

Dynamic term structure models: The best way to enforce the zero lower bound*

Martin M. Andreasen[†]

Andrew Meldrum[‡]

Aarhus University and CREATES

Bank of England

August 28, 2014

Preliminary version

Abstract

This paper studies whether term structure models for US nominal bond yields should enforce the zero lower bound by a quadratic policy rate or a shadow rate. We address the question by estimating quadratic term structure models (QTSMs) and shadow rate models using the sequential regression approach. Our findings suggest that the two models largely provide the same in-sample fit, but loadings from ordinary and risk-adjusted Campbell-Shiller regressions are generally best matched by the shadow rate model. We also find that the shadow rate model performs better than the QTSM when forecasting bond yields out of sample.

Keywords: Bias adjustment, Forecasting study, Quadratic term structure models, Shadow rate models, Sequential regression approach.

JEL: C10, C50, G12.

*We thank Scott Joslin, Peter Hördahl, Hans Dewachter, Oreste Tristani, Tom Engsted, and Thomas Pedersen for useful comments and discussions. Remarks and suggestions from seminar participants at the conference on “Term structure modelling at the zero lower bound” at the Federal Reserve Bank of San Francisco, October 2013, and the workshop on “Staying at zero with affine processes” at the ECB, July 2014, are also much appreciated. M. Andreasen greatly acknowledges financial support from the Danish Center for Scientific Computing (DCSC). He also appreciates financial support to CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. We also note that the views expressed in the present paper are those of the authors and do not necessarily reflect those of the Bank of England or members of the Monetary Policy Committee or Financial Policy Committee.

[†]Department of Economics and Business, Aarhus university, Fuglesangs Allé 4, 8210 Aarhus V, Denmark. Email: mandreasen@econ.au.dk. Telephone number: +45 8716 5982.

[‡]Bank of England, Threadneedle Street, London, EC2R 8AH, UK. Email: andrew.meldrum@bankofengland.co.uk. Telephone number: +44 (0)20 7601 5607.

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 a well-known shortcoming of Gaussian affine term structure models (ATSMs) as they are unable to ensure positive 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 well-known class of quadratic term structure models (QTSMs) studied in Ahn, Dittmar & Gallant (2002), Leippold & Wu (2002), 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 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 on US bond yields.¹ That is, should dynamic term structure models (DTSMs) for US bond yields enforce the ZLB by an appropriately specified quadratic policy rate or by relying on a shadow rate?

The aim of this paper is to address this question by comparing the in- and out-of-sample performance of QTSMs and shadow rate models. Our main focus is devoted to models with three latent pricing factors for comparability with much of the existing literature, but models with two and four factors are also studied when relevant. 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 for bond yields. However, none of the ATSMs satisfying the LPY tests in Dai & Singleton (2002) enforce the ZLB, and it is therefore unclear if DTSMs jointly can enforce the ZLB and match key moments of the physical and risk-neutral distributions for bond yields. To address this question special attention is devoted to the LPY tests when comparing the performance of QTSMs and shadow rate models on US bond yields.

The non-linear DTSMs considered are estimated by the sequential regression (SR) approach of Andreasen & Christensen (2014), where latent pricing factors are obtained by cross-section regressions and model parameters are determined by a three-step moment-based estimation procedure. The

¹Kim & Singleton (2012) explore a similar question on Japanese bond yields in models with two latent pricing factors.

SR approach gives consistent and asymptotically normal estimates and these properties hold under weaker restrictions than typically imposed in likelihood-based inference for DTSMs. For instance, the SR approach allows measurement errors in bond yields to display heteroskedasticity and correlation in both the cross-section and time series dimension. Building on the work of Andreasen & Christensen (2014), we improve the finite sample properties of the SR approach by introducing a bias-adjustment when estimating the physical dynamics of the pricing factors. Apart from these robust econometric properties, the SR approach is also well-suited from a finance perspective, as the QTSMs and shadow rate models considered only differ in their risk-neutral distributions which may be estimated independently of their physical distributions in the SR approach. Hence, the ability of these models to match in-sample bond yields reported below hold for *any* considered functional form of the market price of risk.

The performance of DTSMs on US bond yields is typically studied using either a long sample starting in the 1960s or a shorter sample from around 1990 and onwards. We find it informative to include both samples because bond yields in the long sample attain very high values and display frequent changes in conditional volatility, whereas bond yields in the short sample are lower and/or have fairly stable conditional volatility. Hence, if one believes that the US in the future is likely to experience very high bond yields and frequent changes in the conditional volatilities, the results from our long sample is 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 are likely to be more useful for how to specify the ZLB in DTSMs.

We highlight the following results from our analysis on monthly US bond yields ending in December 2013. First, accounting for the ZLB by either a QTSM and shadow rate model gives largely the same in-sample fit of US bond yields, with both models clearly outperforming the Gaussian ATSM. Second, the three and four factor QTSMs generally struggle to match loadings from the Campbell-Shiller regressions, whereas these moments are matched by the shadow rate models. The shadow rate models are also more successful at matching the loadings from risk-adjusted Campbell-Shiller regressions than the QTSMs, although the latter performs well in the full sample. Third, the fall in conditional volatility of most bond yields when reaching the ZLB is nicely matched by the QTSM and the shadow rate model, although both models struggle to reproduce the observed level of volatility before reaching the ZLB. Fourth, in an out-of-sample forecasting study from January 2005 to December 2013, we find

that the shadow rate model generally performs better than the QTSM, and that models accounting for the ZLB do better than the Gaussian ATSM. The shadow rate model is also found to be more robust and less subject to overfitting than the QTSM, as the forecasts in the shadow rate model generally improves when moving from three to four pricing factors whereas the opposite holds for the QTSM. Importantly, the QTSM and the shadow rate model ensure sensible forecasts as predicted bond yields stay non-negative whereas they easily turn negative in the Gaussian ATSM.

Overall, our findings suggests that the best way to enforce the ZLB for US bond yields is to adopt a shadow rate specification as oppose to considering a quadratic specification for the policy rate.

The rest of the paper is organized as follows. Section 2 presents the DTSMs considered, and we describe how these models may be estimated by the SR approach in Section 3. In-sample results are reported in Section 4 and the out-of-sample performance are presented in Section 5. Concluding comments are provided in Section 6.

2 Dynamic term structure models

We start by describing the Gaussian ATSM in Section 2.1 which serves as our benchmark. The next two subsections present the QTSM and the shadow rate model, respectively. The pricing factors are here assumed to be Gaussian under both the risk-neutral and physical measure as in the benchmark model. That is, we consider 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 with correlated factors as in the $\mathbb{A}_m(m)$ models by Dai & Singleton (2000), although such models also account for the ZLB. We omit these models from our analysis because they are unable to reproduce key moments of term premia in the US, whereas these moments are nicely matched by the Gaussian ATSM as shown in Dai & Singleton (2002).²

²In addition, Kim & Singleton (2012) find that the in-sample fit of the QTSM and the shadow rate model clearly outperforms the $\mathbb{A}_m(m)$ model on Japanese bond yields.

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.

$$r_t = \alpha + \boldsymbol{\beta}' \mathbf{x}_t, \quad (1)$$

where α is a scalar and $\boldsymbol{\beta}$ is an $n_x \times 1$ vector. This specification is typically motivated by referring to a Taylor rule, where the policy rate is determined by a desire to stabilize the inflation and output gap (see for instance Clarida, Gali & Gertler (2000)). 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.

$$\mathbf{x}_{t+1} = \boldsymbol{\Phi} \boldsymbol{\mu} + (\mathbf{I} - \boldsymbol{\Phi}) \mathbf{x}_t + \boldsymbol{\Sigma} \boldsymbol{\varepsilon}_{t+1}^{\mathbb{Q}}, \quad (2)$$

where $\boldsymbol{\varepsilon}_{t+1}^{\mathbb{Q}} \sim \mathcal{NID}(\mathbf{0}, \mathbf{I})$. The mean level of the pricing factors is controlled by $\boldsymbol{\mu}$ of dimension $n_x \times 1$, while the persistence and the conditional volatility of the factors are determined by the $n_x \times n_x$ matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$, respectively. In the absence of arbitrage, the price at time t of an k -period zero-coupon bond is $P_{t,k} = E_t^{\mathbb{Q}}[\exp\{-r_t\} P_{t+1,k-1}]$. Given the assumptions in (1) and (2), bond prices are exponentially affine in the factors, i.e.

$$P_{t,k} = \exp\{A_k + \mathbf{B}'_k \mathbf{x}_t\} \quad (3)$$

for $k = 1, 2, \dots, K$, where the recursive formulae for A_k and \mathbf{B}_k are easily derived.

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 $\boldsymbol{\varepsilon}_{t+1}^{\mathbb{Q}} = \boldsymbol{\varepsilon}_{t+1}^{\mathbb{P}} + \mathbf{f}(\mathbf{x}_t)$, and the factor dynamics under \mathbb{P} are therefore

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

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

$$\mathbf{x}_{t+1} = \boldsymbol{\Phi} \boldsymbol{\mu} + \mathbf{f}_0 + (\mathbf{I} - \boldsymbol{\Phi} + \mathbf{f}_1) \mathbf{x}_t + \boldsymbol{\Sigma} \boldsymbol{\varepsilon}_{t+1}^{\mathbb{P}}. \quad (4)$$

To obtain stationary bond yields with finite first and second unconditional moments, we require the process for \mathbf{x}_t to be stationary, i.e. that all eigenvalues of $\mathbf{I} - \Phi + \mathbf{f}_1$ 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) $\beta = \mathbf{1}$, ii) $\mu = \mathbf{0}$, iii) Φ to be diagonal, and iv) Σ to be triangular.³ This identification scheme constrains the \mathbb{Q} dynamics for the pricing factors whereas the \mathbb{P} dynamics are unrestricted. The latter is convenient when the model is estimated by the SR approach, as explained in Section 3.

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.

$$r_t = \alpha + \beta' \mathbf{x}_t + \mathbf{x}_t' \Psi \mathbf{x}_t, \quad (5)$$

where Ψ is a symmetric $n_x \times n_x$ matrix. This specification may also be motivated from a Taylor rule if it displays time-varying parameters as considered in Ang, Boivin, Dong & Loo-Kung (2011). Following Kim & Singleton (2012), we adopt the decomposition $\Psi = \mathbf{A} \mathbf{D} \mathbf{A}'$, 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) $\alpha \geq \frac{1}{4} \beta' \Psi^{-1} \beta$ and ii) Ψ 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 (5), it is convenient to adopt the same specification for the pricing factors as in (2), because it gives a closed-form solution for zero-coupon bonds, i.e.

$$P_{t,k} = \exp \left\{ \tilde{A}_k + \tilde{\mathbf{B}}_k' \mathbf{x}_t + \mathbf{x}_t' \tilde{\mathbf{C}}_k \mathbf{x}_t \right\} \quad (6)$$

for $k = 1, 2, \dots, K$, with the recursive formulae for \tilde{A}_k , $\tilde{\mathbf{B}}_k$, and $\tilde{\mathbf{C}}_k$ derived in Realdon (2006). Hence, the quadratic terms in (5) imply that all bond yields $y_{t,k} \equiv -\frac{1}{k} \log P_{t,k}$ are quadratic in the pricing factors and bond yields therefore display heteroskedasticity.

³There 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.

For comparability with the Gaussian ATSM, we maintain the affine specification for the market price of risk, meaning that the \mathbb{P} dynamics for the pricing factors in the QTSM are given by (4). As in the Gaussian ATSM, not all parameters are identified in the QTSM with latent pricing factors. We therefore follow Ahn et al. (2002) and impose the restrictions: i) Ψ is symmetric with diagonal elements equal to one, ii) $\mu \geq \mathbf{0}$, iii) $\beta = \mathbf{0}$, iv) Φ is diagonal, and v) Σ is triangular. This normalization scheme implies an unrestricted \mathbb{P} dynamics for the pricing factors and that the ZLB may be enforced by letting $\alpha = 0$.

2.3 The shadow rate model

The ZLB may alternatively be enforced in DTSMs by introducing a shadow interest rate $s(\mathbf{x}_t)$ as suggested by Black (1995).⁴ This shadow rate is unconstrained by the ZLB and may therefore attain negative values. Absent any transaction and storing costs for money, Black (1995) observes that the nominal interest rate cannot be negative because investors may always decide to hold cash. In other words, the nominal interest rate has an option element. This argument leads to the following specification

$$r_t = \max(0, s(\mathbf{x}_t)), \quad (7)$$

where the policy rate r_t is the non-negative part of the shadow rate. 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 $s(\mathbf{x}_t)$ and the considered factor dynamics.

For comparability with the benchmark ATSM, we let the shadow rate be affine in the pricing factors, i.e.

$$s(\mathbf{x}_t) = \alpha + \beta' \mathbf{x}_t, \quad (8)$$

but other specifications may also be considered. For the same reason, we also restrict focus to affine processes for the pricing factors under the \mathbb{Q} and \mathbb{P} measures as in the Gaussian ATSM, but other specifications could be considered. That is, we impose (2) and (4) in the shadow rate model. Finally, the identification conditions for the shadow rate model are identical to those for the Gaussian ATSM in Section 2.1.

Shadow rate models do not attain closed-form expressions for bond prices, except for one-factor

⁴The idea of considering a shadow rate is also briefly mentioned in Rogers (1995).

models with a Gaussian or a square-root process driving $s(\mathbf{x}_t)$ as shown by Gorovoi & Linetsky (2004). Given that one-factor models typically are considered too stylized, numerical approximations are therefore needed when studying multi-factor shadow rate models. We apply the second-order approximation advocated by Priebsch (2013), which delivers a fast and highly accurate approximation to bond yields when $s(\mathbf{x}_t)$ is Gaussian under the \mathbb{Q} measure.⁵

3 The estimation procedure

One way to estimate non-linear DTSMs with latent pricing factors as implied by QTSMs and shadow rate models is to approximate the unknown likelihood function by sequential Monte Carlo methods (see Doucet, de Freitas & Gordon (2001) and Rossi (2004)). This procedure is very time consuming for multi-factor DTSMs and therefore rarely attempted. A computational more feasible alternative is to use a non-linear extension of the Kalman filter and a quasi-maximum likelihood (QML) approach, but its asymptotic properties are generally unknown.⁶ We overcome these difficulties by using the sequential SR approach, which has known asymptotic properties and is even faster to implement than the QML approach (see Andreasen & Christensen (2014)). We also emphasize that the asymptotic properties of the SR approach hold under weaker restrictions than typically considered for likelihood-based inference. This robust nature of the SR approach is particularly attractive in our context because all considered DTSMs differ only in their \mathbb{Q} dynamics which may be estimated independently of the \mathbb{P} dynamics in the SR approach. Hence, the models' abilities to match in-sample bond yields reported below apply for *any* functional form of the market price of risk $\mathbf{f}(\mathbf{x}_t)$.

We next present the SR approach and describe how the latent pricing factors and model parameters are estimated in the models considered.

⁵Other 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 (2013)), and v) ignoring Jensen's inequality term to solve a Gaussian model by a truncated normal distribution (Ichiue & Ueno (2013)).

⁶Recent applications of the procedure in DTSMs enforcing the ZLB may be found in Ichiue & Ueno (2007), Kim & Singleton (2012), Bauer & Rudebusch (2014), Christensen & Rudebusch (2013), and Ichiue & Ueno (2013). For certain ATSMs without the ZLB restriction, the findings by Duan & Simonato (1999) and de Jong (2000) suggest that the bias in a QML approach based on the extended Kalman filter may be small. We refer to Andreasen (2013) for a discussion of the asymptotic properties related to a QML approach when estimating non-linear state space models.

