AARHUS
UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

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

Martin M. Andreasen and Andrew Meldrum

CREATES Research Paper 2014-47

Department of Economics and Business
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

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

November 26, 2014

Abstract

This paper studies whether dynamic term structure models for US nominal bond yields should enforce the zero lower bound by a quadratic policy rate or a shadow rate specification. We address the question by estimating quadratic term structure models (QTSMs) and shadow rate models with at most four pricing factors 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 models. We also find that the shadow rate models perform better than the QTSMs when forecasting bond yields out of sample.

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

JEL: C10, C50, G12.

*We thank Glenn Rudebusch, Don Kim, Gregory R. Duffee, Jean-Paul Renne, 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, the workshop on "Staying at zero with affine processes" at the ECB, July 2014, and the conference "Understanding the yield curve: what has changed with the crisis?" at the ECB, September 2014, are also much appreciated. M. Andreasen greatly acknowledges financial support from the Danish e-Infrastructure Coorporation (DeIC). 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 Steet, 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. That is, should dynamic term structure models (DTSMs) for US bond yields enforce the ZLB by a quadratic policy rate or a shadow rate specification?

The aim of 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.[1] Following Dai & Singleton (2002), the performance of DTSMs is commonly evaluated by their ability to match moments from ordinary and risk-adjusted Campbell-Shiller regressions (the so-called LPY tests), as they capture key features of the physical and risk-neutral distributions of bond yields and hence the implied term premia. However, none of the ATSMs satisfying the LPY tests in Dai & Singleton (2002) enforce the ZLB, and it is therefore unclear if DTSMs jointly can enforce the ZLB and match term premia.[2] Special attention is therefore devoted to the LPY tests when comparing the performance of QTSMs and shadow rate models.

We estimate all the DTSMs considered by the sequential regression (SR) approach of Andreasen & Christensen (Forthcoming), where latent pricing factors and model parameters are obtained from

[1] Kim & Singleton (2012) compare the in-sample fit of QTSMs and shadow rate models with two pricing factors on Japanese bond yields. Apart from the different data set, our work differs from theirs by considering models with three and four pricing factors and by comparing the models' performance when forecasting bond yields out of sample.

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

regressions. The SR approach applies to linear and non-linear models and gives consistent and asymptotically normal estimates under weaker restrictions than typically imposed for likelihood-based inference. For instance, the SR approach allows measurement errors in bond yields to display heteroskedasticity and correlation in both the cross-section and time series dimension. Building on the work of Andreasen & Christensen (Forthcoming), we improve the finite sample properties of the SR approach along two dimension. First, a bias-adjustment is introduced when estimating the physical dynamics of the pricing factors. This extension allows us to explore how small-sample biases affect QTSMs and shadow rate models, which is an unaddressed issue in the literature. Second, a residual-based bootstrap is suggested for the risk-neutral parameters as a refinement to the asymptotic distribution in finite samples.

Apart from these robust econometric properties the SR approach is also attractive from a finance perspective, because 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. Another advantage of the SR approach is its computational simplicity, which allows us to estimate QTSMs and shadow rate models with four pricing factors, whereas all existing work restrict focus to models with at most three factors.

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

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 or shadow rate model gives largely the same in-sample fit of US bond yields, with both models clearly outperforming the Gaussian ATSM. Second, the

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

three- and four-factor QTSMs generally struggle to match loadings from ordinary Campbell-Shiller regressions, whereas these moments are largely matched by the shadow rate models. The shadow rate models are also more successful at reproducing the loadings from risk-adjusted Campbell-Shiller regressions than the QTSMs, although the latter performs well in the long sample. Third, the fall in conditional volatility of most bond yields when reaching the ZLB is nicely captured by the QTSM and the shadow rate model, although both models struggle to generate the increase in volatility just before reaching the ZLB. Fourth, in an extensive out-of-sample forecasting study from January 2005 to December 2013, we find that the shadow rate model generally performs better than the QTSM, and that models accounting for the ZLB outperform 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 *improve* when adding a fourth pricing factors whereas the opposite holds for the QTSM. Somewhat surprising, starting the recursive estimation in 1961 instead of 1990 generally gives more accurate forecasts of bond yields. Any finite sample bias in the physical dynamics for the pricing factors is unlikely to explain this finding as the bias-adjustment is applied in both samples. Instead, the better forecasts is most likely explained by the higher persistence in the pricing factors within the long sample, which we expect improve forecasts given the strong performance of the random walk. Finally, we also study a so-called hybrid model where the ZLB is enforced by a shadow rate that is an unrestricted quadratic function of the pricing factors. This hybrid model fits bond yields marginally better in-sample but struggles to provide a better fit of conditional volatility in bond yields than the other models considered. We also find that the hybrid model delivers less accurate forecasts of medium- and particularly long-term bond yields compared to the QTSM and the shadow rate model.

Overall, our findings suggest that the best way to enforce the ZLB for US bond yields is to adopt a shadow rate specification with the shadow rate being affine in the pricing factors, as opposed to letting the policy rate be a restricted quadratic function of the pricing factors.

The rest of the paper is organized as follows. Section 2 presents the DTSMs considered, and we describe how these models are estimated by the SR approach in Section 3. In-sample results are reported in Section 4 and the out-of-sample performance are presented in Section 5. We finally consider a hybrid version of the QTSM and the shadow rate model in Section 6 and study its performance. Concluding comments are provided in Section 7.

2 Dynamic term structure models

We start by describing the Gaussian ATSM which serves as our benchmark before presenting the QTSM and the shadow rate model. The pricing factors in all these models are assumed to be Gaussian under the risk-neutral and physical probability measures, implying that we adopt an affine specification for the market price of risk. We do not study the multivariate version of the model by Cox, Ingersoll & Ross (1985) with independent pricing factors or its extension by Dai & Singleton (2000) with correlated factors as in the $\mathbb{A}_m(m)$ models. Although these models also account for the ZLB, we do not include them in our analysis because they are generally unable to reproduce key moments of term premia, whereas these moments are nicely matched by the Gaussian ATSM (see Dai & Singleton (2002)).[4]

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.

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

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 Ang & Piazzesi (2003), Hördahl, Tristani & Vestin (2006), Rudebusch & Wu (2008), among others). 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.

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

where $\boldsymbol{\varepsilon}^{\mathbb{Q}}_{t+1} \sim \mathcal{NID}(\mathbf{0}, \mathbf{I})$. 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 in time period t of an j-period zero-coupon bond is $P_{t,j} = E^{\mathbb{Q}}_t \left[\exp\{-r_t\} P_{t+1,j-1}\right]$. Given the assumptions in (1) and (2), bond prices

[4] In 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.

are exponentially affine in the factors, i.e.

$$P_{t,j} = \exp\left\{A_j + \mathbf{B}_j' \mathbf{x}_t\right\}$$

for $j = 1, 2, ..., K$, where the recursive formulae for A_j and $\mathbf{B}_j$ are easily derived.

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

$$\mathbf{x}_{t+1} = \boldsymbol{\Phi}\boldsymbol{\mu} + \mathbf{f_0} + (\mathbf{I} - \boldsymbol{\Phi} + \mathbf{f}_1)\mathbf{x}_t + \boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1}^{\mathbb{P}}. \tag{3}$$

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

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

$$r_t = \alpha + \boldsymbol{\beta}'\mathbf{x}_t + \mathbf{x}_t'\boldsymbol{\Psi}\mathbf{x}_t, \tag{4}$$

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

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

Given the policy rate in (4), 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 bond prices, i.e.

$$\tilde{P}_{t,j} = \exp\left\{\tilde{A}_j + \tilde{\mathbf{B}}_j'\mathbf{x}_t + \mathbf{x}_t'\tilde{\mathbf{C}}_j\mathbf{x}_t\right\}$$

for $j = 1, 2, ..., K$, with the recursive formulae for $\tilde{A}_j$, $\tilde{\mathbf{B}}_j$, and $\tilde{\mathbf{C}}_j$ derived in Realdon (2006). Hence, the quadratic terms in (4) imply that all bond yields $y_{t,j} \equiv -\frac{1}{j}\log\tilde{P}_{t,j}$ are quadratic in the pricing factors and bond yields therefore display heteroskedasticity.

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 (3). 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) $\mathbf{\Psi}$ is symmetric with diagonal elements equal to one, ii) $\boldsymbol{\mu} \geq \mathbf{0}$, iii) $\boldsymbol{\beta} = \mathbf{0}$, iv) $\mathbf{\Phi}$ is diagonal, and v) $\mathbf{\Sigma}$ is triangular.[6] This normalization scheme implies unrestricted $\mathbb{P}$ dynamics for the pricing factors and that the ZLB may be enforced by letting $\alpha = 0$, provided $\mathbf{\Psi}$ is positive semi-definite.

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).[7] This shadow rate is unconstrained by the ZLB and may therefore attain negative values. Absent transaction and storage costs for money, Black (1995) observes that the nomi-

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

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

nal 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

$$r_t = \max\left(0, s\left(\mathbf{x}_t\right)\right),$$

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 $s\left(\mathbf{x}_t\right)$ and the considered factor dynamics.

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

$$s\left(\mathbf{x}_t\right) = \alpha + \boldsymbol{\beta}'\mathbf{x}_t,$$

but other specifications may also be considered as illustrated in Section 6. For the same reason, we also restrict focus to affine processes for the pricing factors under the $\mathbb{Q}$ and $\mathbb{P}$ measures as given in (2) and (3), but other specifications could be considered. 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 models with a Gaussian or a square-root process driving $s\left(\mathbf{x}_t\right)$ as shown by Gorovoi & Linetsky (2004). Given that one-factor models typically are considered too stylized, numerical approximations are therefore required when studying multi-factor shadow rate models. We apply the second-order approximation advocated by Priebsch (2013), which delivers a reasonably fast and highly accurate approximation to bond yields when $s\left(\mathbf{x}_t\right)$ is Gaussian under the $\mathbb{Q}$ measure.[8]

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 (2010)). This procedure is very time consuming for multi-factor DTSMs and therefore rarely attempted. A computational more feasible alternative is

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

to use a non-linear extension of the Kalman filter and a quasi-maximum likelihood (QML) approach, but its asymptotic properties are generally unknown.[9] We overcome these difficulties by using the SR approach, which is faster to implement than the QML approach as shown by Andreasen & Christensen (Forthcoming) and has known asymptotic properties.

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

3.1 The SR approach: The model class

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

$$y_{t,j} = g_j\left(\mathbf{x}_t; \boldsymbol{\theta}_1\right) + v_{t,j}, \tag{5}$$

where the subscript j index the maturity of bond yields. For any time period t, we require the measurement errors $\{v_{t,j}\}_{j=1}^{K}$ to have zero mean and a finite and positive definite co-variance matrix. Apart from the technical regularity conditions in Andreasen & Christensen (Forthcoming), no further assumptions are imposed on $v_{t,j}$, which may display heteroskedasticity and correlation in both the cross-section and time series dimension.

The functional relationship between the pricing factors and bond yields in (5) 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 (5), whereas $\boldsymbol{\theta}_{12}$ may be obtained from (5) and the factor dynamics under the $\mathbb{P}$ measure. For the Gaussian ATSM, the g-function is linear in the pricing factors, i.e. $g_j^{ATSM}\left(\mathbf{x}_t; \boldsymbol{\theta}_1^{ATSM}\right) \equiv -\frac{1}{j}\left(A_j + \mathbf{B}_j'\mathbf{x}_t\right)$, and we have $\boldsymbol{\theta}_{11}^{ATSM} \equiv \begin{bmatrix} \alpha & diag(\boldsymbol{\Phi})' \end{bmatrix}'$ with $\boldsymbol{\theta}_{12}^{ATSM} \equiv \begin{bmatrix} vech\left(\boldsymbol{\Sigma}\right)' \end{bmatrix}'$. The QTSM induces a slightly more complicated expression for bond yields because $g_j^{QTSM}\left(\mathbf{x}_t; \boldsymbol{\theta}_1^{QTSM}\right) \equiv -\frac{1}{j}\left(\tilde{A}_j + \tilde{\mathbf{B}}_j'\mathbf{x}_t + \mathbf{x}_t'\tilde{\mathbf{C}}_j\mathbf{x}_t\right)$, and for this model $\boldsymbol{\theta}_{11}^{QTSM} \equiv$

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

$\begin{bmatrix} \left(\boldsymbol{\theta}_1^{ATSM}\right)' & \boldsymbol{\mu}' & vech\left(\mathbf{A}\right)' \end{bmatrix}'$ with $\boldsymbol{\theta}_{12}^{QTSM} = \boldsymbol{\theta}_{12}^{ATSM}$. In the shadow rate model, $g_j^{SH}\left(\mathbf{x}_t; \boldsymbol{\theta}_1^{SH}\right)$ is an unknown non-linear mapping from the pricing factors to bond yields with $\boldsymbol{\theta}_1^{SH} = \boldsymbol{\theta}_1^{ATSM}$.

The SR approach allows the pricing factors under the $\mathbb{P}$ measure to evolve according to a first-order Markov process of the form $\mathbf{x}_{t+1} = \mathbf{h}\left(\mathbf{x}_t, \boldsymbol{\epsilon}_{t+1}^{\mathbb{P}}; \boldsymbol{\theta}_{11}, \boldsymbol{\theta}_2\right)$. 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. All the DTSMs considered have a linear and unrestricted transition function which we represent by

$$\mathbf{x}_{t+1} = \mathbf{h}_0 + \mathbf{h}_{\mathbf{x}}\mathbf{x}_t + \boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1}^{\mathbb{P}}, \tag{6}$$

where $\mathbf{h}_0 \equiv \boldsymbol{\Phi}\boldsymbol{\mu} + \mathbf{f_0}$, $\mathbf{h}_{\mathbf{x}} \equiv \mathbf{I} - \boldsymbol{\Phi} + \mathbf{f}_1$, and $\boldsymbol{\varepsilon}_{t+1}^{\mathbb{P}} \sim \mathcal{NID}\left(\mathbf{0}, \mathbf{I}\right)$. Hence, given the parametrization of the $\mathbf{h}$-function in (6), we have $\boldsymbol{\theta}_{22} \equiv \begin{bmatrix} \mathbf{h}_0' & vec\left(\mathbf{h}_{\mathbf{x}}\right)' \end{bmatrix}'$ 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.

3.2 The SR approach: Step 1

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

$$\hat{\mathbf{x}}_t\left(\boldsymbol{\theta}_1\right) = \arg\min_{\mathbf{x}_t \in \mathcal{X}_t} Q_t = \frac{1}{2n_{y,t}} \sum_{j=1}^{n_{y,t}} \left(y_{t,j} - g_j\left(\mathbf{x}_t; \boldsymbol{\theta}_1\right)\right)^2 \tag{7}$$

for $t = 1, 2, ..., T$, where $n_{y,t}$ refers to the number of bond yields in time period t. The estimated factors are denoted $\{\hat{\mathbf{x}}_{2,t}\left(\boldsymbol{\theta}_1\right)\}_{t=1}^T$ because they are computed for a given $\boldsymbol{\theta}_1$. These regressions have a closed-form solution for the Gaussian ATSM as g_j^{ATSM} is linear in $\mathbf{x}_t$. For the QTSM and the shadow rate model, the regressions in (7) are non-linear and solved using the Levenberg-Marquardt optimizer with the pricing factors from the previous time period $\hat{\mathbf{x}}_{t-1}\left(\boldsymbol{\theta}_1\right)$ serving as ideal starting values for $t = 2, 3, ..., T$.[10] Although these non-linear regressions converge within a few iterations, some care is needed for the QTSM not to end up in a local optimum. This is illustrated in Figure 1 where we plot the objective functions at four selected dates when filtering out the the pricing factors. To facilitate

[10] The main input for Levenberg-Marquardt optimizer is the jacobian $\partial \mathbf{g}\left(\mathbf{x}_t; \boldsymbol{\theta}_1\right) / \partial \mathbf{x}_t'$ 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 the second-order approximation by Priebsch (2013) is otherwise 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 procedure.

the plotting, we focus on a one-factor QTSM but analogue results apply with multiple factors. When only one bond yield is used (the thin black line), the regressions are not identified as there are two solutions with exactly the same fit. Identification is obtained by including more observations and with 25 bond yields (the wide black line) the negative solution is clearly only a local optimum.

The model parameters $\boldsymbol{\theta}_1$ are obtained by pooling all squared residuals from (7) and minimizing their sum with respect to $\boldsymbol{\theta}_1$, i.e.

$$\hat{\boldsymbol{\theta}}_1^{step1} = \arg\min_{\boldsymbol{\theta}_1 \in \Theta_1} Q_{1:T}^{step1} = \frac{1}{2N} \sum_{t=1}^{T} \sum_{j=1}^{n_{y,t}} \left(y_{t,j} - g_j \left(\hat{\mathbf{x}}_t \left(\boldsymbol{\theta}_1 \right) ; \boldsymbol{\theta}_1 \right) \right)^2, \tag{8}$$

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

$$\sqrt{N} \left(\hat{\boldsymbol{\theta}}_1^{step1} - \boldsymbol{\theta}_1^o \right) \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \left(\mathbf{A}_o^{\boldsymbol{\theta}_1} \right)^{-1} \mathbf{B}_o^{\boldsymbol{\theta}_1} \left(\mathbf{A}_o^{\boldsymbol{\theta}_1} \right)^{-1} \right), \tag{9}$$

where "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 $N \to \infty$. The expected value of the average Hessian matrix $\mathbf{A}_o^{\boldsymbol{\theta}_1}$ may be estimated consistently by

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

where

$$\boldsymbol{\Psi}_{t,j}^{\boldsymbol{\theta}_1} \left(\boldsymbol{\theta}_1 \right) \equiv \frac{\partial \hat{\mathbf{x}}_t' \left(\boldsymbol{\theta}_1 \right)}{\partial \boldsymbol{\theta}_1} \frac{\partial g_j \left(\hat{\mathbf{x}}_t \left(\boldsymbol{\theta}_1 \right) ; \boldsymbol{\theta}_1 \right)}{\partial \mathbf{x}_t \left(\boldsymbol{\theta}_1 \right)} + \frac{\partial g_j \left(\hat{\mathbf{x}}_t \left(\boldsymbol{\theta}_1 \right) ; \boldsymbol{\theta}_1 \right)}{\partial \boldsymbol{\theta}_1}$$

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

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

where

$$\hat{\sigma}_t^2 = \frac{1}{n_{y,t} - n_{x_2}} \sum_{j=1}^{n_{y,t}} \hat{v}_{t,j}^2 \quad \text{for } t = 1, 2, ..., T$$

and $\hat{v}_{t,j} = y_{t,j} - g_j\left(\hat{\mathbf{x}}_t; \hat{\boldsymbol{\theta}}_1^{step1}\right)$. 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.

