A Shadow Rate or a Quadratic Policy Rule? The Best Way to Enforce the Zero Lower Bound in the United States

Martin M. Andreasen and Andrew Meldrum

A Shadow Rate or a Quadratic Policy Rule? The Best Way to Enforce the Zero Lower Bound in the United States

Martin Andreasen and Andrew Meldrum[*]

Abstract

We study whether it is better to enforce the zero lower bound (ZLB) in models of U.S. Treasury yields using a shadow rate model or a quadratic term structure model. We show that the models achieve a similar in-sample fit and perform comparably in matching conditional expectations of future yields. However, when the recent ZLB period is included in the sample, the models' ability to match conditional expectations away from the ZLB deteriorates because the time-series dynamics of the pricing factors change. In addition, neither model provides a reasonable description of conditional volatilities when yields are away from the ZLB.

I. Introduction

The Gaussian affine term structure model (ATSM) has been one of the most popular dynamic term structure models (DTSMs) of the past two decades. However, a well-known shortcoming of the Gaussian ATSM is its inability to ensure non-negative nominal bond yields. The fact that U.S. yields have reached historically low levels during the past several years has therefore generated substantial interest in alternative DTSMs that are consistent with the zero lower bound (ZLB).

Several studies of U.S. yields have extended the ATSM by truncating the short rate using the maximum function, which gives rise to the shadow rate model (SRM) proposed by Black (1995) (see e.g. Kim and Singleton (2012), Christensen and Rudebusch (2015), Bauer and Rudebusch (2016), and Wu and Xia (2016)). The SRM is appealing because it enforces the ZLB while preserving an approximately linear relationship between yields and pricing factors when yields are away from the ZLB. However, it is not the only way of enforcing the ZLB in DTSMs. Another possibility is to let the short rate be a restricted quadratic function of pricing factors, which leads to the quadratic term structure model (QTSM) studied by Ahn, Dittmar, and Gallant (2002), Leippold and Wu (2002), and Realdon (2006) among others.[1] An obvious advantage of the QTSM relative to the SRM is that the QTSM attains closed-form expressions for bond prices, which makes it computationally much more tractable than the (multifactor) SRM, where bond prices are unavailable in closed form.

Although both the SRM and QTSM enforce the ZLB, it is unclear which model performs best when estimated using U.S. data. Unlike the SRM, the QTSM implies a nonlinear relationship between yields and pricing factors when yields are away from the ZLB. Thus, the two models may have quite different implications for the conditional moments of yields. The aim of this article is to increase our understanding of ZLB-consistent DTSMs by analyzing the ability of the SRM and QTSM to match these moments of U.S. yields.

---

[1] The ZLB may also be enforced in ATSMs with square-root processes, as in Cox, Ingersoll, and Ross (1985) and Dai and Singleton (2000). Alternative ways to account for the ZLB have recently been suggested by Feunou, Fontaine, Le, and Lundblad (2015), Filipovic, Larsson, and Trolle (2017), and Monfort, Pegoraro, Renne, and Roussellet (2017).

We highlight the following three main conclusions. First, there is little to distinguish between the ability of the SRM and QTSM in matching conditional expectations of yields. The two models display similar abilities to fit yields in-sample, i.e. to match the conditional expectations of current yields, both away from and at the ZLB. When it comes to matching the conditional expectations of future yields, the SRM and QTSM also display similar performance, e.g. they have similar abilities to match short-rate expectations from surveys and to forecast yields out of sample. A standard 3-factor SRM does appear to offer some small advantages relative to a 3-factor QTSM in satisfying standard tests for conditional expectations of future yields, i.e. Campbell-Shiller regressions, risk-adjusted Campbell-Shiller regressions, and Mincer-Zarnowitz regressions for realized excess returns. However, the differences are not statistically significant at standard confidence levels and are largely eliminated by the addition of a fourth factor to the QTSM. Thus, it is not clear that any small benefits of using the SRM rather than the QTSM are sufficient to outweigh the greater computational complexity involved with estimating the SRM.

Perhaps more noteworthy than the small differences between the SRM and QTSM are their common failings. Indeed, our second main conclusion is that neither the SRM nor the QTSM appears to fully capture the change in the dynamics of U.S. yields that occurred when the short rate reached the ZLB. When the SRM and QTSM are estimated using a sample that ends before the recent ZLB period, both models replicate the well-known ability of the ATSM to satisfy the standard tests referred to above. However, when the sample is extended to cover the ZLB period, neither model continues to satisfy these tests when yields are away from the ZLB (the same is true for the ATSM). The problem seems to be that the time-series dynamics of the pricing factors change when the recent period of low yields is included in the sample, which in turn affects model-implied expectations away from the ZLB. We also show that extending the standard 3-factor version of the SRM and QTSM with a fourth or fifth pricing factor does not resolve this shortcoming.

Our third main conclusion is that neither the SRM nor the QTSM has the flexibility to provide a realistic description of the conditional volatilities of U.S. yields. Both models generate

a compression in the conditional volatilities of short-term yields at the ZLB, similar to what we observe in the data. However, they have counterfactual implications for conditional volatilities when yields are away from the ZLB because the SRM generates conditional volatilities that become approximately constant, while the QTSM generates a tight positive relationship between conditional volatilities and the level of yields.

This study is most closely related to Kim and Singleton (2012), who examined the performance of 2-factor models estimated using Japanese yields from 1995 to 2008. They found that both the SRM and QTSM achieve a reasonable fit to yields and are able to match movements in observed conditional volatilities close to the ZLB. In addition, their results suggest that the two models have similar implications for bond risk premia. However, we should be cautious about extrapolating their results to other countries, including the United States, because Japanese short rates were close to the ZLB for most of their sample period, whereas most studies of DTSMs in the United States consider samples where the short rate is also away from zero for an extended period. This difference makes the comparison of ZLB-consistent models more complicated when using U.S. yields, because we not only need to consider the performance of DTSMs at the ZLB, but also whether they preserve the desirable properties of the ATSM *away from* the ZLB.[2]

The rest of this article is organized as follows: In Section II, we outline the DTSMs and how they are estimated. Sections III, IV, and V respectively explore the models' ability to match the cross section of yields, the conditional expectations of future yields, and the conditional volatilities of future yields. In Section VI, we investigate the implications of the models for Sharpe ratios and return predictability through the lens of the "robust properties" of Sharpe ratios and return predictability documented for ATSMs by Duffee (2010). Here, we show that the SRM and QTSM reproduce these robust properties when yields are away from the ZLB, but that there are some differences between model-implied Sharpe ratios when short rates are close to the ZLB. In Section VII, we discuss our conclusions.

---

[2]Realdon (2016) studies ZLB-consistent models estimated using euro-area sovereign yields, which may provide a more relevant comparison with U.S. yields. However, the version of the SRM he estimates differs from the specification considered previously in the literature.

II. Models, Estimation Method, and Data

In this section we present the ATSM, SRM, and QTSM and explain how we estimate them using U.S. Treasury yields. Although the ATSM does not enforce the ZLB, we include it because it serves as a benchmark for assessing the performance of the ZLB-consistent models.

A. Dynamic Term Structure Models

The ATSM, SRM, and QTSM that we consider are standard models. They all specify yields as functions of a small number of pricing factors, which follow first-order Vector Autoregressions (VARs) under both the real-world and risk-neutral probability measures. In Section II.A.1, we present the common features of the models, while in Section II.A.2 we discuss how the functional forms for the short rate differ between the models, and how we compute bond yields.

1. Pricing Factors

In all three models, yields are driven by an $n_x \times 1$ vector of unobserved pricing factors $\mathbf{x}_t$ that follow a first-order VAR under the physical measure $\mathbb{P}$, i.e.

$$(1) \qquad \mathbf{x}_{t+1} = \mathbf{h}_0 + \mathbf{h_x}\mathbf{x}_t + \boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1},$$

where $\boldsymbol{\varepsilon}_{t+1} \sim \text{NID}(\mathbf{0}, \mathbf{I})$, $\mathbf{h}_0$ is an $n_x \times 1$ vector, and $\mathbf{h_x}$ and $\boldsymbol{\Sigma}$ are $n_x \times n_x$ matrices. In the absence of arbitrage, the time-t price of a j-period zero-coupon bond is $P_{t,j} = \mathrm{E}_t^{\mathbb{Q}}\left[\exp\left(-r_t\right) P_{t+1,j-1}\right]$. Here, r_t denotes the (risk-free) short rate and expectations are formed with respect to the risk-neutral probability measure $\mathbb{Q}$. Following Duffee (2002), the factors also follow a first-order VAR under $\mathbb{Q}$, i.e.