3.1 The SR approach

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

$$y_{t,k} = g_k(\mathbf{x}_t; \boldsymbol{\theta}_1) + v_{t,k}, \quad (9)$$

where the subscript k index the maturity of the bond yields. The functional relationship between the pricing factors and bond yields is parameterized by $\boldsymbol{\theta}_1 \equiv \begin{bmatrix} \boldsymbol{\theta}'_{11} & \boldsymbol{\theta}'_{12} \end{bmatrix}'$ containing the risk-neutral parameters. Elements in $\boldsymbol{\theta}_{11}$ may only be determined from the measurement equations in (9), whereas $\boldsymbol{\theta}_{12}$ may be obtained from (9) and the factor dynamics under the \mathbb{P} measure. For the Gaussian ATSM, the g -function is linear in the pricing factors, i.e. $g_k^{ATSM}(\mathbf{x}_t; \boldsymbol{\theta}_1^{ATSM}) \equiv -\frac{1}{k}(A_k + \mathbf{B}'_k \mathbf{x}_t)$, and we have $\boldsymbol{\theta}_{11}^{ATSM} \equiv \begin{bmatrix} \alpha & \text{diag}(\boldsymbol{\Phi})' \end{bmatrix}'$ with $\boldsymbol{\theta}_{12}^{ATSM} \equiv \begin{bmatrix} \text{vech}(\boldsymbol{\Sigma})' \end{bmatrix}'$. The QTSM induces a slightly more complicated expression for bond yields because $g_k^{QTSM}(\mathbf{x}_t; \boldsymbol{\theta}_1^{QTSM}) \equiv -\frac{1}{k}(\tilde{A}_k + \tilde{\mathbf{B}}'_k \mathbf{x}_t + \mathbf{x}'_t \tilde{\mathbf{C}}_k \mathbf{x}_t)$, and for this model $\boldsymbol{\theta}_{11}^{QTSM} \equiv \begin{bmatrix} (\boldsymbol{\theta}_1^{ATSM})' & \boldsymbol{\mu}' & \text{vech}(\boldsymbol{\Psi})' \end{bmatrix}'$ with $\boldsymbol{\theta}_{12}^{QTSM} = \boldsymbol{\theta}_{12}^{ATSM}$. In the shadow rate model, $g_k^{SH}(\mathbf{x}_t; \boldsymbol{\theta}_1^{SH})$ is an unknown non-linear mapping from the pricing factors to bond yields with $\boldsymbol{\theta}_1^{SH} = \boldsymbol{\theta}_1^{ATSM}$. It is important to stress that the SR approach does not impose any distributional assumptions on the measurement errors $v_{t,k}$, which furthermore may display heteroskedasticity and correlation in both the cross-section and the time series dimension.

The SR approach allows the pricing factors under the \mathbb{P} measure to evolve according to a general Markov process of the form

$$\mathbf{x}_{t+1} = \mathbf{h}(\mathbf{x}_t, \boldsymbol{\epsilon}_{t+1}^{\mathbb{P}}; \boldsymbol{\theta}_{11}, \boldsymbol{\theta}_2). \quad (10)$$

The \mathbf{h} -function may depend on $\boldsymbol{\theta}_{11}$ and $\boldsymbol{\theta}_2 \equiv \begin{bmatrix} \boldsymbol{\theta}'_{22} & \boldsymbol{\theta}'_{12} \end{bmatrix}'$, where $\boldsymbol{\theta}_{22}$ must be determined from the factor dynamics in (10). All the DTSMs considered have a linear and unrestricted transition function which we represent by

$$\mathbf{x}_{t+1} = \mathbf{h}_0 + \mathbf{h}_x \mathbf{x}_t + \boldsymbol{\Sigma} \boldsymbol{\epsilon}_{t+1}^{\mathbb{P}}, \quad (11)$$

where $\mathbf{h}_0 \equiv \boldsymbol{\Phi} \boldsymbol{\mu} + \mathbf{f}_0$, $\mathbf{h}_x \equiv \mathbf{I} - \boldsymbol{\Phi} + \mathbf{f}_1$, and $\boldsymbol{\epsilon}_{t+1}^{\mathbb{P}} \sim \mathcal{NID}(\mathbf{0}, \mathbf{I})$. Hence, given the parametrization of the \mathbf{h} -function in (11), we have $\boldsymbol{\theta}_{22} \equiv \begin{bmatrix} \mathbf{h}'_0 & \text{vec}(\mathbf{h}_x)' \end{bmatrix}'$ for all the models considered.

The subsequent sections describe how the latent pricing factors $\{\mathbf{x}_t\}_{t=1}^T$ and the model parameters

$(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ are estimated in the SR approach using a three-step procedure.

3.1.1 The SR approach: Step 1

The latent pricing factors are estimated by running the cross-section regressions

$$\hat{\mathbf{x}}_t(\boldsymbol{\theta}_1) = \arg \min_{\mathbf{x}_t \in \mathcal{X}_t} Q_t = \frac{1}{2n_{y,t}} \sum_{k=1}^{n_{y,t}} (y_{t,k} - g_k(\mathbf{x}_t; \boldsymbol{\theta}_1))^2 \quad (12)$$

for $t = 1, 2, \dots, T$, where $n_{y,t}$ refers to the number of bond yields in time period t . The estimated factors are denoted $\{\hat{\mathbf{x}}_{2,t}(\boldsymbol{\theta}_1)\}_{t=1}^T$ because they are computed for a given $\boldsymbol{\theta}_1$. These regressions have a closed-form solution for the Gaussian ATSM with g_k^{ATSM} being linear in the pricing factors. For the QTSM and the shadow rate model, the regressions in (12) are non-linear and solved using the Levenberg-Marquardt method with the pricing factors from the previous time period $\hat{\mathbf{x}}_{2,t-1}(\boldsymbol{\theta}_1)$ serving as ideal starting values for $t = 2, 3, \dots, T$.⁷

The model parameters $\boldsymbol{\theta}_1$ are obtained by pooling all squared residuals from (12) and minimizing their sum with respect to $\boldsymbol{\theta}_1$, i.e.

$$\hat{\boldsymbol{\theta}}_1^{step1} = \arg \min_{\boldsymbol{\theta}_1 \in \Theta_1} Q_{1:T}^{step1} = \frac{1}{2N} \sum_{t=1}^T \sum_{k=1}^{n_{y,t}} (y_{t,k} - g_k(\hat{\mathbf{x}}_t(\boldsymbol{\theta}_1); \boldsymbol{\theta}_1))^2, \quad (13)$$

where $N \equiv \sum_{t=1}^T n_{y,t}$. Given standard regularity conditions, Andreasen & Christensen (2014) show consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_1^{step1}$, i.e.

$$\sqrt{N} \left(\hat{\boldsymbol{\theta}}_1^{step1} - \boldsymbol{\theta}_1^o \right) \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \left(\mathbf{A}_o^{\boldsymbol{\theta}_1} \right)^{-1} \mathbf{B}_o^{\boldsymbol{\theta}_1} \left(\mathbf{A}_o^{\boldsymbol{\theta}_1} \right)^{-1} \right), \quad (14)$$

where the superscript "o" denotes the true value. These asymptotic properties are derived by letting the number of bond yields in each time period $n_{y,t}$ tend to infinity, implying $N \rightarrow \infty$. The expected value of the average Hessian matrix $\mathbf{A}_o^{\boldsymbol{\theta}_1}$ may be estimated consistently by

$$\hat{\mathbf{A}}^{\boldsymbol{\theta}_1} = \frac{1}{N} \sum_{t=1}^T \sum_{k=1}^{n_{y,t}} \left(\hat{\boldsymbol{\Psi}}_{t,k}^{\boldsymbol{\theta}_1} \right) \left(\hat{\boldsymbol{\Psi}}_{t,k}^{\boldsymbol{\theta}_1} \right)', \quad (15)$$

⁷The main input for Levenberg-Marquardt optimizer is the jacobian $\partial \mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta}_1) / \partial \mathbf{x}_t'$ which is available in closed form for the QTSM. For the shadow rate model, the jacobian is obtained by numerical differentiation using a first-order approximation as in Ichiue & Ueno (2013) but otherwise the second-order approximation by Priebisch (2013) is applied in the optimizer. Using the second-order approximation to also compute the jacobian in the optimizer gives identical results but is somewhat slower than the adopted our procedure.

where

$$\Psi_{t,k}^{\theta_1}(\theta_1) \equiv \frac{\partial \hat{\mathbf{x}}'_{2,t}(\theta_1)}{\partial \theta_1} \frac{\partial g_k(\hat{\mathbf{x}}_{2,t}(\theta_1); \theta_1)}{\partial \mathbf{x}_{2,t}(\theta_1)} + \frac{\partial g_k(\hat{\mathbf{x}}_{2,t}(\theta_1); \theta_1)}{\partial \theta_1} \quad (16)$$

and $\hat{\Psi}_{t,k}^{\theta_1} \equiv \Psi_{t,k}^{\theta_1}(\hat{\theta}_1^{step1})$. The average of the score function $\mathbf{B}_o^{\theta_1}$ is estimated using an extension of the Newey-West estimator that is robust to heteroskedasticity, cross-section correlation, and autocorrelation in $v_{t,k}$. The most general specification considered in Andreasen & Christensen (2014) is

$$\begin{aligned} \hat{\mathbf{B}}^{\theta_1} &= \frac{1}{N} \sum_{t=1}^T \sum_{k=1}^{n_{y,t}} \sum_{k_T=-w_T}^{w_T} \sum_{k_D=-w_D}^{w_D} \left(1 - \frac{|k_T|}{1+w_T}\right) \left(1 - \frac{|k_D|}{1+w_D}\right) \\ &\quad \times \left(\hat{\Psi}_{t,k}^{\theta_1}\right) \left(\hat{\Psi}_{t+k_T, j+k_D}^{\theta_1}\right)' \hat{v}_{t,k} \hat{v}_{t+k_T, k+k_D}. \end{aligned} \quad (17)$$

Here, w_D is the bandwidth for bond yields in the cross-section dimension when ordered by duration (i.e. maturity) and w_T is the corresponding bandwidth for the time series dimension.

3.1.2 The SR approach: Step 2

We estimate θ_2 in (11) using $\{\hat{\mathbf{x}}_t\}_{t=1}^T$ and moment conditions accounting for the 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. As in Andreasen & Christensen (2014), we modify the standard moment conditions for VAR models to account for uncertainty in $\{\hat{\mathbf{x}}_t\}_{t=1}^T$ and consider

$$\mathbf{q}_T(\theta_2) \equiv \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbf{q}_t(\theta_2) = \mathbf{0}, \quad (18)$$

where

$$\mathbf{q}_t(\theta_2) \equiv \begin{bmatrix} \hat{\mathbf{w}}_{t+1} \\ \text{vec}(\hat{\mathbf{w}}_{t+1} \hat{\mathbf{x}}'_t - \text{Cov}(\mathbf{u}_{t+1}, \mathbf{u}_t) + \mathbf{h}_x \text{Var}(\mathbf{u}_t)) \\ \text{vech} \left(\begin{array}{l} \hat{\mathbf{w}}_{t+1} \hat{\mathbf{w}}'_{t+1} - \text{Var}(\hat{\mathbf{w}}_{t+1}) - \text{Var}(\mathbf{u}_t) - \mathbf{h}_x \text{Var}(\mathbf{u}_t) \mathbf{h}'_x \\ + \text{Cov}(\mathbf{u}_{t+1}, \mathbf{u}_t) \mathbf{h}'_x + \mathbf{h}_x \text{Cov}(\mathbf{u}_t, \mathbf{u}_{t+1}) \end{array} \right) \end{bmatrix}$$

and

$$\hat{\mathbf{w}}_{t+1} \equiv \Sigma \hat{\boldsymbol{\varepsilon}}_{t+1}^{\mathbb{P}} \equiv \hat{\mathbf{x}}_{t+1} - \mathbf{h}_0 - \mathbf{h}_x \hat{\mathbf{x}}_t.$$

Note that $\hat{\varepsilon}_{t+1}^{\mathbb{P}}$ refers to the residuals using the true values of \mathbf{h}_0 and \mathbf{h}_x but the estimated pricing factors $\hat{\mathbf{x}}_t$. Consistent estimators of $Var(\mathbf{u}_t)$, $Cov(\mathbf{u}_{t+1}, \mathbf{u}_t)$, and $Cov(\mathbf{u}_t, \mathbf{u}_{t+1})$ are provided in Andreasen & Christensen (2014) using output from the first estimation step, and $\boldsymbol{\theta}_2$ can therefore be estimated consistently by generalized methods of moments when the number of time periods T tends to infinity. All models considered in the present paper have unrestricted \mathbb{P} dynamics, and the moment conditions in (18) may then be solved in closed form. The solution is obtained by correcting all second moments for estimation uncertainty in $\{\hat{\mathbf{x}}_t\}_{t=1}^T$ and running the regression

$$\begin{aligned} \begin{bmatrix} \hat{\mathbf{h}}_x^{step2} & \hat{\mathbf{h}}_0^{step2} \end{bmatrix} &= \left(\sum_{t=1}^{T-1} \begin{bmatrix} \hat{\mathbf{x}}_{t+1} \hat{\mathbf{x}}_t' - \widehat{Cov}(\mathbf{u}_{t+1}, \mathbf{u}_t) & \hat{\mathbf{x}}_{t+1} \end{bmatrix} \right) \\ &\times \left(\sum_{t=1}^{T-1} \begin{bmatrix} \hat{\mathbf{x}}_t \hat{\mathbf{x}}_t' - \widehat{Var}(\mathbf{u}_t) & \hat{\mathbf{x}}_t \\ & \hat{\mathbf{x}}_t' & 1 \end{bmatrix} \right)^{-1} \end{aligned} \quad (19)$$

and

$$\begin{aligned} \widehat{Var}(\hat{\mathbf{w}}_{t+1})^{step2} &= \frac{1}{T-1-n_x-1} \sum_{t=1}^{T-1} (\hat{\mathbf{w}}_{t+1} (\hat{\mathbf{w}}_{t+1})') \\ &- \frac{1}{T-1} \sum_{t=1}^{T-1} \left(\widehat{Var}(\mathbf{u}_t) + \hat{\mathbf{h}}_x \widehat{Var}(\mathbf{u}_t) \hat{\mathbf{h}}_x' \right) \\ &+ \frac{1}{T-1} \sum_{t=1}^{T-1} \left(\widehat{Cov}(\mathbf{u}_{t+1}, \mathbf{u}_t) \hat{\mathbf{h}}_x' + \hat{\mathbf{h}}_x \widehat{Cov}(\mathbf{u}_t, \mathbf{u}_{t+1}) \right), \end{aligned} \quad (20)$$

with $\hat{\boldsymbol{\Sigma}}^{step2}$ obtained from a Cholesky decomposition of $\widehat{Var}(\hat{\mathbf{w}}_{t+1})^{step2}$. When T tends to infinity, Andreasen & Christensen (2014) show that the asymptotic distribution of $\boldsymbol{\theta}_2$ is

$$\sqrt{T} \left(\boldsymbol{\theta}_2^{step2} - \boldsymbol{\theta}_2^o \right) \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \left(\mathbf{R}_o^{\boldsymbol{\theta}_2} \mathbf{S}_o^{-1} \left(\mathbf{R}_o^{\boldsymbol{\theta}_2} \right)' \right)^{-1} \right) \quad (21)$$

when using the optimal weighting matrix. Here, $\mathbf{R}_o^{\boldsymbol{\theta}_2} \equiv \frac{\partial \mathbf{q}_T(\boldsymbol{\theta}_2^o)'}{\partial \boldsymbol{\theta}_2}$ and $\mathbf{S}_o \equiv \sum_{\nu=-\infty}^{\infty} E[\mathbf{q}_t(\boldsymbol{\theta}_2^o) \mathbf{q}_{t-\nu}(\boldsymbol{\theta}_2^o)']$. We estimate $\mathbf{R}_o^{\boldsymbol{\theta}_2}$ using numerical differentiation and \mathbf{S}_o by the Newey-West estimator.

