* refers to the modified t-test $\left| \frac{\hat{\theta}_{11} - \theta_{11}^o}{SE(\hat{\theta}_{11})^{Boot}} \right| \geq 1.96$ with bootstrapped standard error, and iii) Boot** denotes the t-test $z_{0.025}^* \leq \frac{\hat{\theta}_{11} - \theta_{11}^o}{SE(\hat{\theta}_{11})^{asymp}} \leq z_{0.975}^*$ where z_p^* is the bootstrapped p-percentile.

Table 3: In-sample fit: The objective functions

This table reports $100\sqrt{2 \times Q_{1:T}^{step1}}$ and $100\sqrt{2 \times Q_{1:T}^{step3}}$ from the first and third step in the SR approach. Figures in squared brackets refer to the scaled objective functions from 2009 to 2013 when short-term bond yields are at the ZLB. Figures in bold highlight the best in-sample fit for a given estimation step and sample period.

	Step 1			Step 3		
	ATSM	QTSM	SRM	ATSM	QTSM	SRM
Sample: 1961-2013						
2 factors	9.627	**8.327**	9.414	9.725	**8.780**	11.936
	[11.951]	[**6.866**]	[8.306]	[11.353]	[10.314]	[**5.177**]
3 factors	2.895	**2.702**	2.735	2.896	**2.718**	2.786
	[2.510]	[**1.347**]	[1.404]	[2.517]	[**1.451**]	[1.655]
4 factors	1.057	1.030	**1.013**	1.058	1.034	**1.025**
	[1.007]	[0.557]	[**0.517**]	[0.995]	[0.559]	[**0.558**]
Sample: 1990-2013						
2 factors	7.457	**6.554**	6.818	7.457	**6.650**	6.998
	[7.677]	[**4.457**]	[4.402]	[7.674]	[**4.621**]	[5.601]
3 factors	1.808	**1.613**	1.692	1.829	**1.630**	1.754
	[1.723]	[**1.265**]	[1.399]	[1.804]	[**1.292**]	[1.560]
4 factors	0.801	0.766	**0.749**	0.807	0.772	**0.763**
	[0.760]	[0.592]	[**0.510**]	[0.755]	[**0.571**]	[0.607]

Table 4: Model comparison by quasi maximum likelihood

The quasi log-likelihood functions are evaluated using the CDKF, which simplifies to the Kalman filter for the ATSM. Bold figures denote the best performing model for a given sample and number of pricing factors.

	Data: 1961-2013 $\mathcal{L}_{1:T}^{CDKF}$	Data: 1990-2013 $\mathcal{L}_{1:T}^{CDKF}$
2 factors		
ATSM	667	916
QTSM	**1240**	**1053**
SRM	883	1010
3 factors		
ATSM	3349	2476
QTSM	**3763**	**2565**
SRM	3530	2550
4 factors		
ATSM	4782	2971
QTSM	**5173**	3007
SRM	4937	**3073**

Table 5: Estimation results for three-factor models: sample from 1961-2013

Asymptotic standard errors for $\hat{\boldsymbol{\theta}}_{11}^{step3}$ are given by (13) with $w_D = 5$ and $w_T = 10$. For the Gaussian ATSM and the shadow rate model, bootstrapped standard errors using 1,000 draws for $\hat{\boldsymbol{\theta}}_{11}^{step3}$ are shown in brackets. For the QTSM, 95%th confidence intervals are provided for $\hat{\boldsymbol{\theta}}_{11}^{step3}$ using the 2.5th and 97.5th percentiles of $\hat{\boldsymbol{\theta}}_{11}^{step3}$ in the bootstrap. All bootstraps for $\hat{\boldsymbol{\theta}}_{11}$ use a bias-adjusted AR(1) model (based on 100 draws to account for cross-correlation) and draws of $\hat{\boldsymbol{\theta}}_2$ from its asymptotic distribution. For $\hat{\boldsymbol{\theta}}_2^{step3}$, asymptotic standard errors are given by (9), and bootstrapped standard errors from 5,000 draws are shown in brackets.

	ATSM		QTSM		SRM	
	Estimate	SE	Estimate	CI$_{95\%}$ or SE	Estimate	SE
α	0.0124	0.0016 [0.0003]	-	-	0.0153	0.0037 [0.0071]
A_{12}	-	-	0.9886	[0.9524, 1.0000]	-	-
A_{13}	-	-	0.9915	[0.9651, 1.0000]	-	-
A_{23}	-	-	0.8642	[0.5654, 1.0398]	-	-
Φ_{11}	0.0022	0.0005 [0.0001]	0.0011	[0.0002, 0.0016]	0.0013	0.0005 [0.0004]
Φ_{22}	0.0355	0.0027 [0.0007]	0.0405	[0.0366, 0.0484]	0.0427	0.0040 [0.0054]
Φ_{33}	0.0685	0.0063 [0.0016]	0.0806	[0.0499, 0.0878]	0.0666	0.0069 [0.0072]
μ_1	-	-	0.0231	[0.0000, 0.2929]	-	-
μ_2	-	-	0.0035	[0.0001, 0.1265]	-	-
μ_3	-	-	0.1122	[0.0000, 0.1535]	-	-
$h_0(1,1)$	-1.03×10^{-4}	6.65×10^{-5} [7.91×10^{-5}]	-0.0017	9.27×10^{-4} [0.0011]	-1.67×10^{-4}	8.74×10^{-5} [1.01×10^{-4}]
$h_0(2,1)$	3.83×10^{-4}	2.57×10^{-4} [2.76×10^{-4}]	-0.0093	0.0050 [0.0052]	9.15×10^{-4}	5.00×10^{-4} [5.65×10^{-4}]
$h_0(3,1)$	-4.69×10^{-4}	2.59×10^{-5} [2.50×10^{-4}]	0.0155	0.0050 [0.0052]	-0.0010	5.05×10^{-4} [5.46×10^{-4}]
$h_x(1,1)$	0.9847	0.0079 [0.0094]	0.9733	0.0072 [0.0091]	0.9822	0.0079 [0.0091]
$h_x(1,2)$	0.0252	0.0082 [0.0100]	0.0082	0.0052 [0.0069]	0.0216	0.0078 [0.0094]
$h_x(1,3)$	0.0186	0.0114 [0.0121]	0.0041	0.0094 [0.0090]	0.0188	0.0099 [0.0107]
$h_x(2,1)$	0.0489	0.0289 [0.0327]	0.0677	0.0252 [0.0426]	0.0906	0.0436 [0.0507]
$h_x(2,2)$	0.9668	0.0444 [0.0378]	1.0310	0.0329 [0.0339]	1.0097	0.0678 [0.0565]
$h_x(2,3)$	0.0611	0.0625 [0.0452]	0.1115	0.0481 [0.0438]	0.0866	0.0819 [0.0634]
$h_x(3,1)$	-0.0557	0.0291 [0.0300]	-0.0754	0.0255 [0.0422]	-0.1004	0.0443 [0.0493]
$h_x(3,2)$	-0.0151	0.0391 [0.0338]	-0.0649	0.0338 [0.0335]	-0.0637	0.0625 [0.0544]
$h_x(3,3)$	0.8685	0.0548 [0.0410]	0.8295	0.0470 [0.0442]	0.8452	0.0750 [0.0613]
Σ_{11}	3.56×10^{-4}	1.90×10^{-5} [1.62×10^{-5}]	0.0023	1.25×10^{-4} [1.03×10^{-4}]	3.30×10^{-4}	1.71×10^{-5} [1.47×10^{-5}]
Σ_{21}	-6.22×10^{-4}	8.14×10^{-5} [7.26×10^{-5}]	-0.0030	7.07×10^{-4} [7.09×10^{-4}]	-6.49×10^{-4}	1.18×10^{-4} [1.12×10^{-4}]
Σ_{22}	0.0011	5.88×10^{-5} [5.27×10^{-5}]	0.0101	5.23×10^{-4} [4.62×10^{-4}]	0.0018	8.77×10^{-5} [7.87×10^{-5}]
Σ_{31}	3.95×10^{-4}	7.65×10^{-5} [6.98×10^{-5}]	0.0015	6.92×10^{-4} [7.03×10^{-4}]	4.54×10^{-4}	1.16×10^{-4} [1.11×10^{-4}]
Σ_{32}	-0.0010	5.08×10^{-5} [4.76×10^{-5}]	-0.0101	5.67×10^{-4} [5.06×10^{-4}]	-0.0017	8.00×10^{-5} [7.17×10^{-5}]
Σ_{33}	4.31×10^{-4}	4.23×10^{-5} [3.41×10^{-5}]	0.0029	1.81×10^{-4} [1.58×10^{-4}]	4.40×10^{-4}	4.29×10^{-5} [3.49×10^{-5}]

Table 6: Estimation results for three-factor models: sample from 1990-2013

Asymptotic standard errors for $\hat{\boldsymbol{\theta}}_{11}^{step3}$ are given by (13) with $w_D = 5$ and $w_T = 10$. For the Gaussian ATSM and the shadow rate model, bootstrapped standard errors using 1,000 draws for $\hat{\boldsymbol{\theta}}_{11}^{step3}$ are shown in brackets. For the QTSM, 95% confidence intervals are provided for $\hat{\boldsymbol{\theta}}_{11}^{step3}$ using the 2.5th and 97.5th percentiles of $\hat{\boldsymbol{\theta}}_{11}^{step3}$ in the bootstrap. All bootstraps for $\hat{\boldsymbol{\theta}}_{11}$ use a bias-adjusted AR(1) model (based on 100 draws to account for cross-correlation) and draws of $\hat{\boldsymbol{\theta}}_2$ from its asymptotic distribution. For $\hat{\boldsymbol{\theta}}_2^{step3}$, asymptotic standard errors are given by (9), and bootstrapped standard errors from 5,000 draws are shown in brackets.