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

where $\boldsymbol{\varepsilon}_{t+1}^{\mathbb{Q}} \sim \text{NID}(\mathbf{0}, \mathbf{I})$, $\boldsymbol{\mu}$ is an $n_x \times 1$ vector, and $\boldsymbol{\Phi}$ is an $n_x \times n_x$ matrix.

We impose standard restrictions on equations (1) and (2) to identify the models. In all of the models, we require $\mathbf{\Sigma}$ to be lower triangular and let $\mathbf{\Phi}$ be a Jordan matrix with diagonal elements $\phi_{11} \leq \phi_{22} \leq ... \leq \phi_{n_x n_x}$.[3] In addition, we require $\boldsymbol{\mu} = \mathbf{0}$ in the ATSM and SRM, whereas it is sufficient in the QTSM to restrict $\mathbf{\Phi}\boldsymbol{\mu} \geq \mathbf{0}$ (as shown by Realdon (2016)). The parameters in $\mathbf{h}_0$ and $\mathbf{h_x}$ are unrestricted in all three models.

As is standard in most recent studies, we mainly focus on models with 3 pricing factors. However, some previous studies of ATSMs have argued that the inclusion of a fourth or fifth pricing factor can help to predict future yields, even though these additional factors have little explanatory power for the current cross section of yields (see e.g. Cochrane and Piazzesi (2008), Duffee (2011b), and Adrian, Crump, and Moench (2013)). We therefore also consider whether ZLB-consistent models with 4 or 5 factors deliver substantially different results than the benchmark 3-factor model. As far as we are aware, this issue has been neglected in previous studies of ZLB-consistent DTSMs.

2. Short Rate Equations

The models are closed by different functional forms for the short rate. Here, we adopt the general specification that $r_t = f(s_t)$, where $f(\cdot)$ is some function and s_t is a "short rate factor" that is affine in the pricing factors, i.e. $s_t = \alpha + \boldsymbol{\beta}'\mathbf{x}_t$, where α is a scalar and $\boldsymbol{\beta}$ is an $n_x \times 1$ vector. In the remainder of this section, we discuss the different functional forms for $f(\cdot)$.

ATSM The short rate in the ATSM is simply given by the short rate factor, i.e. $r_t = s_t$. We follow Joslin et al. (2011) and impose the identifying condition that $\boldsymbol{\beta} = \mathbf{1}$ (in addition to the restrictions on the $\mathbb{Q}$ dynamics mentioned in Section II.A.1). The standard 3-factor ATSM therefore has 22 free parameters (three in $\mathbf{h}_0$, nine in $\mathbf{h_x}$, six in $\mathbf{\Sigma}$, three in $\mathbf{\Phi}$, and one in α), with 4- and 5-factor ATSMs having 35 and 51 free parameters, respectively. The yield on a j-period zero-coupon bond is affine in the pricing factors, i.e. $y_{t,j}^{ATSM} = -(1/j)\left(A_j + \mathbf{B}_j'\mathbf{x}_t\right)$, where the recursive formulae for A_j and $\mathbf{B}_j$ are easily derived.[4]

---

[3]We restrict $\mathbf{\Phi}$ to have real eigenvalues. Strictly speaking, a maximally-flexible model would allow for complex eigenvalues (see Joslin, Singleton, and Zhu (2011)).

[4]An appendix providing further details on bond pricing in all three models is available on request.

SRM The short-rate in the SRM is given by $r_t = \max\{0, s_t\}$, which ensures that r_t cannot be negative. We apply the same restriction that $\boldsymbol{\beta} = \mathbf{1}$ as in the ATSM, meaning that the SRM has the same number of free parameters as the ATSM. Closed-form expressions for yields are not available for the SRM with multiple factors, and we therefore approximate yields using the second-order approximation of Priebsch (2013).

QTSM The short rate in the QTSM is given by $r_t = s_t^2$, which also ensures that r_t cannot be negative. We impose the parameter restrictions that $\alpha = 0$ and $\boldsymbol{\beta} = \mathbf{1}$, leaving a 3-factor QTSM with 24 free parameters (three in $\mathbf{h_0}$, nine in $\mathbf{h_x}$, six in $\boldsymbol{\Sigma}$, three in $\boldsymbol{\mu}$, and three in $\boldsymbol{\Phi}$), with 4- and 5-factor QTSMs having 38 and 55 free parameters, respectively. The yield on a j-period zero-coupon bond is quadratic in the pricing factors, i.e.
$y_{t,j}^{QTSM} = -(1/j)\left(\widetilde{A}_j + \widetilde{\mathbf{B}}_j'\mathbf{x}_t + \mathbf{x}_t'\widetilde{\mathbf{C}}_j\mathbf{x}_t\right)$, where the recursive formulae for $\widetilde{A}_j$, $\widetilde{\mathbf{B}}_j$, and $\widetilde{\mathbf{C}}_j$ are derived in Realdon (2006). The existence of closed-form bond prices means that the QTSM is computationally more tractable than the SRM, e.g. with 3 pricing factors and 1 period corresponding to 1 month, it takes around 1,000 times longer to compute yields with maturities up to 10 years in the SRM than in the QTSM.[5]

We should stress that this version of the QTSM is not maximally-flexible. To understand why this is the case, note first that the short rate in a maximally-flexible QTSM is given by

(3) $$r_t = \delta_0 + \boldsymbol{\delta}_x'\mathbf{x}_t + \mathbf{x}_t'\boldsymbol{\Delta}_{xx}\mathbf{x}_t,$$

where δ_0 is a scalar, $\boldsymbol{\delta}_x$ is an $n_x \times 1$ vector, and $\boldsymbol{\Delta}_{xx}$ is an $n_x \times n_x$ matrix. Together with the restrictions given in Section II.A.1, identification is achieved when $\boldsymbol{\delta}_x = \mathbf{0}$ and $\boldsymbol{\Delta}_{xx}$ is symmetric with diagonal elements equal to 1 (see Ahn et al. (2002) and Realdon (2006)). The ZLB may then be enforced by imposing the additional restrictions that $\delta_0 = 0$ and $\boldsymbol{\Delta}_{xx}$ is positive semi-definite.

The version of the QTSM that we consider imposes both the normalizing and ZLB

---

[5]For this comparison we use Matlab 2017 code running on a PC with a 3.40 GHz Intel Core i7-6700 processor and 16 GB RAM. It takes about 0.01 seconds to compute yields for a single time period in the SRM and about 10^{-5} seconds in the QTSM.

restrictions. However, it also imposes that the *off-diagonal* elements of $\mathbf{\Delta}_{xx}$ are equal to 1 (because $\mathbf{\Delta}_{xx} = \boldsymbol{\beta}\boldsymbol{\beta}' = \mathbf{1}_{n_x \times n_x}$). These further restrictions on $\mathbf{\Delta}_{xx}$ are convenient, because they imply that the short rate in the QTSM can also be written as a function of a linear combination of the pricing factors, as in the ATSM and SRM. Moreover, unreported results show that this convenient feature of our QTSM is obtained with hardly any loss of flexibility when fitting yields, because the off-diagonal elements in $\mathbf{\Delta}_{xx}$ are essentially equal to 1 if we freely estimate them using our sample of U.S. yields from 1990–2016. We therefore only report results from our over-identified QTSM throughout this article, with one exception discussed in Section V.

To understand why the off-diagonal elements in $\mathbf{\Delta}_{xx}$ turn out to be close to 1 when freely estimated, consider the decomposition $\mathbf{\Delta}_{xx} = \mathbf{A}\mathbf{D}\mathbf{A}'$, where $\mathbf{A}$ is a lower triangular matrix with diagonal elements equal to 1 and $\mathbf{D}$ is a diagonal matrix.[6] We can then rotate the factors to $\widetilde{\mathbf{x}}_t = \mathbf{A}'\mathbf{x}_t$, which implies that the short rate can be re-written as $r_t = \widetilde{x}_{1,t}^2 + d_{22}\widetilde{x}_{2,t}^2 + ... + d_{n_x n_x}\widetilde{x}_{n_x,t}^2$. As pointed out by Kim and Singleton (2012), if $\mathbf{\Delta}_{xx}$ is positive definite (i.e. $d_{ii} > 0$ for all i) then all of the rotated factors must be simultaneously equal to 0 if the short rate is equal to 0, and thus longer-term yields are constant when the short rate is at the ZLB, which is counter to empirical evidence. If instead $\mathbf{\Delta}_{xx} = \mathbf{1}_{n_x \times n_x}$, such that the rank of $\mathbf{\Delta}_{xx}$ is equal to 1, then only 1 eigenvalue of $\mathbf{\Delta}_{xx}$ is different from 0, and hence only the first rotated factor appears in the short rate. The remaining factors are therefore free to match longer-term yields even when the short rate is at the ZLB.[7]

B.   Estimation Method and Data

1.   The Sequential Regression Approach

In this section, we provide a brief description of how we estimate the models using the sequential regression (SR) approach of Andreasen and Christensen (2015). Previous studies have

---

[6]The just-identifying restrictions imply that the diagonal elements of $\mathbf{D}$ are given by $d_{ii} = 1 - \sum_{j=1}^{i-1} a_{ij}^2 d_{jj}$ for $i = 1, 2, ..., n_x$. Note that Kim and Singleton (2012) use a different (but invariant) normalization scheme, in which they restrict $\mathbf{\Sigma} = 0.1 \times \mathbf{I}$ but allow the diagonal elements of $\mathbf{\Psi}$ (and hence $\mathbf{D}$) to be free parameters.