For persistent processes it is well-known that the standard moment conditions for estimating VAR models, extended in (18) to account for generated regressors, induce a negative bias for $\hat{\mathbf{h}}_x$ in finite samples (see for instance Sherman and Stine (1988), UPDATE REF). Bauer, Rudebusch & Wu (2012) show that this bias may be substantial for Gaussian ATSMs and have sizeable effects on the estimated

term premium. A popular method to reduce this bias is to apply a bootstrap procedure. 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 of $\hat{\mathbf{h}}_{\mathbf{x}}$ in the bootstrap, and the bias-adjusted estimate is then given by $\hat{\mathbf{h}}_{\mathbf{x}}^{adj} = \hat{\mathbf{h}}_{\mathbf{x}} - (\bar{\mathbf{h}}_{\mathbf{x}} - \hat{\mathbf{h}}_{\mathbf{x}})$. Appendix A outlines how the bootstrap procedure for VAR models can be extended to account for generated regressors as implied by the SR approach.

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

$$\hat{\mathbf{h}}_{\mathbf{x}}^{adj,B}(\delta) = \hat{\mathbf{h}}_{\mathbf{x}} - \delta \times (\bar{\mathbf{h}}_{\mathbf{x}} - \hat{\mathbf{h}}_{\mathbf{x}}) \quad (22)$$

until all eigenvalues of $\hat{\mathbf{h}}_{\mathbf{x}}^{adj,B}(\delta_i)$ are inside the unit circle. This is a simple method to induce stationarity and also the one adopted in Bauer et al. (2012). It should be noted, however, that the size of the recursive reduction in δ_i is not derived from any optimality conditions or a data-driven selection criteria. Moreover, the largest eigenvalue of $\hat{\mathbf{h}}_{\mathbf{x}}^{adj,B}(\delta_i)$ may be made arbitrarily close to one by changing the grid for δ_i appropriately.

Although Killian's method to induce stationarity may have minor effects on conditional moments in VAR models, as used for impulse response functions in Kilian (1998) and term premia in Bauer et al. (2012), it has substantial effects on any unconditional moments. To realize this, suppose we select the grid for δ_i such that the length of the largest eigenvalue of $\hat{\mathbf{h}}_{\mathbf{x}}^{adj,B}(\delta_i)$ is arbitrarily close to one. Given this alternative grid, the process for \mathbf{x}_t approaches a non-stationarity VAR model with infinite unconditional second moments. In other words, Killian's method implies that unconditional moments in the VAR model depend on an arbitrary grid for δ_i and are in this way not uniquely determined.

As a supplement to Killian's method, we therefore suggest a data-driven method to determine δ . Our method is based on the observation that the standard estimator of the unconditional variance in $x_{i,t}$, i.e. $\sigma_{i,Data}^2 = 1/(T-1) \sum_{t=1}^T (x_{i,t} - \bar{x}_i)^2$ with $\bar{x}_i = 1/T \sum_{t=1}^T x_{i,t}$, is unbiased when x_t is Gaussian. We therefore suggest to determine δ in (22) by minimizing the distance between $\sigma_{i,Data}^2$ and the variance of $x_{i,t}$ in the VAR model across all variables, i.e. for $i = 1, 2, \dots, n_x$. The latter estimate

is computed for a given value of δ and is therefore denoted $\sigma_{i,VAR}^2(\delta)$. More formally, we let

$$\hat{\delta} = \arg \min_{\delta \in [\delta_{lower}, 1]} \sum_{i=1}^{n_x} \left(\frac{\sigma_{i,VAR}^2(\delta) - \hat{\sigma}_{i,Data}^2}{\hat{\sigma}_{i,Data}^2} \right)^2. \quad (23)$$

Monte Carlo evidence in Table 1 suggest 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 (23). The better performance is related to $\hat{\delta}$, which in the former procedure tends to be larger than the scaling parameter related to $\hat{\mathbf{h}}_x^{adj,B}(\delta)$, implying that more of the bias-adjustment is preserved in the former procedure. For instance, when using the Gaussian ATSM from the full 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 $\hat{\mathbf{h}}_x$, whereas the average of $\hat{\delta}$ falls to 0.6950 when only the bias is down-scaled. Hence, we prefer the adjustment

$$\hat{\mathbf{h}}_x^{adj,*}(\delta) = \delta \times \left(\hat{\mathbf{h}}_x - \left(\bar{\mathbf{h}}_x - \hat{\mathbf{h}}_x \right) \right), \quad (24)$$

and determine δ using (23). As expected, the Monte Carlo study in Table 1 also shows that the data-driven methods to determine δ give smaller bias in the unconditional standard deviations of the pricing factors compared to Killian's method. Another advantage of considering (24) is that it always ensures stationarity of VAR models, contrary to the specification in (22). Our method to induce stationarity is summarized in Appendix B, which also describes how to account for measurement errors in \mathbf{x}_t . Unless stated otherwise, we use the bias-adjustment in (24) throughout the paper.

< Table 1 about here >

3.1.3 The SR approach: Step 3

The elements in Σ appear in θ_1 and θ_2 and are therefore estimated in both the first and second estimation step. These estimates may be combined in an optimal way to reduce any potential efficiency loss from the use of sequential identification as shown by Andreasen & Christensen (2014). We generally find that $\hat{\Sigma}^{step1}$ is estimated very inaccurately compared to $\hat{\Sigma}^{step2}$, meaning that the time series estimate $\hat{\Sigma}^{step2}$ cannot be improved by adding cross-section information from $\hat{\Sigma}^{step1}$.⁸ Hence, the adopted estimate of Σ after the first two steps is given by $\hat{\Sigma}^{step2}$.

⁸Adopting the notation in Andreasen & Christensen (2014), we generally find $\Lambda \approx \mathbf{0}$.

Based on the more accurate estimate of Σ from the second step, it is natural to re-estimate θ_{11} when conditioned on $\hat{\Sigma}^{step2}$. That is

$$\hat{\theta}_{11}^{step3} = \arg \min_{\theta_{11} \in \Theta_{11}} Q_{1:T}^{step3} = \frac{1}{2N} \sum_{t=1}^T \sum_{k=1}^{n_{y,t}} \left(y_{t,k} - g_k \left(\hat{\mathbf{x}}_t \left(\theta_{11}, \hat{\Sigma}^{step2} \right); \theta_{11}, \hat{\Sigma}^{step2} \right) \right)^2. \quad (25)$$

Andreasen & Christensen (2014) show consistency and asymptotic normality of $\hat{\theta}_{11}^{step3}$ with

$$\widehat{Var} \left(\hat{\theta}_{11}^{step3} \right) = \frac{\hat{\mathbf{V}}_{\theta_{11}}^{step3} \left(\hat{\Sigma}^{step2} \right)}{N} + \hat{\mathbf{K}} \widehat{Var} \left(\hat{\Sigma}^{step2} \right) \hat{\mathbf{K}}'. \quad (26)$$

The first term $\hat{\mathbf{V}}_{\theta_{11}}^{step3} \left(\hat{\Sigma}^{step2} \right) / N$ is given by (14) when used on the subset of θ_1 corresponding to θ_{11} . The second term in (26) corrects for estimation uncertainty in $\hat{\Sigma}^{step2}$ with $\mathbf{K} \equiv \partial \hat{\theta}_{11}^{step3} (\Sigma) / \partial \text{vech} (\Sigma)'$. We estimate \mathbf{K} as suggested in Andreasen & Christensen (2014) and refer to their paper for further details.

Given the estimated pricing factors $\left\{ \hat{\mathbf{x}}_t \left(\theta_{11}^{step3}, \hat{\Sigma}^{step2} \right) \right\}_{t=1}^T$ from (25), we finally update our estimates of θ_2 using (19) and (20). These estimates are denoted $\hat{\mathbf{h}}_0^{step3}$, $\hat{\mathbf{h}}_{\mathbf{x}}^{step3}$, and $\hat{\Sigma}^{step3}$ and we refer to them as θ_2^{step3} .

4 Empirical results: In-sample performance

This section estimates the Gaussian ATSM, the QTSM, and the shadow rate model on US data. Our analysis is structured as follows. Section 4.1 presents the data, and our estimates for models with three pricing factors are discussed in Section 4.2. The two subsequent sections explore how well the models match various moments not included in the estimation.

4.1 Data

We use end-of-month nominal bond yields in the US from July 1961 to December 2013 as provided by Gürkaynak, Sack & Wright (2007). The SR approach is constructed for a setting where many observables are available each time period, and we therefore include more bond yields than typically used in the literature when taking DTSMs to the data. Simulation results for the SR approach by

Andreasen & Christensen (2014) suggest that about 15 bond yields are sufficient and that any efficiency loss of the SR approach compared to Maximum Likelihood may be small with 25 bond yields. Given our interest in the 10-year term structure, we include bond yields in the 0.5-3.0 year maturity range at maturities three months apart, whereas bond yields in the remaining segment of the term structure are included at maturities six months apart.⁹ In total, we thus have 25 points on the yield curve in each time period. Due to a lack of long-term Treasury notes before September 1971, bonds yields in the 7-10 year maturity range are not available before this date. We address this problem by explicitly accounting for missing values in the SR approach.

As mentioned above, we test the performance of the DTSMs considered on a long and a short sample for US bond yields. The full sample is from July 1961 to December 2013 ($T = 631$), whereas the short sample starts in January 1990 and ends in December 2013 ($T = 288$).

4.2 Model estimates

A preliminary estimation of the three-factor models suggests that they are badly identified in the SR approach given the standard normalization restrictions listed above. To realize this for the Gaussian ATSM, recall that bond yields are given by $y_{t,k} = -\frac{1}{k} (A_k + \mathbf{B}'_k \mathbf{x}_t)$ for $k = 1, 2, \dots, K$. The well-known solution to the Gaussian ATSM with our normalization is

$$A_k = -\alpha + A_{k-1} + \frac{1}{2} \mathbf{B}'_{k-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}' \mathbf{B}_{k-1} \approx -\alpha + A_{k-1}, \quad (27)$$

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

$$\mathbf{B}'_k = -\mathbf{1}' + \mathbf{B}'_{k-1} (\mathbf{I} - \boldsymbol{\Phi}). \quad (28)$$

Hence, to estimate the latent factors \mathbf{x}_t by the regression filter it follows from (28) that all eigenvalues of $\boldsymbol{\Phi}$ must be distinct. Moreover, given that $\boldsymbol{\Sigma}$ is badly identified from the cross-section dimension of bond yields due to (27), the ordering of the factors is therefore also badly identified.¹⁰ That is, we obtain nearly identical values for the objective functions in first and third step of the SR approach by changing the order of the eigenvalues in $\boldsymbol{\Phi}$. To eliminate this identification issue we therefore require

⁹These bond yields are computed using the estimated parametric form for the yield curve in Gürkaynak et al. (2007). Ongoing work explores the robustness of our results when using non-parametric estimation methods to extract the yield curves from coupon bonds.

¹⁰A similar finding is reported in Ait-Sahalia & Kimmel (2010) using likelihood inference.

that all eigenvalues of Φ are strictly increasing.¹¹ A similar lack of identification is observed for the QTSM and the shadow rate model, and we therefore also require that the eigenvalues of Φ are strictly increasing in these models.

The estimation results for the three-factor models in the full sample are reported in Table 2. The Gaussian ATSM displays the usual properties with stationary and highly persistent factors under both the \mathbb{Q} and \mathbb{P} measure as all diagonal elements of Φ are positive and the largest eigenvalue of \mathbf{h}_x is 0.9914. Similar properties hold for the pricing factors in the QTSM, where Ψ enforces the ZLB by having strictly positive eigenvalues of 0.0008, 0.0099, and 2.9893, although the smallest eigenvalue is very close to zero.¹² The estimates in the shadow rate model are very similar to those in the Gaussian ATSM, but we also observe some differences, particularly for \mathbf{h}_x . To quantify these differences, the Gaussian ATSM implies that the unconditional correlation between the first and second pricing factors is 0.21 and -0.88 between the second and third pricing factors. The corresponding correlations in the shadow rate model are 0.36 and -0.96 , respectively, and hence somewhat larger in absolute terms.

< Table 2 about here >

Table 3 reveals that the pricing factors for all models in the short sample are slightly less persistent than in the full sample when measured by the largest eigenvalue of \mathbf{h}_x . In the QTSM, the estimates of Ψ imply eigenvalues of 0.0017, 0.0202, and 2.9780, meaning that the short rate is primarily controlled by one pricing factor as in the long sample, given our normalization with $\beta = \mathbf{0}$. We also see that the estimates of \mathbf{h}_x and Σ for the Gaussian ATSM differ substantially from the corresponding estimates in the shadow rate model, contrary to the findings in Bauer & Rudebusch (2014). Hence, our results suggests that one should be cautious of directly using parameters from the Gaussian ATSM in the shadow rate model to explore the implications of the ZLB.

< Table 3 about here >

¹¹This empirical observation is related to Hamilton & Wu (2012), showing that eigenvalues under the \mathbb{Q} measure must be decreasing in Gaussian ATSMs with observed factors to ensure identification.

¹²This implies that the estimates of the \mathbb{Q} parameters are close to the boundary of their domain, implying that the conventional asymptotic distribution cannot be applied for inference. Ongoing work aims to compute these standard errors using a bootstrap procedure.

4.3 Goodness of in-sample fit

This section studies the in-sample fit of the three models. We start by focusing on the objective functions from the first step which for convenience is reported as $\tilde{Q}_{1:T}^{step1} \equiv 100\sqrt{2 \times Q_{1:T}^{step1}}$. This implies that the scaled objective functions denote the standard deviation in basis points of all residuals in the sample. The top part of Table 4 to the left shows that $\tilde{Q}_{1:T}^{step1} = 2.90$ for the Gaussian ATSM, meaning that average pricing errors for this model is 2.90 basis points. Accounting for the ZLB by the shadow rate model delivers a sizable improvement in the fit with $\tilde{Q}_{1:T}^{step1} = 2.74$. A further improvement is seen for the QTSM which marginally provides the best fit in the full sample with $\tilde{Q}_{1:T}^{step1} = 2.70$. This is in line with our prior expectations, given that the three-factor QTSM has five additional parameters compared to the Gaussian ATSM and the shadow rate model.

For the short sample starting in 1990, Table 4 also shows that all three-factor models provide a better fit to bond yields. In this shorter sample, the shadow rate model with $\tilde{Q}_{1:T}^{step1} = 1.69$ clearly outperforms the Gaussian ATSM where $\tilde{Q}_{1:T}^{step1} = 1.81$, and the best in-sample fit is once again obtained by the QTSM having $\tilde{Q}_{1:T}^{step1} = 1.61$.

The right part of Table 4 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 Σ 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 larger than $\tilde{Q}_{1:T}^{step1}$, meaning that the in-sample fit of bond yields is almost unaffected by the alternative estimator of Σ . It is therefore reasonable to believe that the dependence on the \mathbb{P} dynamics through Σ is minimal in our case and that results in the third step of the SR approach largely remain robust to the chosen functional form of $\mathbf{f}(\mathbf{x}_t)$. Unreported results show that the in-sample fit is also robust to omitting the bias-adjustment of θ_2 , partly because Σ is badly identified from the cross-section dimension of bond yields, and partly because the bias-adjustment in $\hat{\Sigma}^{step2}$ typically is small.