	ATSM		QTSM		SRM	
	Estimate	SE	Estimate	CI$_{95\%}$ or SE	Estimate	SE
α	0.0093	0.0007 [0.0003]	-	-	0.0099	0.0008 [0.0004]
A_{12}	-	-	0.9725	[0.9491, 0.9998]	-	-
A_{13}	-	-	0.9861	[0.9681, 0.9983]	-	-
A_{23}	-	-	0.6846	[0.5691, 1.1857]	-	-
Φ_{11}	0.0043	0.0004 [0.0002]	0.0028	[0.0021, 0.0032]	0.0035	0.0004 [0.0002]
Φ_{22}	0.0487	0.0020 [0.0035]	0.0459	[0.0424, 0.0541]	0.0475	0.0024 [0.0035]
Φ_{33}	0.0518	0.0033 [0.0045]	0.0724	[0.0589, 0.0768]	0.0558	0.0050 [0.0039]
μ_1	-	-	0.0040	[0.0000, 0.0504]	-	-
μ_2	-	-	0.0028	[0.0000, 0.1080]	-	-
μ_3	-	-	0.0977	[0.0000, 0.1079]	-	-
$h_0(1,1)$	-2.81×10^{-4}	1.36×10^{-4} [2.23×10^{-4}]	-2.23×10^{-4}	0.0013 [0.0016]	-3.32×10^{-4}	1.54×10^{-4} [2.32×10^{-4}]
$h_0(2,1)$	0.0078	0.0035 [0.0066]	-0.0135	0.0081 [0.0098]	0.0037	0.0015 [0.0028]
$h_0(3,1)$	-0.0079	0.0034 [0.0065]	0.0181	0.0081 [0.0097]	-0.0039	0.0015 [0.0027]
$h_x(1,1)$	0.9530	0.0214 [0.0302]	0.9427	0.0225 [0.0242]	0.9474	0.0234 [0.0297]
$h_x(1,2)$	-0.0216	0.0255 [0.0310]	-0.0095	0.0105 [0.0150]	-0.0230	0.0235 [0.0290]
$h_x(1,3)$	-0.0237	0.0273 [0.0317]	-0.0218	0.0194 [0.0178]	-0.0285	0.0280 [0.0307]
$h_x(2,1)$	1.2792	0.5281 [0.8856]	0.2415	0.0967 [0.1451]	0.5884	0.2268 [0.3573]
$h_x(2,2)$	2.4472	0.7791 [0.8986]	1.1202	0.0688 [0.0916]	1.5625	0.2837 [0.3454]
$h_x(2,3)$	1.5652	0.8202 [0.9186]	0.2371	0.1026 [0.1077]	0.6813	0.3231 [0.3651]
$h_x(3,1)$	-1.3025	0.5172 [0.8726]	-0.2611	0.0965 [0.1427]	-0.6150	0.2167 [0.3468]
$h_x(3,2)$	-1.5152	0.7694 [0.8864]	-0.1584	0.0702 [0.0907]	-0.6272	0.2764 [0.3364]
$h_x(3,3)$	-0.6356	0.8093 [0.9063]	0.7070	0.1013 [0.1070]	0.2476	0.3132 [0.3556]
Σ_{11}	3.92×10^{-4}	3.55×10^{-5} [2.92×10^{-5}]	0.0028	2.49×10^{-4} [2.08×10^{-4}]	3.73×10^{-4}	3.21×10^{-5} [2.67×10^{-5}]
Σ_{21}	-0.0049	0.0016 [0.0014]	-0.0033	0.0020 [0.0020]	-0.0018	5.73×10^{-4} [5.21×10^{-4}]
Σ_{22}	0.0108	0.0007 [0.0006]	0.0169	0.0012 [0.0011]	0.0042	2.49×10^{-4} [2.39×10^{-4}]
Σ_{31}	0.0046	0.0015 [0.0014]	0.0007	0.0019 [0.0019]	0.0014	5.53×10^{-4} [5.08×10^{-4}]
Σ_{32}	-0.0108	0.0007 [0.0006]	-0.0170	0.0012 [0.0011]	-0.0042	2.52×10^{-4} [2.42×10^{-4}]
Σ_{33}	1.66×10^{-5}	1.28×10^{-5} [1.21×10^{-5}]	0.0023	2.39×10^{-4} [2.41×10^{-4}]	1.69×10^{-4}	1.30×10^{-5} [1.24×10^{-5}]

BANK OF ENGLAND

Table 7: Model summary: Unconditional moments

An "Q" indicates that the moments across maturity are approximately matched by the QTSM. An "S" indicates that the moments across maturity are approximately matched by the SRM.

	Means	Standard deviations	LPY(i)	LPY(ii)
Sample: 1961-2013				
2 factors	S	S	S	S
3 factors	Q,S	Q,S	S	Q,S
4 factors	Q,S	Q,S	S	Q,S
Sample: 1990-2013				
2 factors	Q,S	Q	Q,S	Q
3 factors	Q,S	Q	Q,S	S
4 factors	Q,S	Q,S	Q,S	S

Table 8: Conditional volatility of bond yields

This table reports the slope coefficient and R^2 of regressing volatility in the data on a constant and model-implied volatility. In the left part of the table, conditional volatility in the data is obtained using a rolling standard deviation of daily bond yields in the past six months, denoted $\sigma_t^{Rolling}$. In the right part of the table, conditional volatility in the data is obtained by a GARCH(1,1) model for changes in monthly bond yields, denoted σ_t^{GARCH}. The model-implied conditional volatilities one-month ahead in time period t are computed from a local linearization of bond yields at $\hat{\mathbf{x}}_{t-1}$. Bold figures indicate the preferred model for a given measure of volatility and for a given sample.

	Data: $\sigma_t^{Rolling}$				Data: σ_t^{GARCH}			
	QTSM		SRM		QTSM		SRM	
	Slope	R^2	Slope	R^2	Slope	R^2	Slope	R^2
Sample: 1961-2013								
0.5-year bond yield	**1.42**	**0.28**	1.21	0.07	**1.26**	**0.33**	0.92	0.06
2-year bond yield	**1.25**	**0.30**	1.19	0.09	**1.21**	**0.38**	1.00	0.09
5-year bond yield	**0.96**	**0.24**	0.95	0.07	**0.85**	**0.34**	0.72	0.07
10-year bond yield	**0.75**	**0.14**	0.59	0.02	**0.51**	**0.24**	0.37	0.02
Sample: 1990-2013								
0.5-year bond yield	0.45	0.09	**2.69**	**0.17**	0.21	0.06	**1.47**	**0.16**
2-year bond yield	0.49	0.15	**1.99**	**0.24**	0.30	0.17	**1.49**	**0.41**
5-year bond yield	0.34	0.09	**1.47**	**0.18**	0.18	0.11	**0.98**	**0.35**
10-year bond yield	0.02	0.00	**0.44**	**0.01**	-0.04	0.01	0.18	0.01

Table 9: Average forecasting results

The figure reports the average root mean squared prediction errors (RMSPEs) across all bond yields in the forecasting study from January 2005 to December 2013. The RMSPEs are generated from models estimated recursively from 1961 or 1990 to the month prior to the forecast. The forecasted bond yields in the SRMs are computed by Monte Carlo integration using 10,000 draws. For a given number of pricing factors and a given starting point for the model estimation, bold figures indicate the model with the lowest RMSPEs. Figures in boxes denote the lowest RMSPEs for a given model when comparing part $\mathcal{A}$ and $\mathcal{B}$ of the table.

| | Part $\mathcal{A}$: Model estimated from 1961 | | | | Part $\mathcal{B}$: Model estimated from 1990 | | | |
| | Forecasting horizon | | | | Forecasting horizon | | | |
	1 mth	3 mths	6 mths	12 mths	1 mth	3 mths	6 mths	12 mths
Random walk	25.87	49.66	72.54	94.76	25.87	49.66	72.54	94.76
2 factor models								
ATSM	41.50	59.97	78.40	104.92	41.04	62.83	87.41	128.27
QTSM	**27.92**	**51.78**	76.81	106.47	27.61	55.41	86.43	122.09
SRM	39.27	55.68	**76.01**	**99.98**	27.17	52.74	**81.12**	**119.98**
3 factor models								
ATSM	40.51	60.86	80.48	**108.02**	40.50	62.83	88.79	133.05
QTSM	**26.62**	53.00	79.46	110.32	**26.49**	**53.49**	**83.02**	**123.55**
SRM	26.70	**52.33**	**78.09**	109.41	27.30	54.32	84.48	126.35
4 factor models								
ATSM	40.20	59.71	77.85	104.32	41.33	64.73	90.10	128.61
QTSM	27.50	56.01	89.04	131.55	**27.10**	55.59	85.30	124.66
SRM	**26.32**	**50.73**	**74.65**	**102.61**	30.26	**51.37**	**76.08**	**110.10**