3.3 The SR approach: Step 2

We estimate $\boldsymbol{\theta}_2$ in (6) 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 (Forthcoming), we modify the standard moment conditions for VAR models to account for uncertainty in $\{\hat{\mathbf{x}}_t\}_{t=1}^T$ and consider

$$\mathbf{q}_T\left(\boldsymbol{\theta}_2\right) \equiv \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbf{q}_t\left(\boldsymbol{\theta}_2\right) = \mathbf{0}, \tag{10}$$

where

$$\mathbf{q}_t\left(\boldsymbol{\theta}_2\right) \equiv \begin{bmatrix} \hat{\mathbf{w}}_{t+1} \\ vec\left(\hat{\mathbf{w}}_{t+1}\hat{\mathbf{x}}_t' - Cov\left(\mathbf{u}_{t+1}, \mathbf{u}_t\right) + \mathbf{h}_{\mathbf{x}} Var\left(\mathbf{u}_t\right)\right) \\ vech\left(\begin{array}{c} \hat{\mathbf{w}}_{t+1}\hat{\mathbf{w}}_{t+1}' - Var\left(\hat{\mathbf{w}}_{t+1}\right) - Var\left(\mathbf{u}_t\right) - \mathbf{h}_{\mathbf{x}} Var\left(\mathbf{u}_t\right)\mathbf{h}_{\mathbf{x}}' \\ + Cov\left(\mathbf{u}_{t+1}, \mathbf{u}_t\right)\mathbf{h}_{\mathbf{x}}' + \mathbf{h}_{\mathbf{x}} Cov\left(\mathbf{u}_t, \mathbf{u}_{t+1}\right) \end{array} \right) \end{bmatrix}$$

and $\hat{\mathbf{w}}_{t+1} \equiv \boldsymbol{\Sigma}\hat{\boldsymbol{\varepsilon}}_{t+1}^{\mathbb{P}} \equiv \hat{\mathbf{x}}_{t+1} - \mathbf{h}_0 - \mathbf{h}_{\mathbf{x}}\hat{\mathbf{x}}_t$. Note that $\hat{\boldsymbol{\varepsilon}}_{t+1}^{\mathbb{P}}$ refers to the residuals using the true values of $\mathbf{h}_0$ and $\mathbf{h}_{\mathbf{x}}$ but the estimated pricing factors $\hat{\mathbf{x}}_t$. Consistent estimators of $Var\left(\mathbf{u}_t\right)$, $Cov\left(\mathbf{u}_{t+1}, \mathbf{u}_t\right)$, and $Cov\left(\mathbf{u}_t, \mathbf{u}_{t+1}\right)$ are provided in Andreasen & Christensen (Forthcoming) 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 (10) 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

$$\begin{aligned}\left[\begin{array}{cc}\hat{\mathbf{h}}_{\mathbf{x}}^{step2} & \hat{\mathbf{h}}_{0}^{step2}\end{array}\right] &= \left(\sum_{t=1}^{T-1}\left[\begin{array}{cc}\hat{\mathbf{x}}_{t+1}\hat{\mathbf{x}}_{t}'-\widehat{Cov}\left(\mathbf{u}_{t+1},\mathbf{u}_{t}\right) & \hat{\mathbf{x}}_{t+1}\end{array}\right]\right) \\ &\times\left(\sum_{t=1}^{T-1}\left[\begin{array}{cc}\hat{\mathbf{x}}_{t}\hat{\mathbf{x}}_{t}'-\widehat{Var}\left(\mathbf{u}_{t}\right) & \hat{\mathbf{x}}_{t}\\ \hat{\mathbf{x}}_{t}' & 1\end{array}\right]\right)^{-1},\end{aligned} \tag{11}$$

and by letting

$$\begin{aligned}\widehat{Var}\left(\hat{\mathbf{w}}_{t+1}\right)^{step2} &= \frac{1}{T-1-n_x-1}\sum_{t=1}^{T-1}(\hat{\widehat{\mathbf{w}}}_{t+1}\left(\hat{\widehat{\mathbf{w}}}_{t+1}\right)' \\ &-\frac{1}{T-1}\sum_{t=1}^{T-1}\left(\widehat{Var}\left(\mathbf{u}_{t}\right)+\hat{\mathbf{h}}_{\mathbf{x}}\widehat{Var}\left(\mathbf{u}_{t}\right)\hat{\mathbf{h}}_{\mathbf{x}}'\right) \\ &+\frac{1}{T-1}\sum_{t=1}^{T-1}\left(\widehat{Cov}\left(\mathbf{u}_{t+1},\mathbf{u}_{t}\right)\hat{\mathbf{h}}_{\mathbf{x}}'+\hat{\mathbf{h}}_{\mathbf{x}}\widehat{Cov}\left(\mathbf{u}_{t},\mathbf{u}_{t+1}\right)\right),\end{aligned} \tag{12}$$

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

$$\sqrt{T}\left(\boldsymbol{\theta}_{2}^{step2}-\boldsymbol{\theta}_{2}^{o}\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\left(\mathbf{R}_{o}^{\boldsymbol{\theta}_2}\mathbf{S}_{o}^{-1}\left(\mathbf{R}_{o}^{\boldsymbol{\theta}_2}\right)'\right)^{-1}\right) \tag{13}$$

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 $\mathbf{S}_o\equiv\sum_{\nu=-\infty}^{\infty}E\left[\mathbf{q}_t\left(\boldsymbol{\theta}_2^o\right)\mathbf{q}_{t-\nu}\left(\boldsymbol{\theta}_2^o\right)'\right]$. We estimate $\mathbf{R}_{o}^{\boldsymbol{\theta}_2}$ using numerical differentiation and $\mathbf{S}_o$ by the Newey-West estimator.

3.3.1 The SR approach: Bias-adjusting step 2

For persistent processes it is well-known that the standard moment conditions for estimating VAR models (extended in (10) to account for generated regressors) give biased estimates of $\hat{\mathbf{h}}_{\mathbf{x}}$ in finite samples where the eigenvalues of $\hat{\mathbf{h}}_{\mathbf{x}}$ have a downward bias (see for instance Yamamoto & Kunitomo (1984)). Bauer, Rudebusch & Wu (2012) show that this bias may be substantial for Gaussian ATSMs and have sizeable effects on model-implied term premium. A popular method to reduce 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 $\hat{\mathbf{h}}_{\mathbf{x}}^{adj}=\hat{\mathbf{h}}_{\mathbf{x}}-\left(\bar{\mathbf{h}}_{\mathbf{x}}-\hat{\mathbf{h}}_{\mathbf{x}}\right)$. We cannot directly apply the standard bootstrap for VAR models in our setting due to the presence of

generated regressors, and we therefore generalize it in Appendix A to account for this feature and hence make it applicable for the SR approach.

Given the persistent nature of the pricing factors in DTSMs, the bias-adjusted estimate $\hat{\mathbf{h}}_{\mathbf{x}}^{adj}$ is unfortunately often pushed into the non-stationary region. To induce stationarity, Kilian (1998) therefore suggests 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

$$\hat{\mathbf{h}}_{\mathbf{x}}^{adj,B}(\delta) = \hat{\mathbf{h}}_{\mathbf{x}} - \delta \times \left(\bar{\mathbf{h}}_{\mathbf{x}} - \hat{\mathbf{h}}_{\mathbf{x}}\right) \tag{14}$$

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 consider a sequence of grids for δ_i constructed such that the length of the largest eigenvalue of $\hat{\mathbf{h}}_{\mathbf{x}}^{adj,B}(\delta_i)$ converges to one. This implies that the process for $\mathbf{x}_t$ converges to a non-stationary VAR model with infinite unconditional second moments. In other words, 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 procedure 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 (14) 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 $i = 1, 2, ..., n_x$. The latter estimate is computed for a given value of δ and is therefore denoted $\sigma_{i,VAR}^2(\delta)$. More formally, we let

$$\hat{\delta} = \arg \min_{\delta \in [\delta_{lower}, 1]} \sum_{i=1}^{n_x} \left(\frac{\sigma_{i,VAR}^2(\delta) - \hat{\sigma}_{i,Data}^2}{\hat{\sigma}_{i,Data}^2} \right)^2. \tag{15}$$

Monte Carlo evidence in Table 1 suggests that down-scaling the bias *and* the initial estimate of $\mathbf{h}_{\mathbf{x}}$ gives slightly lower bias than only down-scaling the estimated bias when δ is determined using (15). The better performance is related to $\hat{\delta}$, which tends to be larger when also down-scaling $\hat{\mathbf{h}}_{\mathbf{x}}$, implying that more of the bias-adjustment is preserved. For instance, when using the estimated factor dynamics from the Gaussian ATSM in the long sample and $T = 250$ in our Monte Carlo study, the average of $\hat{\delta}$ across all draws is 0.9921 when down-scaling the bias and $\hat{\mathbf{h}}_{\mathbf{x}}$, whereas the average of $\hat{\delta}$ falls to 0.6950 when only down-scaling the bias. Hence, we prefer the adjustment

$$\hat{\mathbf{h}}_{\mathbf{x}}^{adj,*}(\delta) = \delta \times \left(\hat{\mathbf{h}}_{\mathbf{x}} - \left(\bar{\mathbf{h}}_{\mathbf{x}} - \hat{\mathbf{h}}_{\mathbf{x}}\right)\right) \tag{16}$$

and determine δ using (15). 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 $\mathbf{x}_t$ compared to Killian's method. Another advantage of considering (16) is that it always ensures stationarity of VAR models, contrary to the specification in (14). Our method to induce stationarity is summarized in Appendix B, which also describes how to account for measurement errors in $\mathbf{x}_t$ as implied by the SR approach. Unless stated otherwise, we use the bias-adjustment in (16) throughout the paper.

< Table 1 about here >

3.4 The SR approach: Step 3

The elements in $\boldsymbol{\Sigma}$ appear in $\boldsymbol{\theta}_{12}$ which are estimated in both the first and second estimation step. Andreasen & Christensen (Forthcoming) suggest a linear combination of the two estimators, i.e.

$$\hat{\boldsymbol{\theta}}_{12}^{step3} = \boldsymbol{\Lambda}\hat{\boldsymbol{\theta}}_{12}^{step1} + (\mathbf{I} - \boldsymbol{\Lambda})\,\hat{\boldsymbol{\theta}}_{12}^{step2},$$

where $\boldsymbol{\Lambda}$ is determined to minimize the variance of $\hat{\boldsymbol{\theta}}_{12}^{step3}$ and hence the efficiency loss from sequential identification. We generally find that $\hat{\boldsymbol{\Sigma}}^{step1}$ is estimated very inaccurately compared to $\hat{\boldsymbol{\Sigma}}^{step2}$, meaning that the time series estimate $\hat{\boldsymbol{\Sigma}}^{step2}$ cannot be improved by adding cross-section information from $\hat{\boldsymbol{\Sigma}}^{step1}$, i.e. $\boldsymbol{\Lambda} \approx \mathbf{0}$.[11] Hence, the adopted estimate of $\boldsymbol{\Sigma}$ after the first two steps is simply $\hat{\boldsymbol{\Sigma}}^{step2}$.

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

Based on the more accurate estimate of $\boldsymbol{\Sigma}$, it is natural to re-estimate $\boldsymbol{\theta}_{11}$ when conditioned on $\hat{\boldsymbol{\Sigma}}^{step2}$. That is,

$$\hat{\boldsymbol{\theta}}_{11}^{step3} = \arg\min_{\boldsymbol{\theta}_{11}\in\Theta_{11}} Q_{1:T}^{step3} = \frac{1}{2N}\sum_{t=1}^{T}\sum_{j=1}^{n_{y,t}}\left(y_{t,j} - g_j\left(\hat{\mathbf{x}}_t\left(\boldsymbol{\theta}_{11}, \hat{\boldsymbol{\Sigma}}^{step2}\right); \boldsymbol{\theta}_{11}, \hat{\boldsymbol{\Sigma}}^{step2}\right)\right)^2. \tag{17}$$

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

$$\widehat{Var}\left(\hat{\boldsymbol{\theta}}_{11}^{step3}\right) = \frac{\hat{\mathbf{V}}_{\boldsymbol{\theta}_{11}}^{step3}\left(\hat{\boldsymbol{\Sigma}}^{step2}\right)}{N} + \hat{\mathbf{K}}\widehat{Var}\left(\hat{\boldsymbol{\Sigma}}^{step2}\right)\hat{\mathbf{K}}'. \tag{18}$$

The first term $\hat{\mathbf{V}}_{\boldsymbol{\theta}_{11}}^{step3}\left(\hat{\boldsymbol{\Sigma}}^{step2}\right)/N$ is given by (9) when used on the subset of $\boldsymbol{\theta}_1$ corresponding to $\boldsymbol{\theta}_{11}$. The second term in (18) corrects for estimation uncertainty in $\hat{\boldsymbol{\Sigma}}^{step2}$ with $\mathbf{K} \equiv \partial\hat{\boldsymbol{\theta}}_{11}^{step3}(\boldsymbol{\Sigma})/\partial vech(\boldsymbol{\Sigma})'$. We estimate $\mathbf{K}$ as suggested in Andreasen & Christensen (Forthcoming) and refer to their paper for further details. Given the estimated pricing factors $\left\{\hat{\mathbf{x}}_t\left(\boldsymbol{\theta}_{11}^{step3}, \hat{\boldsymbol{\Sigma}}^{step2}\right)\right\}_{t=1}^{T}$ from (17), we finally update our estimates of $\boldsymbol{\theta}_2$ using (11) and (12).

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

Although the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{11}^{step3}$ performs well in finite samples with just 15 to 25 bond yields according to Andreasen & Christensen (Forthcoming), one may nevertheless be hesitant to use it for inference given the small number of observations in the cross-section dimension. To address this potential concern, we next describe another approximation to the distribution of $\hat{\boldsymbol{\theta}}_{11}^{step3}$ by means of a bootstrap. To present the bootstrap, recall that the SR approach considers the non-linear regression in (5), which we re-write in stacked form as

$$\mathbf{Y}_j = \mathbf{G}_j\left(\mathbf{x}_{t:T}; \boldsymbol{\theta}_1\right) + \mathbf{v}_j, \tag{19}$$

where

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

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

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

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

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

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

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

Although the residual-based bootstrap is well-known, it is useful to highlight a few details specific to the SR approach. First, in the absence of an intercept in (19), it is necessary to re-center the residuals in Step A to ensure they have zero mean and hence that the bootstrap samples in Step C are from a model correctly specified for the conditional mean. Second, we follow Bühlmann (1997) and use an AR(p) model to account for cross-correlation in the residuals when ordered by maturity.[13] Third, by resampling the entire vector of $\hat{\boldsymbol{\epsilon}}_j$ in Step C, the variance and co-variance structure in the residuals is preserved (see MacKinnon (2009)), meaning that the bootstrap accounts for time-variation in the variances of $v_{t,j}$ and auto-correlation in $v_{t,j}$. Fourth, by drawing from $\hat{\boldsymbol{\theta}}_{12}^{step2,(b)}$ in Step D we condition on the distribution of $\hat{\boldsymbol{\theta}}_{12}^{step2}$ and incorporate this source of uncertainty in the bootstrap. The draws

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

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

for $\hat{\boldsymbol{\theta}}_{12}^{step2,(b)}$ may be obtained from the bootstrap in Step 2 of the SR approach or from the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{12}^{step2}$ in (13).

Finally, when estimating the AR(p) model in Step B, Appendix C shows that OLS is biased in finite samples due to estimation error in the latent factors. It is, however, straightforward to correct for this bias by running a preliminary bootstrap with $B_1 < B$ draws, where we in Step D use OLS to estimate the AR(p) model $\left\{\hat{\phi}_{j,OLS}^{(b)}\right\}_{j=1}^{p}$ on the residuals in the bootstrap sample, i.e. on $\hat{\mathbf{v}}_j^{*,(b)} = \mathbf{Y}_j^{*,(b)} - \mathbf{G}_j\left(\hat{\mathbf{x}}_{1:T}^{step3,(b)}; \hat{\boldsymbol{\theta}}_{11}^{step3,(b)}\right)$ for $j = 1, 2, ..., K$. The bias-adjusted estimates are then give by $\hat{\phi}_i = 2\hat{\phi}_{i,OLS} - \bar{\phi}_i$ for $i = 1, 2, ..., p$, where $\bar{\phi}_i = \frac{1}{B_1}\sum_{b=1}^{B_1}\hat{\phi}_{i,OLS}^{(b)}$ and $\hat{\phi}_{i,OLS}$ denotes the initial OLS estimate in Step B.[14]

We briefly explore the finite sample properties of the bootstrap in a Monte Carlo study, using a one-factor Gaussian ATSM to reduce the computational burden (see Appendix D for the remaining details of the Monte Carlo study). Our results in Table 2 may be summarized as follows. The asymptotic distribution of $\boldsymbol{\theta}_{11} \equiv \begin{bmatrix} \alpha & \Phi_{11} \end{bmatrix}$ serves as a useful approximation in finite samples with standard errors and rejection probabilities in the ball park of the desired values. We emphasize that these results hold even when measurement errors are auto-correlated (Case II), display time-varying heteroskedasticity (Case III), are cross-sectionally correlated (Case IV), or when the three features are combined (Case V). The bootstrap generally provides an refinement to asymptotic inference, as bootstrapped standard errors and rejection probabilities in most cases outperform those from the asymptotic distribution. The largest improvement appears with cross-correlation in the measurement errors (i.e. Case IV and V), where the bootstrap corrects the positive bias in the asymptotic standard errors which otherwise generate too low rejection probabilities. The satisfying performance of the bootstrap in these two cases is closely related to the estimated AR(1) model for cross-correlation, where the bias-adjustment returns nearly unbiased estimates of $\phi_1 = 0.40$ with $\hat{\phi}_1^{Case\ IV} = 0.395$ and $\hat{\phi}_1^{Case\ V} = 0.396$ using just 100 draws in the preliminary bootstrap. The corresponding averages of the unadjusted OLS estimates in the Monte Carlo study are $\hat{\phi}_{1,OLS}^{Case\ IV} = 0.336$ and $\hat{\phi}_{1,OLS}^{Case\ V} = 0.335$, respectively.