< Table 4 about here >

A more careful inspection of the in-sample fit for the full sample is provided in Figure 1, where the first chart shows recursively computed objective functions for all three-factor models, i.e. $\left\{ \tilde{Q}_{1:t}^{step3} \right\}_{t=1}^T$. We find that the fit of all models deteriorates during the 1970s and improves afterwards, and that the relative performance of the models is fairly stable throughout the sample. To study the performance

of these models in greater detail when bond yields are close to the ZLB, the second chart in Figure 1 displays the recursively computed objective functions from January 2005 and onwards.¹³ The Gaussian ATSM delivers the best fit going into the financial crisis in 2008, but its performance deteriorates steadily compared to the shadow rate model and the QTSM after November 2008, where the policy rate approaches the ZLB. The corresponding plot in Figure 2 shows the same pattern for the Gaussian ATSM when the sample starts in 1990. This finding indicates that the three-factor Gaussian ATSM struggles to match bond yields close to the ZLB.

Another way to explore the in-sample fit of bond yields is provided in the third chart of Figure 1 and 2, showing the standard deviation of the residuals by maturity, i.e. $\sigma_k = 100\sqrt{\frac{1}{T} \sum_{t=1}^T v_{t,k}^2}$ for $k = 1, 2, \dots, K$. For both samples, all three-factor models deliver relative low standard errors between 2 and 3 basis points, but when compared to intermediate bond yields we also observe minor spikes in the pricing errors at the short and long end of the term structure with $\sigma_{0.25y} = 7$ and $\sigma_{10y} = 4$ basis points in the long sample and $\sigma_{0.25y} = 4$ and $\sigma_{10y} = 3$ basis points in the short sample

< Figures 1 and 2 about here >

An obvious way to improve the in-sample fit of these models is to add a fourth pricing factors.¹⁴ The bottom part of Table 4 shows that including a fourth pricing factors roughly reduces the objective function and hence the in-sample fit by more than 50% for all models across both samples. Importantly, the shadow rate model now marginally outperforms the QTSM in the two samples and hence provides the best in-sample fit. We consider this a somewhat surprising finding, given that the QTSM with four pricing factors has nine additional parameters compared to the four-factor shadow rate model. The last charts in the second row of Figure 1 and 2 show that the better in-sample performance mainly is due to a closer fit of short and long-term bond yields, where the standard deviation of the pricing errors now are within 2 basis points. In other words, a fourth pricing factors helps to match short-term and long-term bond yields in all models.

Based on these findings we conclude that accounting for the ZLB by either QTSMs and shadow rate models clearly gives a better in-sample fit of US bond yields when compared to the Gaussian

¹³The Federal Open Market Committee has since December 2008 set a target range of 0 to 0.25% for the effective Federal Funds Rate.

¹⁴The estimated model parameters for these four factor models are available on request.

ATSM. With three pricing factors, the QTSM marginally provides the best in-sample fit, whereas the shadow rate model marginally outperforms the QTSM with four pricing factors. We therefore conclude that the two mechanisms to enforce the ZLB largely provide the same in-sample fit of US bond yields. However, the shadow rate specification is more parsimonious than the quadratic policy function and could therefore be preferred for this reason.

4.4 Matching key moments for bond yields

This section tests the models on their ability to match moments not directly 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 run the regressions

$$y_{t+1,k-1} - y_{t,k} = \delta_k + \frac{\phi_k}{k-1} (y_{t,k} - r_t) + u_{t,k} \quad (29)$$

for $k = 2, \dots, K$ where $u_{t,k} \sim \mathcal{IID}(0, \text{Var}(u_{t,k}))$.¹⁵ We then explore if the DTSMs can reproduce the empirical pattern in $\{\phi_k\}_{k=2}^K$ 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 the Campbell-Shiller regressions in (29) is given by

$$y_{t+1,k-1} - y_{t,k} - (c_{t+1,k-1} - c_{t,k-1}) + \frac{1}{k-1} \theta_{t,k-1} = \delta_k^{\mathbb{Q}} + \frac{\phi_k^{\mathbb{Q}}}{k-1} (y_{t,k} - r_t) + u_{t,k}^{\mathbb{Q}} \quad (30)$$

where $u_{t,k}^{\mathbb{Q}} \sim \mathcal{IID}\left(0, \text{Var}\left(u_{t,k}^{\mathbb{Q}}\right)\right)$, $c_{t,k} \equiv y_{t,k} - \frac{1}{k} \sum_{i=0}^{k-1} E_t[r_{t+i}]$ is the spot term premium, and $\theta_{t,k} \equiv f_{t,k} - E_t[r_{t+k}]$ is the forward term premium with $f_{t,k} \equiv -\log(P_{t,k+1}/P_{t,k})$.¹⁶ If term premia are correctly specified by the DTSMs, then $\phi_k^{\mathbb{Q}} = 1$ for $k = 2, 3, \dots, K$. The ability of DTSMs to match these moments is the LPY(ii) test and studies if the models can capture key moments of the \mathbb{Q} dynamics for bond yields.

The ability of the three-factor models to match these four sets of unconditional moments in the

¹⁵In practice we run the regressions $y_{t+m,k-m} - y_{t,k} = \delta_k + \phi_k \frac{m}{k-m} (y_{t,k} - y_{t,m}) + u_{t,k}$ with $m = 6$, i.e. the regressions are done for bi-annually excess returns. We run these regressions on empirical bond yields and on simulated data from each of the models to obtain the model-implied regression loadings.

¹⁶As for (29), in practice we run the regressions $y_{t+m,k-m} - y_{t,k} - (c_{t+m,k-m} - c_{t,k-m}) + \frac{m}{k-m} \theta_{t,k-m} = \delta_k^{\mathbb{Q}} + \phi_k^{\mathbb{Q}} \frac{m}{k-m} (y_{t,k} - y_{t,m}) + u_{t,k}^{\mathbb{Q}}$ with $m = 6$, i.e. the regressions are done for bi-annually risk-adjusted excess returns. We run these regressions using empirical bond yields and model-implied estimates of term premia obtained at $\{\hat{\mathbf{x}}_t\}_{t=1}^T$.

full sample is examined in Figure 3. To explore the impact of the bias-adjustment in θ_2 , charts to the left report the model-implied moments using the unadjusted estimates of θ_2 , whereas the adjustment is imposed in charts to the right. The first row in Figure 3 shows that all models underestimate the average level of the yield curve when θ_2 is not bias-adjusted, whereas these moments are matched when correcting for the bias in θ_2 . The unconditional standard deviations of bond yields are also matched by the Gaussian ATSM and the shadow rate model, but not by the QTSM. We further observe that only the Gaussian ATSM and the shadow rate model reproduce the downward sloping pattern in $\{\phi_k\}_{k=2}^K$, whereas the estimated QTSM cannot match this aspect of bond yields and hence pass the LPY(i) test. The QTSM is more successful at satisfying the LPY(ii) test with $\phi_k^{\mathbb{Q}}$ close to one along all maturities, whereas the two other models imply slightly larger deviations of $\phi_k^{\mathbb{Q}}$ from one.

< Figure 3 about here >

Figure 4 explores how well the three models match the same set of moments for the short sample starting in 1990. Due to the bias-adjustment in θ_2 , all models match the average level of the yield and we also see that they pass the LPY(i) test. The unconditional standard deviations of bond yields are slightly underestimated in the Gaussian ATSM and the shadow rate model, whereas these moments are matched by the QTSM. The last row in Figure 4 suggests that the Gaussian ATSM and the shadow rate model are able to pass the LPY(ii) with $\phi_k^{\mathbb{Q}}$ close to one, whereas the QTSM shows clear deviations from one.

< Figure 4 about here >

We next examine if models with four pricing factors are more successful at matching the moments considered. To conserve space, focus is here devoted to moments from models estimated with the bias-adjustment in θ_2 . For the full sample in Figure 5, we see marginal improvements for the Gaussian ATSM and the shadow rate model in matching LPY(i) and LPY(ii), whereas the performance of the QTSM is largely unaffected. Figure 6 shows that the fourth pricing factor has also minor effects in the short sample, as this additional factor only helps the Gaussian ATSM and the shadow rate model to match the unconditional standard deviations of bond yields.

< Figure 5 and 6 about here >

These findings lead us to the following conclusions. First, the three and four factor QTSMs generally struggle to match loadings from the Campbell-Shiller regressions, whereas these moments are matched by the shadow rate models. Second, the shadow rate models are generally also more successful at passing the LPY(ii) test than the QTSMs, although the latter performs well in the full sample. From a methodological perspective, we document that bias-adjusting θ_2 has a significant impact on unconditional moments of bond yields, and our results therefore supplements those of Bauer et al. (2012), focusing on conditional moments and term premia.

4.5 Matching conditional volatilities in 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 when these yields are close to the ZLB and when this bound is not binding. The shadow rate model 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 conditional volatility of bond yields, and this section therefore studies how well the QTSM and the shadow rate model with three pricing factors match this feature of the data.¹⁷

We use two measures of conditional volatility in the data. The first is the rolling standard deviation of bond yields (denoted $\sigma_{t,k}^{Rolling}$) computed from daily observations with a six month lookback.¹⁸ As a supplement to these non-parametric estimates we also provide the conditional volatility from a GARCH(1,1) model when applied to changes in monthly bond yields (denoted $\sigma_{t,k}^{GARCH}$). Figure 7 shows these estimates for four selected maturities and the model-implied volatilities in the full sample. Overall, the two measures of volatility in the data are fairly similar, although $\sigma_{t,k}^{Rolling}$ is more noisy than $\sigma_{t,k}^{GARCH}$. The QTSM captures most of the gradual increase in volatility during the 1960s and 1970s but does not match the elevated levels in the early 1980s. 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 yield just before approaching the ZLB, as emphasized by the second part of Figure 7 focusing on volatility after

¹⁷The model-implied estimates of conditional volatility in bond yields in the corresponding four-factor models are nearly identical to those from the three factor models and therefore not reported.

¹⁸These daily bond yields are also computed using the estimated parametric form for the yield curve in Gürkaynak et al. (2007).

2005. The shadow rate model predicts constant volatility when the policy rate is far from the ZLB, and the model is therefore unable to capture the overall changes in volatility before 2008. Volatility becomes time-varying when the policy rate approaches the ZLB and the shadow rate model is here able to reproduce the lower volatility level.

< Figure 7 about here >

For the short sample starting in 1990, the QTSM is generally less successful in matching volatility according to Figure 8. To see why, observe that volatility in the QTSM is closely related the level factor and hence the short rate. As shown in Figure 7, 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 shadow rate model, 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 8 about here >

To summarize the relative performance of two models, we regress volatility in the data on a constant and the model-implied volatility. Table 5 confirms our impression from above that the QTSM provides the best fit in the full sample but not in the short sample where the shadow rate model dominates. The low R^2 in these regressions also suggests that both models generally struggle to capture the 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 dynamics of the pricing factors should display heteroskedasticity, for instance induced by stochastic volatility.

< Table 5 about here >

5 Empirical results: performance out-of-sample

This section studies the ability of the models considered 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. We focus on models with three and four pricing factors as above, but two-factor models are also considered

because parsimonious models often perform well out of sample. The forecasting study is carried out by recursively re-estimating all nine models every month to forecasts bond yields up to 12 months ahead. We do so when starting the sample in 1961 and in 1990.¹⁹

Figure 9 reports the root mean squared prediction errors (RMSPE) by maturity when the estimation is started in 1961. Columns in Figure 9 refer to the number of pricing factors and rows refer to 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 roughly similar performance when forecasting 6 and 12 months ahead. The two-factor shadow rate model delivers even better forecasts for short- and medium-term bond yields at the 3, 6, and 12 month horizons, but the model struggles when predicting long-term bond yields. Turning to three-factor models, the QTSM and the shadow rate model have very similar forecasting abilities and dominate the Gaussian ATSM for nearly all maturities and forecast horizons. Importantly, the forecasts of the shadow rate model generally *improves* when including a fourth pricing factor whereas the opposite applies for the QTSM. This suggests that the parsimonious mechanism in shadow rate models to enforce the ZLB is more robust and less likely to overfitting the data than the quadratic specification. A careful inspection of Figure 9 reveals that the three-factor QTSM and the three- and four-factor shadow rate models outperform the random walk for short-term bond yields at all forecast horizons. For medium and long-term bond yields, the models are unable to out-forecast the random walk which is a common finding in the literature. Note also that the forecasting ability of all models generally decline with the forecast horizon, which is qualitatively in line with the findings of Pooter, Ravazzolo & van Dijk (2010).

< Figure 9 about here >

The forecasting results when the estimation is started in 1990 are provided in Figure 10. The overall results are very similar to those obtained in Figure 9 and we therefore only highlight the following. First, the two-factor shadow rate model generally benefits from the shorter estimation window as its RMSPEs are lower than the two other models or very close to the best performing model. Second, the QTSM and the shadow rate model with three pricing factors display roughly similar performance.

¹⁹Given 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. Such an extensive forecasting study is very demanding to carry out with conventional estimation methods, but easily done in our case due to the computational efficiency of the SR approach.

Third, forecasts generally *improve* in the shadow rate model when adding a fourth factor whereas the opposite holds for the QTSM. Finally, regardless of the considered number of pricing factors, the QTSM and the shadow rate model outperform the Gaussian ATSM at nearly all maturities and forecasts horizons.

< Figure 10 about here >

In addition to providing more accurate forecasts than the Gaussian ATSM, the QTSM and the shadow rate model 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 11 by showing forecasts for the 0.5-year bond yield on two occasions. The first is the end of December 2008 where the policy rate reached the ZLB. Predicted bond yields in the three-factor Gaussian ATSM barely stay positive at the considered forecast horizons but not in the four-factor version, where the 0.5-year bond yield is predicted to turn negative after 5 months, i.e. after the end of May 2009. The second row of Figure 11 for the end of May 2010 shows that negative forecasts in the Gaussian ATSM may happen with two, three, and four pricing factors and even when the policy rate has been at the ZLB for several years. The shortcoming of the Gaussian ATSM is even more severe when considering density forecasts, as a substantial part of its predictive distribution is in the negative domain according to Figure 12. Note also that probabilities above 50% in this figure denote negative level forecasts, which happens frequently when the estimation is started in 1961 but less often when starting in 1990.

< Figure 11 about here >

< Figure 12 about here >

We summarize the forecasting performance of the three models in Table 6 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 color 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 shadow rate model has 16 bold figures, the QTSM has 8, and the Gaussian ATSM has none. Based on this finding and the results in Figure 9 and 10 we conclude that the shadow rate model generally performs best out of sample, and that both models accounting for the ZLB do better than the Gaussian ATSM.

Our second color coding scheme in Table 6 applies blue to 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 A and B of Table 6. We surprisingly find that starting the estimation in 1961 leads to the most accurate forecasts, as part A of Table 6 has 25 blue figures whereas part B only has 11. That is, the best forecasts are generally obtained by using a long sample for the estimation, particularly for the shadow rate model. Any finite sample bias in the estimated \mathbb{P} parameters is unlikely to explain this finding as we bias-adjust $\hat{\theta}_2$ regardless of the starting point for the estimation. 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 shorter sample (see Section 4.2) and this is likely to improve forecasts, given the strong forecasting performance of the random walk on bond yields. 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 6 about here >

6 Conclusion

This paper studies the performance of QTSMs and shadow rate models on post-war US bond yields. Accounting for the ZLB by either a QTSM and shadow rate model gives largely the same in-sample fit of US bond yields, with both models clearly outperforming the Gaussian ATSM. The three and four factor QTSMs generally struggle to match loadings from the Campbell-Shiller regressions, whereas these moments are matched by the shadow rate models. In an out-of-sample forecasting study from January 2005 to December 2013, we find that the shadow rate model generally performs better than the QTSM, and that models accounting for the ZLB do better than the Gaussian ATSM. The shadow rate model is also found to be more robust and less subject to overfitting than the QTSM, as the forecasts in the shadow rate model generally improves when moving from three to four pricing factors whereas the opposite holds for the QTSM. Importantly, the QTSM and the shadow rate model ensure sensible forecasts as predicted bond yields stay non-negative whereas they easily turn negative in the Gaussian ATSM.