Robust standard errors for elements in $\hat{\boldsymbol{\theta}}_{11}^{step3}$ are computed using (13) with $w_D = 5$ and $w_T = 10$. For elements in $\hat{\boldsymbol{\theta}}_2^{step3}$, robust standard errors are computed using (9). The standard errors in the short sample from 1990-2013 for the three-factor model are approximated by fixing μ_2 to zero and treating this parameter as known.

| | Data: 1961-2013 | | | | Data: 1990-2013 | | | |
| | 2 factors | | 3 factors | | 2 factors | | 3 factors | |
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
α	-0.0076	5.57×10^{-4}	1.50×10^{-4}	0.0002	1.82×10^{-5}	6.57×10^{-5}	1.68×10^{-4}	0.0016
A_{12}	0.3615	0.0055	2.5659	0.0161	1.0237	0.0156	1.0836	0.0043
A_{13}	-	-	2.9113	0.0133	-	-	1.1422	0.0111
A_{23}	-	-	1.1520	0.0058	-	-	1.2615	0.0175
Φ_{11}	0.0024	4.14×10^{-8}	0.0011	0.0003	0.0010	0.0021	0.0028	1.98×10^{-4}
Φ_{22}	0.0491	0.0022	0.0419	0.0031	0.0258	0.0033	0.0442	0.0017
Φ_{33}	-	-	0.0816	0.0087	-	-	0.0787	0.0063
μ_1	0.1233	0.0022	0.0664	0.0150	1.68×10^{-4}	0.0224	0.0918	0.0174
μ_2	0.0193	8.71×10^{-6}	0.0369	0.0082	0.0800	0.0225	0.0000	-
μ_3	-	-	0.0034	0.0074	-	-	0.0119	0.0039
$h_0(1,1)$	0.0018	6.68×10^{-4}	1.62×10^{-4}	0.0002	-0.0017	0.0011	-0.0149	0.0070
$h_0(2,1)$	-0.0018	0.0023	-0.0021	0.0015	0.0023	0.0011	-0.2481	0.1124
$h_0(3,1)$	-	-	0.0032	0.0015	-	-	0.4045	0.1950
$h_x(1,1)$	0.9826	0.0067	0.9761	0.0073	0.9587	0.0190	1.0175	0.0924
$h_x(1,2)$	0.0034	0.0022	0.0056	0.0035	0.0094	0.0174	0.0398	0.0333
$h_x(1,3)$	-	-	0.0032	0.0051	-	-	-0.0313	0.0170
$h_x(2,1)$	0.0179	0.0228	0.0854	0.0336	0.0284	0.0187	-4.0149	1.6836
$h_x(2,2)$	0.9817	0.0089	1.0129	0.0332	0.9776	0.0170	0.1381	0.5268
$h_x(2,3)$	-	-	0.0750	0.0441	-	-	0.5685	0.2330
$h_x(3,1)$	-	-	-0.0866	0.0332	-	-	6.5232	2.9065
$h_x(3,2)$	-	-	-0.0430	0.0364	-	-	1.3900	0.8786
$h_x(3,3)$	-	-	0.8760	0.0508	-	-	0.0581	0.3745
Σ_{11}	0.0015	5.48×10^{-5}	0.0013	6.94×10^{-5}	0.0067	4.18×10^{-4}	0.0031	4.90×10^{-4}
Σ_{21}	-0.0011	4.44×10^{-4}	-0.0022	5.75×10^{-4}	-0.0069	5.63×10^{-4}	-0.0131	0.0130
Σ_{22}	0.0062	6.17×10^{-4}	0.0091	4.53×10^{-4}	0.0039	3.84×10^{-4}	0.0165	0.0142
Σ_{31}	-	-	6.38×10^{-4}	6.23×10^{-4}	-	-	0.0177	0.0216
Σ_{32}	-	-	-0.0093	6.38×10^{-4}	-	-	-0.0247	0.0279
Σ_{33}	-	-	0.0033	3.06×10^{-4}	-	-	0.0131	0.0020

Table 11: The three-factor hybrid model: Conditional volatility of bond yields

This table reports the slope coefficient and R^2 of regressing volatility in the data on a constant and model-implied volatility. Conditional volatility in the data is either obtained using a rolling standard deviation of daily bond yields in the past six months, denoted $\sigma_t^{Rolling}$, or a GARCH(1,1) model for changes in monthly bond yields, denoted σ_t^{GARCH}. The model-implied conditional volatilities one-month ahead in time period t are computed from a local linearization of bond yields at $\hat{\mathbf{x}}_{t-1}$. Figures marked by a box indicate that the R^2 for the hybrid model is larger than the R^2 for both the QTSM and the SRM in Table 8.

| | Sample: 1961-2013 | | | | Sample: 1990-2013 | | | |
| | Data: $\sigma_t^{Rolling}$ | | Data: σ_t^{GARCH} | | Data: $\sigma_t^{Rolling}$ | | Data: σ_t^{GARCH} | |
	Slope	R^2	Slope	R^2	Slope	R^2	Slope	R^2
0.5-year bond yield	1.14	0.32	1.04	0.40	0.38	0.08	0.18	0.06
2-year bond yield	1.23	0.32	1.21	0.43	0.58	0.15	0.35	0.16
5-year bond yield	1.04	0.24	0.93	0.34	0.41	0.10	0.21	0.11
10-year bond yield	0.80	0.09	0.55	0.16	0.02	0.00	−0.05	0.01

Figure 1: The QTSM: Non-linear filtering

The objective function for filtering out x_t in a QTSM with one pricing factor. The risk-neutral parameters are $\mu_1 = 0.0790$, $\Phi_{11} = 0.0072$, and $\Sigma_{11} = 0.0066$, which are the optimal values in the sample from 1990 to 2013.
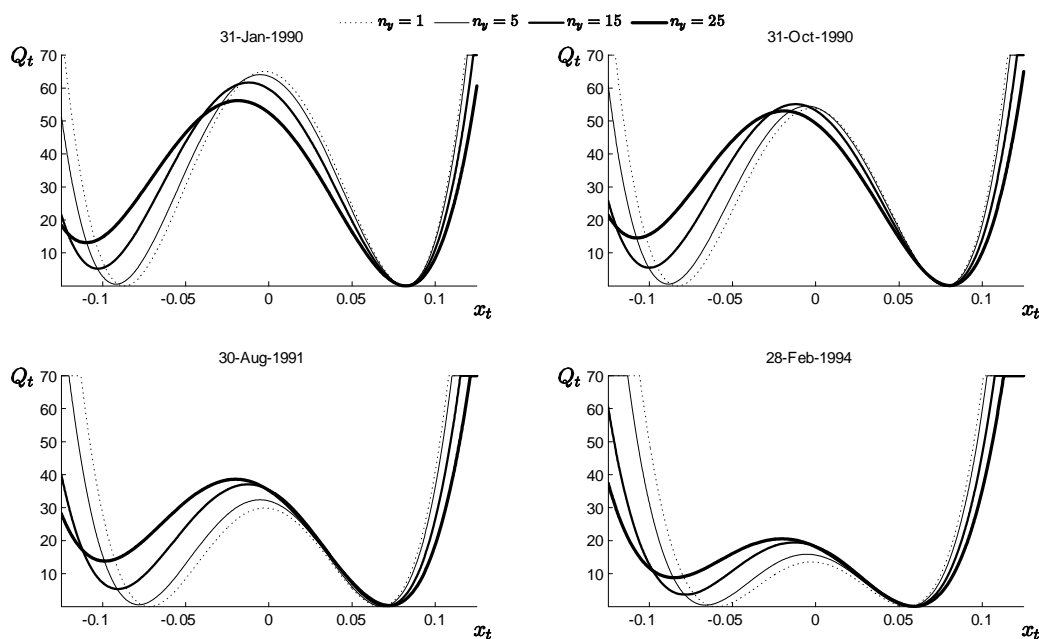
Figure 2: Cross-section Fit: RMSEs by maturity
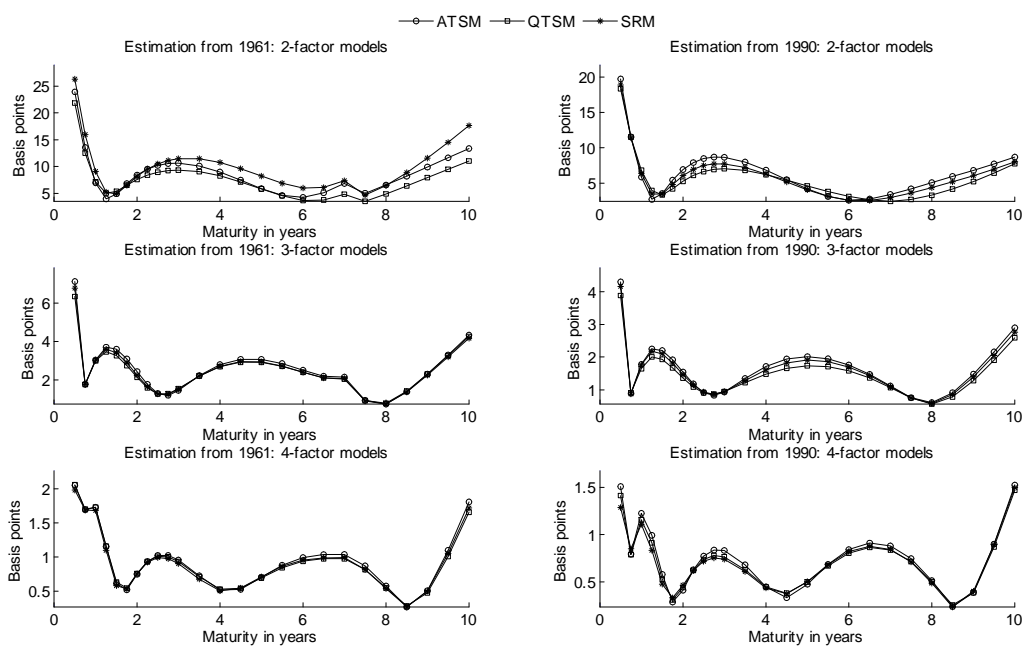
Figure 3: The SRM: Accuracy of approximated bond yields