[14] Although less likely, if the bias-adjusted estimates in the AR(p) model violate the stationarity requirement, it may be imposed using the same principle as in Section 3.3.1, that is by down-scaling $\left\{\hat{\phi}_i\right\}_{i=1}^{p}$ by a constant δ which we determine by minimizing the distance between the unconditional variance in the AR(p) model and $\frac{1}{N-Tn_{x_2}}\sum_{t=1}^{T}\sum_{j=1}^{n_{y,t}}\hat{v}_{t,j}^2$.

< Table 2 about here >

4 Empirical results: In-sample performance

This section estimates the Gaussian ATSM, the QTSM, and the shadow rate model. Section 4.1 presents the data, and we discuss our estimates of models with three pricing factors 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 June 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 (Forthcoming) 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 to 3.0 year maturity range at maturities three months apart, whereas bond yields in the remaining segment of the 10-year term structure are included at maturities six months apart.[15] 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 to 10 year maturity range are not available before this date. We address this problem by explicitly accounting for missing observations in the SR approach.

As mentioned above, we test the performance of the DTSMs considered on a long and a short sample. The long sample is from June 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

[15] These bond yields are computed using the estimated parametric form for the yield curves in Gürkaynak et al. (2007).

ATSM, recall that bond yields are given by $y_{t,j} = -\frac{1}{j}\left(A_j + \mathbf{B}_j'\mathbf{x}_t\right)$ for $j = 1, 2, ..., K$. The well-known solution to the Gaussian ATSM with our normalization is

$$A_j = -\alpha + A_{j-1} + \frac{1}{2}\mathbf{B}_{j-1}'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{B}_{j-1} \approx -\alpha + A_{j-1}, \tag{20}$$

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

$$\mathbf{B}_j' = -\mathbf{1}' + \mathbf{B}_{j-1}'\left(\mathbf{I} - \boldsymbol{\Phi}\right). \tag{21}$$

Eq. (21) shows that all eigenvalues of $\boldsymbol{\Phi}$ must be distinct to estimate the latent factors by the regression filter. Moreover, given that $\boldsymbol{\Sigma}$ is badly identified from the cross-section dimension of bond yields due to (20), the ordering of the factors is therefore also badly identified.[16] That is, we obtain nearly identical values for the objective functions in the first and third step of the SR approach by changing the order of the eigenvalues in $\boldsymbol{\Phi}$. To eliminate this identification issue we therefore require that all eigenvalues of $\boldsymbol{\Phi}$ are strictly increasing.[17] A similar lack of identification is observed for the QTSM and the shadow rate model, and we therefore also require that the eigenvalues of $\boldsymbol{\Phi}$ are strictly increasing in these models.

The estimation results for the three-factor models in the long sample are reported in Table 3. The Gaussian ATSM displays the usual properties with stationary and highly persistent factors under the $\mathbb{Q}$ and $\mathbb{P}$ measures as all diagonal elements of $\hat{\Phi}$ are positive and the largest eigenvalue of $\hat{\mathbf{h}}_{\mathbf{x}}$ is 0.9914. Similar properties hold for the pricing factors in the QTSM, where $\hat{\boldsymbol{\Psi}}$ enforces the ZLB by having eigenvalues of 0.0000, 0.0134, and 2.9866. The requirement of non-negative eigenvalues in $\hat{\boldsymbol{\Psi}}$ is equivalent to imposing $\hat{A}_{21}^2 \leq 1$ and $\hat{A}_{31}^2 + \left(1 - \hat{A}_{21}^2\right)\hat{A}_{32}^2 \leq 1$, implying that the absolute value of $\hat{A}_{12}$ and $\hat{A}_{31}$ cannot exceed one. The presence of a zero eigenvalue means that we are at the boundary (as $\hat{A}_{31}^2 + \left(1 - \hat{A}_{21}^2\right)\hat{A}_{32}^2 \approx 1$), and the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{11}^{step3}$ is therefore unlikely to perform well in this case. The provided bootstrapped confidence intervals should be more reliable, although we acknowledge that the bootstrap may also be inaccurate in this particular case, because the bootstrap method is known to be inconsistent when parameters are at the boundary of their domain (see Andrews (2000)). Subject to this qualification, we find that the 95% confidence intervals in Table 3 for $\hat{A}_{23}$ and

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

[17] This 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.

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

The estimates in the shadow rate model are very similar to those in the Gaussian ATSM, and in both models are the asymptotical standard errors close to those from the bootstrap (provided in squared parenthesis in Table 3). Despite the strong similarities between the two models we do observe some differences in the estimates, particularly for $\mathbf{h}_{\mathbf{x}}$. 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 3 about here >

Table 4 reveals that the pricing factors for all models in the short sample are slightly less persistent than in the long sample when measured by the largest eigenvalue of $\hat{\mathbf{h}}_{\mathbf{x}}$. In the QTSM, the estimates of $\boldsymbol{\Psi}$ imply eigenvalues of 0.0014, 0.0287, and 2.9699, meaning that the short rate is primarily controlled by one pricing factor as in the long sample, given our normalization with $\boldsymbol{\beta} = \mathbf{0}$. We also see that $\hat{\mathbf{h}}_{\mathbf{x}}$ and $\hat{\boldsymbol{\Sigma}}$ for the Gaussian ATSM differ substantially from the corresponding estimates in the shadow rate model. 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 4 about here >

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 are reported as $\tilde{Q}_{1:T}^{step1} \equiv 100\sqrt{2 \times Q_{1:T}^{step1}}$ so they denote standard deviations of all residuals in the sample. The top part of Table 5 to the left shows that $\tilde{Q}_{1:T}^{step1} = 2.90$ for the Gaussian ATSM, meaning that average pricing errors in this model is 2.90 basis points. Accounting for the ZLB by the shadow rate model improves 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 long sample with $\tilde{Q}_{1:T}^{step1} = 2.70$. This is in line with our 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 5 also shows that all three-factor models provide a better fit to bond yields. The shadow rate model with $\tilde{Q}_{1:T}^{step1} = 1.69$ clearly outperforms the Gaussian ATSM having $\tilde{Q}_{1:T}^{step1} = 1.81$, and the best in-sample fit is once again obtained by the QTSM where $\tilde{Q}_{1:T}^{step1} = 1.61$.

The right part of Table 5 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 $\mathbf{\Sigma}$ is estimated from the time series dimension instead of the cross-section dimension as in the first step. For all models and in both samples, $\tilde{Q}_{1:T}^{step3}$ is only marginally larger than $\tilde{Q}_{1:T}^{step1}$, meaning that the in-sample fit of bond yields is almost unaffected by the alternative estimator of $\mathbf{\Sigma}$. It is therefore reasonable to believe that the dependence on the $\mathbb{P}$ dynamics through $\mathbf{\Sigma}$ 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 $\boldsymbol{\theta}_2$, partly because $\mathbf{\Sigma}$ is badly identified from the cross-section dimension of bond yields, and partly because the bias-adjustment in $\hat{\mathbf{\Sigma}}^{step2}$ is small.

< Table 5 about here >

A more careful inspection of the in-sample fit for the long sample is provided in Figure 2, 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. To study the performance of these models in greater detail when bond yields are close to the ZLB, the second chart in Figure 2 displays the recursively computed objective functions from January 2005 and onwards. 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 December 2008, where the policy rate hits the ZLB.[18] The corresponding plot in Figure 3 shows the same pattern for the Gaussian ATSM when the sample starts in 1990.[19] This finding indicates that the three-factor Gaussian ATSM struggles to match bond yields close to the ZLB.

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

[19] For US data from 1985 to 2012, Christensen & Rudebusch (2013) also find that a shadow rate model with three pricing factors outperforms the three-factor Gaussian ATSM when measured by in-sample fit. Christensen & Rudebusch (2013) consider non-canonical models with factor loadings restricted to mimic the Nelson-Siegel specification as in Christensen, Diebold & Rudebusch (2011), whereas our models are fully flexible.

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

< Figures 2 and 3 about here >

An obvious way to improve the in-sample fit of these models is to include a fourth pricing factor.[20] The bottom part of Table 5 shows that including a fourth pricing factor roughly improves the in-sample fit by more than 50% for all models across both samples. Adrian, Crump & Moench (2013) also provide evidence for more than three pricing factors in the Gaussian ATSM, and our results suggest that the same applies for the QTSM and the shadow rate model. Importantly, with four pricing factors 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 2 and 3 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 all pricing errors now are within 2 basis points. In other words, a fourth pricing factor helps to match the short and long end of the yield curve in all models.

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

[20] The estimated model parameters for these four-factor models are available on request.

4.4 Matching key moments for bond yields

We next test the models' 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 also run the ordinary Campbell-Shiller regressions

$$y_{t+1,j-1} - y_{t,j} = \delta_j + \frac{\phi_j}{j-1}\left(y_{t,j} - r_t\right) + u_{t,j}, \tag{22}$$

where $u_{t,j} \sim \mathcal{IID}\left(0, Var\left(u_{t,j}\right)\right)$.[21] We then explore if the DTSMs can reproduce the empirical pattern in $\left\{\phi_j\right\}_{j=2}^{K}$ and hence capture key moments of the $\mathbb{P}$ dynamics for bond yields, also known as the LPY(i) test. Following Dai & Singleton (2002), a risk-adjusted version of the Campbell-Shiller regressions in (22) is given by

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

where $u_{t,j}^{\mathbb{Q}} \sim \mathcal{IID}\left(0, Var\left(u_{t,j}^{\mathbb{Q}}\right)\right)$, $c_{t,j} \equiv y_{t,j} - \frac{1}{j}\sum_{i=0}^{j-1} E_t\left[r_{t+i}\right]$ is the spot term premium, and $\theta_{t,j} \equiv f_{t,j} - E_t\left[r_{t+j}\right]$ is the forward term premium with $f_{t,j} \equiv -\log\left(P_{t,j+1}/P_{t,j}\right)$.[22] If the $\mathbb{Q}$ dynamics are correctly specified by the DTSMs (equivalently to a well-specified term premia and $\mathbb{P}$ dynamics), then $\phi_j^{\mathbb{Q}} = 1$ for $j = 2, 3, ..., K$. The ability of DTSMs to match these moments is the LPY(ii) test and studies whether the models 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 long sample is examined in Figure 4. To explore the impact of the bias-adjustment in $\boldsymbol{\theta}_2$, charts to the left report the model-implied moments using the unadjusted estimates of $\boldsymbol{\theta}_2$, whereas the adjustment is imposed in charts to the right. The first row in Figure 4 shows that all models underestimate the average level of the yield curve when $\boldsymbol{\theta}_2$ is not bias-adjusted, whereas these moments are perfectly matched when correcting for the bias in $\boldsymbol{\theta}_2$. 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

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

[22] As for (22), in practice we run the regressions $y_{t+m,j-m} - y_{t,j} - \left(c_{t+m,j-m} - c_{t,j-m}\right) + \frac{m}{j-m}\theta_{t,j-m} = \delta_j^{\mathbb{Q}} + \phi_j^{\mathbb{Q}} \frac{m}{j-m}\left(y_{t,j} - y_{t,m}\right) + u_{t,j}^{\mathbb{Q}}$ with $m = 6$, i.e. the regressions are done for biannual risk-adjusted excess returns. We compute these regressions using empirical bond yields and model-implied estimates of term premia obtained at $\{\hat{\mathbf{x}}_t\}_{t=1}^{T}$.

observe that only the Gaussian ATSM and the shadow rate model reproduce the downward sloping pattern in $\left\{\phi_j\right\}_{j=2}^{K}$, whereas the estimated QTSM cannot match this aspect of bond yields and hence pass the LPY(i) test. We acknowledge that the empirical loadings in the ordinary Campbell-Shiller regressions are likely to display some instability across subsamples (see Rudebusch & Tao (2007)), and the models are therefore not expected to match these loadings perfectly but only to capture their overall pattern. The LPY(ii) test does not suffer from the same instability issues as the desired regression loadings of one for $\phi_j^{\mathbb{Q}}$ must hold in all subsamples, making the LPY(ii) test potentially more informative. Figure 4 shows that the QTSM is successful at satisfying the LPY(ii) test, whereas the two other models imply slightly larger deviations of $\phi_j^{\mathbb{Q}}$ from one.

< Figure 4 about here >

Figure 5 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 $\boldsymbol{\theta}_2$, all models match the average level of the yield and 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 5 suggests that the Gaussian ATSM and the shadow rate model are able to pass the LPY(ii) with $\phi_j^{\mathbb{Q}}$ close to one, whereas the QTSM shows clear deviations from one.

< Figure 5 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 $\boldsymbol{\theta}_2$. For the long sample in Figure 6, 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 7 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 6 and 7 about here >

These findings lead us to the following conclusions. First, the three and four factor QTSMs generally struggle to match loadings from the ordinary Campbell-Shiller regressions, whereas these

moments are largely matched by the shadow rate models. Second, the shadow rate models are also more successful at passing the LPY(ii) test than the QTSMs, although the latter performs well in the long sample. From a methodological perspective, we document that bias-adjusting $\boldsymbol{\theta}_2$ has a significant impact on unconditional moments of bond yields, and our results therefore supplement 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 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 volatility of bond yields, and this section therefore studies how well the QTSM and the shadow rate model with three pricing factors match these moments.[23]

We use two measures of conditional volatility in the data. The first is the rolling standard deviation of bond yields (denoted $\sigma_{t,j}^{Rolling}$) computed from daily observations with a six month lookback.[24] 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 $\sigma_{t,j}^{GARCH}$). Figure 8 shows these estimates at four selected maturities and the model-implied volatilities in the long sample. Overall, the two measures of volatility in the data are fairly similar, although $\sigma_{t,j}^{Rolling}$ is more noisy than $\sigma_{t,j}^{GARCH}$. The QTSM captures most of the gradual increase in volatility during the 1960s and 1970s but does not match the elevated levels in the early 1980s. A similar finding is reported in Ahn et al. (2002). The gradual fall in volatility from the end of 2008 when the policy rate approaches the ZLB is also largely matched by the QTSM. However, the model is unable to reproduce the increase in volatility for the 0.5-, 2-, and 5-year bond yield just before approaching the ZLB, as emphasized by the second part of Figure 8 focusing on volatility after 2005. The shadow rate model predicts constant volatility when the policy rate is far from zero, and the model is therefore unable to capture the overall

[23] The 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.

[24] These daily bond yields are computed using the estimated parametric form for the yield curves in Gürkaynak et al. (2007).

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 8 about here >

For the short sample starting in 1990, the QTSM is generally less successful in matching volatility according to Figure 9. To see why, observe that volatility in the QTSM is closely related to the level factor and hence the short rate. As shown in Figure 8, 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 9 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 6 confirms our impression from above that the QTSM provides the best fit in the long 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 conditional volatility of bond yields. This may indicate that a more flexible functional form for the policy rate is required in models with Gaussian pricing factors or that the dynamics of the pricing factors should display heteroskedasticity, for instance induced by stochastic volatility. We return to this issue in Section 6 where the first extension is considered.

< Table 6 about here >

5 Empirical results: performance out-of-sample

This section studies the models' ability to predict future bond yields from January 2005 to December 2013. This forecasting sample is particularly challenging as it contains bond yields i) far from zero, ii) when hitting the ZLB, and iii) a prolonged period at the lower bound. 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.[25]

Figure 10 reports the root mean squared prediction errors (RMSPE) by maturity when the estimation is started in 1961. Columns in Figure 10 refer to the number of pricing factors and rows denote the forecast horizon of 1, 3, 6, and 12 months, respectively. Starting with the two-factor models, the QTSM clearly outperforms the Gaussian ATSM at the 1- and 3-month forecast horizons for all maturities, whereas the two models display 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 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 from the shadow rate model generally *improve* when including a fourth pricing factor, whereas the opposite applies for the QTSM. This suggests that the parsimonious mechanism in shadow rate model to enforce the ZLB is more robust and less subject to overfitting than the quadratic specification. A careful inspection of Figure 10 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.

< Figure 10 about here >

The forecasting results when the estimation is started in 1990 are provided in Figure 11. The overall results are very similar to those obtained in Figure 10 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. Third, forecasts again 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

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

all maturities and forecast horizons.[26]

< Figure 11 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 12 for the long sample by showing forecasts for the 0.5-year bond yield on two occasions. The first is the end of December 2008, when the policy rate reached the ZLB. Predicted bond yields in the three-factor Gaussian ATSM barely stay positive at the considered forecast horizons but 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 12 for the end of May 2010 shows that negative forecasts in the Gaussian ATSM occur with two, three, and four pricing factors and even when the policy rate has been at the ZLB for several years. 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 as shown in Figure 13. Note also that probabilities above 50% in this figure denote negative mean forecasts, which happens frequently when the estimation is started in 1961 but less often when started in 1990.