[7]This follows from the fact that the $\mathbb{Q}$ dynamics of the rotated factors are $\widetilde{\mathbf{x}}_{t+1} = \mathbf{A}'\mathbf{\Phi}\boldsymbol{\mu} + \left(\mathbf{I} - \mathbf{A}'\mathbf{\Phi}\left(\mathbf{A}'\right)^{-1}\right)\widetilde{\mathbf{x}}_t + \mathbf{A}'\mathbf{\Sigma}\boldsymbol{\varepsilon}_{t+1}^{\mathbb{Q}}$. Since $\mathbf{A}$ is lower triangular, $\mathbf{I} - \mathbf{A}'\mathbf{\Phi}\left(\mathbf{A}'\right)^{-1}$ is upper triangular, and hence all factors affect the expected value of $\widetilde{x}_{1,t}$ under $\mathbb{Q}$ and therefore longer-term yields because $r_t = \widetilde{x}_{1,t}^2$ when $\mathbf{\Delta}_{xx} = \mathbf{1}_{n_x \times n_x}$.

typically estimated non-linear DTSMs by quasi-maximum likelihood using a non-linear extension of the Kalman filter. However, the asymptotic properties of this estimator are unknown, and the joint optimization across many parameters makes the estimation computationally challenging, even for 3-factor models. The SR approach overcomes these limitations because it provides consistent and asymptotically normal estimates and is computationally straightforward to implement. This computational simplicity greatly facilitates our comparison of DTSMs with up to 5 pricing factors.

Before we describe the SR approach, it is convenient to define two vectors containing partly overlapping sub-sets of parameters. First, $\boldsymbol{\theta}_1$ denotes the "risk-neutral parameters" that determine the relationship between the factors and yields, while $\boldsymbol{\theta}_2$ denotes the "time-series parameters" that determine the $\mathbb{P}$ dynamics in equation (1). Because $\boldsymbol{\Sigma}$ appears in both $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, it is convenient to further partition these vectors as $\boldsymbol{\theta}_1 \equiv \begin{bmatrix} \boldsymbol{\theta}'_{11} & vech\left(\boldsymbol{\Sigma}\right)' \end{bmatrix}'$ and $\boldsymbol{\theta}_2 \equiv \begin{bmatrix} \boldsymbol{\theta}'_{22} & vech\left(\boldsymbol{\Sigma}\right)' \end{bmatrix}'$. The vector $\boldsymbol{\theta}_{11}$ is given by $\begin{bmatrix} \alpha & diag(\boldsymbol{\Phi})' \end{bmatrix}$ in the ATSM and SRM, and by $\begin{bmatrix} \boldsymbol{\mu}' & diag(\boldsymbol{\Phi})' \end{bmatrix}'$ in the QTSM. Finally, the vector $\boldsymbol{\theta}_{22} = \begin{bmatrix} \mathbf{h}'_0 & vec\left(\mathbf{h}_{\mathbf{x}}\right)' \end{bmatrix}'$ in all three models.

Suppose in period t that we observe $n_{y,t}$ yields with maturities $m_1, m_2, ..., m_{n_{y,t}}$. The observed yield with maturity m_j at time t is given by $y_{t,m_j} = g_{m_j}\left(\mathbf{x}_t; \boldsymbol{\theta}_1\right) + v_{t,m_j}$, where $g_{m_j}\left(\mathbf{x}_t; \boldsymbol{\theta}_1\right)$ is the model-specific function that relates the pricing factors to the cross section of yields and v_{t,m_j} is a measurement error. We assume that these measurement errors have means equal to 0 and finite, positive-definite covariance matrices.

The SR approach has three steps. At Step 1, we jointly estimate $\boldsymbol{\theta}_1$ and the factors using cross-sectional regressions. For a given value of $\boldsymbol{\theta}_1$, we can estimate the factors in period t as

$$\hat{\mathbf{x}}_t\left(\boldsymbol{\theta}_1\right) = \arg\min_{\mathbf{x}_t \in \mathbb{R}^{n_x}} \frac{1}{2n_{y,t}} \sum_{j=1}^{n_{y,t}} \left(y_{t,m_j} - g_{m_j}\left(\mathbf{x}_t; \boldsymbol{\theta}_1\right)\right)^2.$$

To estimate $\boldsymbol{\theta}_1$ we pool the squared residuals from these regressions and minimize their sum with

respect to $\boldsymbol{\theta}_1$, i.e.

$$(4) \qquad \hat{\boldsymbol{\theta}}_1^{step1} = \arg \min_{\boldsymbol{\theta}_1 \in \Theta_1} \frac{1}{2N} \sum_{t=1}^{T} \sum_{j=1}^{n_{y,t}} \left(y_{t,m_j} - g_{m_j} \left(\hat{\mathbf{x}}_t \left(\boldsymbol{\theta}_1 \right); \boldsymbol{\theta}_1 \right) \right)^2.$$

Here, $\hat{\boldsymbol{\theta}}_1^{step1}$ denotes the Step 1 estimate of $\boldsymbol{\theta}_1$, $N \equiv \sum_{t=1}^{T} n_{y,t}$, and Θ_1 is the feasible domain of $\boldsymbol{\theta}_1$. Given standard regularity conditions, $\hat{\boldsymbol{\theta}}_1^{step1}$ is consistent and asymptotically normal when $n_{y,t} \to \infty$ for all t.[8]

At Step 2 of the SR approach, we estimate $\boldsymbol{\theta}_2$ using the estimated factors $\left\{ \hat{\mathbf{x}}_t \left(\hat{\boldsymbol{\theta}}_1^{step1} \right) \right\}_{t=1}^{T}$. As shown in Andreasen and Christensen (2015), when $\boldsymbol{\theta}_2$ is unrestricted, we can estimate equation (1) simply by running a modified regression with all second moments corrected for estimation uncertainty in the factors.

At Step 3 of the SR approach, we combine the estimates of $\boldsymbol{\Sigma}$ from Step 1 and Step 2 ($\hat{\boldsymbol{\Sigma}}^{step1}$ and $\hat{\boldsymbol{\Sigma}}^{step2}$, respectively) optimally and re-estimate $\boldsymbol{\theta}_{11}$ conditional on the optimal estimate of $\boldsymbol{\Sigma}$.

Preliminary analysis revealed that $\hat{\boldsymbol{\Sigma}}^{step1}$ tends to be estimated very inaccurately compared to $\hat{\boldsymbol{\Sigma}}^{step2}$, meaning that the time-series estimate $\hat{\boldsymbol{\Sigma}}^{step2}$ cannot be improved by adding information from the cross section of yields.[9] We therefore simply condition on $\hat{\boldsymbol{\Sigma}}^{step2}$ and re-estimate $\boldsymbol{\theta}_{11}$ as

$$\hat{\boldsymbol{\theta}}_{11}^{step3} = \arg \min_{\boldsymbol{\theta}_{11} \in \Theta_{11}} \sum_{t=1}^{T} \sum_{j=1}^{n_{y,t}} \left(y_{t,m_j} - g_{m_j} \left(\hat{\mathbf{x}}_t \left(\boldsymbol{\theta}_{11}, \hat{\boldsymbol{\Sigma}}^{step2} \right); \boldsymbol{\theta}_{11}, \hat{\boldsymbol{\Sigma}}^{step2} \right) \right)^2,$$

where $\hat{\boldsymbol{\theta}}_{11}^{step3}$ denotes the Step 3 estimate of $\boldsymbol{\theta}_{11}$ and Θ_{11} is the feasible domain of $\boldsymbol{\theta}_{11}$. We finally update our estimate of $\boldsymbol{\theta}_2$ by re-running Step 2 using the estimated factors $\left\{ \hat{\mathbf{x}}_t \left(\hat{\boldsymbol{\theta}}_{11}^{step3}, \hat{\boldsymbol{\Sigma}}^{step2} \right) \right\}_{t=1}^{T}$.