The pricing errors when evaluating bond yields in the three-factor SRM by the second-order approximation at $\{\hat{\mathbf{x}}_t\}_{t=1}^{T}$ and the estimated parameters from Table 5 and 6, respectively. The true solution to bond yields is approximated using a Monte Carlo method with 100,000 draws and anti-thetic sampling.
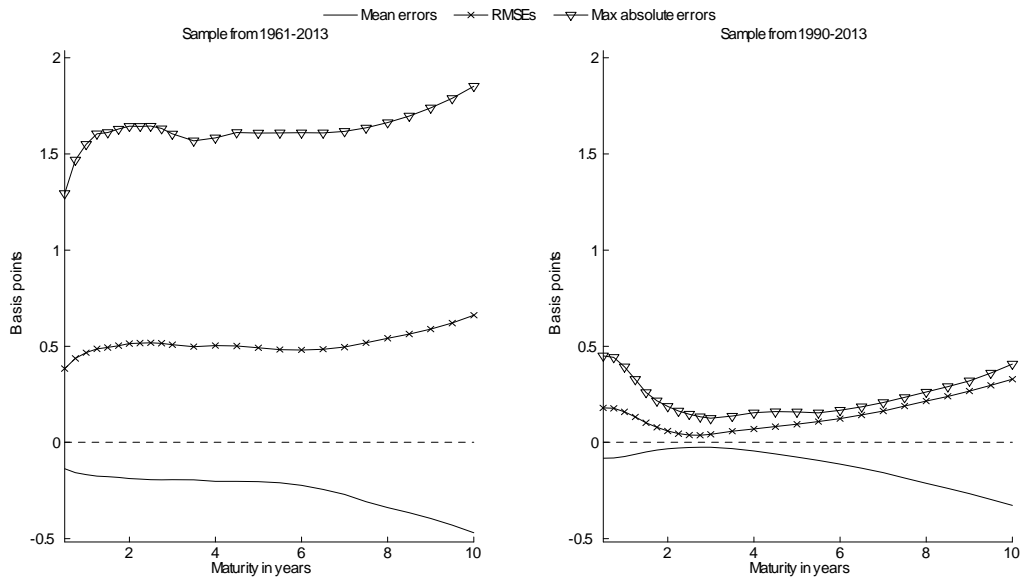
Figure 4: Three-factor models: Unconditional moments in sample from 1961-2013
All model-implied moments are computed from a simulated time series of 100,000 observations. Empirical moments are computed from September 1971 to December 2013 to avoid missing observations for long bond yields.
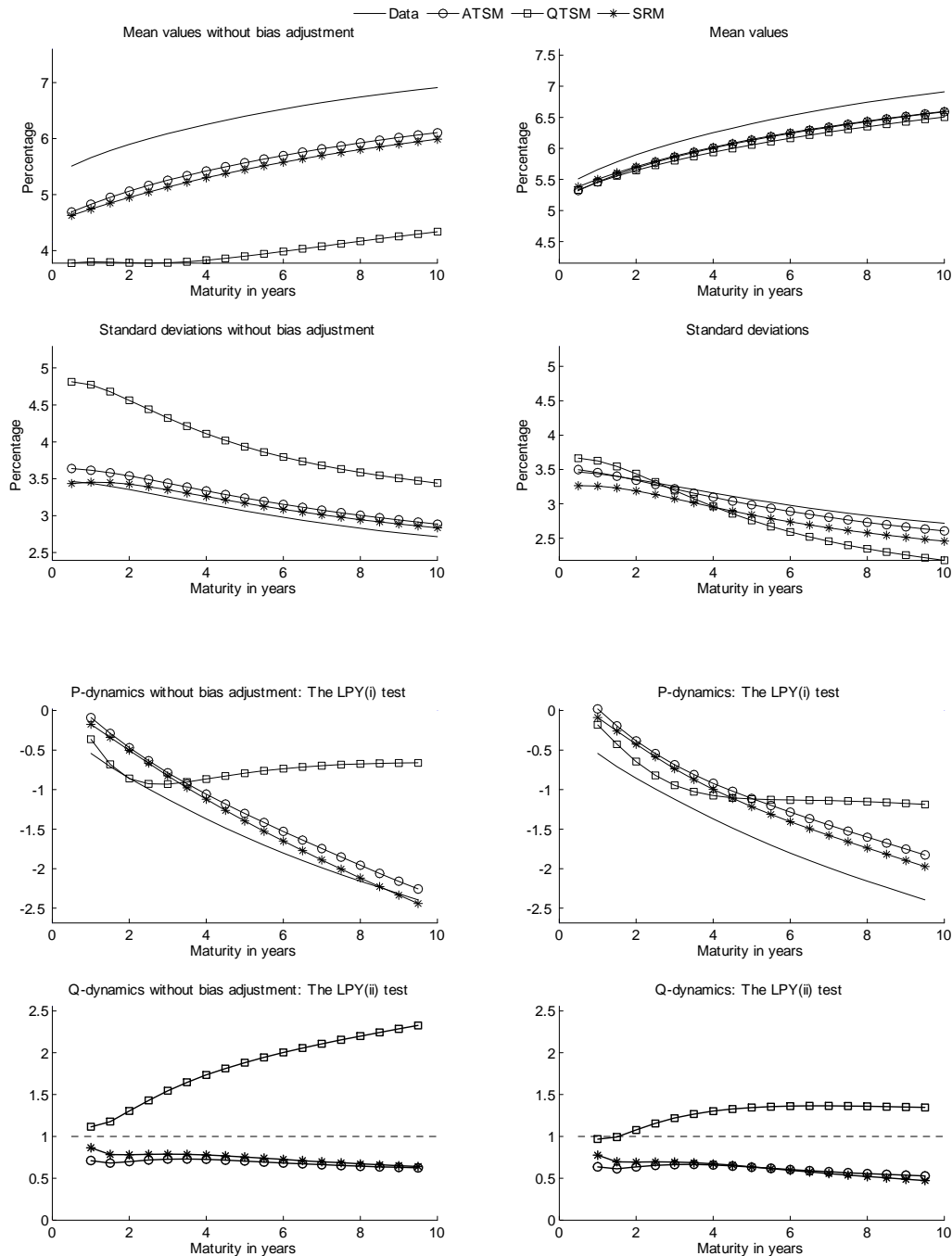
Figure 5: Three-factor models: Unconditional moments in sample from 1990-2013
All model-implied moments are computed from a simulated time series of 100,000 observations. Empirical moments are computed from January 1990 to December 2013.
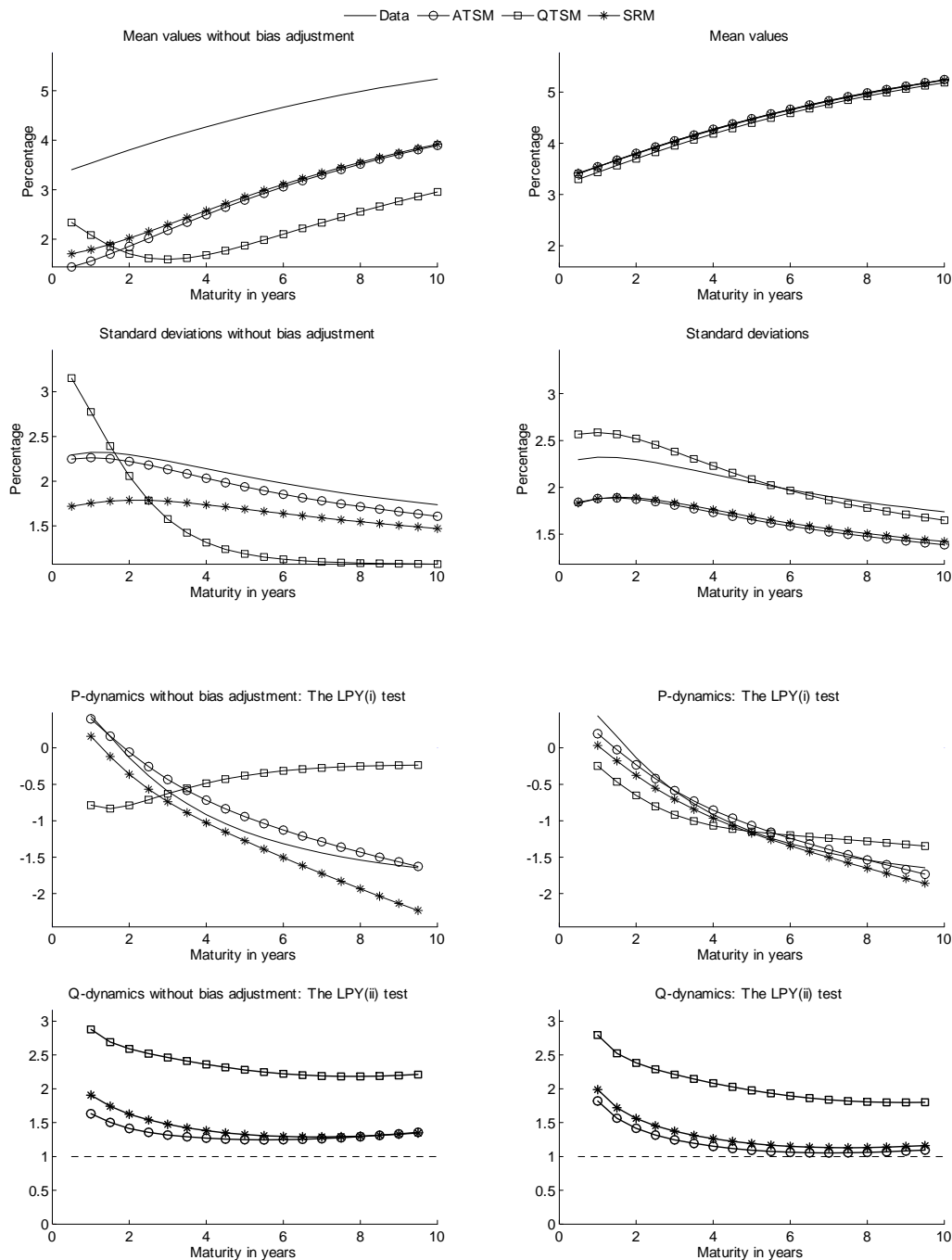
Figure 6: Two-factor models: Unconditional moments in sample from 1961-2013