< Figure 12 about here >

< Figure 13 about here >

We summarize the forecasting performance of the three models in Table 7 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 10 and 11 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.

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

Our second color coding scheme in Table 7 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 $\mathcal{A}$ and $\mathcal{B}$ of Table 7. We surprisingly find that starting the estimation in 1961 generally gives the most accurate forecasts, as part $\mathcal{A}$ of Table 7 has 25 blue figures whereas part $\mathcal{B}$ only has 11. That is, the best forecasts are in general obtained by using a long sample for the estimation, particularly for the shadow rate model. Any finite sample bias in the estimated $\mathbb{P}$ dynamics is unlikely to explain this finding as we bias-adjust $\hat{\boldsymbol{\theta}}_2$ regardless of the starting point for the sample. Instead, the better forecasting performance from using a long sample is likely to be driven by two features. First, the pricing factors and hence bond yields are more persistent in the long sample compared to the short sample (see Section 4.2) and this is likely to improve forecasts, given the strong performance of the random walk. Second, bond yields in the 1960s were fairly low compared to their average level, meaning that the long sample includes bond yields closer to the levels seen after 2008 than a sample starting in 1990.

< Table 7 about here >

6 A hybrid model

The quadratic terms in the QTSM serve a dual purpose as they enforce the ZLB and generate time-varying conditional volatility in bond yields. Our results in Section 4.2 suggests that the estimates of $\boldsymbol{\Psi}$ are constrained by the non-negativity condition requiring $\boldsymbol{\Psi}$ to be positive semi-definite. Hence, there is a trade-off within the QTSM between enforcing the ZLB and matching the conditional volatility of bond yields. This section explores the potential benefit of eliminating this trade-off by considering an extended shadow rate model, where the shadow rate is an unrestricted quadratic function of the pricing factors. This type of model was first considered by Kim & Singleton (2012) with two pricing factors and extended below to include a third factor. Given that this extended shadow rate model merges the QTSM and the shadow rate model considered above, we refer to it as the hybrid model.

This section is structured as follows. We present the hybrid model in Section 6.1 and discuss the estimated parameters and in-sample fit in Section 6.2. The ability of the hybrid model to match the LPY tests and conditional volatility is examined in Section 6.3. We finally study the ability of the hybrid model to forecast bond yields out of sample in Section 6.4.

6.1 The hybrid model

As in the shadow rate model in Section 2.3, we let $r(\mathbf{x}_t) = \max(0, s(\mathbf{x}_t))$ but $s(\mathbf{x}_t)$ is now quadratic in the pricing factors, i.e.

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

where $\boldsymbol{\Psi} = \mathbf{ADA}'$ as in the QTSM. The non-negativity of the policy rate in the hybrid model is enforced by the shadow rate mechanism and no restrictions are therefore imposed on $\boldsymbol{\Psi}$. This gives the model greater flexibility in matching the level and conditional volatility of bond yields than any of the models considered previously. As in the QTSM and the shadow rate model, we assume affine processes for the pricing factors under the $\mathbb{Q}$ and $\mathbb{P}$ measures, i.e. (2) and (3) are imposed. The identification conditions for the hybrid model are therefore identical to those for the QTSM in Section 2.2.

In the absence of arbitrage, the price in time period t of an j-period zero-coupon bond is

$$P_{t,j}^{Hybrid} = E_t^{\mathbb{Q}}\left[\exp\left\{-\sum_{i=0}^{j-1} r(\mathbf{x}_{t+i})\right\}\right]$$

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

6.2 Model estimates and in-sample fit

Bond yields in the hybrid model are approximated using just 500 draws in the MC method. The top left chart in Figure 14 shows that this approximation is very accurate as the largest RMSE is just 0.77

basis points when evaluating bond yields at $\{\hat{\mathbf{x}}_t\}_{t=1}^T$ for the estimated parameters in the long sample. Without anti-control sampling, the largest RMSE increases to 4.63 basis points as shown in the top right chart of Figure 14, documenting the benefit of using anti-control sampling. The bottom row of Figure 14 shows that our MC approximation is even more accurate in the short sample, with the largest RMSE being only 0.07 basis points.

The estimation results for the hybrid model are provided in Table 8.[27] Without the ZLB restriction on $\mathbf{\Psi}$, the estimated elements in $\mathbf{A}$ now exceed one, although most predominantly in the long sample. This implies that $\mathbf{\Psi}$ is indefinite with eigenvalues of $\{-2.403, -0.028, 5.431\}$ in the long sample, meaning that the short rate is controlled by two pricing factors instead of one in the QTSM. In the short sample, $\mathbf{\Psi}$ is also indefinite with eigenvalues of $\{-0.153, -0.010, 3.163\}$. The positive eigenvalue is here substantially larger than the two negative eigenvalues, implying that the policy rate mainly is controlled by one pricing factor as in the QTSM within the short sample. We also note that many of the standard errors are fairly wide for the hybrid model in the short sample, indicating that it may be hard to accurately estimate this flexible model when starting the sample in 1990.[28]

The hybrid model nests the QTSM which delivers the best in-sample fit with three pricing factor models as shown in Section 4.3. Hence, the hybrid model should at least fit bond yields as well as the QTSM. Indeed, in the long sample the objective functions are $\tilde{Q}_{1:T}^{step1} = 2.677$ and $\tilde{Q}_{1:T}^{step3} = 2.709$ for the hybrid model, which is a marginal improvement compared to the QTSM having $\tilde{Q}_{1:T}^{step1} = 2.704$ and $\tilde{Q}_{1:T}^{step3} = 2.719$. For the short sample, the hybrid model implies $\tilde{Q}_{1:T}^{step1} = 1.600$ and $\tilde{Q}_{1:T}^{step3} = 1.618$, which again is a marginal improvement compared to the QTSM where $\tilde{Q}_{1:T}^{step1} = 1.614$ and $\tilde{Q}_{1:T}^{step3} = 1.632$.

A more careful inspection of the in-sample fit is provided in Figure 15, where charts in the first column show the recursively computed objective functions from January 2005 and onwards. In the long sample, the hybrid model clearly outperforms the two other models from the end of 2008 where the short rate approaches the ZLB. A somewhat smaller improvement is found in the short sample, which seems consistent with our finding that relaxing the constraint on $\mathbf{\Psi}$ has only minor effects on

[27] The estimation of the hybrid model is computationally demanding even with just 500 draws in the MC approximation to bond yields. Furthermore, the objective function in the first step of the SR approach has several local optima. We address these challenges by using the CMA-ES optimizer of Hansen, Müller & Koumoutsakos (2003), capable of optimizing multi-model objective functions and implemented with multiprocessing in FORTRAN on a computer cluster to make the estimation feasible. The estimation is carried out with multiple starting values and with 60 CPUs per optimization.

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

A in this case. Another way to draw the same conclusion is to look at $\hat{\sigma}_t$, showing the in-sample fit of the entire yield curve in a given time period. This measure of in-sample fit is provided in the second column of Figure 15 and shows that the hybrid model in general provides a better fit of the yield curve from the end of 2008 than any of the two other models.

< Figure 15 about here >

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

6.3 The LPY tests and conditional volatility

Figure 16 shows that the hybrid model preserves the ability of the QTSM to match the mean level and the unconditional volatility of bond yields in the two samples. As for the LPY tests, the hybrid model also behaves very much like the QTSM, i.e. it struggles to match LPY(i) but not LPY(ii) in the long sample, whereas the opposite holds in the short sample.

< Figure 16 about here >

The ability of the hybrid model to match conditional volatility in bond yields is summarized in Table 9 where we run the volatility regressions from Section 4.5. In the long sample, we see a small improvement for the 0.5- and 2-year bond yield as the R^2 increases from 0.33 and 0.38 in the QTSM to 0.40 and 0.43 in the hybrid model, respectively, when using the GARCH measure of volatility in the data. This is highlighted by the blue figures in Table 9. For bond yields at the 5- and 10-year maturity, we do not find an improvement compared to the QTSM. In the short sample, we somewhat surprisingly do not find that the hybrid model provides a better fit of volatility than the QTSM and the shadow rate model. Unreported results show that volatility in the hybrid model and the QTSM are very similar and closely liked to the short rate, which is less successful at predicting volatility after 1990 as argued in Section 4.5.

< Table 9 about here >

6.4 Performance out-of-sample

We finally explore the forecasting performance of the hybrid model in Figure 17, where columns index the starting point for the estimation and rows refer to the forecast horizon. When starting the estimation in 1961 (charts to the left), forecasts in the hybrid model are very similarly to those in the shadow rate model and the QTSM for bond yields within the 0.5- to 7-year maturity range. Beyond the 7-year maturity, the performance of the hybrid model gradually deteriorates, particularly at the 1 and 3 months forecast horizons. We see the same pattern when starting the estimation in 1990 (charts to the right), except the deteriorating performance of the hybrid model starts already from the 3-year maturity. Hence, the hybrid model delivers less accurate forecasts of medium- and particularly long-term bond yields compared to the QTSM and the shadow rate model. To understand this finding, recall that the hybrid model only differs from the QTSM by having a more flexible $\mathbb{Q}$ dynamics. This makes overfitting of the policy rate more like in the hybrid model and its effects are gradually propagated through the yield curve by the no-arbitrage pricing, thereby generating less accurate forecasts of medium- and long-term bond yields.

< Figure 17 about here >

7 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 or 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 ordinary and risk-adjusted Campbell-Shiller regressions, whereas these moments are better 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 outperforms 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 forecasts in the shadow rate model generally improve when including a fourth pricing factor whereas the opposite holds in 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.

In an attempt to improve the performance of the QTSM, we also study an extended shadow rate model where the policy rate is an unrestricted quadratic function of the pricing factors. This hybrid model fits bond yields marginally better in-sample than the other models but is outperformed by the QTSM and the shadow rate model when forecasting bond yields out of sample. The hybrid model also struggles to provide a better fit of conditional volatility compared to the QTSM, at least when these models are estimated solely on bond yields. It is likely that the ability of these models to match conditional volatility of bond yields could be improved by also including this time series in the estimation, as done for instance in Monfort, Pegoraro, Renne & Roussellet (2014). This extension would be straightforward for the QTSM but not for the hybrid model due to the absence of a closed-form solution for bond yields. Another way to improve the fit of conditional volatility in bond yields would be to maintain the affine specification for the shadow rate and instead introduce heteroskedasticity in the dynamics of the pricing factors. Given the strong performance of the affine shadow rate, this extension seems particularly promissing and deserves attention in future research.

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

The standard bootstrap for a VAR model without measurement errors in the pricing factors generates the sampling distribution for moments involving $\hat{\mathbf{x}}_t$ and $\widehat{\hat{\mathbf{w}}}_{t+1}$ in (11) and (12). The variability in the remaining moments in (11) and (12) related to the measurement errors is accounted for by resampling with replace from $\left\{\widehat{Var}\left(\mathbf{u}_t\right)\right\}_{t=1}^{T}$,$\left\{\widehat{Cov}\left(\mathbf{u}_{t+1},\mathbf{u}_t\right)\right\}_{t=1}^{T-1}$, and $\left\{\widehat{Cov}\left(\mathbf{u}_t,\mathbf{u}_{t+1}\right)\right\}_{t=1}^{T-1}$. The suggested bootstrap for a VAR model with measurement errors in the pricing factors is therefore:

Step A: Use (11) and (12) to obtain $\hat{\boldsymbol{\theta}}_2$. Compute the residuals, i.e. $\widehat{\hat{\mathbf{w}}}_{t+1} = \hat{\mathbf{x}}_{t+1} - \hat{\mathbf{h}}_0 - \hat{\mathbf{h}}_{\mathbf{x}}\hat{\mathbf{x}}_t$ for $t = 1, 2, ..., T-1$. Let $b = 1$.

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

$$\mathbf{x}_{t+1}^{*} = \hat{\mathbf{h}}_0 + \hat{\mathbf{h}}_{\mathbf{x}}\mathbf{x}_t^{*} + \widehat{\hat{\mathbf{w}}}_{t+1}^{*} \quad \text{for } t = 1, 2, ..., T-1. \tag{23}$$

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

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

Step D: Use the draws from Step B and C in (11) and (12) to obtain $\hat{\mathbf{h}}_0^{(b)}$, $\hat{\mathbf{h}}_{\mathbf{x}}^{(b)}$, and $\hat{\boldsymbol{\Sigma}}^{(b)}$.

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

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

$$\hat{\mathbf{h}}_{\mathbf{x}}^{adj} = \hat{\mathbf{h}}_{\mathbf{x}} - \left(\bar{\mathbf{h}}_{\mathbf{x}} - \hat{\mathbf{h}}_{\mathbf{x}}\right) = 2\hat{\mathbf{h}}_{\mathbf{x}} - \bar{\mathbf{h}}_{\mathbf{x}}, \tag{24}$$

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

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

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

$$\widehat{\hat{\mathbf{w}}}_{t+1}^{adj} = \hat{\mathbf{x}}_{t+1} - \hat{\mathbf{h}}_{\mathbf{0}}^{adj} - \hat{\mathbf{h}}_{\mathbf{x}}^{adj}\hat{\mathbf{x}}_t \quad \text{for } t = 1, 2, ..., T-1 \tag{25}$$

and a direct modification of (12), i.e.

$$\begin{aligned}\widehat{Var}\left(\mathbf{w}_{t+1}\right)^{adj} &= \frac{1}{T-1-n_x-1}\sum_{t=1}^{T-1}\widehat{\hat{\mathbf{w}}}_{t+1}^{adj}\left(\widehat{\hat{\mathbf{w}}}_{t+1}^{adj}\right)' \\ &-\frac{1}{T-1}\sum_{t=1}^{T-1}\left(\widehat{Var}\left(\mathbf{u}_t\right)+\hat{\mathbf{h}}_{\mathbf{x}}^{adj}\widehat{Var}\left(\mathbf{u}_t\right)\left(\hat{\mathbf{h}}_{\mathbf{x}}^{adj}\right)'\right) \\ &+\frac{1}{T-1}\sum_{t=1}^{T-1}\left(\widehat{Cov}\left(\mathbf{u}_{t+1},\mathbf{u}_t\right)\left(\hat{\mathbf{h}}_{\mathbf{x}}^{adj}\right)'+\hat{\mathbf{h}}_{\mathbf{x}}^{adj}\widehat{Cov}\left(\mathbf{u}_t,\mathbf{u}_{t+1}\right)\right),\end{aligned} \tag{26}$$

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

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

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

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

and

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

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

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

and estimate the variance of the innovations by

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

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

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

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

To compute the model-independent unconditional variances of the factors as implied by $\{\hat{\mathbf{x}}_t\}_{t=1}^T$, the unconditional mean of the i'th pricing factor is estimated by $\hat{E}\left[\hat{x}_{i,t}\right]=1/T\sum_{t=1}^T\hat{x}_{i,t}$. We also

have

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

for $i=1,2,...,n_x$, where the last line follows by considering $u_{i,t}^2$ as a point estimate of $Var\left(u_{i,t}\right)$. A similar argument is used when computing standard errors robust to heteroskedasticity. Clearly, $\frac{1}{T-1}\sum_{t=1}^{T}\left(x_{i,t}^o-\hat{E}\left[\hat{x}_i\right]\right)^2\xrightarrow{p}Var\left(x_{i,t}^o\right)$ as $T\longrightarrow\infty$. We also have for $T\longrightarrow\infty$, that

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

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

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

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

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

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

$$\hat{\delta}=\arg\min_{\delta\in[\delta_{lower},1]}\sum_{i=1}^{n_x}\left(\frac{\sigma_{i,VAR}^2\left(\delta\right)-\hat{\sigma}_{i,Data}^2}{\hat{\sigma}_{i,Data}^2}\right)^2 \tag{27}$$

where $\delta_{lower}>0$. The constraint on the domain of δ is imposed because at $\delta=0$, we have $\hat{\mathbf{h}}_{\mathbf{x}}^{adj}\left(\delta=0\right)=\mathbf{0}$ and $\hat{\mathbf{h}}_{\mathbf{0}}^{adj}\left(\delta=0\right)=\hat{E}\left[\hat{\mathbf{x}}_t\right]$, meaning that the two estimators of the unconditional variances in (27) nearly coincides as they only differ by $\frac{1}{T}\sum_{t=1}^{T}Var\left(u_{i,t}\right)$.

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

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

where $u_{t,j}^v$ denotes the estimation error in $v_{t,j}$. Hence,

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

Moreover, we have also have

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

Analyzing the moment condition for the OLS estimator, we get

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

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

D Details for Monte Carlo study in Section 3.5

The data generating process for the Monte Carlo study is a one-factor Gaussian ATSM where we let $\alpha = 0.008$, $\Phi_{11} = 0.01$, $\Sigma_{11} = 5.5 \times 10^{-4}$, $h_0 = -0.0002$, and $h_x = 0.96$. This calibration ensures that the one-factor model roughly matches the level and variability of the US yield curve from 1990 to 2013.

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

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

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

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

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

according to

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

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

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

To apply the SR approach for this one-factor model, we let $\boldsymbol{\theta}_{11} \equiv \left[\begin{array}{cc} \alpha & \Phi_{11} \end{array}\right]'$, $\boldsymbol{\theta}_{12} \equiv \left[\begin{array}{c} \Sigma_{11} \end{array}\right]$, and $\boldsymbol{\theta}_{22} \equiv \left[\begin{array}{cc} h_0 & h_x \end{array}\right]$. All the risk-neutral parameters are estimated in step 1 using (8), and $\boldsymbol{\theta}_2$ is obtained in step 2 using the bias-adjustment described in Section 3.3.1. In step 3, we let $\boldsymbol{\Lambda} = \mathbf{0}$ and re-estimate $\boldsymbol{\theta}_{11}$ using (17).