---

[8]These regularity conditions are stated by Andreasen and Christensen (2015). No further assumptions are imposed on the measurement errors, i.e. they may be correlated across both maturity and time.

[9]This finding is consistent with the results of Joslin et al. (2011) for ATSMs, because their estimates of $\boldsymbol{\Sigma}$ from the time-series dynamics of their (observed) factors hardly change when taking account of the cross section of yields.

2. Data

Our data set consists of end-month U.S. nominal zero-coupon Treasury yields computed using the method of Fama and Bliss (1987). Our sample starts in Jan. 1990 and ends in Dec. 2016. The starting point of this sample is broadly representative of recent studies of SRMs using U.S. data and is consistent with the findings of Rudebusch and Wu (2007), who argue there was a structural break in U.S. yields during the middle or late 1980s.[10]

The SR approach is constructed for settings with large cross sections, and we therefore include more yields than are typically used when estimating DTSMs. Specifically, we represent the yield curve by 27 points, using the 1-month yield, yields in the 3-month to 3-year range at 3-month intervals, and yields in the 3- to 10-year range at 6-month intervals.

III. Matching Conditional Expectations of Current Yields

This section explores how well the models match the cross section of yields, i.e. the conditional expectations of current yields. In Section III.A, we show that the 3-factor models give a very similar average in-sample fit. In Section III.B, we explain this result by showing that it is possible to rotate the factors in the three models such that they are approximately the same and have broadly similar loadings on yields. That said, the loadings in the SRM and QTSM do vary over time, which has potential consequences for the ability of the models to match the conditional expectations and volatilities of future yields, which we explore in Sections IV and V. Finally, in Section III.C, we show that adding a fourth and fifth factor makes only a modest difference to the in-sample fit of the models.

A. In-Sample Fit

We start by considering the fit of the 3-factor models to the short rate (i.e. the 1-month yield). Figure 1 shows the plot of the model-implied short rate factor (s_t) along the horizontal

---

[10]In an appendix that is available on request, we assess the robustness of our main conclusions to extending the sample back to June 1961. We find that most of our main conclusions continue to hold, although the inclusion of the ZLB period in the sample has a smaller effect than when the models are estimated using post-1990 data.

axes against both the model-implied short rate (black dots) and the observed 1-month yield in the data (gray stars). The figure illustrates the different assumptions regarding the functional form of the short rate, i.e. a linear mapping in the ATSM, the so-called "hockey stick" of the maximum function in the SRM, and a smooth quadratic function in the QTSM. However, the differences between the models' ability to fit the data are most apparent if we focus on the ZLB period from 2009–2016, as shown in the right-hand column in Figure 1.[11] The top, right chart highlights occasionally negative fitted short rates in the ATSM. The middle, right chart shows that the SRM avoids negative short rates by construction but produces fitted short rates that are occasionally too low. This is particularly evident when the shadow rate is negative and the fitted short rate is exactly 0%, whereas observed 1-month rates actually remain slightly positive throughout most of our sample. This means that the 3-factor QTSM (shown in the bottom, right chart) achieves a slightly closer fit to the short rate during the ZLB period than the SRM. Nonetheless, as reported in Table 1, the differences between the three models are small. The root mean squared error (RMSE) for the short rate implied by the ATSM is 8.8 basis points over the ZLB period, compared with 7.1 basis points by the SRM and 5.6 basis points in the QTSM.

The differences in fit are also small for longer-maturity yields. Table 1 also shows RMSEs for yields with selected maturities, in addition to the overall RMSE computed across all 27 considered maturities. The 3-factor models achieve similar average fit away from the ZLB and at the ZLB, e.g. the differences between the RMSEs implied by the SRM and QTSM are less than 2 basis points at almost all maturities (including those not reported in the table).

B. Factor Loadings

To explain why the 3-factor models give broadly the same in-sample fit, we next show that their factors may be rotated such that they are approximately the same in each of the three models and have broadly similar loadings on yields. More specifically, we apply invariant linear

---

[11]Throughout this paper, we refer to a "pre-ZLB period" ending in Dec. 2007 and a "ZLB period" from Jan. 2009–Dec. 2016. Although the target for the federal funds rate remained at 4.25% at the end of 2007 and did not fall to a target range of 0–0.25% until Dec. 2008, ending the pre-ZLB period in 2007 is consistent with the sample periods chosen by some recent studies of Gaussian ATSMs to avoid near-zero yields (for example, Bauer, Rudebusch, and Wu (2012)). Although the target range for the federal funds rate did rise to 0.25–0.5% in Dec. 2015, it stayed at this low level until Dec. 2016 (i.e. the final month in our sample), when it rose to 0.5–0.75%.

Figure 1: In-Sample Fit of the Short Rate

This figure plots the short rate factor (s_t) against the short rates implied by 3-factor models and observed 1-month yields. Charts to the left cover the full sample from Jan. 1990–Dec. 2016, while charts to the right focus on the ZLB period from Jan. 2009–Dec. 2016.
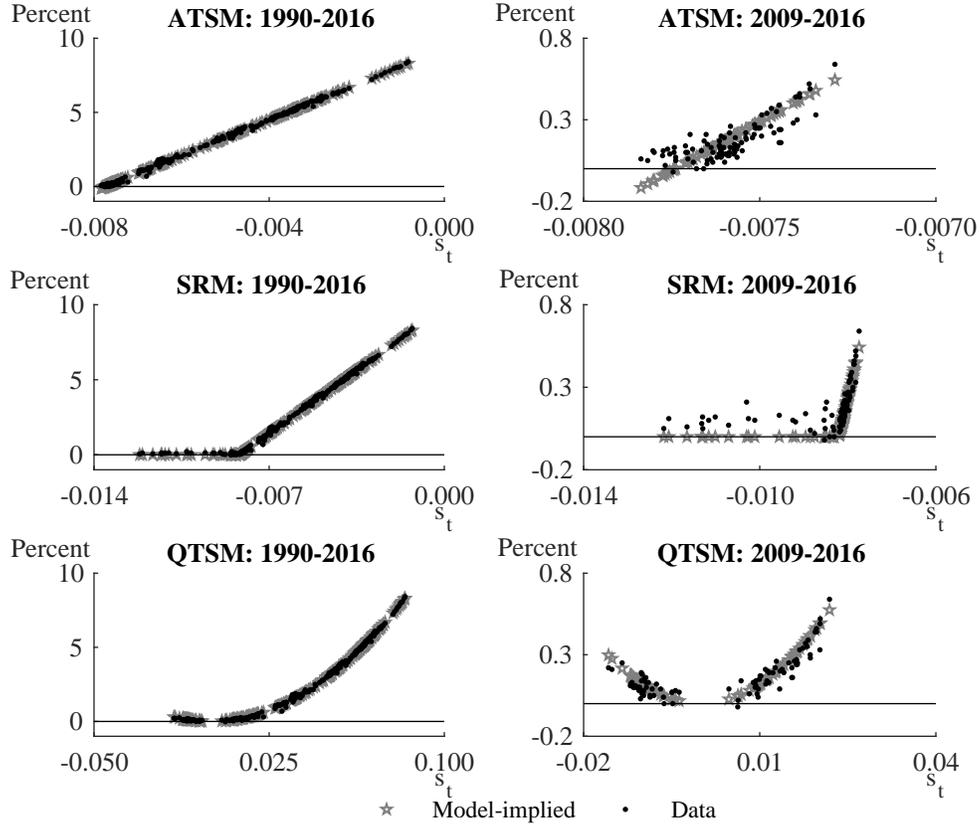
transformations to obtain rotated factors that may be interpreted as representing the level, slope, and curvature of the yield curve.