All model-implied moments are computed from a simulated time series of 100,000 observations.
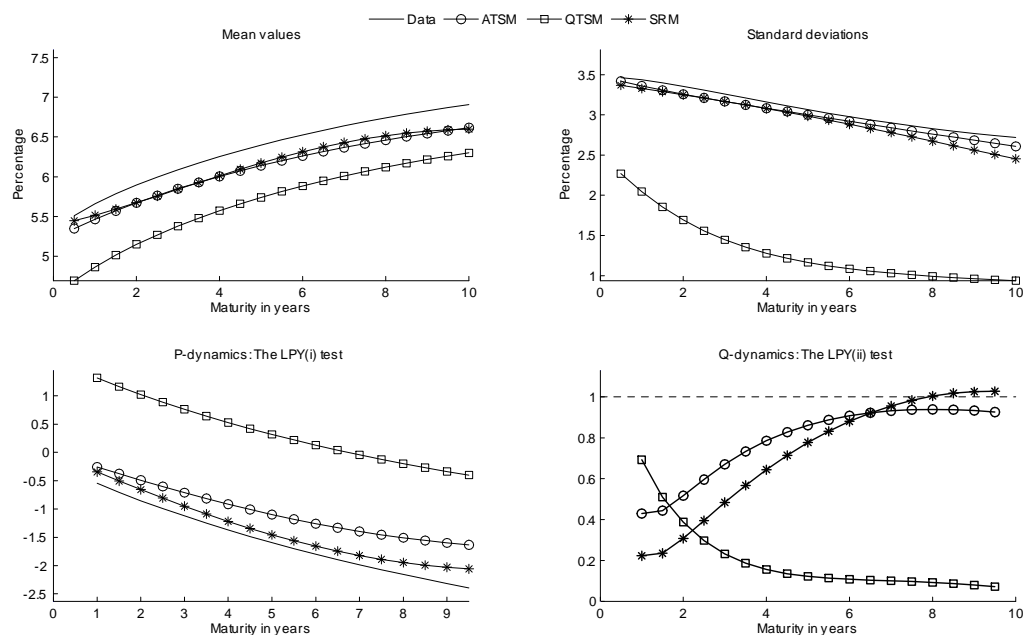
Figure 7: Two–factor models: Unconditional moments in sample from 1990-2013

All model-implied moments are computed from a simulated time series of 100,000 observations.
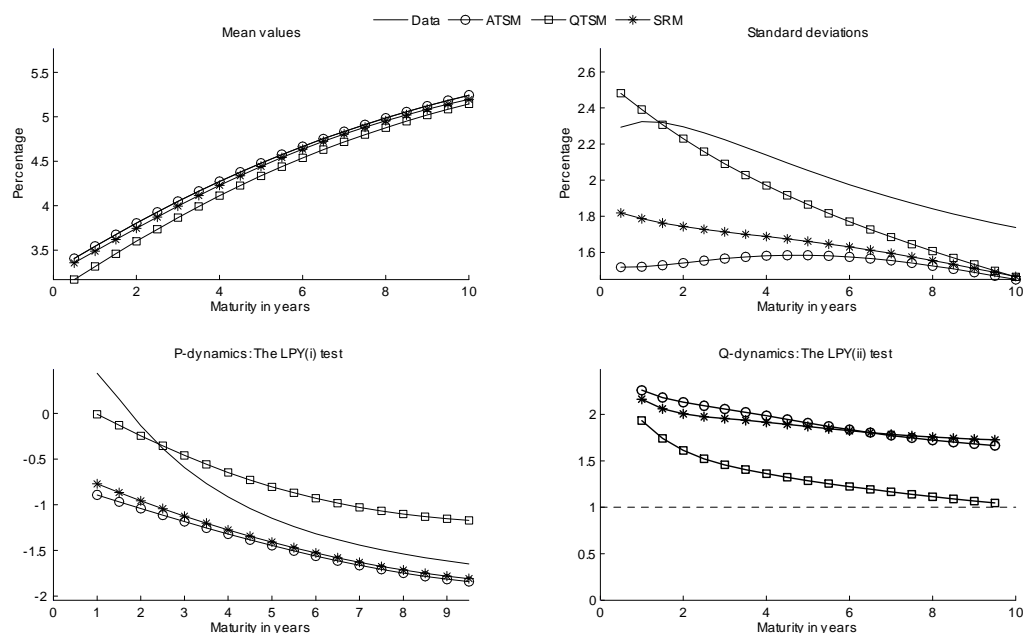
Figure 8: Four-factor models: Unconditional moments in sample from 1961-2013
All model-implied moments are computed from a simulated time series of 100,000 observations.
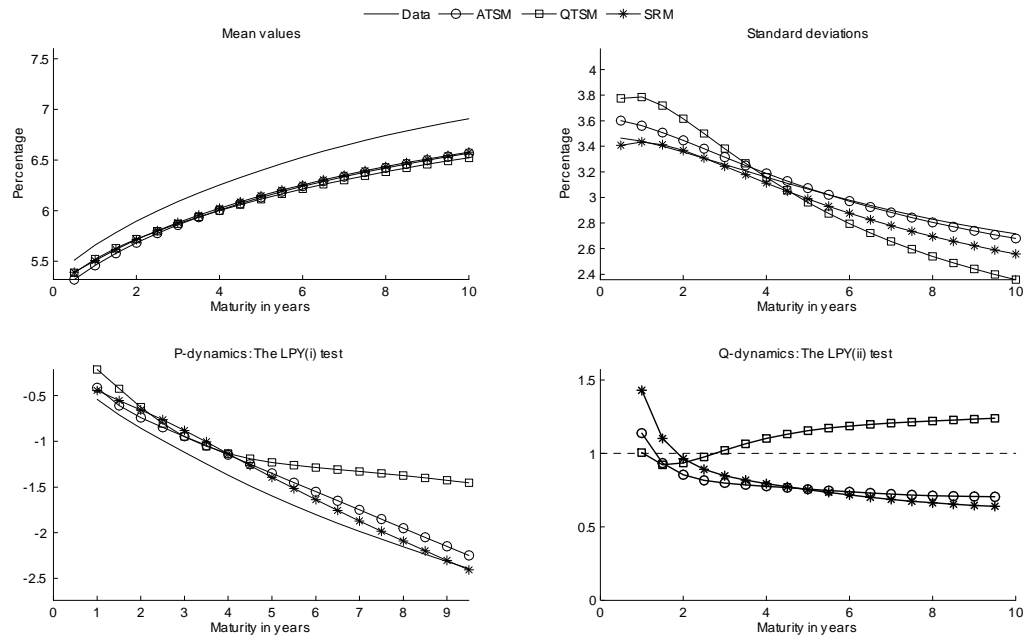
Figure 9: Four-factor models: Unconditional moments in sample from 1990-2013
All model-implied moments are computed from a simulated time series of 100,000 observations.
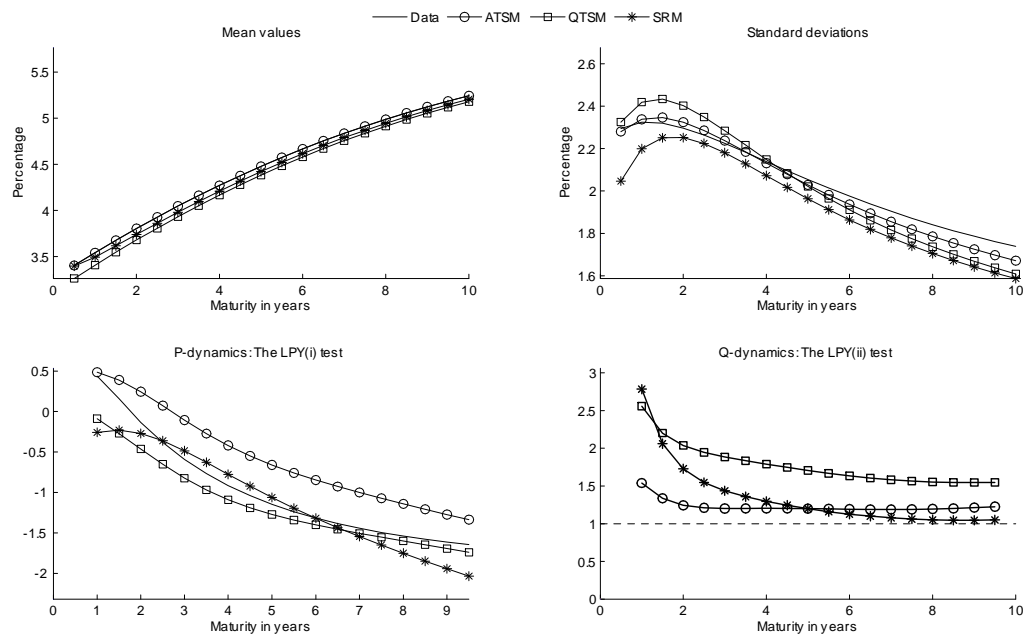
Figure 10: The 10-year term premium: Sample from 1990-2013