E The hybrid model: Monte Carlo approximation to bond yields

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

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

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

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

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

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

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

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

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

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

implying

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

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

References

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

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

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

Andreasen, M. M. (2013), 'Non-linear DSGE models and the central difference kalman filter', *Journal of Applied Econometrics* **28**, 929–955.

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

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

Ang, A., Boivin, J., Dong, S. & Loo-Kung, R. (2011), 'Monetary policy shifts and the term structure', *Review of Economic Studies* .

Ang, A. & Piazzesi, M. (2003), 'A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables', *Journal of Monetary Economics* **50**, 745–787.

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.

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

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

Christensen, J. H. E., Diebold, F. X. & Rudebusch, G. D. (2011), 'The affine arbitrage-free class of Nelson-Siegel term structure models', *Journal of Econometrics* **164**, 4–20.

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

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

Christensen, J. H. E. & Rudebusch, G. D. (2014), 'Estimating shadow-rate term structure models with near-zero yields', *Journal of Financial Econometrics* **0**(0), 1–34.

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.

Gagnon, J., Raskin, M., Rernache, J. & Sack, B. (2011), 'Large-Scale Asset Purchases by the Federal Reserve: Did They Work?', *Working Paper* .

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

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

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

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

Hördahl, P., Tristani, O. & Vestin, D. (2006), 'A joint econometric model of macroeconomic and term-structure dynamics', *Journal of Econometrics* **131**, 404–444.

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

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

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

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

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

Kim, D. H. & 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.

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

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

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

Priebsch, 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. (2010), 'Maximum likelihood estimation of the Cox-Ingersoll-Ross model using particle filters', *Computational Economics* **36**(1), 1–16.

Rudebusch, G. D. & Tao (2007), 'Accounting for a shift in term structure behavior with no-arbitrage and macro-finance models', *Journal of Money, Credit and Banking* **39**(2-3), 395–422.

Rudebusch, G. D. & Wu, T. (2008), 'A macro-finance model of the term structure, monetary policy, and the economy', *The Economic Journal* **118**, 906–926.

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

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

The Monte Carlo study is implemented without measurement errors in the pricing factors and with $M = 5.000$ draws, where each bootstrap adjustment is computed with $B = 5.000$ bootstrap replications. The data generating processes (DGP) are the estimated VAR models for the pricing factors under physical measure in the Gaussian ATSM reported in Table 3 and 4. The notation Bias($\mathbf{h}_0$) indicates the total absolute bias for $\mathbf{h}_0$ and similarly for the other rows. When computing the total absolut bias in the unconditional standard deviation in the pricing factors, denoted 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($\mathbf{\Sigma} \times 100$)	0.0012	0.0006	0.0006	0.0007	0.0007
	Bias$\left(\{\sigma_{x_i}\}_{i=1}^{n_x}\right)$	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($\mathbf{\Sigma} \times 100$)	0.0005	0.0002	0.0003	0.0003	0.0003
	Bias$\left(\{\sigma_{x_i}\}_{i=1}^{n_x}\right)$	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($\mathbf{\Sigma} \times 100$)	0.0129	0.0045	0.0062	0.0077	**0.0068**
	Bias$\left(\{\sigma_{x_i}\}_{i=1}^{n_x}\right)$	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($\mathbf{\Sigma} \times 100$)	0.0037	0.0017	0.0018	0.0019	0.0019
	Bias$\left(\{\sigma_{x_i}\}_{i=1}^{n_x}\right)$	0.0050	0.0193	0.0853	0.0192	**0.0187**
	Pct of nonstationary draws	0.00	5.28	0.00	0.00	0.00

Table 2: Monte Carlo study: Bootstrapping the third step of the SR approach

The Monte Carlo study is carried out for $T = 250$ and $n_y = 25$ using the same maturities as selected in Section 4.1 for our empirical analysis. We use $M = 5,000$ replications and $B = 1,000$ in the bootstrap. The asymptotical standard errors are computed using (18) with $w_T = 10$ when $\rho_{Time} = 0.6$ and $w_D = 5$ when $\rho_{Cross} = 0.4$, otherwise $w_T = 0$ and $w_D = 0$. Standard resampling is used, unless $\phi_1 = 0.4$, where we let $p = 1$ and estimate ϕ_1 using a preliminary bootstrap with 100 draws. Bold figures indicate the best method for computing standard errors and rejection probabilities.

		Standard errors ($\times 10^{-4}$)			Reject probabilities at the 5% level		
		True	Asymp	Boot	Asymp	Boot*	Boot**
Case I: IID errors	α	0.777	**0.775**	0.781	6.02	5.80	**5.32**
	Φ_{11}	0.618	**0.610**	0.583	**4.74**	6.80	6.44
Case II: autocorrelated errors ($\rho_{Time} = 0.6$)	α	0.920	1.005	**0.925**	3.90	**5.12**	4.78
	Φ_{11}	1.398	1.673	**1.320**	2.22	6.96	**6.38**
Case III: autocorrelated and heteroskedastic errors ($\rho_{Time} = 0.6, \rho_{Rv} = 0.98$)	α	1.094	1.246	**1.098**	3.04	**5.12**	4.46
	Φ_{11}	2.061	2.437	**1.934**	2.12	7.22	**6.38**
Case IV: cross-correlated errors ($\phi_1 = 0.4$)	α	0.824	1.502	**0.835**	0.12	5.28	**5.18**
	Φ_{11}	0.938	1.346	**0.909**	0.96	**6.30**	6.72
Case V: autocorrelated, heteroskedastic, and cross-correlated errors ($\rho_{Time} = 0.6, \rho_{Rv} = 0.98$, $\phi_1 = 0.4$)	α	1.444	2.044	**1.488**	1.20	4.84	**4.86**
	Φ_{11}	3.205	3.511	**3.100**	3.68	**6.06**	7.16

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

Table 3: Estimation results for three-factor models: sample from 1961-2013

Asymptotical robust standard errors for elements in $\hat{\boldsymbol{\theta}}_{11}^{step3}$ are computed using (18) with $w_D = 5$ and $w_T = 10$. For the Gaussian ATSM and the shadow rate model, bootstrapped standard errors using 1,000 draws for $\hat{\boldsymbol{\theta}}_{11}^{step3}$ are shown in brackets. For the QTSM, 95 percent confidence intervals are provided for elements in $\hat{\boldsymbol{\theta}}_{11}^{step3}$, computed using the 2.5 and 97.5 percentiles of $\hat{\boldsymbol{\theta}}_{11}^{step3}$ in the bootstrap. All bootstraps use a bias-adjusted AR(1) model (based on 100 draws) to account for cross-correlation and use draws for $\hat{\boldsymbol{\theta}}_2$ from its asymptotic distribution to make the bootstrap inference for $\hat{\boldsymbol{\theta}}_{11}$ comparable to the asymptotic inference. For elements in $\hat{\boldsymbol{\theta}}_2^{step3}$, robust standard errors are computed using (13) with $\mathbf{S}$ obtained by the Newey-West estimator for a bandwidth of 5.

	ATSM		QTSM		Shadow rate	
	Estimate	SE	Estimate	$\text{CI}_{95\%}$ or SE	Estimate	SE
α	0.0124	0.0016 [0.0003]	-	-	0.0153	0.0037 [0.0071]
A_{12}	-	-	0.9886	$[0.9524, 1.0000]$	-	-
A_{13}	-	-	0.9915	$[0.9651, 1.0000]$	-	-
A_{23}	-	-	0.8642	$[0.5654, 1.0398]$	-	-
Φ_{11}	0.0022	0.0005 [0.0001]	0.0011	$[0.0002, 0.0016]$	0.0013	0.0005 [0.0004]
Φ_{22}	0.0355	0.0027 [0.0007]	0.0405	$[0.0366, 0.0484]$	0.0427	0.0040 [0.0054]
Φ_{33}	0.0685	0.0063 [0.0016]	0.0806	$[0.0499, 0.0878]$	0.0666	0.0069 [0.0072]
μ_1	-	-	0.0231	$[0.0000, 0.2929]$	-	-
μ_2	-	-	0.0035	$[0.0001, 0.1265]$	-	-
μ_3	-	-	0.1122	$[0.0000, 0.1535]$	-	-
$h_0\,(1,1)$	-1.03×10^{-4}	6.57×10^{-5}	-0.0017	9.30×10^{-4}	-1.67×10^{-4}	8.60×10^{-5}
$h_0\,(2,1)$	3.83×10^{-4}	1.92×10^{-4}	-0.0093	0.0046	9.15×10^{-4}	3.65×10^{-4}
$h_0\,(3,1)$	-4.69×10^{-4}	2.09×10^{-5}	0.0155	0.0046	-0.0010	3.87×10^{-4}
$h_x\,(1,1)$	0.9847	0.0077	0.9733	0.0079	0.9822	0.0077
$h_x\,(1,2)$	0.0252	0.0084	0.0082	0.0056	0.0216	0.0081
$h_x\,(1,3)$	0.0186	0.0121	0.0041	0.0100	0.0188	0.0104
$h_x\,(2,1)$	0.0489	0.0223	0.0677	0.0234	0.0906	0.0329
$h_x\,(2,2)$	0.9668	0.0420	1.0310	0.0314	1.0097	0.0620
$h_x\,(2,3)$	0.0611	0.0546	0.1115	0.0438	0.0866	0.0714
$h_x\,(3,1)$	-0.0557	0.0236	-0.0754	0.0243	-0.1004	0.0345
$h_x\,(3,2)$	-0.0151	0.0376	-0.0649	0.0334	-0.0637	0.0578
$h_x\,(3,3)$	0.8685	0.0488	0.8295	0.0431	0.8452	0.0661
Σ_{11}	3.56×10^{-4}	2.37×10^{-5}	0.0023	1.63×10^{-4}	3.30×10^{-4}	2.14×10^{-5}
Σ_{21}	-6.22×10^{-4}	9.83×10^{-5}	-0.0030	7.86×10^{-4}	-6.49×10^{-4}	1.37×10^{-4}
Σ_{22}	0.0011	7.31×10^{-5}	0.0101	6.94×10^{-4}	0.0018	1.07×10^{-4}
Σ_{31}	3.95×10^{-4}	8.88×10^{-5}	0.0015	7.70×10^{-4}	4.54×10^{-4}	1.30×10^{-4}
Σ_{32}	-0.0010	5.60×10^{-5}	-0.0101	7.78×10^{-4}	-0.0017	9.40×10^{-5}
Σ_{33}	4.31×10^{-4}	5.45×10^{-5}	0.0029	2.26×10^{-4}	4.40×10^{-4}	5.52×10^{-5}

Table 4: Estimation results for three-factor models: sample from 1990-2013

Asymptotical robust standard errors for elements in $\hat{\boldsymbol{\theta}}_{11}^{step3}$ are computed using (18) with $w_D = 5$ and $w_T = 10$. For the Gaussian ATSM and the shadow rate model, bootstrapped standard errors using 1,000 draws for $\hat{\boldsymbol{\theta}}_{11}^{step3}$ are shown in brackets. For the QTSM, 95 percent confidence intervals are provided for elements in $\hat{\boldsymbol{\theta}}_{11}^{step3}$, computed using the 2.5 and 97.5 percentiles of $\hat{\boldsymbol{\theta}}_{11}^{step3}$ in the bootstrap. All bootstraps use a bias-adjusted AR(1) model (based on 100 draws) to account for cross-correlation and use draws for $\hat{\boldsymbol{\theta}}_2$ from its asymptotic distribution to make the bootstrap inference for $\hat{\boldsymbol{\theta}}_{11}$ comparable to the asymptotic inference. For elements in $\hat{\boldsymbol{\theta}}_2^{step3}$, robust standard errors are computed using (13) with $\mathbf{S}$ obtained by the Newey-West estimator for a bandwidth of 5.

	ATSM		QTSM		Shadow rate	
	Estimate	SE	Estimate	$\text{CI}_{95\%}$ or SE	Estimate	SE
α	0.0093	0.0007 [0.0003]	-	-	0.0099	0.0008 [0.0004]
A_{12}	-	-	0.9725	$[0.9491, 0.9998]$	-	-
A_{13}	-	-	0.9861	$[0.9681, 0.9983]$	-	-
A_{23}	-	-	0.6846	$[0.5691, 1.1857]$	-	-
Φ_{11}	0.0043	0.0004 [0.0002]	0.0028	$[0.0021, 0.0032]$	0.0035	0.0004 [0.0002]
Φ_{22}	0.0487	0.0020 [0.0035]	0.0459	$[0.0424, 0.0541]$	0.0475	0.0024 [0.0035]
Φ_{33}	0.0518	0.0033 [0.0045]	0.0724	$[0.0589, 0.0768]$	0.0558	0.0050 [0.0039]
μ_1	-	-	0.0040	$[0.0000, 0.0504]$	-	-
μ_2	-	-	0.0028	$[0.0000, 0.1080]$	-	-
μ_3	-	-	0.0977	$[0.0000, 0.1079]$	-	-
$h_0\,(1,1)$	-2.81×10^{-4}	1.50×10^{-4}	-2.23×10^{-4}	0.0012	-3.32×10^{-4}	1.72×10^{-4}
$h_0\,(2,1)$	0.0078	0.0040	-0.0135	0.0073	0.0037	0.0018
$h_0\,(3,1)$	-0.0079	0.0040	0.0181	0.0073	-0.0039	0.0018
$h_x\,(1,1)$	0.9530	0.0242	0.9427	0.0248	0.9474	0.0265
$h_x\,(1,2)$	-0.0216	0.0272	-0.0095	0.0107	-0.0230	0.0259
$h_x\,(1,3)$	-0.0237	0.0290	-0.0218	0.0197	-0.0285	0.0306
$h_x\,(2,1)$	1.2792	0.6300	0.2415	0.0994	0.5884	0.2723
$h_x\,(2,2)$	2.4472	0.8675	1.1202	0.0663	1.5625	0.3283
$h_x\,(2,3)$	1.5652	0.9067	0.2371	0.0989	0.6813	0.3676
$h_x\,(3,1)$	-1.3025	0.6166	-0.2611	0.0989	-0.6150	0.2589
$h_x\,(3,2)$	-1.5152	0.8566	-0.1584	0.0696	-0.6272	0.3189
$h_x\,(3,3)$	-0.6356	0.8946	0.7070	0.0943	0.2476	0.3556
Σ_{11}	3.92×10^{-4}	4.58×10^{-5}	0.0028	0.0003	3.73×10^{-4}	4.12×10^{-5}
Σ_{21}	-0.0049	0.0020	-0.0033	0.0024	-0.0018	7.01×10^{-4}
Σ_{22}	0.0108	0.0007	0.0169	0.0014	0.0042	2.67×10^{-4}
Σ_{31}	0.0046	0.0019	0.0007	0.0023	0.0014	6.74×10^{-4}
Σ_{32}	-0.0108	0.0007	-0.0170	0.0015	-0.0042	2.72×10^{-4}
Σ_{33}	1.66×10^{-5}	1.56×10^{-5}	0.0023	0.0003	1.69×10^{-4}	1.57×10^{-5}

Table 5: In-sample fit: The objective functions

This table reports $100\sqrt{2 \times Q_{1:T}^{step1}}$ and $100\sqrt{2 \times Q_{1:T}^{step3}}$ from the first and third step in the SR approach. Figures in 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	**2.702**	2.735	2.896	**2.718**	2.786
Sample: 1990-2013	1.808	**1.613**	1.692	1.829	**1.630**	1.754
Four pricing factors						
Sample: 1961-2013	1.057	1.030	**1.013**	1.058	1.034	**1.025**
Sample: 1990-2013	0.801	0.766	**0.749**	0.807	0.772	**0.763**

Table 6: 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 in time period t are computed from a local linearization of bond yields at $\hat{\mathbf{x}}_{t-1}$. Bold figures indicate the preferred model for a given measure of volatility and for a given sample.

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

Table 7: Average forecasting results

The figure reports the average root mean squared prediction errors (RMSPEs) across all bond yields in the forecasting study from January 2005 to December 2013. The RMSPEs are generated from models estimated recursively from 1961 or 1990 to the month prior to the forecast. The forecasted bond yields in the 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 $\mathcal{A}$ and $\mathcal{B}$ of the table.

	Part $\mathcal{A}$: Model estimation from 1961				Part $\mathcal{B}$: 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**

Table 8: Estimation results for three-factor hybrid model

Robust standard errors for elements in $\hat{\boldsymbol{\theta}}_{11}^{step3}$ are computed using (18) with $w_D = 5$ and $w_T = 10$. For elements in $\hat{\boldsymbol{\theta}}_{2}^{step3}$, robust standard errors are computed using (13) with $\mathbf{S}$ obtained by the Newey-West estimator for a bandwidth of 5. The standard errors in the short sample from 1990-2013 are approximated by fixing μ_2 to zero and treating this parameter as known.