To obtain these rotated factors, we consider a vector $\mathbf{y}_t = \mathbf{g}\left(\mathbf{x}_t; \boldsymbol{\theta}_1\right)$ of model-implied yields with maturities from 6 months to 10 years at 6-month intervals. To a first-order approximation, the unconditional covariance matrix of these model-implied yields is given by

$$\boldsymbol{\Omega} \approx \text{var}\left[\mathbf{y}_t\right] = \mathbf{g}_\mathbf{x}\left(\bar{\mathbf{x}}; \boldsymbol{\theta}_1\right)\text{var}[\mathbf{x_t}; \boldsymbol{\theta_2}]\,\mathbf{g}_\mathbf{x}\left(\bar{\mathbf{x}}; \boldsymbol{\theta}_1\right)',$$

where $\mathbf{g}_\mathbf{x}\left(\cdot\right)$ denotes the first derivative of $\mathbf{g}\left(\cdot\right)$ evaluated at the unconditional mean of the factors $\bar{\mathbf{x}}$ and $\text{var}[\mathbf{x_t}; \boldsymbol{\theta_2}]$ is the variance of $\mathbf{x}_t$. We can decompose this covariance matrix as $\boldsymbol{\Omega} = \mathbf{V}\mathbf{D}\mathbf{V}'$, where $\mathbf{V}$ is an $n_y \times n_x$ matrix of eigenvectors and $\mathbf{D}$ is an $n_x \times n_x$ diagonal matrix of eigenvalues. The rotated factors $\mathbf{z}_t = \mathbf{V}'\mathbf{g}_\mathbf{x}\left(\bar{\mathbf{x}}; \boldsymbol{\theta}_1\right)\mathbf{x}_t$ are then approximately equal to the first

Table 1: In-Sample Fit of Bond Yields Across Maturity

This table reports the root mean squared errors (RMSEs) in annualized basis points between actual yields at selected maturities and the fitted values from the models with 3, 4, and 5 factors estimated using data from Jan. 1990–Dec. 2016. Panel $\mathcal{A}$ refers to the pre-ZLB period from Jan. 1990–Dec. 2007, while Panel $\mathcal{B}$ refers to the ZLB period from Jan. 2009–Dec. 2016. The final column of each panel reports the RMSE computed across all 27 considered maturities.

Maturity (months):	$\mathcal{A}$: 1990–2007							$\mathcal{B}$: 2009–2016						
	1	6	12	24	60	120	All	1	6	12	24	60	120	All
(a) ATSM														
$n_x=3$	9.3	6.4	7.5	3.0	4.0	5.3	4.9	8.8	2.7	7.0	4.1	7.6	13.0	6.4
$n_x=4$	4.6	5.3	3.3	3.1	2.4	4.3	3.4	3.6	3.3	3.6	2.6	3.7	7.4	4.0
$n_x=5$	3.6	4.9	2.7	3.6	2.4	5.3	3.5	2.4	3.5	2.2	2.9	2.9	6.5	3.7
(b) SRM														
$n_x=3$	8.7	6.5	7.0	2.9	3.7	5.2	4.7	7.1	4.6	5.0	3.3	6.0	11.1	5.3
$n_x=4$	4.5	5.1	2.8	3.0	2.5	4.4	3.3	2.9	3.4	3.2	2.3	3.7	6.2	3.8
$n_x=5$	3.2	4.4	2.6	2.5	2.5	2.5	2.8	2.9	3.1	2.5	2.5	3.4	5.0	3.4
(c) QTSM														
$n_x=3$	8.8	6.3	6.9	2.9	3.7	4.9	4.7	5.6	3.3	5.8	3.3	5.6	9.7	5.0
$n_x=4$	4.0	4.4	2.7	2.7	2.5	4.5	3.2	2.6	3.2	3.3	2.3	3.7	7.3	3.9
$n_x=5$	2.4	3.5	3.1	2.3	2.4	3.5	2.8	1.4	2.8	2.4	2.4	3.2	5.1	3.5

n_x principal components of model-implied yields in the SRM and QTSM (the decomposition is exact in the ATSM). We can then compute yields as a function of the rotated factors, i.e.

$$\mathbf{y}_t = \mathbf{f}\left(\mathbf{z}_t\right) \equiv \mathbf{g}\left(\left(\mathbf{V}'\mathbf{g_x}\left(\bar{\mathbf{x}};\boldsymbol{\theta}_1\right)\right)^{-1}\mathbf{z}_t;\boldsymbol{\theta}_1\right),$$

and obtain approximate factor loadings for the rotated factors.

The first row of Figure 2 shows the rotated factors normalized to have means equal to 0 and standard deviations equal to 1. As we might expect, the correlations between these factors across the three models are high, particularly for the first two factors, which explain most of the variation in yields.

The remaining rows in Figure 2 show the rotated factor loadings in Mar. 1990, which is the month with the highest observed short rate, and July 2012, when the short rate is at the ZLB and yields are low at all maturities. We first consider the ATSM (which has time-invariant

Figure 2: Factor Loadings

This figure plots the rotated factors and the associated factor loadings from 3-factor models estimated using a sample from Jan. 1990–Dec. 2016. For each model, we compute a first-order approximation to the unconditional covariance matrix of yields when the unrotated pricing factors are equal to their unconditional mean. We then decompose this covariance matrix, which allows us to rotate the factors to be approximately equal to the principal components of model-implied yields. The first row reports the rotated factors, normalized to have means equal to 0 and standard deviations equal to 1. The second and third rows report loadings of yields on the normalized rotated factors in Mar. 1990 and July 2012, respectively.

14
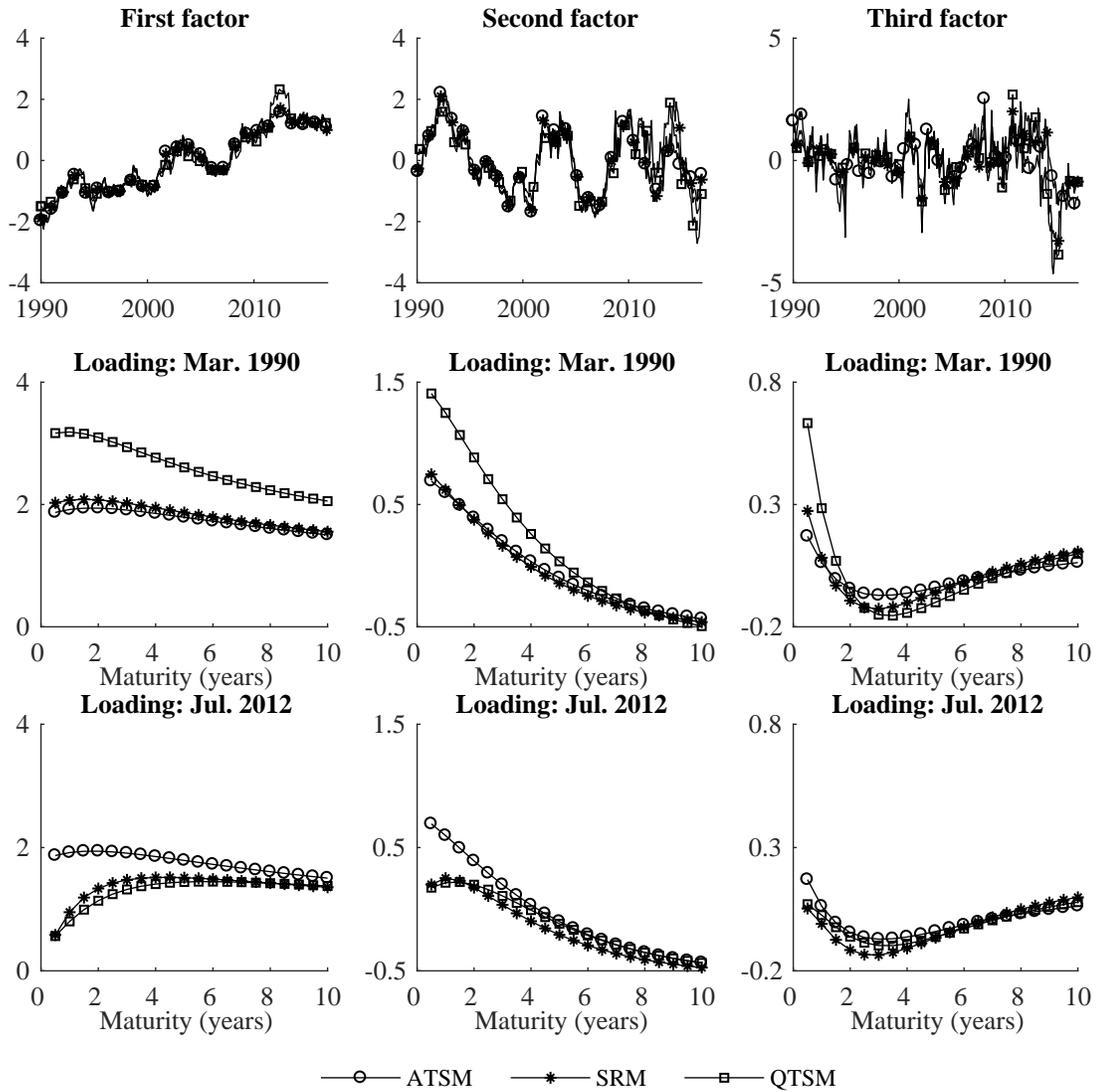
loadings). The first rotated factor has nearly constant loadings with maturity, meaning that it has the usual "level" interpretation. The loadings on the second and third rotated factors display the usual "slope" and "curvature" interpretations. We see broadly the same pattern for the SRM and QTSM at medium and long-term maturities. The fact that we can rotate the factors to be roughly equal and have similar loadings for medium and long-term yields explains why the three models display roughtly similar in-sample fit at these maturities.