For a given number of pricing factors, charts in the first column report the 10-year term premium in the ATSM when bias-adjusting and and not bias-adjusting $\hat{\boldsymbol{\theta}}_2$. For a given number of pricing factors, charts in columns two and three show the differences in the 10-year term premium between the ATSM and the QTSM and between the ATSM and the SRM, respectively. The bias-adjustment of $\hat{\boldsymbol{\theta}}_2$ is used for all models in columns two and three.
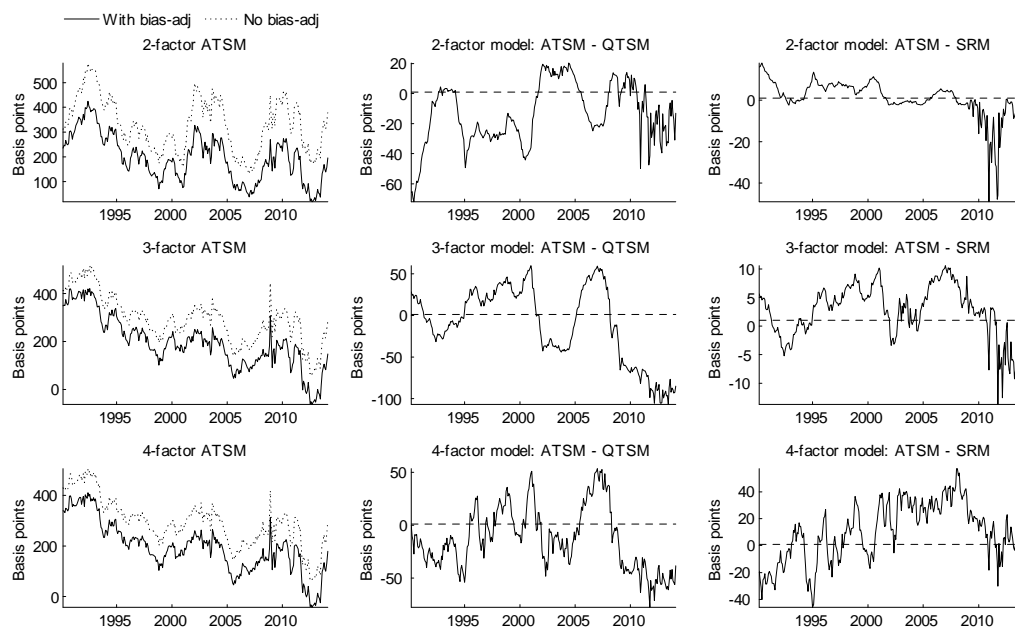
Figure 11: Conditional volatilities of bond yields: Sample from 1961-2013

The model-implied volatilities refer to the one-step-ahead conditional volatilities in the QTSM and the SRM, respectively, where the volatility in time period t is computed from a local linearization of bond yields at $\hat{\mathbf{x}}_{t-1}$.
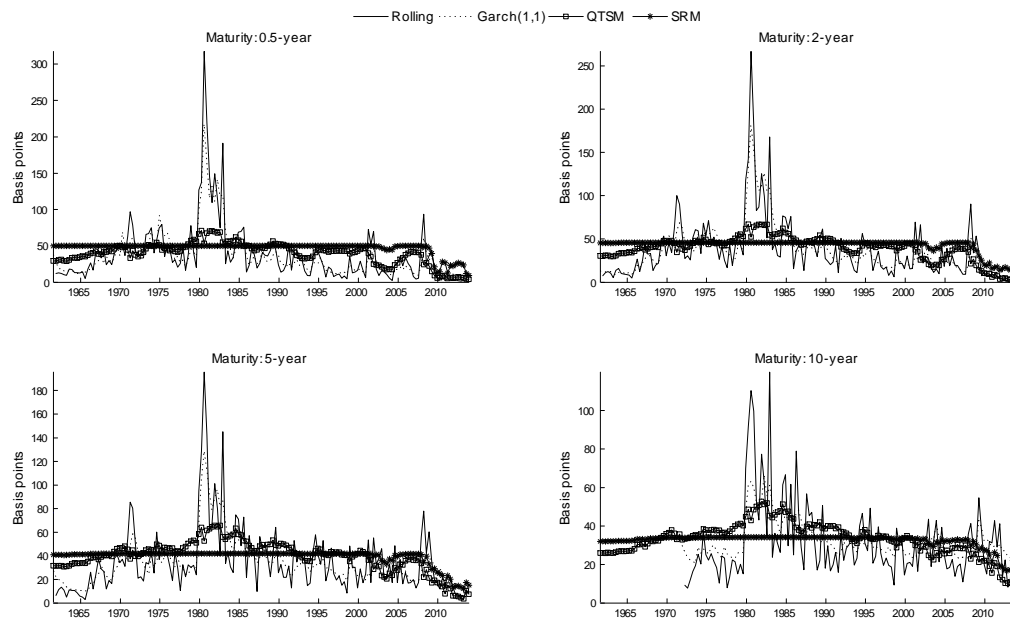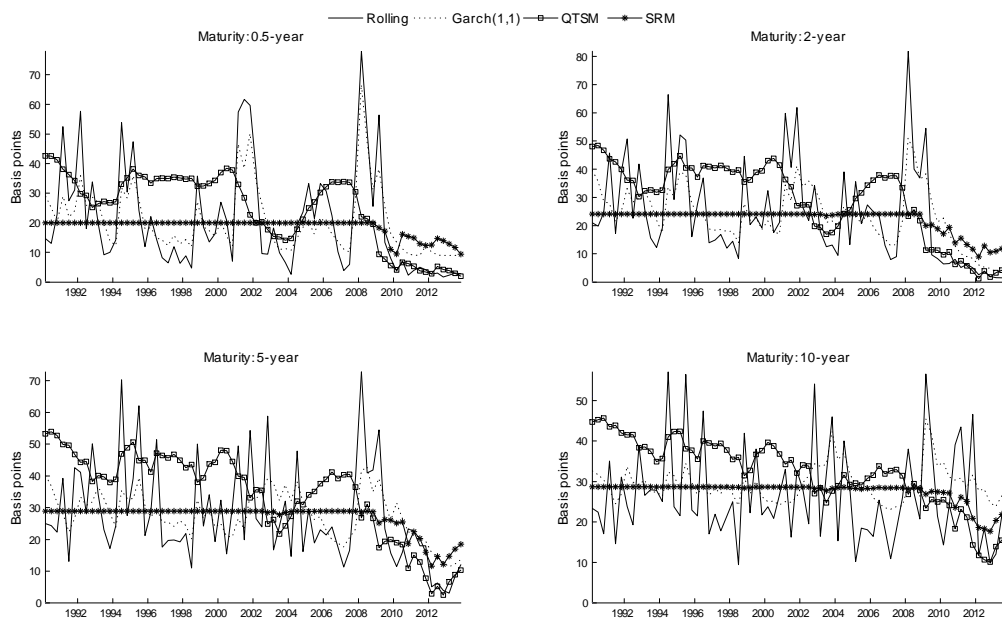

Figure 12: Conditional volatilities of bond yields: Sample from 1990-2013

The model-implied volatilities refer to the one-step-ahead conditional volatilities in the QTSM and the SRM, respectively, where the volatility in time period t is computed from a local linearization of bond yields at $\hat{\mathbf{x}}_{t-1}$.