A Bias-adjusting the second step in the SR approach: A bootstrap procedure

The standard bootstrap procedure for a VAR model without measurement errors in the pricing factors generates the sampling distribution for moments involving $\hat{\mathbf{x}}_t$ and $\hat{\mathbf{w}}_{t+1}$ in (19) and (20). The variability in the remaining moments in (19) and (20) related to the measurement errors is accounted for by resampling with replace from $\left\{\widehat{Var}(\mathbf{u}_t)\right\}_{t=1}^T$, $\left\{\widehat{Cov}(\mathbf{u}_{t+1}, \mathbf{u}_t)\right\}_{t=1}^{T-1}$, and $\left\{\widehat{Cov}(\mathbf{u}_t, \mathbf{u}_{t+1})\right\}_{t=1}^{T-1}$. The suggested bootstrap procedure for a VAR model with measurement errors in the pricing factors is therefore:

1. Use (19) and (20) to obtain $\hat{\boldsymbol{\theta}}_2$. Compute the residuals, i.e. $\hat{\mathbf{w}}_{t+1} = \hat{\mathbf{x}}_{t+1} - \hat{\mathbf{h}}_0 - \hat{\mathbf{h}}_{\mathbf{x}}\hat{\mathbf{x}}_t$ for $t = 1, 2, \dots, T - 1$. Let $b = 1$.
2. Generate a bootstrap sample of length $T - 1$ by resampling with replacement from $\left\{\hat{\mathbf{w}}_{t+1}\right\}_{t=1}^{T-1}$. The bootstrap sample is generated as

$$\mathbf{x}_{t+1}^* = \hat{\mathbf{h}}_0 + \hat{\mathbf{h}}_{\mathbf{x}}\mathbf{x}_t^* + \hat{\mathbf{w}}_{t+1}^* \quad \text{for } t = 1, 2, \dots, T - 1. \quad (31)$$

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

3. Generate $\left\{\widehat{Var}(\mathbf{u}_t)^*\right\}_{t=1}^T$, $\left\{\widehat{Cov}(\mathbf{u}_{t+1}, \mathbf{u}_t)^*\right\}_{t=1}^{T-1}$, and $\left\{\widehat{Cov}(\mathbf{u}_t, \mathbf{u}_{t+1})^*\right\}_{t=1}^{T-1}$ by resampling with replacement from $\left\{\widehat{Var}(\mathbf{u}_t)\right\}_{t=1}^T$, $\left\{\widehat{Cov}(\mathbf{u}_{t+1}, \mathbf{u}_t)\right\}_{t=1}^{T-1}$, and $\left\{\widehat{Cov}(\mathbf{u}_t, \mathbf{u}_{t+1})\right\}_{t=1}^{T-1}$.
4. Use the draws from step 2 and 3 in (19) and (20) to obtain $\hat{\mathbf{h}}_0^{(b)}$, $\hat{\mathbf{h}}_{\mathbf{x}}^{(b)}$, and $\hat{\boldsymbol{\Sigma}}^{(b)}$.
5. If $b < B$, then $b = b + 1$ and go to step 2.
6. Calculate the average of the estimates across all bootstrap samples, i.e.

$$\bar{\mathbf{h}}_{\mathbf{x}} = \frac{1}{B} \sum_{b=1}^B \hat{\mathbf{h}}_{\mathbf{x}}^{(b)} \quad (32)$$

7. Calculate the bootstrap bias-adjusted estimates as

$$\hat{\mathbf{h}}_{\mathbf{x}}^{adj} = \hat{\mathbf{h}}_{\mathbf{x}} - \left(\bar{\mathbf{h}}_{\mathbf{x}} - \hat{\mathbf{h}}_{\mathbf{x}}\right) = 2\hat{\mathbf{h}}_{\mathbf{x}} - \bar{\mathbf{h}}_{\mathbf{x}} \quad (33)$$

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

$$\hat{\mathbf{h}}_0^{adj} = \left(\mathbf{I} - \hat{\mathbf{h}}_{\mathbf{x}}^{adj}\right) \hat{E}[\hat{\mathbf{x}}_t],$$

where $\hat{E}[\hat{\mathbf{x}}_t] \equiv 1/T \sum_{t=1}^T \hat{\mathbf{x}}_t$ remains an unbiased estimator of the sample mean as $E[\mathbf{u}_t] = \mathbf{0}$, given a sufficiently large cross-section panel of bond prices 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

$$\hat{\mathbf{w}}_{t+1}^{adj} = \hat{\mathbf{x}}_{t+1} - \hat{\mathbf{h}}_0^{adj} - \hat{\mathbf{h}}_{\mathbf{x}}^{adj} \hat{\mathbf{x}}_t \quad \text{for } t = 1, 2, \dots, T - 1 \quad (34)$$

and a direct modification of (20), i.e.

$$\begin{aligned}\widehat{Var}(\mathbf{w}_{t+1}) &= \frac{1}{T-1-n_x-1} \sum_{t=1}^{T-1} \widehat{\mathbf{w}}_{t+1}^{adj} \left(\widehat{\mathbf{w}}_{t+1}^{adj} \right)' \\ &\quad - \frac{1}{T-1} \sum_{t=1}^{T-1} \left(\widehat{Var}(\mathbf{u}_t) + \widehat{\mathbf{h}}_{\mathbf{x}}^{adj} \widehat{Var}(\mathbf{u}_t) \left(\widehat{\mathbf{h}}_{\mathbf{x}}^{adj} \right)' \right) \\ &\quad + \frac{1}{T-1} \sum_{t=1}^{T-1} \left(\widehat{Cov}(\mathbf{u}_{t+1}, \mathbf{u}_t) \left(\widehat{\mathbf{h}}_{\mathbf{x}}^{adj} \right)' + \widehat{\mathbf{h}}_{\mathbf{x}}^{adj} \widehat{Cov}(\mathbf{u}_t, \mathbf{u}_{t+1}) \right)\end{aligned}\tag{35}$$

where we have imposed the standard degree of freedom adjustment. Hence, $\widehat{\Sigma}$ is then obtained from a Cholesky decomposition of $\widehat{Var}(\mathbf{w}_{t+1})$. The standard errors for the bias-adjusted estimates are computed using output from the bootstrap in a standard way, i.e. from $\left\{ \widehat{\mathbf{h}}_0^{(b)}, \widehat{\mathbf{h}}_{\mathbf{x}}^{(b)}, \widehat{\Sigma}^{(b)} \right\}_{b=1}^B$.

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 $\{\widehat{\mathbf{x}}_t\}_{t=1}^T$ and the unconditional variances implied by the VAR model. To compute the variance of $x_{i,t}$ by the bias-adjusted VAR model, we consider

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

and

$$\widehat{\mathbf{h}}_0^{adj}(\delta) = \left(\mathbf{I} - \widehat{\mathbf{h}}_{\mathbf{x}}^{adj}(\delta) \right) \widehat{E}[\widehat{\mathbf{x}}_t].$$

For given values of $\widehat{\mathbf{h}}_{\mathbf{x}}^{adj}(\delta)$ and $\widehat{\mathbf{h}}_0^{adj}(\delta)$, we may then compute the residuals as

$$\widehat{\mathbf{w}}_{t+1}^{adj}(\delta) = \widehat{\mathbf{x}}_{t+1} - \widehat{\mathbf{h}}_0^{adj}(\delta) - \widehat{\mathbf{h}}_{\mathbf{x}}^{adj}(\delta) \widehat{\mathbf{x}}_t \quad \text{for } t = 1, 2, \dots, T-1$$

and estimate the variance of the innovations by

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

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

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

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

To compute the model-independent unconditional variances of the factors as implied by $\{\widehat{\mathbf{x}}_t\}_{t=1}^T$,

the unconditional mean of the i 'th pricing factor is estimated by $\hat{E}[\hat{x}_{i,t}]$. We also have

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

for $i = 1, 2, \dots, n_x$, where the last line follows by considering $u_{i,t}^2$ as a point estimate of $\text{Var}(u_{i,t})$. A similar argument is used when computing White's heteroskedastic consistent standard errors. Clearly, $\frac{1}{T-1} \sum_{t=1}^T \left(x_{i,t}^o - \hat{E}[\hat{x}_i] \right)^2 \xrightarrow{p} \text{Var}(x_{i,t}^o)$ as $T \rightarrow \infty$. We also have for $T \rightarrow \infty$, that

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

as $E[u_{i,t}] = 0$ for $i = 1, 2, \dots, n_x$. 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 \mathbf{x}_t . Hence, $E[x_{i,t}^o u_{i,t}] = 0$, at least up to a first-order approximation. Thus,

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

This implies that the unconditional variance of i 'th pricing factor from the sample may be estimated by

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

We then suggest to let the scaling parameter δ be given by

$$\hat{\delta} = \arg \min_{\delta \in [\delta_{lower}, 1]} \sum_{i=1}^{n_x} \left(\frac{\sigma_{i,VAR}^2(\delta) - \hat{\sigma}_{i,Data}^2}{\hat{\sigma}_{i,Data}^2} \right)^2 \tag{36}$$

where $\delta_{lower} > 0$. The constraint on the domain of δ is imposed because at $\delta = 0$, we have $\hat{\mathbf{h}}_{\mathbf{x}}^{adj}(\delta = 0) = \mathbf{0}$ and $\hat{\mathbf{h}}_{\mathbf{0}}^{adj}(\delta = 0) = E[\mathbf{x}_t]$, meaning that the two estimators of the unconditional variances in (36) coincide.

C Bias-adjusting the first and third step in the SR approach: A bootstrap procedure

[To be completed]

References

- 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. (2013), ‘Non-linear DSGE models and the central difference kalman filter’, *Journal of Applied Econometrics* **28**, 929–955.
- Andreasen, M. M. & Christensen, B. J. (2014), ‘The SR approach: A new estimation procedure for non-linear and non-Gaussian dynamic term structure models’, *Working Paper* .
- Ang, A., Boivin, J., Dong, S. & Loo-Kung, R. (2011), ‘Monetary policy shifts and the term structure’, *Review of Economic Studies* .
- Bauer, M. D. & Rudebusch, G. D. (2014), ‘Monetary policy expectations at the zero lower bound’, *Federal Reserve Bank of San Francisco Working 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.
- 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. (2013), ‘Estimating shadow-rate term structure models with near-zero yields’, *Federal Reserve Bank of San Francisco: Working Paper 2013-07* .
- Clarida, R., Gali, J. & Gertler, M. (2000), ‘Monetary policy rules and macroeconomic stability: Evidence and some theory’, *The Quarterly Journal of Economics* **115**, 147–180.
- 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.
- de Jong, F. (2000), ‘Time series and cross-section information in affine term-structure models’, *Journal of Business and Economic Statistics* **18**, 300–314.
- Doucet, A., de Freitas, N. & Gordon, N. (2001), ‘Sequential monte carlo methods in practice’, *Springer* .
- Duan, J.-C. & Simonato, J.-G. (1999), ‘Estimating and testing exponential-affine term structure models by Kalman filter’, *Review of Quantitative Finance and Accounting* **13**, 111–135.
- 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.

- 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.
- Ichihue, H. & Ueno, Y. (2007), ‘Equilibrium interest rate and the yield curve in a low interest rate environment’, *Bank of Japan Working Paper* **07-E-18**.
- Ichihue, 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.
- Kilian, L. (1998), ‘Small-sample confidence intervals for impulse response functions’, *The review of economics and statistics* **80**, 218–230.
- 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.
- Pooter, M. D., Ravazzolo, F. & van Dijk, D. (2010), ‘Term structure forecasting using macro factors and forecast combination’, *Board of Governors of the Federal Reserve System, International Finance Discussion Paper* **993**.
- Pribsch, M. A. (2013), ‘Computing arbitrage-free yields in multi-factor gaussian shadow-rate term structure models’, *Finance and Economics Discussion Series, Federal Reserve Board, Washington, D.C.* .
- Realdon, M. (2006), ‘Quadratic term structure models in discrete time’, *Finance Research Letters* **3**, 277–289.
- Rogers, L. C. G. (1995), ‘Which model for term-structure of interest rates should one use?’, *Mathematical Finance, IMA* **65**, 93–116.
- Rossi, G. D. (2004), ‘Maximum likelihood estimation of the cox-ingersoll-ross model using particle filters’, *Working Paper* .

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 correction is computed with $B = 5.000$ bootstrap replications. The data generating processes (DGP) are the estimated VAR models for the pricing factors under physical measure in the benchmark ATSM reported in Table 2 and Table 3. 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 in 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 | Killian'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 | | | | | | |
| $T = 250$ | Bias(\mathbf{h}_0) | 0.0004 | 0.0002 | 0.0002 | 0.0003 | 0.0003 |
| | Bias($\mathbf{h}_{\mathbf{x}}$) | 0.1563 | 0.0547 | 0.0642 | 0.0850 | 0.0747 |
| | Bias($\Sigma \times 100$) | 0.0012 | 0.0006 | 0.0006 | 0.0007 | 0.0007 |
| | 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 |
| $T = 500$ | Bias(\mathbf{h}_0) | 0.0002 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |
| | Bias($\mathbf{h}_{\mathbf{x}}$) | 0.0676 | 0.0115 | 0.0152 | 0.0234 | 0.0190 |
| | Bias($\Sigma \times 100$) | 0.0005 | 0.0002 | 0.0003 | 0.0003 | 0.0003 |
| | 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 | | | | | | |
| $T = 250$ | Bias(\mathbf{h}_0) | 0.0086 | 0.0027 | 0.0032 | 0.0043 | 0.0031 |
| | Bias($\mathbf{h}_{\mathbf{x}}$) | 3.7685 | 1.2012 | 1.4260 | 1.8965 | 1.2938 |
| | Bias($\Sigma \times 100$) | 0.0129 | 0.0045 | 0.0062 | 0.0077 | 0.0068 |
| | 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 |
| $T = 500$ | Bias(\mathbf{h}_0) | 0.0035 | 0.0006 | 0.0006 | 0.0007 | 0.0006 |
| | Bias($\mathbf{h}_{\mathbf{x}}$) | 1.5484 | 0.2233 | 0.2394 | 0.2949 | 0.2334 |
| | Bias($\Sigma \times 100$) | 0.0037 | 0.0017 | 0.0018 | 0.0019 | 0.0019 |
| | 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: Estimation results for three-factor models: sample from 1961-2013

Robust standard errors for elements in $\hat{\theta}_{11}^{step3}$ are computed using (26) and (17) with $w_D = 10$ and $w_T = 10$. For elements in $\hat{\theta}_2^{step3}$, robust standard errors are computed using (21) with \mathbf{S} obtained by the Newey-West estimator with a bandwidth of 5. Estimates with one or two stars denote significance at the 5 percent and 1 percent level, respectively.