However, there are some more notable differences between the rotated factor loadings at short maturities. When yields are far from the ZLB, the factor loadings in the SRM are (unsuprisingly) close to those in the ATSM. However, at the ZLB, the loadings in the SRM for short-term yields on the first and second factor decline. The QTSM produces a similar compression in these loadings at the ZLB, but it also generates loadings for the level and slope factors that are substantially larger than in the ATSM and SRM when yields are high.

The fact that the factor loadings in the ZLB-consistent models vary depending on the level of yields means that those models imply a trade-off between matching current yields and matching conditional expectations of future yields. To understand why, recall that we essentially estimate the $\mathbb{P}$ dynamics of the models by minimizing the squared differences between 1-month-ahead expectations of the estimated factors and subsequent realizations. In the ATSM, this is essentially equivalent to achieving the best possible 1-step-ahead predictions of yields because there is a linear mapping between factors and yields. In the ZLB-consistent models, however, the conditional expectations of yields also depend on the volatilities of the pricing factors. Thus, the estimated time-series dynamics of the factors in the ZLB-consistent models do not necessarily provide the best possible 1-step-ahead predictions of yields, which may hinder the ability of those models to match conditional expectations of future yields. This concern seems likely to be more acute for the QTSM, which implies strong non-linearities even when yields are far from the ZLB, unlike the SRM. In Section IV, we therefore compare the ability of the three models to match the conditional expectations of future yields.

The variation in the factor loadings in the SRM and QTSM also imply that the conditional volatilities of yields vary over time, even with pricing factors that have constant

conditional volatilities. We therefore also explore the implications of the models for conditional volatilities in Section V.

C.   Increasing the Number of Factors

We conclude our analysis of the in-sample fit of the models by considering whether it can be materially improved by moving beyond the standard 3 pricing factors. The results reported in Table 1 show that the addition of a fourth factor does indeed reduce the average fitting errors, with the largest gains coming for the 1-month yield. However, the improvements are not economically significant. For example, during the ZLB period, the RMSE for the 1-month rate falls nearly 4 basis points in the SRM and about 3 basis points in the QTSM with the addition of a fourth factor. Adding a fifth factor results in even smaller reductions in fitting errors, and unreported results show that repeating the exercise in Section III.B with a 5-factor model produces loadings for the fifth factor that are negligible. These results are consistent with the findings of Duffee (2010), who shows that the fourth and fifth factors in ATSMs estimated using pre-ZLB data have only modest effects on the cross section of yields. However, he also shows that these factors may significantly affect expectations of future yields, and we therefore consider this possibility in the following section.

IV.   Matching Conditional Expectations of Future Yields

We now turn to the models' ability to match conditional expectations of future yields. Because those conditional expectations cannot be observed directly, we use four alternative approaches to evaluate the models' performance. In Section IV.A, we consider the two "linear projections of yields" (LPY) tests of Dai and Singleton (2002), which examine whether the models can replicate the desired slope coefficients from standard and risk-adjusted Campbell-Shiller regressions. In Section IV.B, we report Mincer-Zarnowitz regressions of observed excess returns on model-implied predicted excess returns. In Section IV.C, we consider how well model-implied short-rate expectations match the corresponding expectations implied by

surveys of professional forecasters. Finally, in Section IV.D, we evaluate the out-of-sample forecasting performance of the models.

A. Campbell-Shiller Regressions

1. The LPY Tests of Dai and Singleton (2002)

The two LPY tests proposed by Dai and Singleton (2002) provide a standard approach for testing the ability of DTSMs to match the conditional mean of future yields. The first test (LPY(i)) examines whether DTSMs imply population slope coefficients from Campbell and Shiller (1991) regressions that match those estimated in the data, i.e. the loadings ϕ_j in

$$(5) \qquad y_{t+m,j-m} - y_{t,j} = \delta_j + \phi_j \frac{m}{j-m} \left(y_{t,j} - y_{t,m} \right) + u_{t,j},$$

with $u_{t,j} \sim \text{IID}(0, \text{var}\,(u_{t,j}))$ for $j = m+1, m+2, ..., K$. The second test (LPY(ii)) examines whether yields within the observed sample obey the expectations hypothesis once adjusted for model-implied term premia. This corresponds to testing whether the loadings $\phi_j^{\mathbb{Q}}$ are equal to 1 in the risk-adjusted version of equation (5)

$$(6) \quad y_{t+m,j-m} - y_{t,j} - \left(c_{t+m,j-m} - c_{t,j-m} \right) + \frac{m}{j-m} \theta_{t,j-m} = \delta_j^{\mathbb{Q}} + \phi_j^{\mathbb{Q}} \frac{m}{j-m} \left(y_{t,j} - y_{t,m} \right) + u_{t,j}^{\mathbb{Q}}.$$

Here, $c_{t,j} \equiv y_{t,j} - (1/j) \sum_{i=0}^{j-1} \text{E}_t \left[r_{t+i} \right]$ is the term premium in the j-period yield, $\theta_{t,j} \equiv f_{t,j} - \text{E}_t \left[r_{t+j} \right]$ is the term premium in the forward rate $f_{t,j} \equiv -\log\left(P_{t,j+1}/P_{t,j} \right)$, and $u_{t,j}^{\mathbb{Q}} \sim \text{IID}\left(0, \text{var}\,\left(u_{t,j}^{\mathbb{Q}} \right) \right)$.

The possibility that the linear relationships in equations (5) and (6) may change as yields approach the ZLB is of particular relevance for this study. One option would be to examine the models' ability to produce the desired loadings both in periods away from the ZLB and in periods where the ZLB is binding. However, the slope coefficients in the data for the LPY(i) test are estimated imprecisely when only using the short ZLB period, which means that such a comparison is not particularly informative. Moreover, while there is no uncertainty attached to

the desired slope coefficients of 1 in the LPY(ii) tests, the model-implied coefficients during the short ZLB period will be imprecisely estimated and hence also not particularly informative.

We therefore employ an alternative procedure. We first evaluate the models' ability to satisfy the LPY tests when yields are away from the ZLB and the models are estimated using only pre-ZLB data. Next, we consider whether extending the sample for the model estimation to include the ZLB period affects the models' ability to match the LPY tests when yields are away from the ZLB. If a model specifies the ZLB correctly, including yields at the ZLB in the model estimation should not adversely affect the properties of the model when yields are away from the ZLB.

2. LPY(i) Test Results

For the LPY(i) test the details of our two-step procedure are as follows. In the first step, we carry out the LPY(i) test for models estimated using a pre-ZLB sample, conditioning on yields being away from the ZLB. Specifically, we estimate the models using a sample ending in Dec. 2007. Given these estimates, we then simulate 1,000 samples of the same length as in the data (216 months), discarding any simulated paths where any yield is below 1% in any period. For each of the simulated sample paths, we estimate equation (5) using a horizon (m) of 6 months and compute the mean estimate of ϕ_j. We compare these model-implied loadings conditional on yields being above 1% with the loadings in the data for the period ending in Dec. 2007.

In the second step, we carry out the same test for models estimated on a sample that includes the ZLB period. Specifically, we estimate the models using a sample ending in Dec. 2016. Given these estimates, we repeat the above simulation procedure, conditioning on yields being above 1%. We again compare the resulting model-implied loadings conditional on yields being above 1% to the loadings in the data for the period ending in Dec. 2007.

We start by analyzing the performance of the standard 3-factor models. The left column in Figure 3 shows results for the models estimated using a sample ending in 2007. The heavy black lines indicate the estimated Campbell-Shiller loadings (ϕ_j) in the data. The loadings

implied by the ATSM, SRM, and QTSM with 3 factors (shown using triangular markers) generally match the Campbell-Shiller loadings in the data closely, which is consistent with previous findings for Gaussian ATSMs.