	Data: 1961-2013		Data: 1990-2013	
	Estimate	SE	Estimate	SE
α	1.50×10^{-4}	0.0002	1.68×10^{-4}	0.0007
A_{12}	2.5659	0.0123	1.0836	0.0037
A_{13}	2.9113	0.0126	1.1422	0.0123
A_{23}	1.1520	0.0083	1.2615	0.0309
Φ_{11}	0.0011	0.0004	0.0028	0.0005
Φ_{22}	0.0419	0.0032	0.0442	0.0027
Φ_{33}	0.0816	0.0091	0.0787	0.0029
μ_1	0.0664	0.0157	0.0918	0.0369
μ_2	0.0369	0.0158	0.0000	–
μ_3	0.0034	0.0127	0.0119	0.0040
$h_0(1,1)$	1.62×10^{-4}	0.0002	−0.0123	0.0106
$h_0(2,1)$	−0.0021	0.0015	−0.2793	0.1840
$h_0(3,1)$	0.0032	0.0017	0.4608	0.3247
$h_x(1,1)$	0.9761	0.0077	1.0404	0.1624
$h_x(1,2)$	0.0056	0.0033	0.0420	0.0552
$h_x(1,3)$	0.0032	0.0050	−0.0320	0.0243
$h_x(2,1)$	0.0854	0.0304	−4.2888	2.9977
$h_x(2,2)$	1.0129	0.0334	0.1167	0.9035
$h_x(2,3)$	0.0751	0.0434	0.5871	0.3443
$h_x(3,1)$	−0.0867	0.0311	7.0391	5.1560
$h_x(3,2)$	−0.0430	0.0391	1.4376	1.5053
$h_x(3,3)$	0.8759	0.0502	0.0227	0.5547
Σ_{11}	0.0013	0.0001	0.0031	0.0008
Σ_{21}	−0.0022	0.0006	−0.0127	0.0223
Σ_{22}	0.0091	0.0006	0.0154	0.0258
Σ_{31}	6.38×10^{-4}	0.0007	0.0169	0.0373
Σ_{32}	−0.0093	0.0009	−0.0233	0.0507
Σ_{33}	0.0033	0.0004	0.0130	0.0032

Table 9: The hybrid model: 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. Conditional volatility in the data is either obtained using a rolling standard deviation of daily bond yields in the past six months, denoted $\sigma_t^{Rolling}$, or a GARCH(1,1) model for changes in monthly bond yields, denoted σ_t^{GARCH}. The model-implied conditional volatilities one-month ahead in time period t are computed from a local linearization of bond yields at $\hat{\mathbf{x}}_{t-1}$. Figures marked by blue indicate that the R^2 for the hybrid model is larger than the R^2 for both the QTSM and the shadow rate model in Table 6.

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

Figure 1: The QTSM: Non-linear filtering

The objective function for filtering out x_t in a QTSM with one pricing factor. The risk-neutral parameters are $\mu_1 = 0.0790$, $\Phi_{11} = 0.0072$, and $\Sigma_{11} = 0.0066$, which are the optimal values in the short sample.

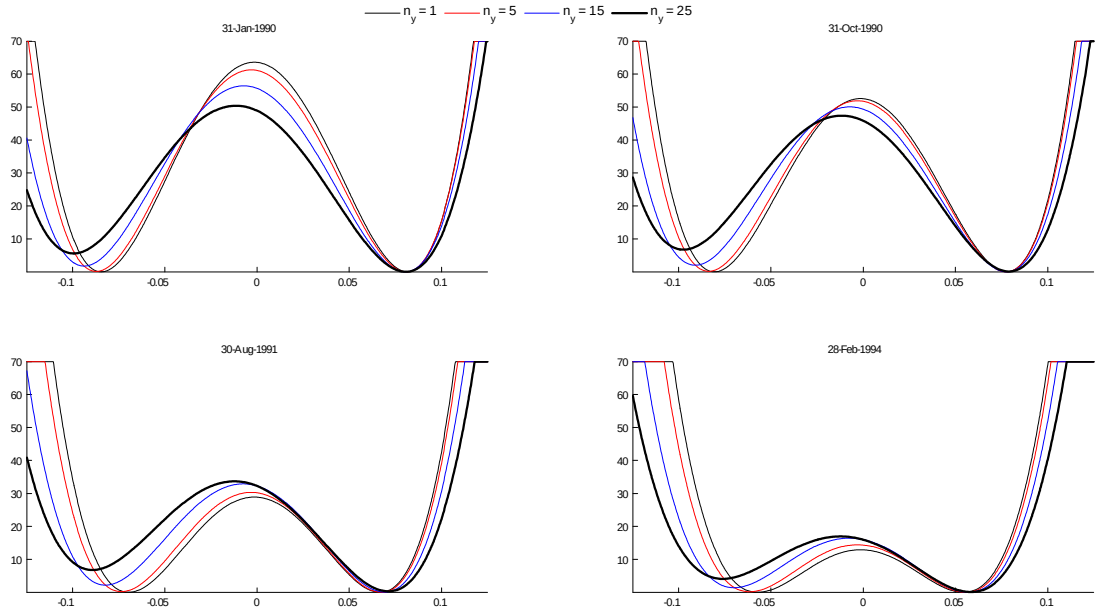


Figure 2: Sample from 1961-2013: In-sample fit for three- and four-factor models

Charts in the first column report $100\sqrt{2 \times Q^{step3}_{1:T}}$. Charts in the second column report $100\sqrt{2 \times Q^{step3}_{2005:T}}$ and the final column reports σ_j for j=1,2,...,K estimated using residuals from 1961-2013.

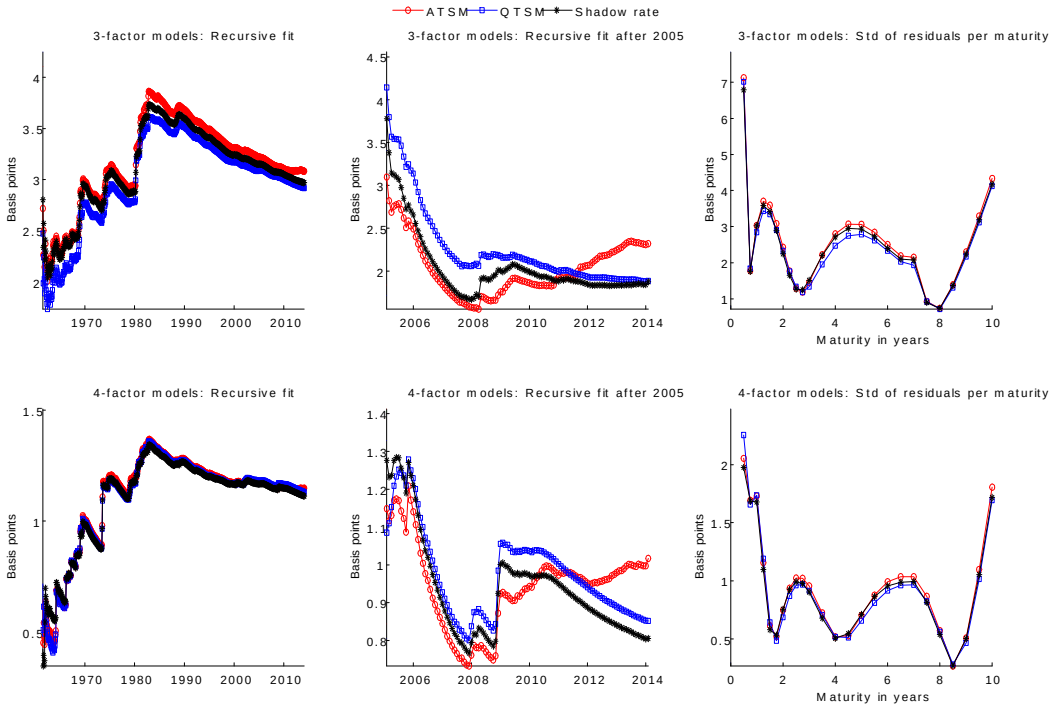


Figure 3: Sample from 1990-2013: In-sample fit for three- and four-factor models

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

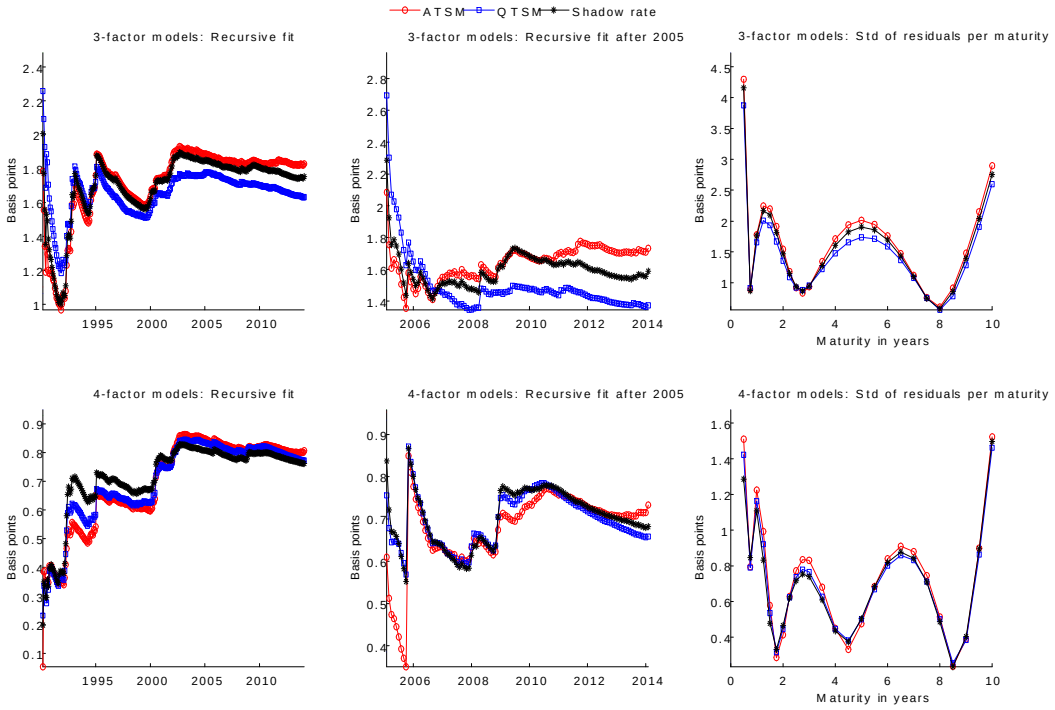


Figure 4: Sample from 1961-2013: Unconditional moments in three-factor models
All model-based moments are obtained from simulated time series of 100,000 observations. Empirical moments are computed from September 1971 to December 2013 to avoid missing observations for long bond yields.

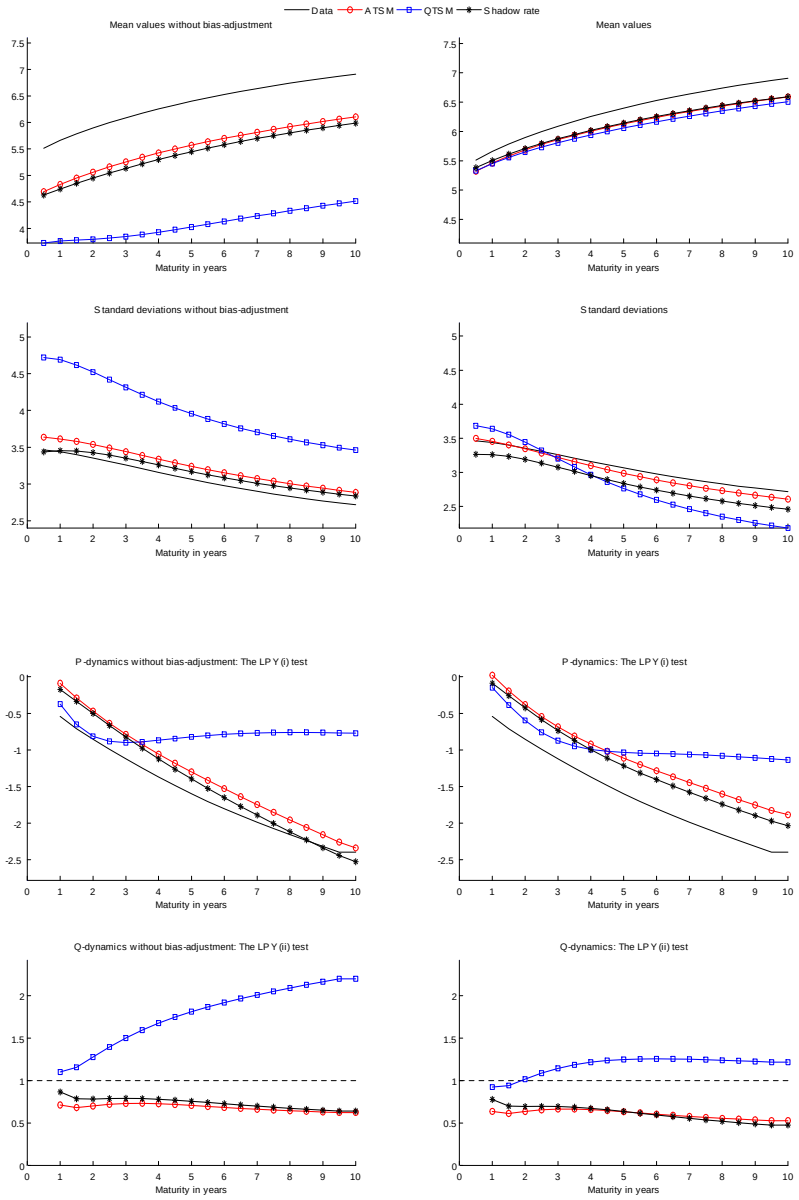


Figure 5: Sample from 1990-2013: Unconditional moments in three-factor models
All model-based moments are obtained from simulated time series of 100,000 observations. Empirical moments are computed from January 1990 to December 2013.

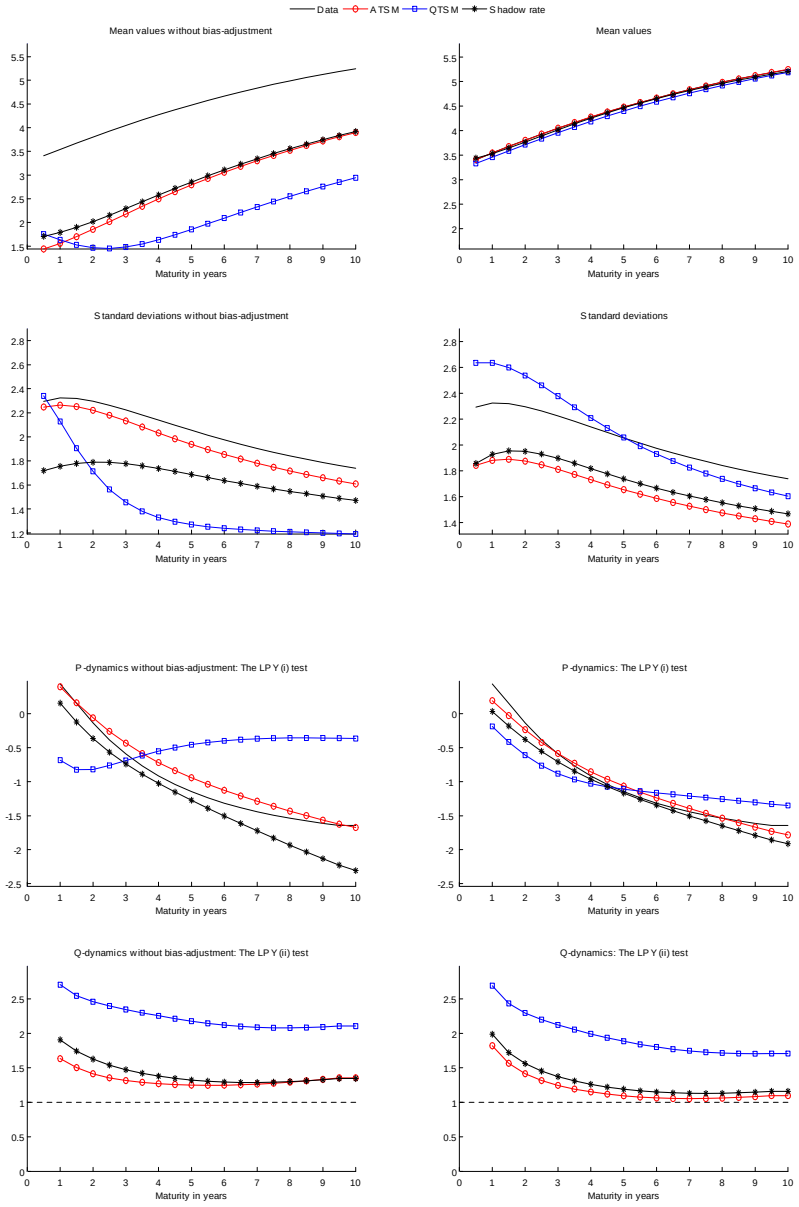


Figure 6: Sample from 1961-2013: Unconditional moments in four-factor models
All model-based moments are obtained from simulated time series of 100,000 observations.

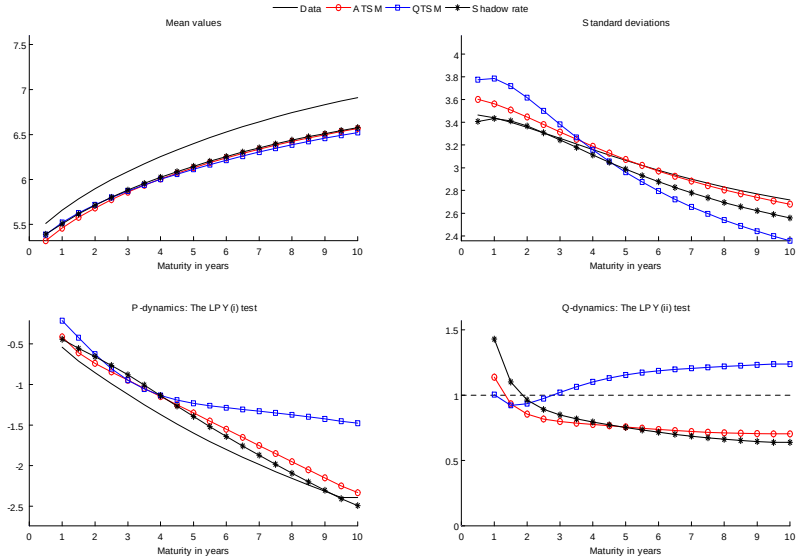


Figure 7: Sample from 1990-2013: Unconditional moments in four-factor models
All model-based moments are obtained from simulated time series of 100,000 observations.