Figure 13: Forecasting results by maturity: model estimation starting in 1961

The RMSPEs for out-of-sample forecasts from January 2005 to December 2013, generated from models estimated recursively from 1961 to the month prior to the forecast. The forecasted bond yields in the SRM are computed by Monte Carlo integration using 10,000 draws.
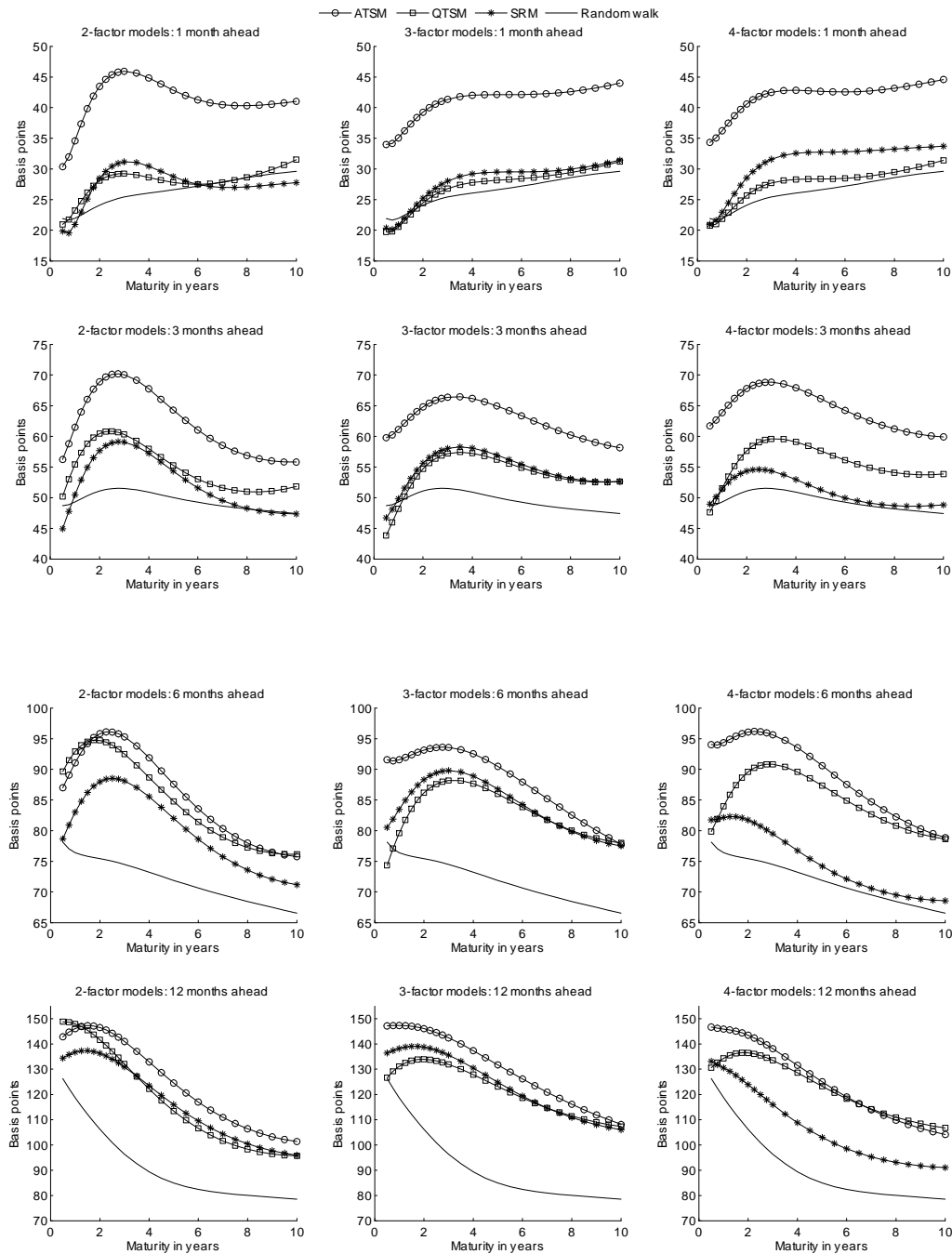
BANK OF ENGLAND
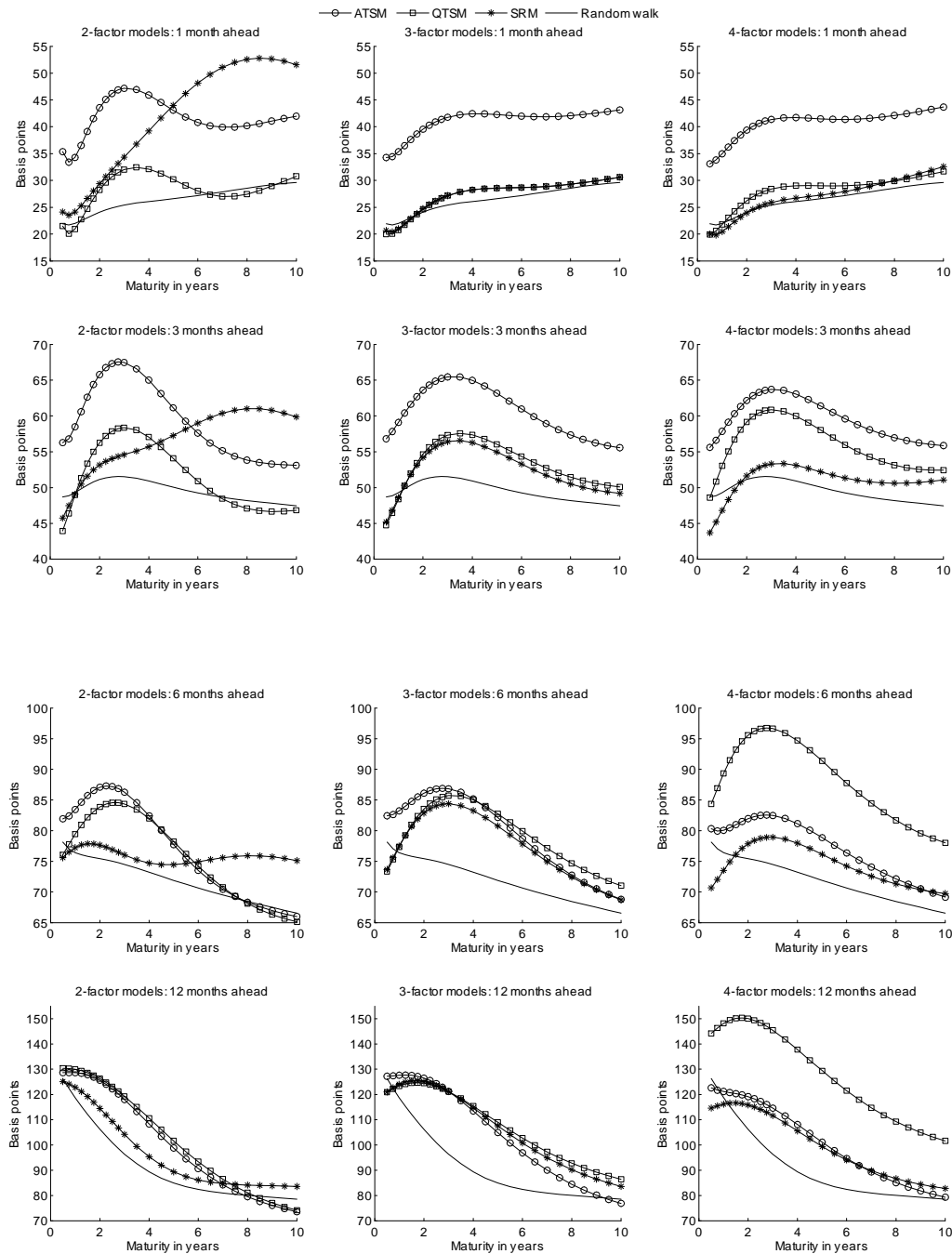
Figure 14: Forecasting results by maturity: model estimation starting in 1990

The RMSPEs for out-of-sample forecasts from January 2005 to December 2013, generated from models estimated recursively from 1990 to the month prior to the forecast. The forecasted bond yields in the SRM are computed by Monte Carlo integration using 10,000 draws.