However, extending the sample for estimating the the model estimation to 2016 distorts the ability of the 3-factor models to pass the LPY(i) test when yields are away from the ZLB. This is indicated by the charts in the right column in Figure 3, which show the loadings implied by models estimated using the sample ending in 2016. All of the 3-factor models now produce Campbell-Shiller loadings conditional on yields being away from the ZLB that fall too quickly with maturity, with the loadings at long maturities outside the 95% confidence interval for the estimates of ϕ_j in the data.

To examine the source of this deterioration against the LPY(i) test once the ZLB period is included in the sample, we take the estimated risk-neutral parameters $\widehat{\boldsymbol{\theta}}_{11}$ and $\hat{\boldsymbol{\Sigma}}^{step2}$ from the 3-factor models estimated using the sample ending in 2016, but exclude the period after 2007 when estimating the $\mathbb{P}$ dynamics at Step 3 of the SR approach. Unreported results show that the resulting model-implied LPY(i) loadings are similar to those that we obtain when we estimate $\widehat{\boldsymbol{\theta}}_{11}$ and $\hat{\boldsymbol{\Sigma}}^{step2}$ using the sample ending in 2007. This result suggests that the observed deterioration against the LPY(i) test when the ZLB period is included is due to changes in the $\mathbb{P}$ dynamics (i.e. $\mathbf{h_0}$ and $\mathbf{h_x}$) at the ZLB, rather than changes in the short rate equation or the $\mathbb{Q}$ dynamics.

We also note that the estimated Campbell-Shiller loadings in the data are lower for long-term yields when the sample is extended to 2016, e.g. the coefficients for the 10-year bond are -1.2 and -1.6 for the samples ending in 2007 and 2016, respectively. During the ZLB period, bond returns of a given size are therefore associated with a smaller slope of the yield curve at the start of the holding period, perhaps because the short rate cannot fall below the ZLB. While the inclusion of yields at the ZLB should not affect the ability of a correctly-specified model to match the desired Campbell-Shiller loadings when yields are away from the ZLB, in the considered models the relationship between the slope of the yield curve and subsequent bond returns during the ZLB period appears to have a material effect on conditional expectations of future yields away from the ZLB.

Figure 3: Campbell-Shiller Loadings Away from the ZLB

This figure reports Campbell-Shiller loadings implied by the models and the data. The loadings in the data are estimated using a sample from Jan. 1990–Dec. 2007. The 95% confidence intervals for these estimates are computed using a block bootstrap applied jointly to the regressand and the regressor in the Campbell-Shiller regressions (in the data) with a block length of 189 months and 5,000 repetitions. The model-implied loadings are the mean loadings from running 1,000 Campbell-Shiller regressions on simulated samples of 216 months, conditional on all yields being above 1% at all points in the simulated sample. In the left column, all models are estimated using a sample from Jan. 1990–Dec. 2007. In the right column, all models are estimated using a sample from Jan. 1990–Dec. 2016.
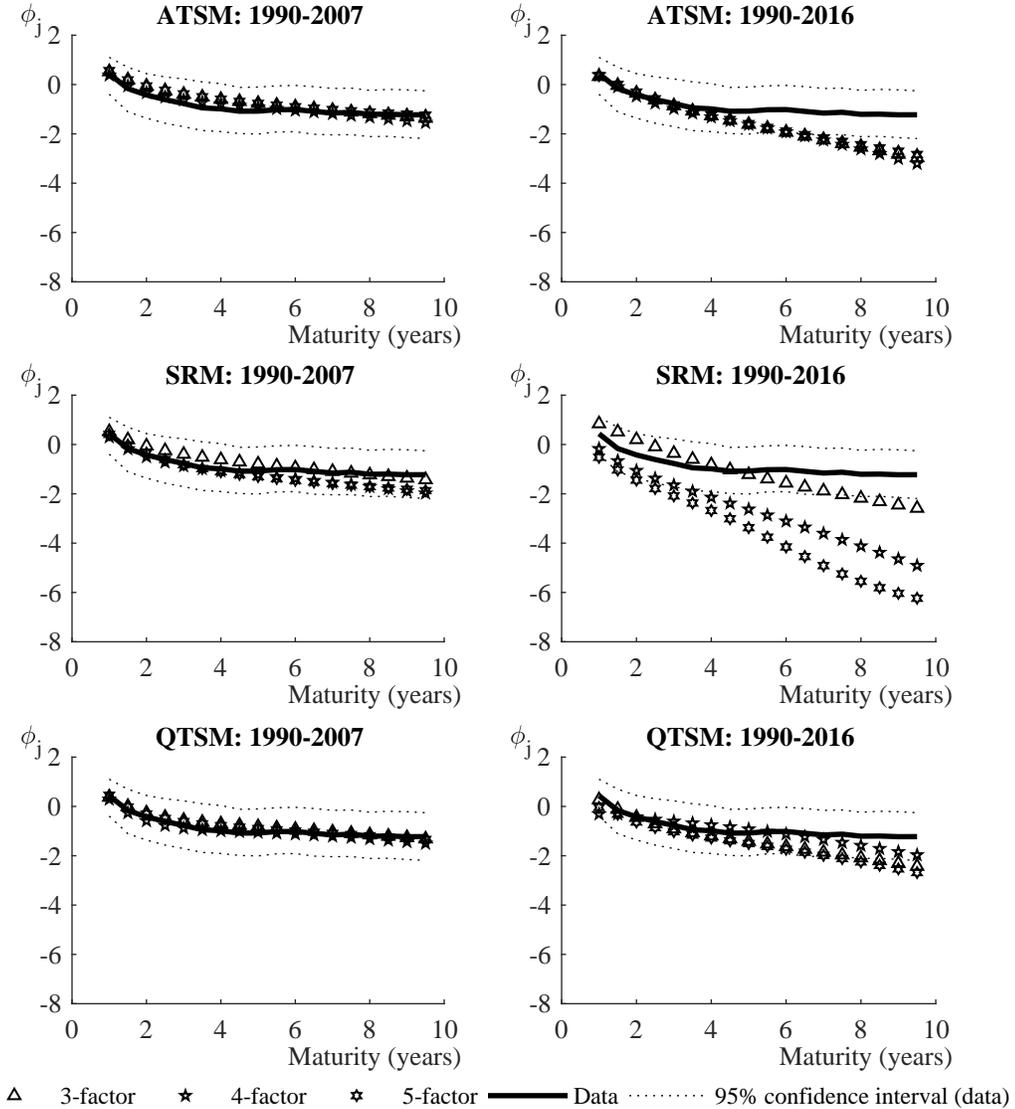

Figure 3 also shows the equivalent results for models with 4 and 5 factors. In short, the conclusions for the 3-factor models are broadly robust to the inclusion of additional factors. Adding a fourth or even fifth factor makes only small differences when the models are estimated over the pre-ZLB period. When the sample is extended to 2016, the 4-factor QTSM actually

performs a little better than the 3-factor version (with loadings that are just within the 95% confidence interval for the loadings in the data), although the 5-factor QTSM performs slightly worse. In contrast, adding factors to the SRM results in a substantial deterioration in performance against the LPY(i) test.

3. LPY(ii) Test Results

When implementing the LPY(ii) test we again adopt a two-step procedure. In the first step, we estimate the models using data from 1990–2007 and estimate equation (6) using model-implied term premia and a horizon (m) of 6 months. In the second step, we reestimate the models using the full sample ending in 2016 but again estimate equation (6) for 1990–2007, i.e. before reaching the ZLB.

The left column of Figure 4 shows the risk-adjusted Campbell-Shiller loadings implied by the 3-factor models estimated using a sample ending in 2007 (shown in triangular markers), with the dotted lines indicating the 95% confidence interval for these estimates. The estimated slope coefficients implied by the 3-factor ATSM and SRM are close to the desired value of 1 at all maturities. While the point estimates from the 3-factor QTSM are slightly above 1, the desired value nevertheless falls within the confidence interval at all maturities. We therefore conclude that all of the 3-factor models satisfy the LPY(ii) test when they are estimated on a sample of yields that is away from the ZLB.

However, the inclusion of the ZLB period in the sample to estimate the model parameters causes problems for all of the 3-factor models in matching the LPY(ii) test away from the ZLB, although the wide confidence intervals around the model-implied estimates mean that it is difficult to be conclusive. This is shown in the right column of Figure 4, which presents estimates of $\phi_j^{\mathbb{Q}}$ for 1990–2007 when the model parameters are estimated using the full sample from 1990–2016. The 3-factor ATSM and SRM continue to perform well for short- and medium-term yields, but display somewhat larger deviations from 1 for long-term yields. For the 3-factor QTSM, we see notable deviations from 1 at shorter maturities and even larger departures from 1 for long-term yields, although these differences are in general not statistically significant.