| | ATSM | | QTSM | | Shadow rate | |
|---------------|-----------------------------|-----------------------|------------------------|----|-----------------------------|-----------------------|
| | Estimate | SE | Estimate | SE | Estimate | SE |
| α | 0.0124** | 0.0016 | - | - | 0.0153** | 0.0027 |
| A_{12} | - | - | 0.9917 | | - | - |
| A_{13} | - | - | 0.9931 | | - | - |
| A_{23} | - | - | 0.8638 | | - | - |
| Φ_{11} | 0.0022** | 1.30×10^{-4} | 0.0011 | | 0.0013** | 7.10×10^{-5} |
| Φ_{22} | 0.0355** | 0.0027 | 0.0402 | | 0.0427** | 0.0028 |
| Φ_{33} | 0.0685** | 0.0034 | 0.0811 | | 0.0666** | 0.0030 |
| μ_1 | - | - | 0.0251 | | - | - |
| μ_2 | - | - | 6.39×10^{-11} | | - | - |
| μ_3 | - | - | 0.1090 | | - | - |
| $h_0(1, 1)$ | -1.03×10^{-4} | 7.91×10^{-5} | -8.51×10^{-4} | | -1.67×10^{-4} | 1.01×10^{-4} |
| $h_0(2, 1)$ | 3.83×10^{-4} | 2.53×10^{-4} | -0.0085 | | $9.15^* \times 10^{-4}$ | 5.45×10^{-4} |
| $h_0(3, 1)$ | $-4.69^{**} \times 10^{-4}$ | 9.16×10^{-5} | 0.0136 | | -0.0010^{**} | 1.29×10^{-4} |
| $h_x(1, 1)$ | 0.9847** | 0.0029 | 0.9722 | | 0.9822** | 0.0021 |
| $h_x(1, 2)$ | 0.0252** | 0.0094 | 0.0082 | | 0.0216* | 0.0088 |
| $h_x(1, 3)$ | 0.0186** | 0.0060 | 0.0042 | | 0.0188** | 0.0032 |
| $h_x(2, 1)$ | 0.0489** | 0.0100 | 0.0726 | | 0.0906** | 0.0124 |
| $h_x(2, 2)$ | 0.9668** | 0.0312 | 1.0284 | | 1.0097** | 0.0500 |
| $h_x(2, 3)$ | 0.0611** | 0.0188 | 0.1074 | | 0.0866** | 0.0169 |
| $h_x(3, 1)$ | -0.0557^{**} | 0.0034 | -0.0829 | | -0.1004^{**} | 0.0028 |
| $h_x(3, 2)$ | -0.0151 | 0.0113 | -0.0615 | | -0.0637^{**} | 0.0119 |
| $h_x(3, 3)$ | 0.8685** | 0.0071 | 0.8328 | | 0.8452** | 0.0042 |
| Σ_{11} | $3.56^{**} \times 10^{-4}$ | 1.61×10^{-5} | 0.0021 | | $3.30^{**} \times 10^{-4}$ | 1.46×10^{-5} |
| Σ_{21} | $-6.22^{**} \times 10^{-4}$ | 5.85×10^{-5} | -0.0025 | | $-6.49^{**} \times 10^{-4}$ | 9.46×10^{-5} |
| Σ_{22} | 0.0011** | 4.97×10^{-5} | 0.0096 | | 0.0018** | 7.37×10^{-5} |
| Σ_{31} | $3.95^{**} \times 10^{-4}$ | 2.51×10^{-5} | 0.0011 | | $4.54^{**} \times 10^{-4}$ | 2.50×10^{-5} |
| Σ_{32} | -0.0010^{**} | 3.41×10^{-5} | -0.0094 | | -0.0017^{**} | 3.40×10^{-5} |
| Σ_{33} | $4.31^{**} \times 10^{-4}$ | 2.65×10^{-5} | 0.0029 | | $4.40^{**} \times 10^{-4}$ | 2.66×10^{-5} |

Table 3: Estimation results for three-factor models: sample from 1990-2013

Robust standard errors for elements in $\hat{\theta}_{11}^{step3}$ are computed using (26) and (17) with $w_D = 10$ and $w_T = 10$. For elements in $\hat{\theta}_2^{step3}$, robust standard errors are computed using (21) with \mathbf{S} obtained by the Newey-West estimator with a bandwidth of 5. Estimates with one or two stars denote significance at the 5 percent and 1 percent level, respectively.

| | ATSM | | QTSM | | Shadow rate | |
|---------------|----------------------------|-----------------------|------------------------|----|------------------------|-----------------------|
| | Estimate | SE | Estimate | SE | Estimate | SE |
| α | 0.0093** | 0.0008 | - | - | 0.0099** | 7.10×10^{-4} |
| A_{12} | - | - | 0.9807 | | - | - |
| A_{13} | - | - | 0.9897 | | - | - |
| A_{23} | - | - | 0.6829 | | - | - |
| Φ_{11} | 0.0043** | 0.0003 | 0.0029 | | 0.0035** | 1.41×10^{-4} |
| Φ_{22} | 0.0488** | 0.0025 | 0.0460 | | 0.0475** | 0.0047 |
| Φ_{33} | 0.0517** | 0.0022 | 0.0724 | | 0.0558** | 0.0022 |
| μ_1 | - | - | 0.0073 | | - | - |
| μ_2 | - | - | 6.70×10^{-10} | | - | - |
| μ_3 | - | - | 0.0971 | | - | - |
| $h_0(1, 1)$ | -2.81×10^{-4} | 2.23×10^{-4} | 6.23×10^{-4} | | -3.32×10^{-4} | 2.32×10^{-4} |
| $h_0(2, 1)$ | 0.0078 | 0.0060 | -0.0111 | | 0.0037 | 0.0026 |
| $h_0(3, 1)$ | -0.0079** | 8.04×10^{-5} | 0.0159 | | -0.0039** | 9.00×10^{-5} |
| $h_x(1, 1)$ | 0.9530** | 0.0071 | 0.9430 | | 0.9474** | 0.0070 |
| $h_x(1, 2)$ | -0.0216 | 0.0226 | -0.0063 | | -0.0230 | 0.0207 |
| $h_x(1, 3)$ | -0.0237** | 9.89×10^{-4} | -0.0231 | | -0.0285** | 0.0024 |
| $h_x(2, 1)$ | 1.2792** | 0.2267 | 0.1616 | | 0.5884** | 0.0921 |
| $h_x(2, 2)$ | 2.4472** | 0.6265 | 1.0599 | | 1.5625** | 0.2385 |
| $h_x(2, 3)$ | 1.5652** | 0.0278 | 0.1724 | | 0.6813** | 0.0282 |
| $h_x(3, 1)$ | -1.3025** | 0.0023 | -0.1840 | | -0.6150** | 0.0025 |
| $h_x(3, 2)$ | -1.5152** | 0.0084 | -0.0972 | | -0.6272** | 0.0082 |
| $h_x(3, 3)$ | -0.6356** | 3.65×10^{-4} | 0.7619 | | 0.2476** | 9.27×10^{-4} |
| Σ_{11} | $3.92^{**} \times 10^{-4}$ | 2.88×10^{-5} | 0.0028 | | 3.73×10^{-4} | 2.64×10^{-5} |
| Σ_{21} | -0.0049** | 9.40×10^{-4} | -0.0028 | | -0.0018** | 3.70×10^{-4} |
| Σ_{22} | 0.0108** | 5.32×10^{-4} | 0.0118 | | 0.0042** | 1.97×10^{-4} |
| Σ_{31} | 0.0046** | 9.36×10^{-6} | 1.82×10^{-4} | | 0.0014** | 9.62×10^{-6} |
| Σ_{32} | -0.0108** | 1.09×10^{-5} | -0.0117 | | -0.0042** | 1.14×10^{-5} |
| Σ_{33} | 1.66×10^{-5} | 1.10×10^{-5} | 0.0024 | | 1.69×10^{-4} | 1.14×10^{-4} |

Table 4: In-sample fit: The objective functions

This table reports $100\sqrt{2 \times Q_{1:T}^{step1}}$ and $100\sqrt{2 \times Q_{1:T}^{step3}}$ for the first and third step in the SR approach. Figures in bold highlight the best in-sample fit for a given estimation step.

| | Step 1 | | | Step 3 | | |
|-----------------------|--------|--------------|--------------|--------|--------------|--------------|
| | ATSM | QTSM | Shadow rate | ATSM | QTSM | Shadow rate |
| Three pricing factors | | | | | | |
| Sample: 1961-2013 | 2.895 | 2.704 | 2.735 | 2.896 | 2.719 | 2.786 |
| Sample: 1990-2013 | 1.807 | 1.614 | 1.692 | 1.829 | 1.632 | 1.754 |
| Four pricing factors | | | | | | |
| Sample: 1961-2013 | 1.057 | 1.030 | 1.013 | 1.058 | 1.034 | 1.025 |
| Sample: 1990-2013 | 0.802 | 0.766 | 0.749 | 0.806 | 0.772 | 0.763 |

Table 5: Conditional volatility of bond yields

This table reports the slope 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 at time t are computed from a local linearization of bond yields at \hat{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 | | Shadow rate | | QTSM | | Shadow rate | |
| | Slope | R^2 | Slope | R^2 | Slope | R^2 | Slope | R^2 |
| Sample: 1961-2013 | | | | | | | | |
| 0.5-year bond yield | 1.37 | 0.28 | 1.21 | 0.07 | 1.21 | 0.33 | 0.92 | 0.06 |
| 2-year bond yield | 1.20 | 0.29 | 1.19 | 0.09 | 1.16 | 0.38 | 1.00 | 0.09 |
| 5-year bond yield | 0.94 | 0.24 | 0.95 | 0.07 | 0.84 | 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 | 0.44 | 0.09 | 2.69 | 0.17 | 0.21 | 0.06 | 1.47 | 0.16 |
| 2-year bond yield | 0.48 | 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 | -0.05 | 0.01 | 0.18 | 0.01 |

Table 6: 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 before the forecast. The forecasted bond yields in the shadow rate models 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 marked by blue denote the lowest RMSPEs for a given model when comparing part A and B of the table.

| | Part A: Model estimation from 1961 | | | | Part B: Model estimation 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 |
| Shadow rate | 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 |
| Shadow rate | 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 |
| Shadow rate | 26.32 | 50.73 | 74.65 | 102.61 | 30.26 | 51.37 | 76.08 | 110.10 |

Figure 1: Sample from 1961-2013: In-sample fit for three- and four-factor models

Charts in the first column report $100\sqrt{2 \times Q_{1:T}^{step3}}$. Charts in the second column report $100\sqrt{2 \times Q_{2005:T}^{step3}}$ and the final column reports σ_k estimated using residuals from 1961-2013.

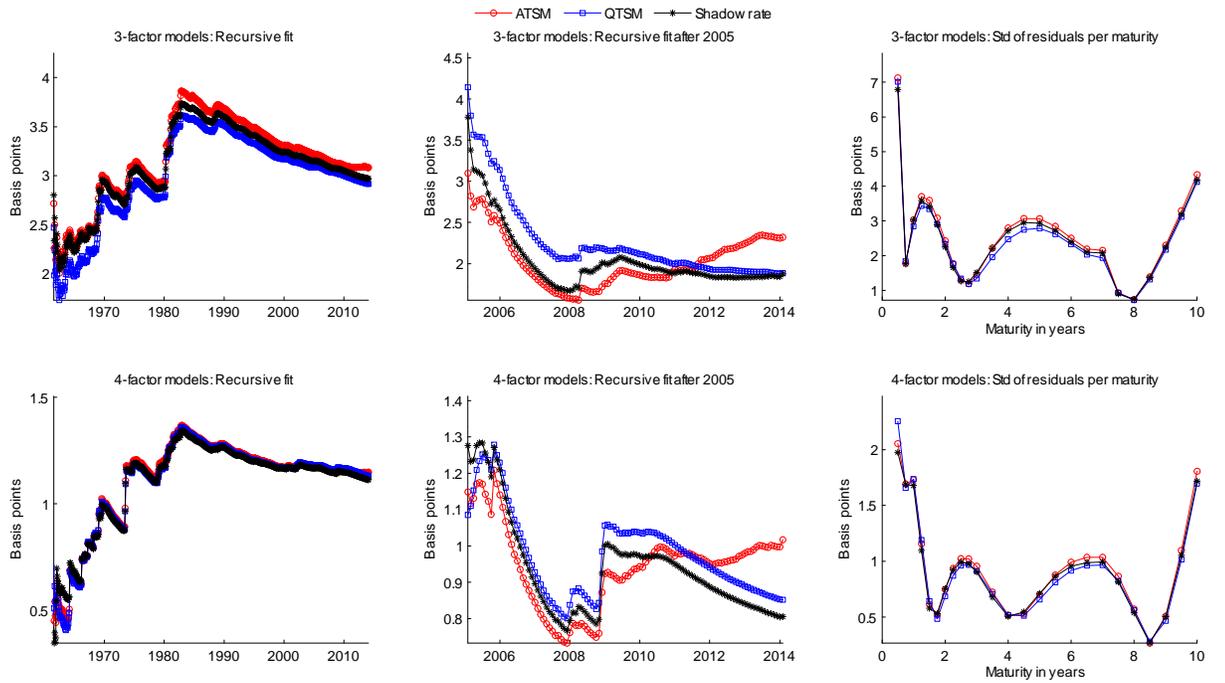


Figure 2: Sample from 1990-2013: In-sample fit for three- and four-factor models

Charts in the first column report $100\sqrt{2 \times Q_{1:T}^{step3}}$. Charts in the second column report $100\sqrt{2 \times Q_{2005:T}^{step3}}$ and the final column reports σ_k estimated using residuals from 1990-2013.

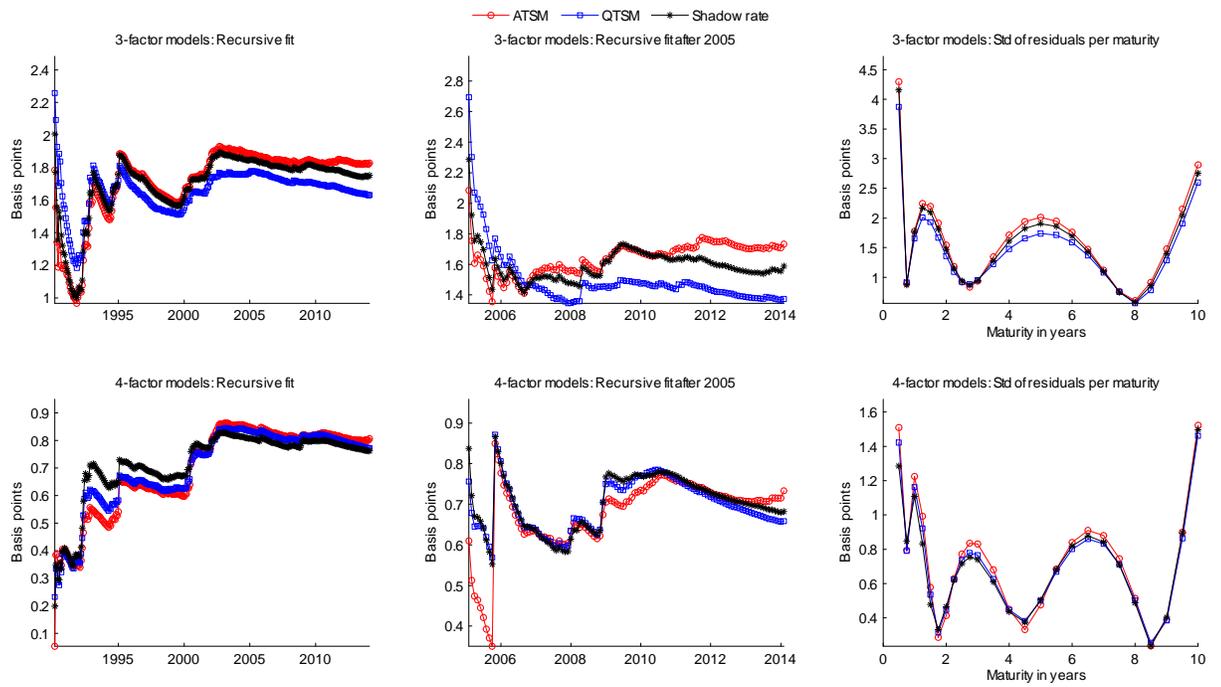


Figure 3: Sample from 1961-2013: Unconditional moments in three-factor models
 All model-based moments are obtained from simulated time series of 100,000 observations. Empirical moments are computed from September 1971 to December 2013 to avoid missing observations for long bond yields.