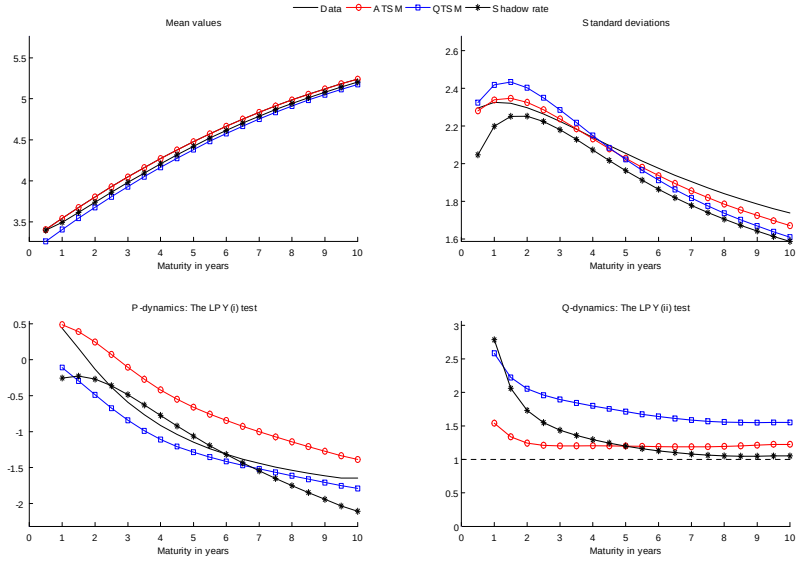


Figure 8: Sample from 1961-2013: Conditional volatilities for bond yields

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 in time period t is computed from a local linearization of bond yields at $\hat{\mathbf{x}}_{t-1}$.

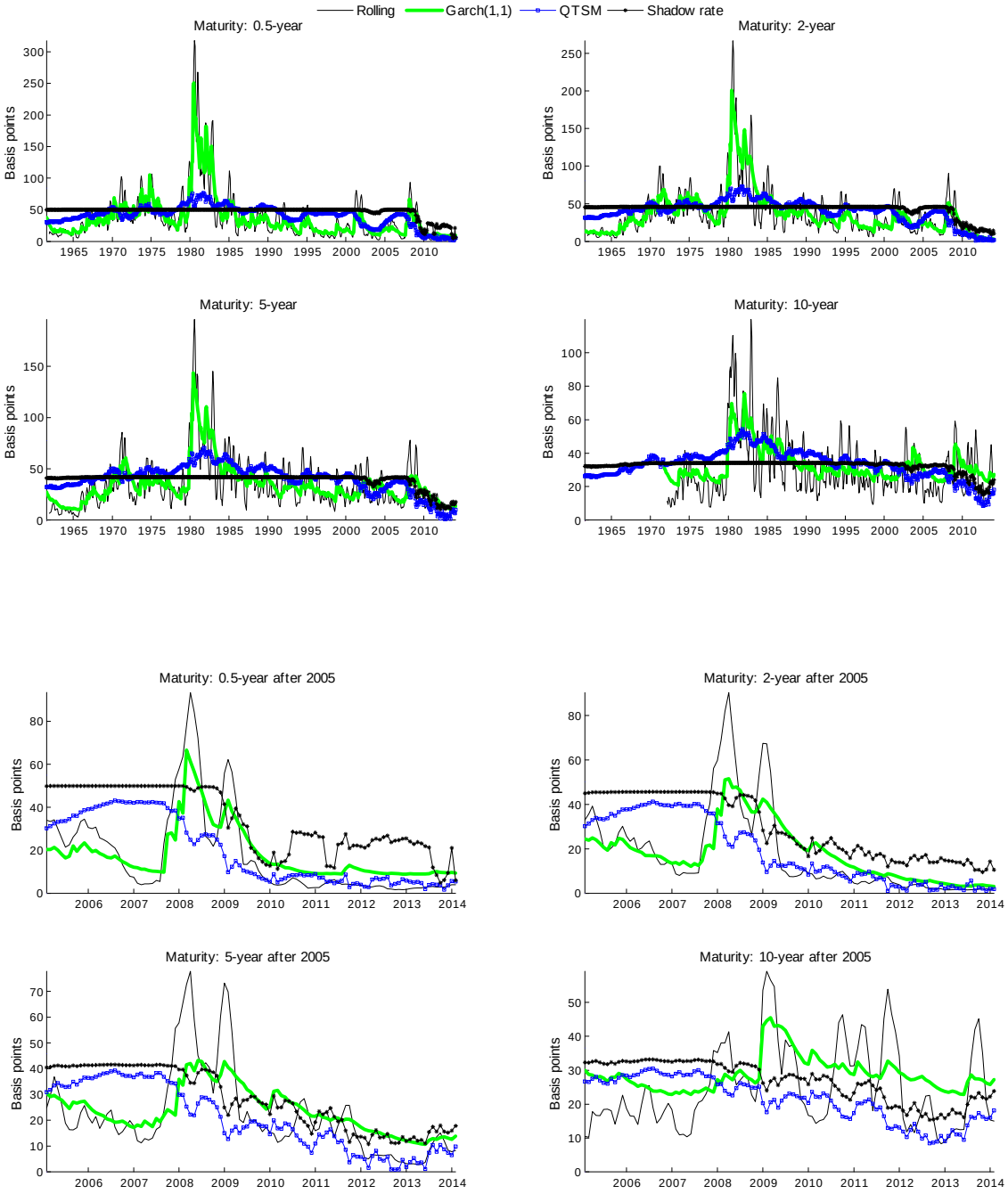


Figure 9: Sample from 1990-2013: Conditional volatilities for bond yields

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 in time period t is computed from a local linearization of bond yields at $\hat{\mathbf{x}}_{t-1}$.

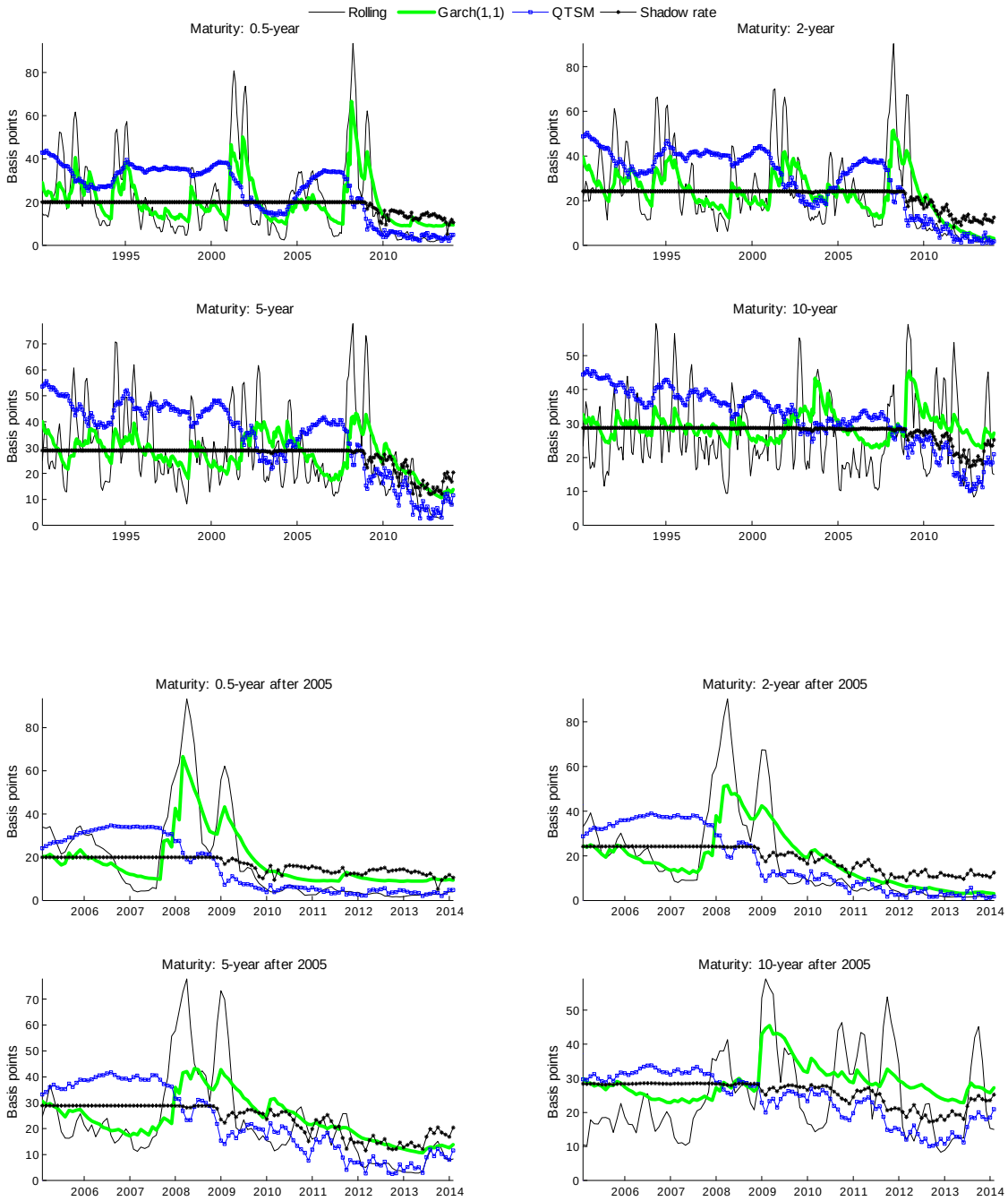


Figure 10: Forecasting results by maturity: model estimation starting in 1961
This figure reports the root mean squared prediction errors (RMSPEs) for out-of-sample forecasts from January 2005 to December 2013. The RMSPEs are generated from models estimated recursively from 1961 to the month prior to the forecast. The forecasted bond yields in the shadow rate models are computed by Monte Carlo integration using 10.000 draws.

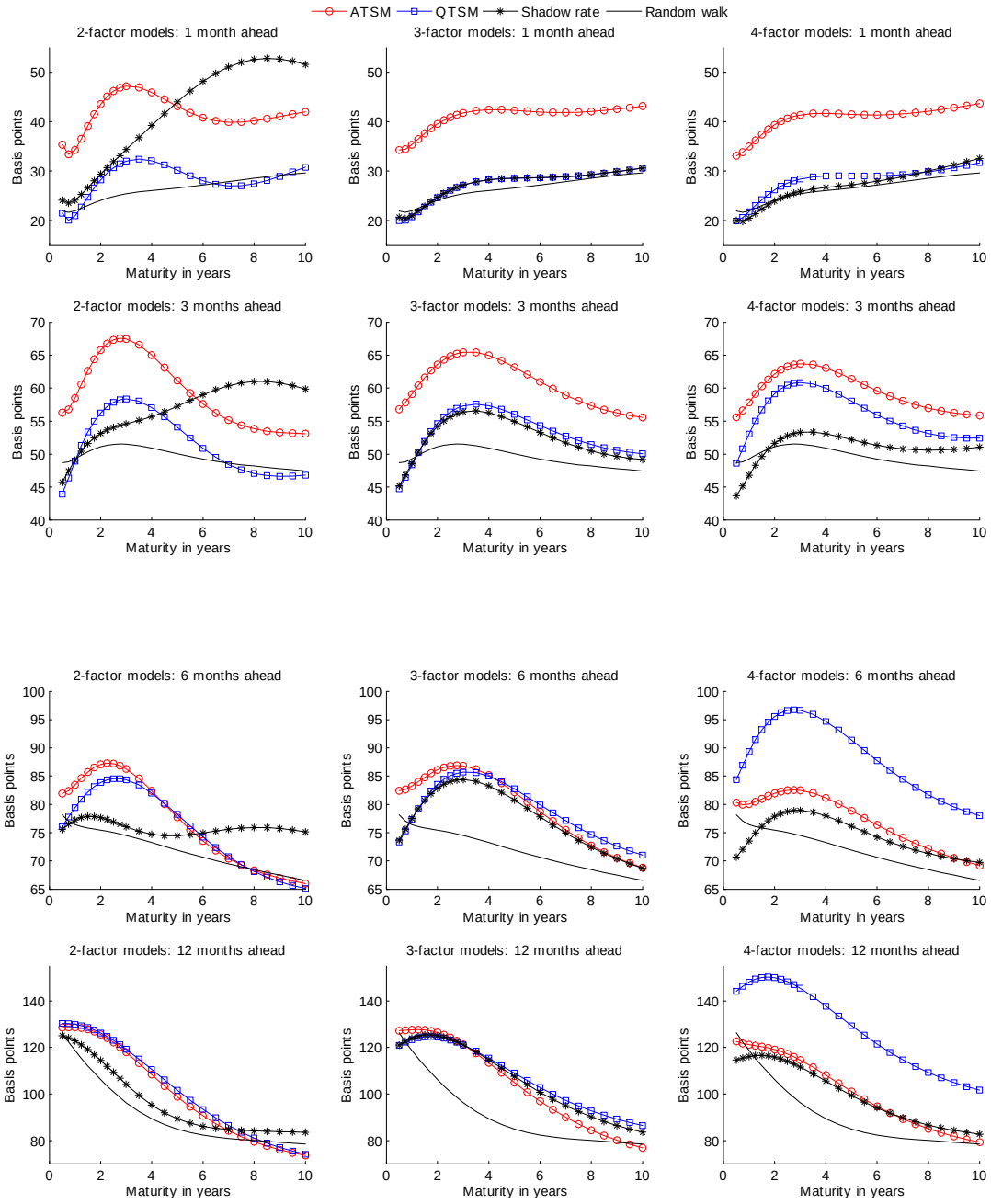


Figure 11: Forecasting results by maturity: model estimation starting in 1990
This figure reports the root mean squared prediction errors (RMSPEs) for out-of-sample forecasts from January 2005 to December 2013. The RMSPEs are generated from models estimated recursively from 1990 to the month prior to the forecast. The forecasted bond yields in the shadow rate models are computed by Monte Carlo integration using 10.000 draws.

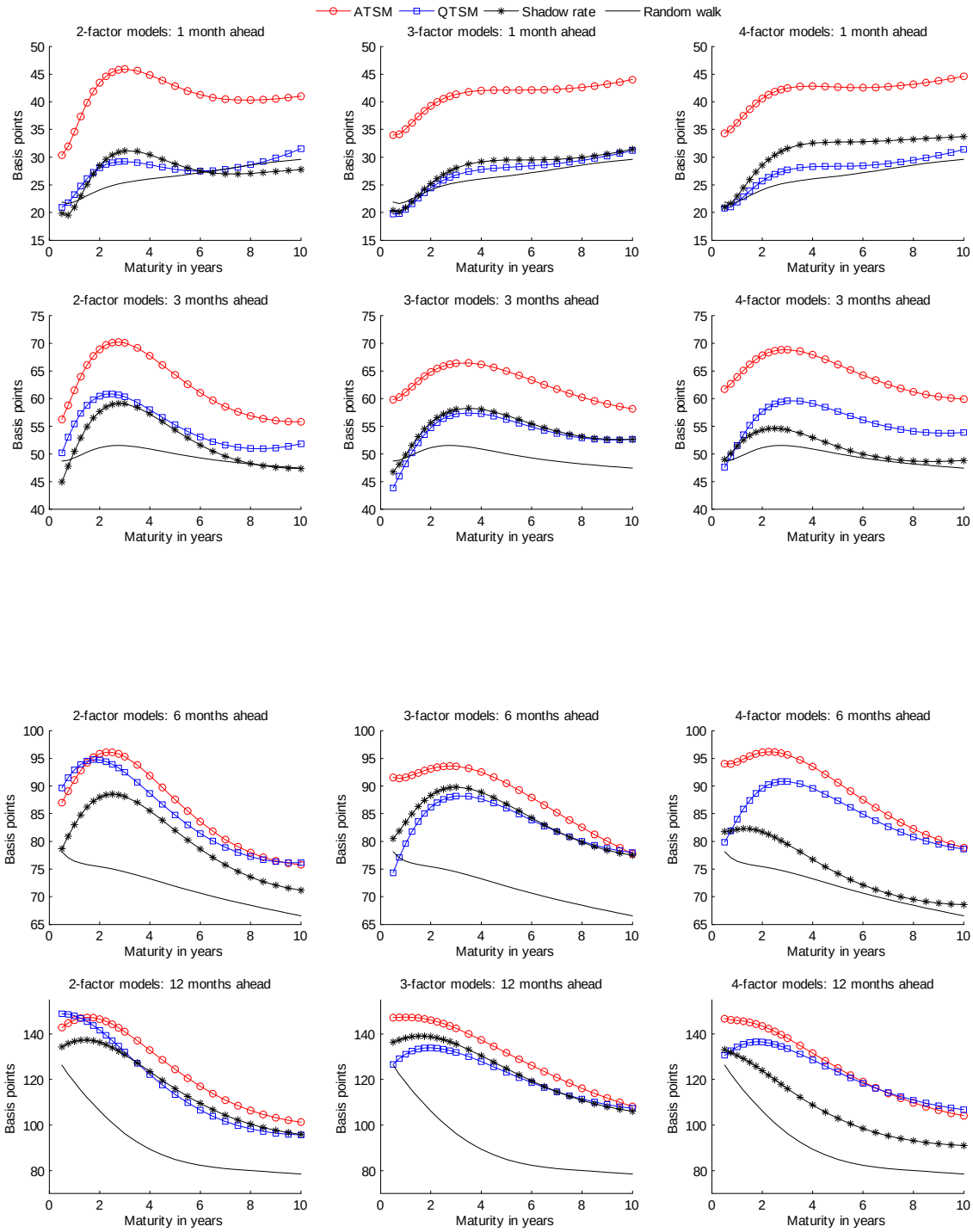


Figure 12: Forecasting illustration for the 0.5-year bond yield

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

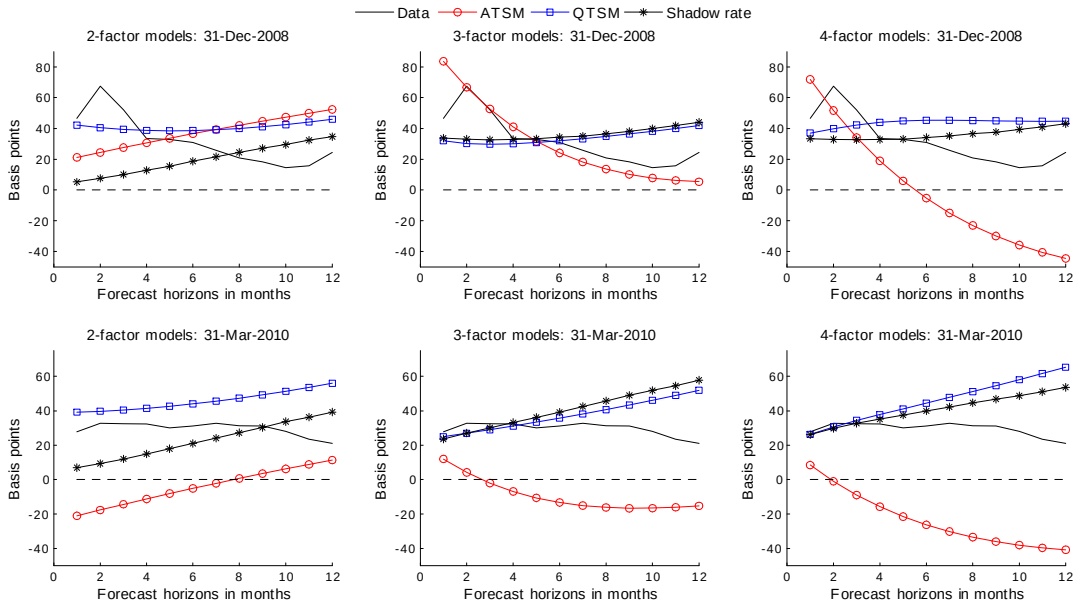


Figure 13: ATSM: Fraction of forecast the distribution below zero

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.