Figure 15: Forecasting illustration for the 0.5-year bond yield

The forecasts are generated from models estimated recursively from 1961 to the month prior to the forecast. The forecasted bond yields in the SRMs are computed by Monte Carlo integration using 10,000 draws.
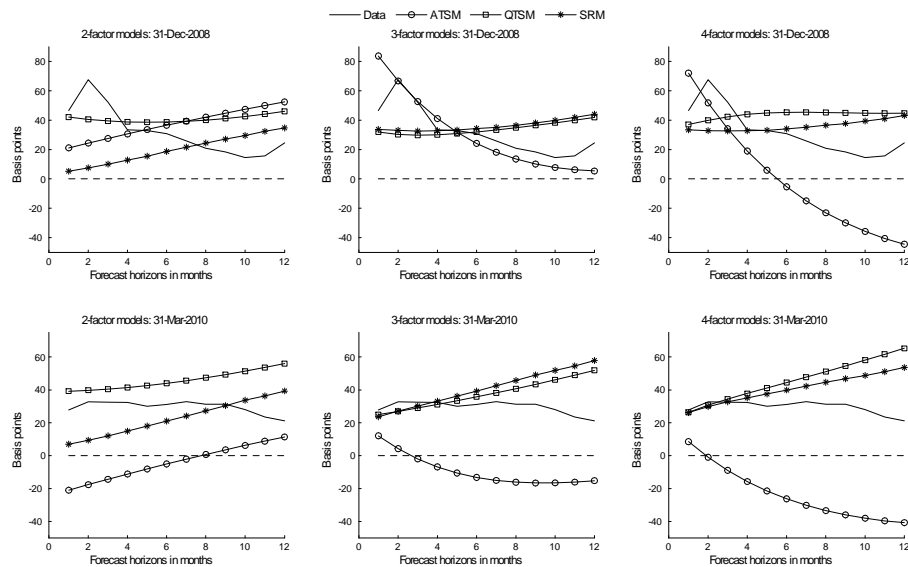
Figure 16: Three-factor hybrid model: Accuracy of approximated bond yields

These charts report the pricing errors for the three-factor hybrid model when evaluating bond yields at $\{\hat{\mathbf{x}}_t\}_{t=1}^{T}$ for the estimated parameters in the long and short samples using 500 draws in the MC method. The true solution is approximated using a MC method with 100,000 draws and anti-thetic sampling.
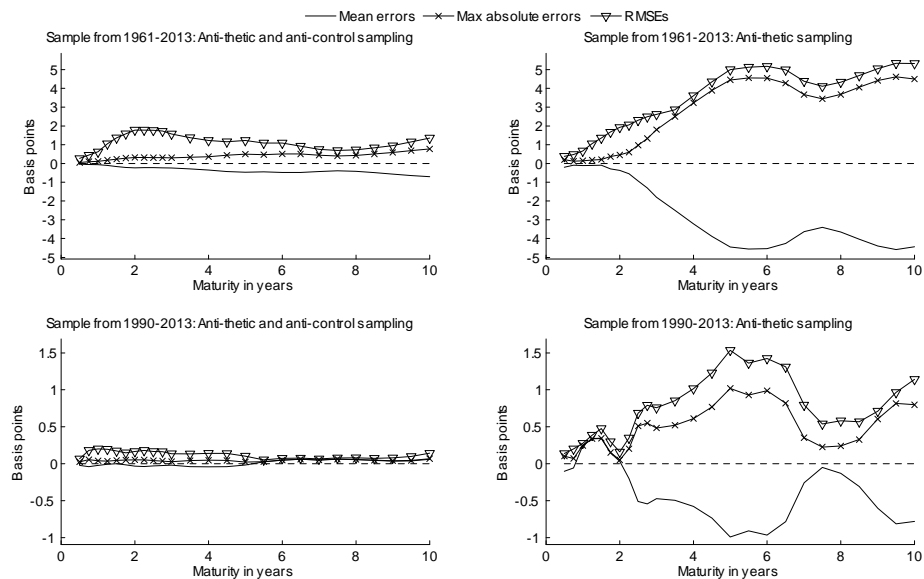
Figure 17: The hybrid model: Unconditional moments

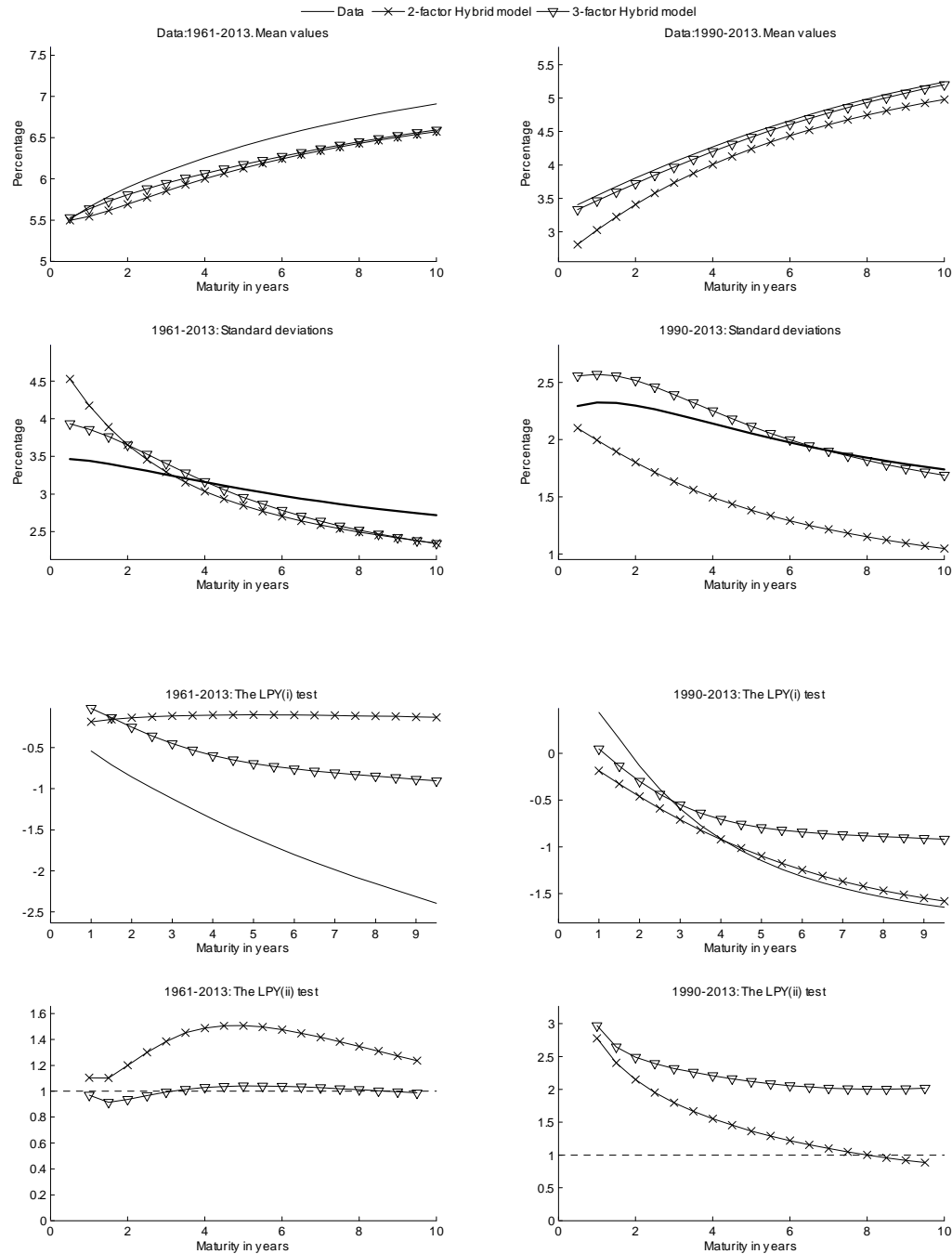All moments per model are obtained from a simulated time series of 100,000 observations.

Figure 18: Forecasting results for hybrid model: model estimation starting in 1961
The RMSPEs for out-of-sample forecasts from January 2005 to December 2013, generated from models estimated recursively from 1961 to the month prior to the forecast. Forecasted in the SRMs and the hybrid models are computed using MC integration with 10,000 draws.
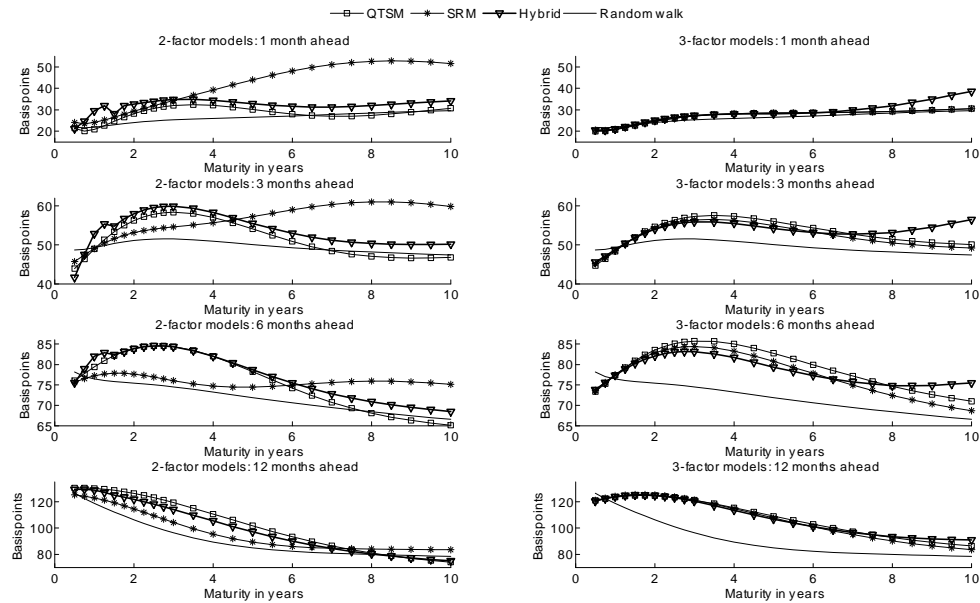

Figure 19: Forecasting results for hybrid model: model estimation starting in 1990
The RMSPEs for out-of-sample forecasts from January 2005 to December 2013, generated from models estimated recursively from 1990 to the month prior to the forecast. Forecasted in the SRMs and the hybrid models are computed using MC integration with 10,000 draws.