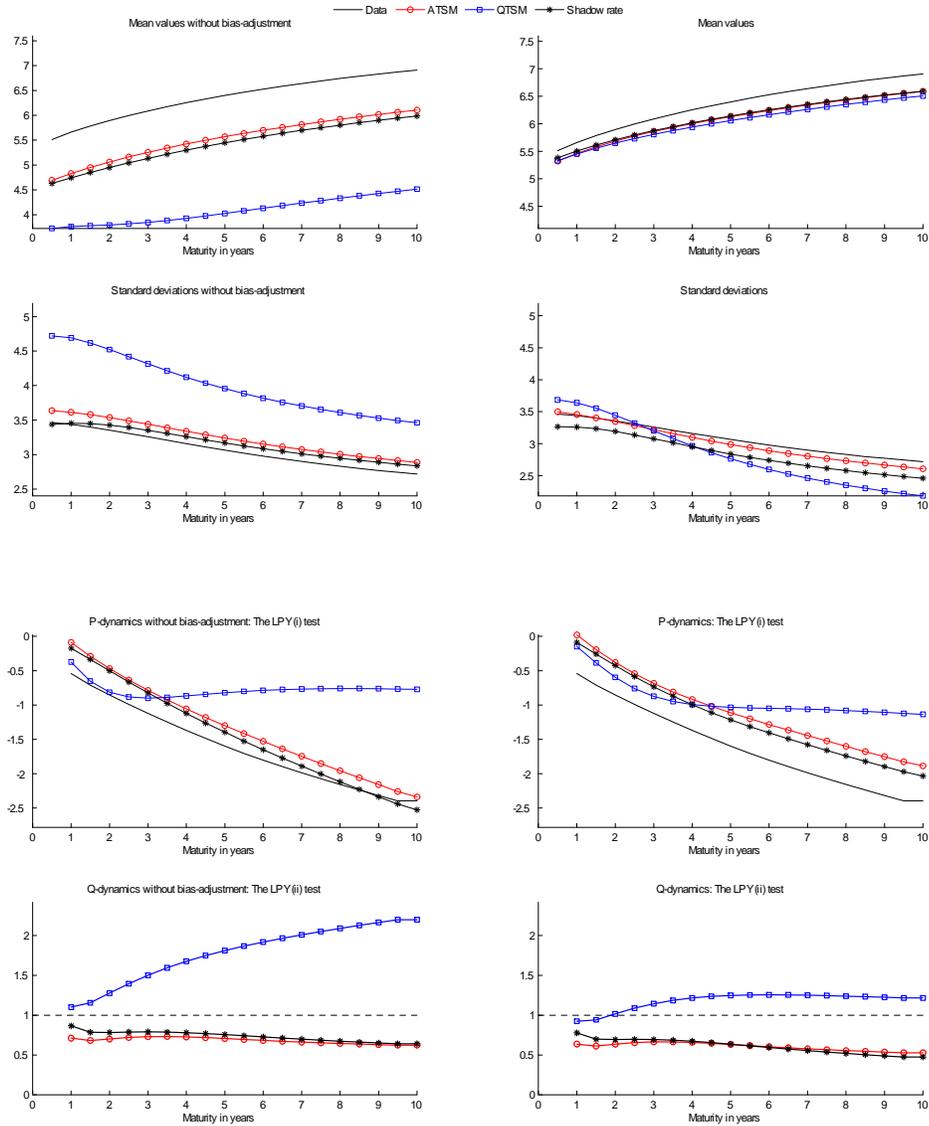


Figure 4: Sample from 1990-2013: Unconditional moments in three-factor models
 All model-based moments are obtained from simulated time series of 100,000 observations. Empirical moments are computed from January 1990 to December 2013.

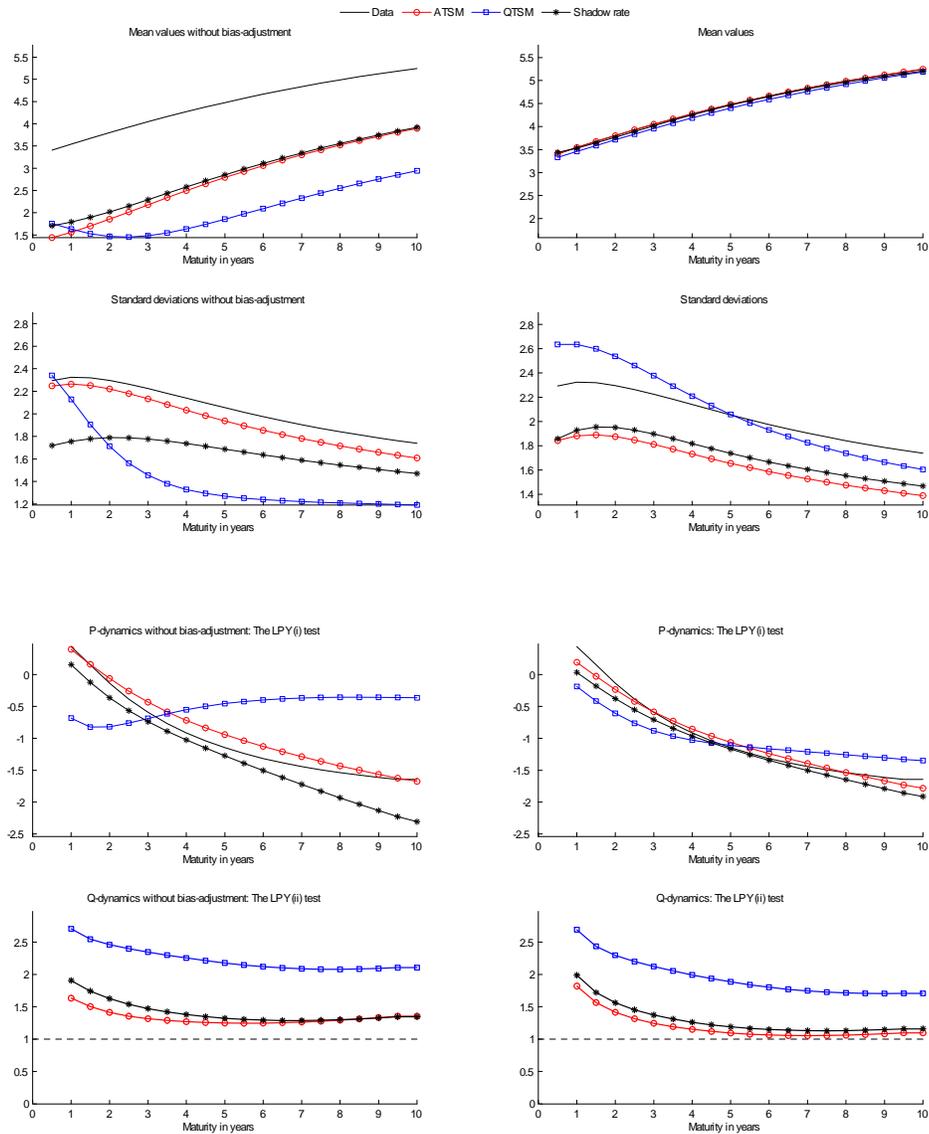


Figure 5: Sample from 1961-2013: Unconditional moments in four-factor models
 All model-based moments are obtained from simulated time series of 100,000 observations using the estimates from Tables 1 to 3.

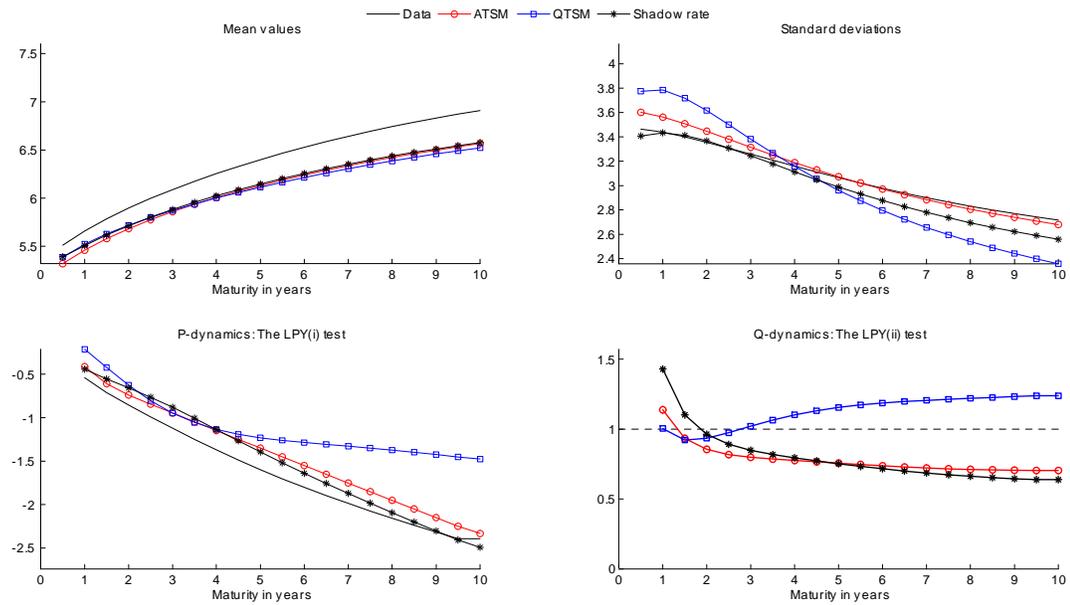


Figure 6: Sample from 1990-2013: Unconditional moments in four-factor models
 All model-based moments are obtained from simulated time series of 100,000 observations using the estimates from Tables 1 to 3.

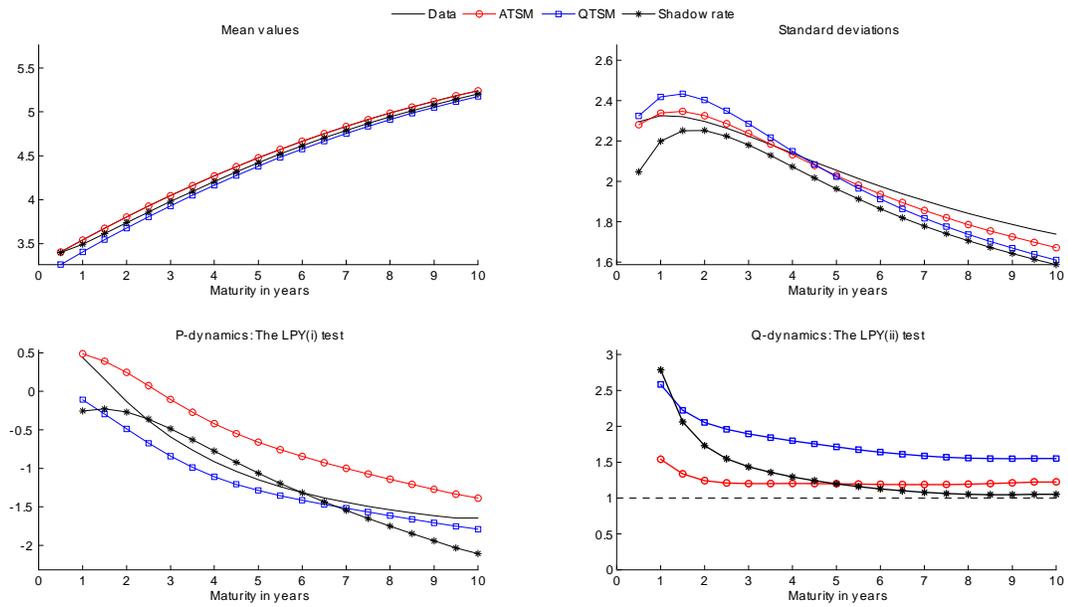


Figure 7: Sample from 1961-2013: Conditional volatilities for bond yields

Black lines denote the rolling standard deviation of bond yields computed from daily observations with a six month lookback. Green lines refer to the conditional volatilities from a GARCH(1,1) model applied to changes in monthly bond yields. Blue lines with squares and black lines with stars denote the one step ahead conditional volatilities in the QTSM and the shadow rate model, respectively, where the volatility at time t is computed from a local linearization of bond yields at \hat{x}_{t-1} .

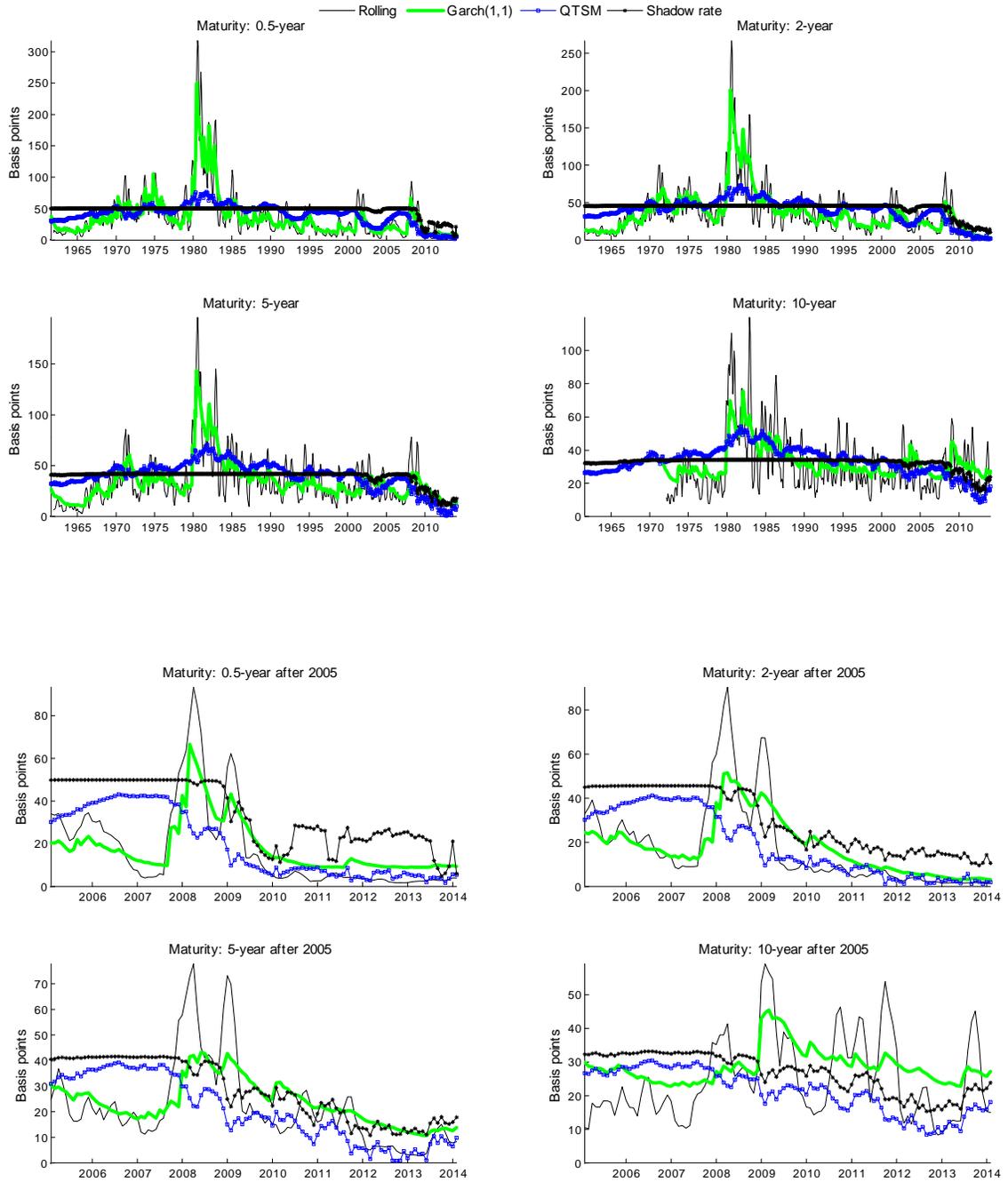


Figure 8: Sample from 1990-2013: Conditional volatilities for bond yields

Black lines denote the rolling standard deviation of bond yields computed from daily observations with a six month lookback. Green lines refer to the conditional volatilities from a GARCH(1,1) model applied to changes in monthly bond yields. Blue lines with squares and black lines with stars denote the one step ahead conditional volatilities in the QTSM and the shadow rate model, respectively, where the volatility at time t is computed from a local linearization of bond yields at \hat{x}_{t-1} .