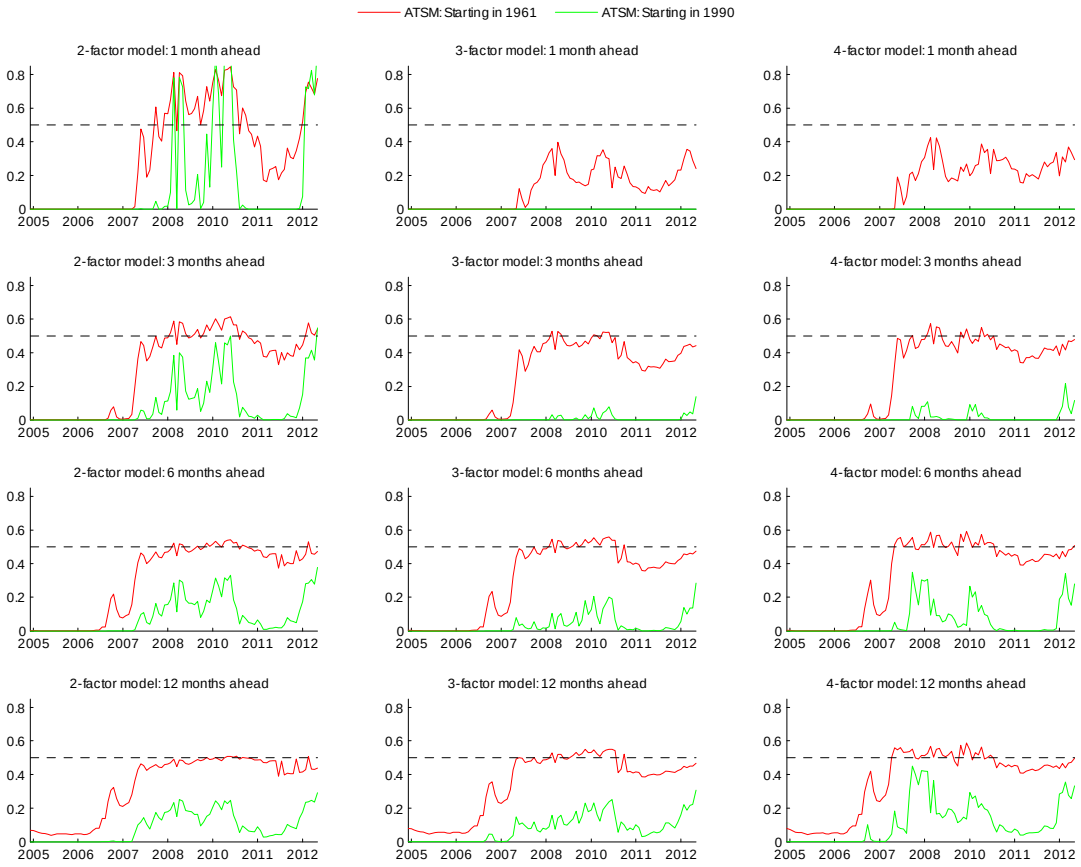


Figure 14: The hybrid model: Accuracy of the MC approximation to bond yields
These charts report the pricing errors when evaluating bond yields at $\{\hat{\mathbf{x}}_t\}_{t=1}^{T}$ for the estimated parameters in the long and short sample using 500 draws in the MC method. The true solution is approximated by the MC method using 100,000 draws and anti-thetic sampling.

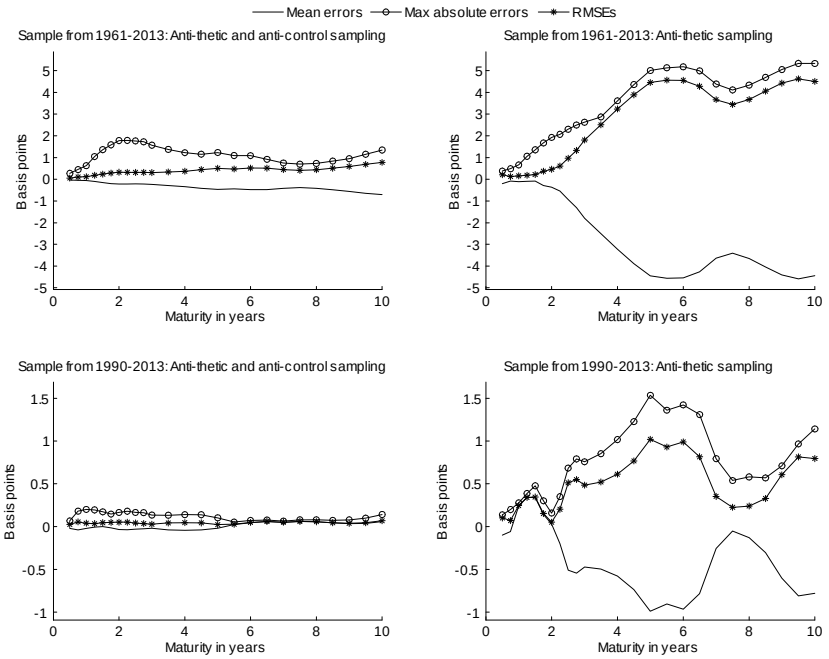


Figure 15: The hybrid model: In-sample fit

Charts in the first column report $100\sqrt{2 \times Q^{step3}_{2005:T}}$. Charts in the second column plots the estimated values of σ_t, measuring the in-sample fit of the yield curve at a given point in time.

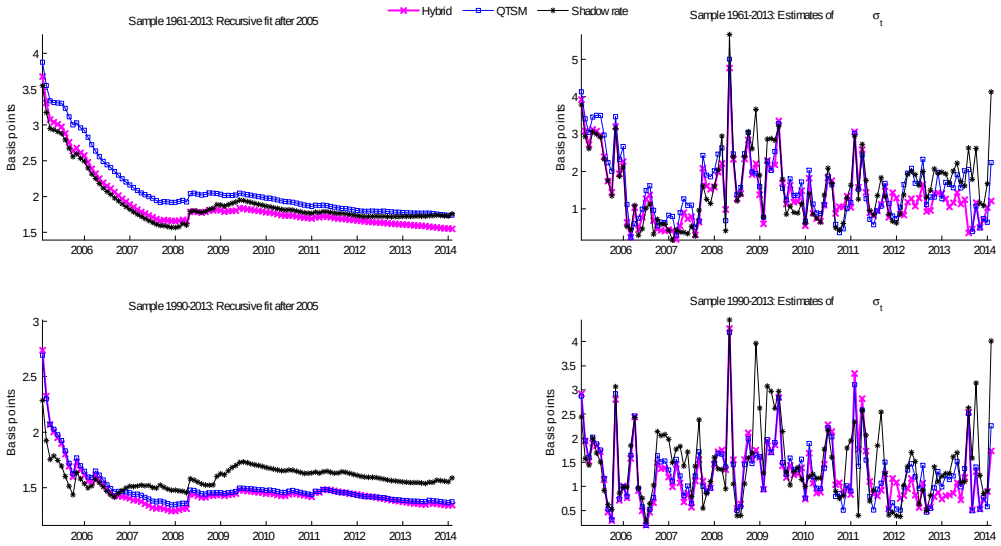


Figure 16: The hybrid model: Unconditional moments

All model-based moments are obtained from simulated time series of 100,000 observations.

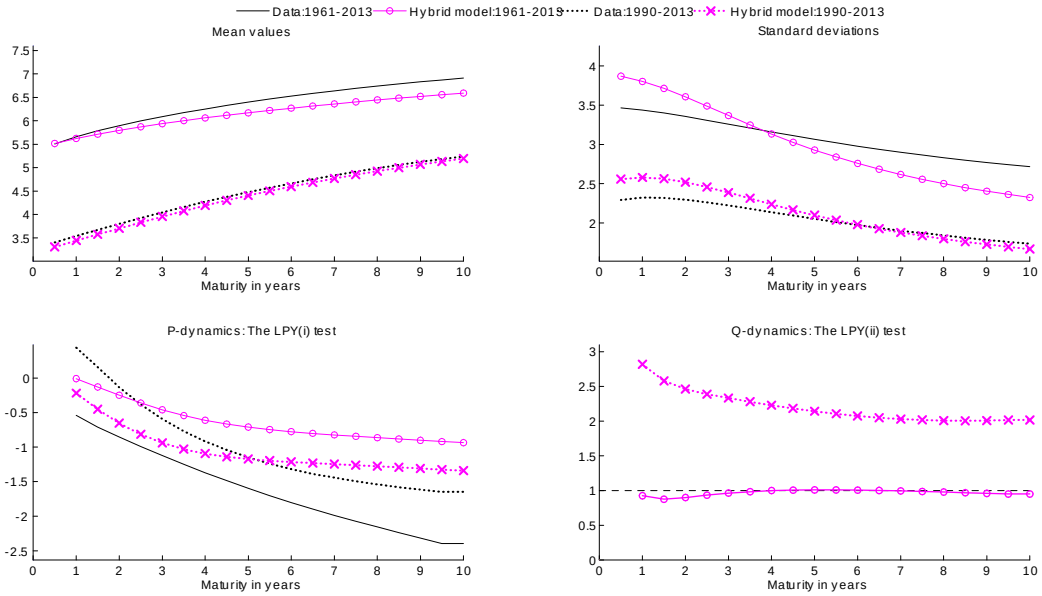
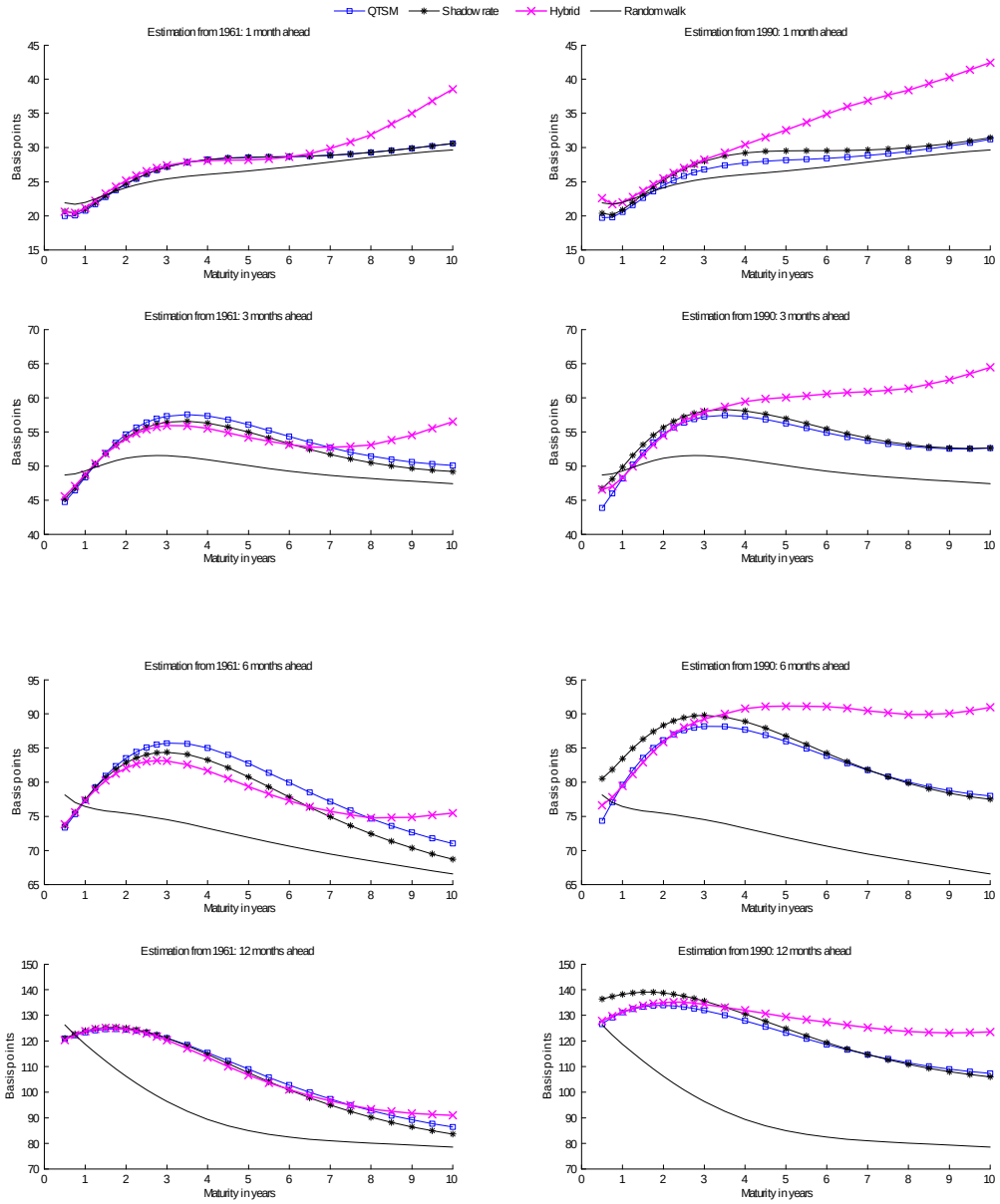


Figure 17: The hybrid model: Forecasting results by maturity

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 in the first column or from 1990 in the second column up to the month prior to the forecast. The forecasted bond yields in the shadow rate model and the hybrid model are computed by Monte Carlo integration using 10.000 draws.



Research Papers 2013

2014-30: Michael Creel and Dennis Kristensen: ABC of SV: Limited Information Likelihood Inference in Stochastic Volatility Jump-Diffusion Models

2014-31: Peter Christoffersen, Asger Lunde and Kasper V. Olesen: Factor Structure in Commodity Futures Return and Volatility

2014-32: Ulrich Hounyo: The wild tapered block bootstrap

2014-33: Massimiliano Caporin, Luca Corazzini and Michele Costola: Measuring the Behavioral Component of Financial Fluctuations: An Analysis Based on the S&P 500

2014-34: Morten Ørregaard Nielsen: Asymptotics for the conditional-sum-of-squares estimator in multivariate fractional time series models

2014-35: Ulrich Hounyo: Bootstrapping integrated covariance matrix estimators in noisy jump-diffusion models with non-synchronous trading

2014-36: Mehmet Caner and Anders Bredahl Kock: Asymptotically Honest Confidence Regions for High Dimensional

2014-37: Gustavo Fruet Dias and George Kapetanios: Forecasting Medium and Large Datasets with Vector Autoregressive Moving Average (VARMA) Models

2014-38: Søren Johansen: Times Series: Cointegration

2014-39: Søren Johansen and Bent Nielsen: Outlier detection algorithms for least squares time series regression

2014-40: Søren Johansen and Lukasz Gatarek: Optimal hedging with the cointegrated vector autoregressive model

2014-41: Laurent Callot and Johannes Tang Kristensen: Vector Autoregressions with Parsimoniously Time Varying Parameters and an Application to Monetary Policy

2014-42: Laurent A. F. Callot, Anders B. Kock and Marcelo C. Medeiros: Estimation and Forecasting of Large Realized Covariance Matrices and Portfolio Choice

2014-43: Paolo Santucci de Magistris and Federico Carlini: On the identification of fractionally cointegrated VAR models with the F(d) condition

2014-44: Laurent Callot, Niels Haldrup and Malene Kallestrup Lamb: Deterministic and stochastic trends in the Lee-Carter mortality model

2014-45: Nektarios Aslanidis, Charlotte Christiansen, Neophytos Lambertides and Christos S. Savva: Idiosyncratic Volatility Puzzle: Infl▫uence of Macro-Finance Factors

2014-46: Alessandro Giovannelli and Tommaso Proietti: On the Selection of Common Factors for Macroeconomic Forecasting

2014-47: Martin M. Andreasen and Andrew Meldrum: Dynamic term structure models: The best way to enforce the zero lower bound