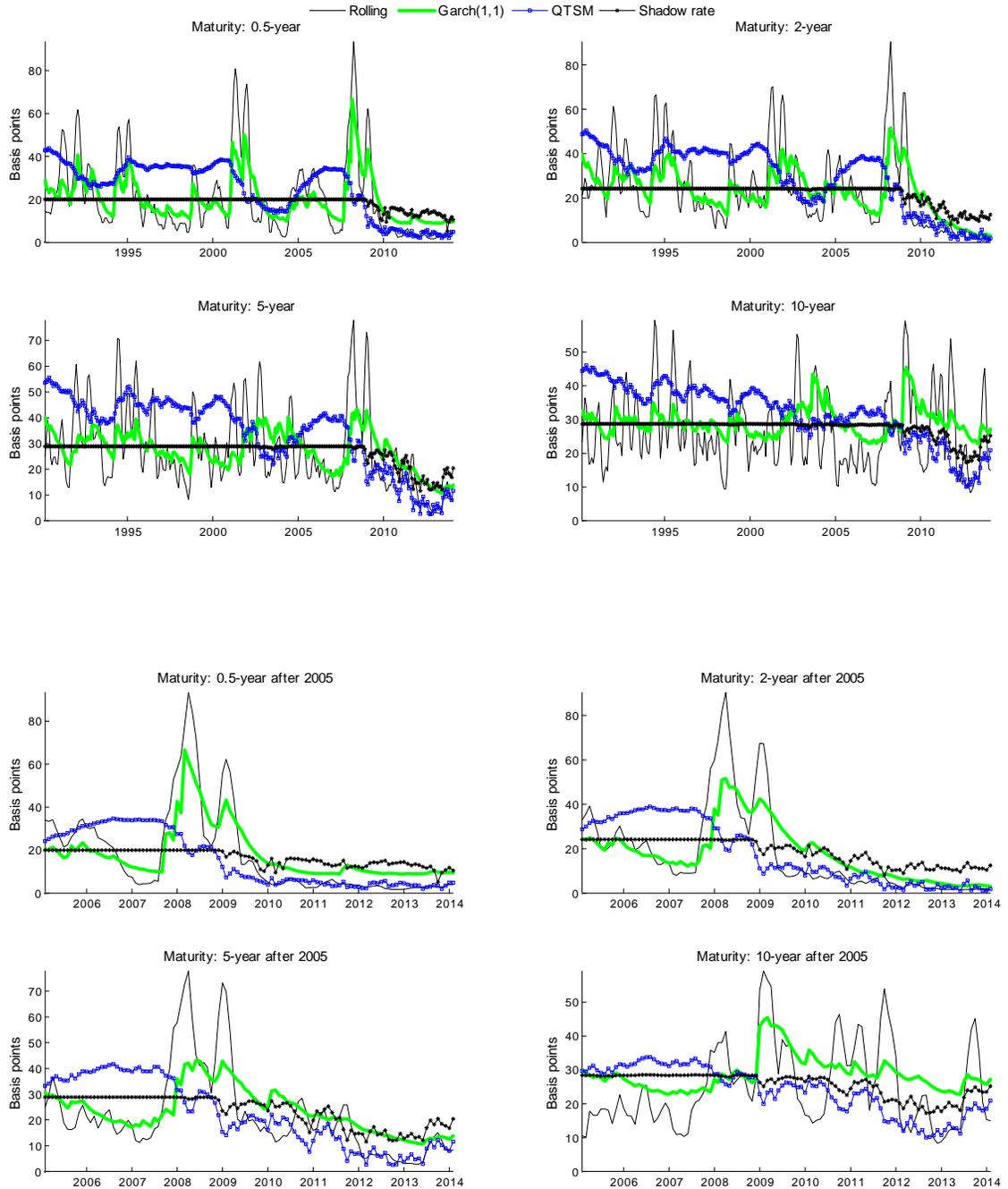


Figure 9: Forecasting results by maturity: model estimation starting in 1961

This figure reports the root mean squared prediction errors (RMSPEs) for out-of-sample forecasts from January 2005 to December 2013. The RMSPEs are generated from models estimated recursively from 1961 to the month prior to the forecast. The forecasted bond yields in the shadow rate models are computed by Monte Carlo integration using 10,000 draws.

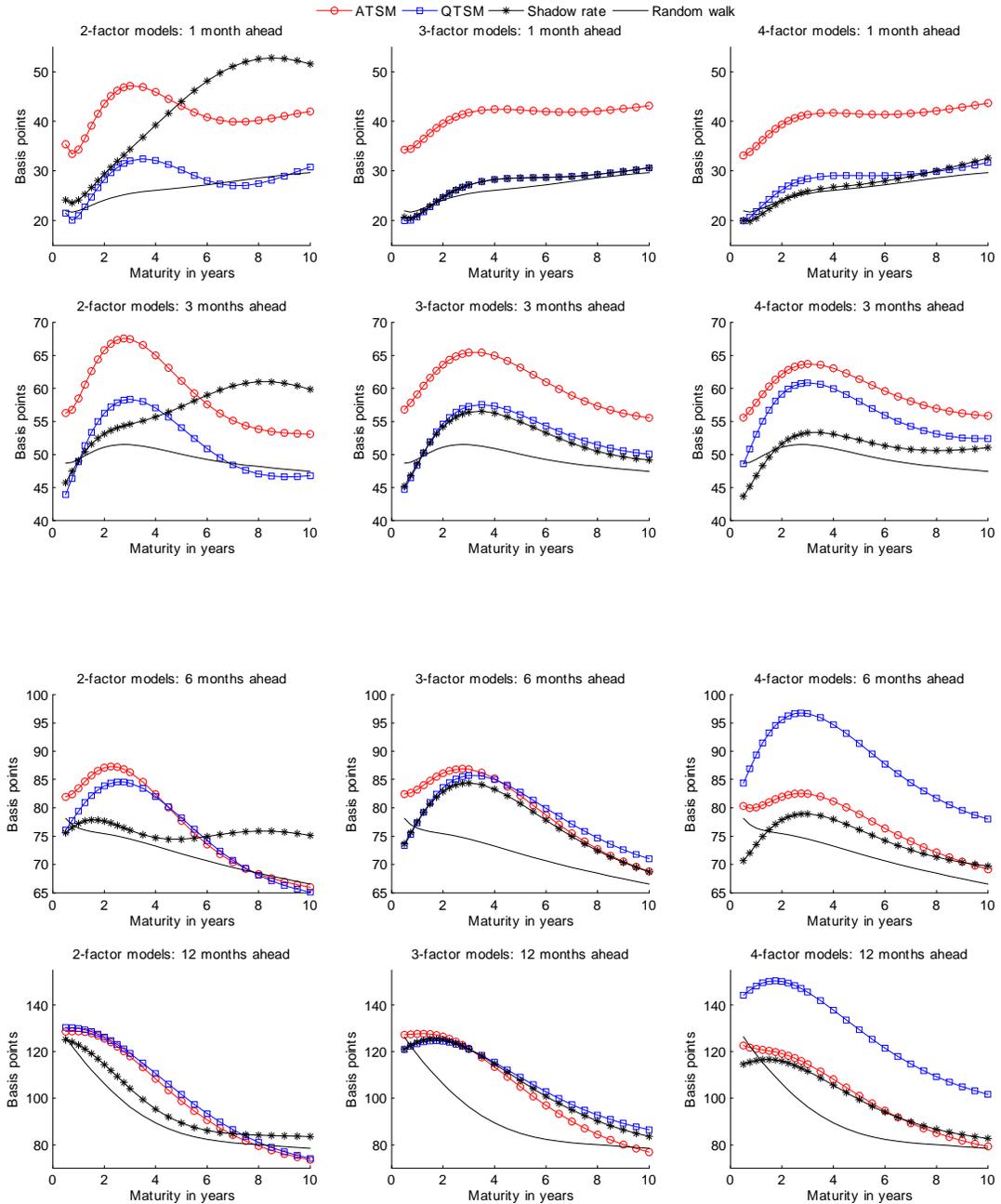


Figure 10: Forecasting results by maturity: model estimation starting in 1990

This figure reports the root mean squared prediction errors (RMSPEs) for out-of-sample forecasts from January 2005 to December 2013. The RMSPEs are generated from models estimated recursively from 1990 to the month prior to the forecast. The forecasted bond yields in the shadow rate models are computed by Monte Carlo integration using 10,000 draws.

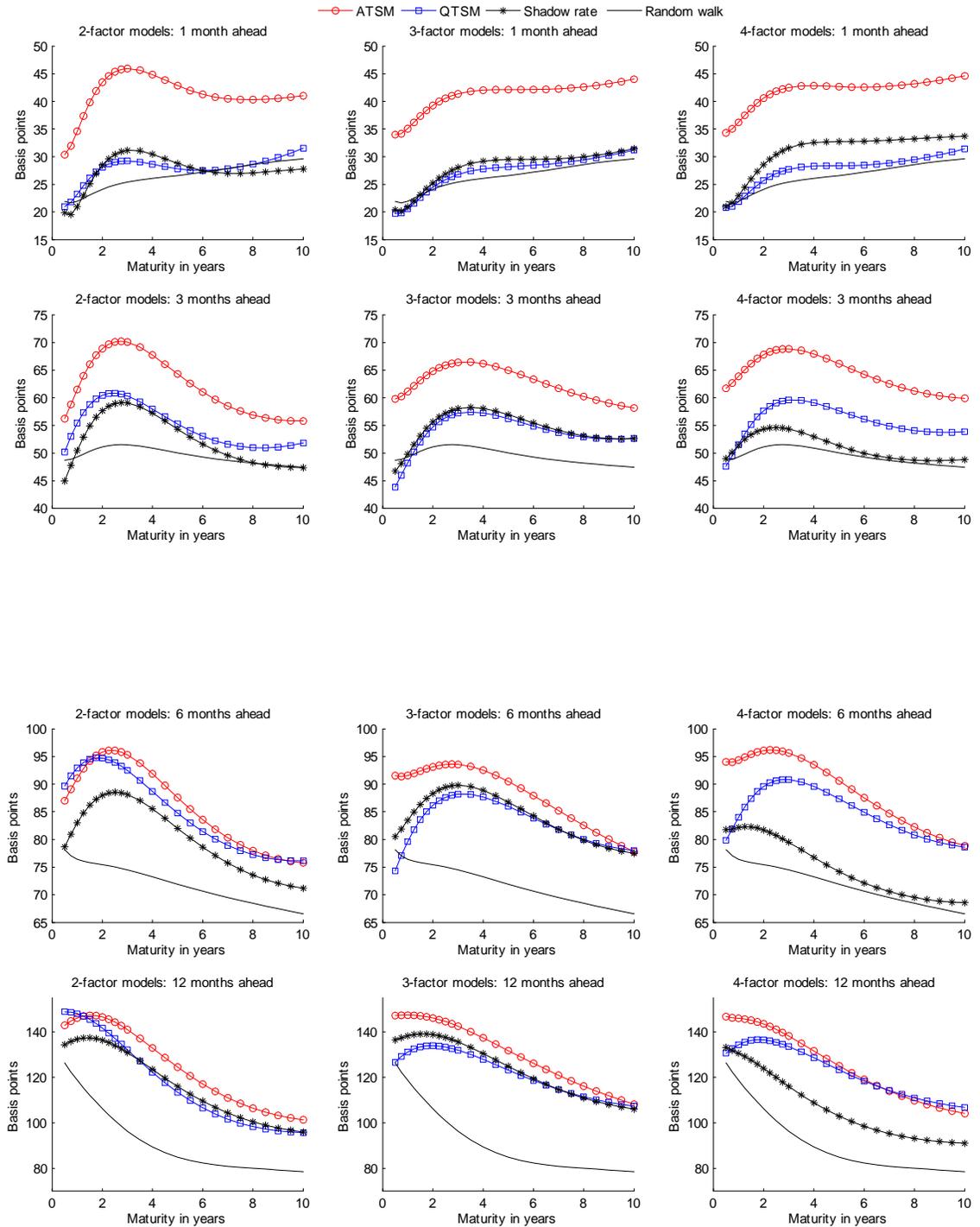


Figure 11: Forecasting illustration for the 0.5-bond yield

The forecasts generated from models estimated recursively from 1961 to the month prior to the forecast. The forecasted bond yields in the shadow rate models are computed by Monte Carlo integration using 10.000 draws.

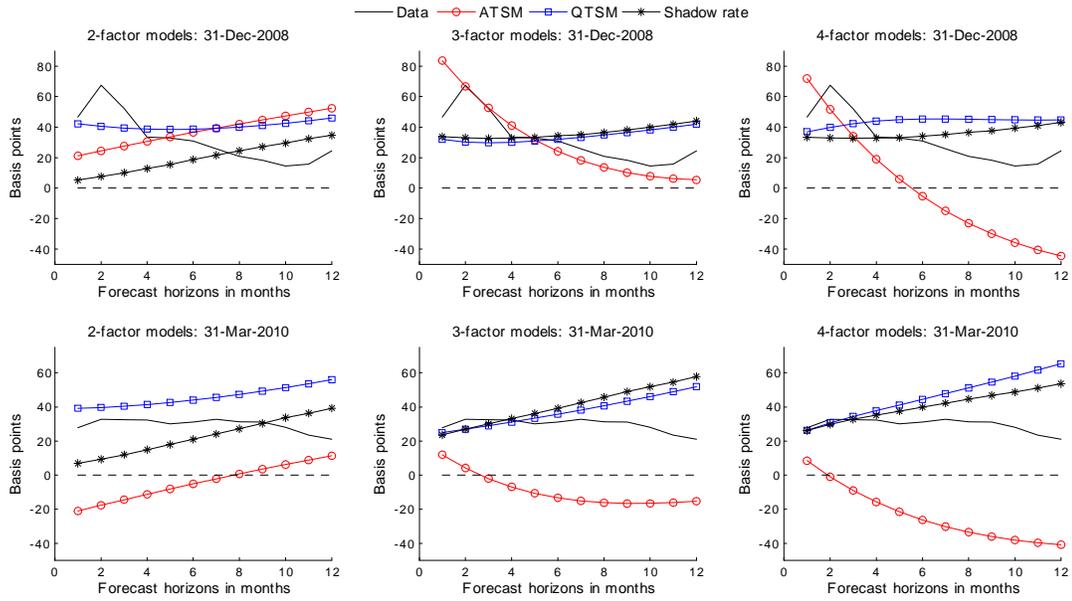


Figure 12: ATSM: Fraction of forecast distribution below zero

The charts report the fraction of the forecast distribution for the 0.5-year bond yield which are below zero. All forecast distributions are computed using recursively estimated parameters, starting in 1961 or 1990.