Figure 4: Risk-Adjusted Campbell-Shiller Loadings Away from the ZLB

This figure reports risk-adjusted Campbell-Shiller loadings from Jan. 1990–Dec. 2007, where term premia are obtained from models estimated using data from Jan. 1990–Dec. 2007 (left column) and from Jan. 1990–Dec. 2016 (right column). A well-specified model should return loadings equal to 1, which is highlighted using the heavy solid line. Conditional on the model estimates of term premia, the 95% confidence intervals for the risk-adjusted Campbell-Shiller loadings from the 3-factor models are computed using a block bootstrap applied jointly to the regressand and the regressor in the risk-adjusted Campbell-Shiller regressions with a block length of 189 months and 5,000 repetitions.
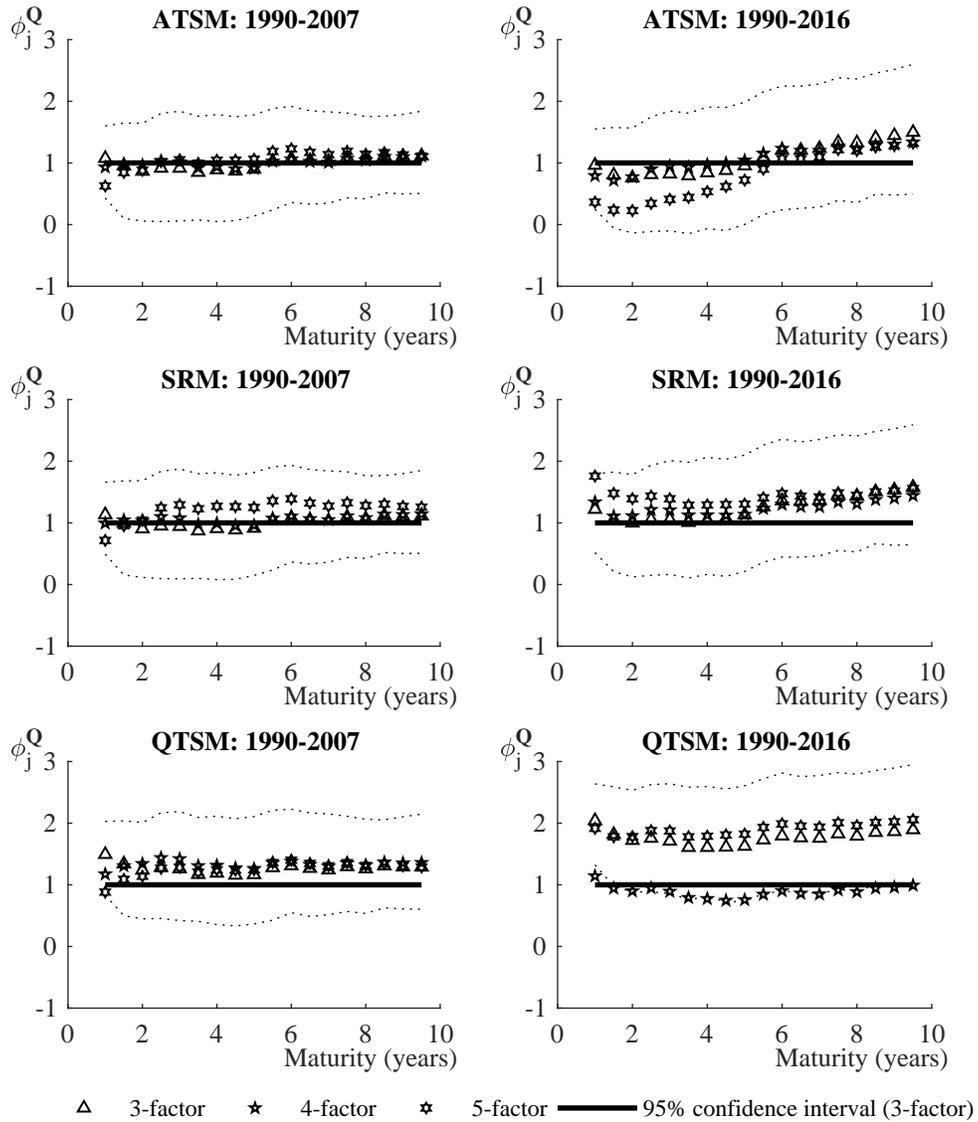
To shed light on what is causing the relatively large change in the loadings for the 3-factor QTSM, we repeat the additional exercise reported for the LPY(i) test. We take the estimated risk-neutral parameters $\widehat{\boldsymbol{\theta}}_{11}$ and $\hat{\boldsymbol{\Sigma}}^{step2}$ for the QTSM estimated using a sample ending in 2016, but exclude the period after 2007 when estimating the $\mathbb{P}$ dynamics at Step 3 of the SR approach. Unreported results show that the model-implied LPY(ii) loadings for the 1990–2007 period are similar to those that we obtain when we estimate $\widehat{\boldsymbol{\theta}}_{11}$ and $\hat{\boldsymbol{\Sigma}}^{step2}$ using the sample ending in 2007. This result suggests that the deterioration in the performance against the LPY(ii) test is due to a change in the $\mathbb{P}$ dynamics (i.e. $\mathbf{h}_0$ and $\mathbf{h_x}$) at the ZLB, rather than a change in the short-rate equation or the $\mathbb{Q}$ dynamics.

The addition of a fourth pricing factor again improves the performance of the QTSM, such that it achieves risk-adjusted Campbell-Shiller loadings close to the desired value of 1 even when the model is estimated using the full sample period. In contrast, adding a fourth factor to the SRM or a fifth factor to either the SRM or the QTSM again results in a deterioration in performance, as it does for the LPY(i) test.

B.   Mincer-Zarnowitz Regressions for Excess Bond Returns

Another test of a DTSM's ability to match conditional expectations uses Mincer and Zarnowitz (1969) regressions of realized m-period excess returns on model-implied expectations, i.e.

(7) $$rx_{t+m,m,j} = \alpha_{0,j} + \alpha_{1,j}\mathrm{E}_t\left[rx_{t,m,j}\right] + u_{t+m,m,j},$$

where $rx_{t+m,m,j} \equiv -\left(j-m\right)y_{t+m,j-m} + jy_{t,j} - my_{t,m}$. For a correctly specified model, the intercept $\alpha_{0,j}$ should equal zero and the slope coefficient $\alpha_{1,j}$ should equal 1. Because the results for the LPY tests raise the question of whether the models' ability to match the loadings in equation (7) depends on whether the ZLB period is included in the sample for estimating the model, we again adopt a two-step procedure. In the first step, we estimate the three models using a sample ending in 2007 and estimate equation (7) for model-implied expected returns

with a horizon (m) of 6 months. In the second step, we reestimate the model parameters using a sample ending in 2016, but again estimate equation (7) for excess returns from 1990–2007.

The charts in the left column of Figure 5 show the slope coefficients from equation (7) when the models are estimated using a sample ending in 2007, while the dotted lines report the associated 95% confidence intervals. Although the 3-factor models (shown with triangular markers) imply slope coefficients that are slightly larger than the desired value of 1, the desired value generally falls well within the confidence intervals. We therefore conclude that the 3-factor models can broadly satisfy the test posed by Mincer-Zarnowitz regressions when these models are estimated using a sample of yields that are away from the ZLB.

However, the performance of the 3-factor models again deteriorates when they are estimated using a sample that includes the ZLB period. The right column of Figure 5 shows estimates of the slope coefficients from equation (7), with the sample for estimating the models extended to 2016. The 3-factor models now imply slope coefficients for the pre-ZLB period that are above 1 at short maturies and below 1 at long maturities.[12]

We can understand what generates this pattern in the slope coefficients by examining the volatilities of excess returns in the data. The volatilities of excess returns on short-term yields are compressed during the ZLB period (as we discuss further in Section V). The ATSM implies that conditional volatilities are constant, meaning that it is unable to capture this volatility compression. When the sample is extended to include the ZLB period, the overall best fit is therefore obtained by lowering the volatilities of excess returns on short-term bonds in all periods, which leaves the volatilities of excess returns during the pre-ZLB period too low, and hence generates values of $\alpha_{1,j}$ above 1 for short maturities. However, at longer maturities the volatilities of returns do not fall by as much in the ZLB period. Instead, the pattern observed for the Campbell-Shiller loadings in Section IV.A.1 dominates, meaning that the relationship between the slope of the yield curve and bond returns becomes more negative at the ZLB. As we show in Section IV.A.1, the ATSM is unable to capture this effect, and the best fit is obtained by

---

[12]Unreported results reveal that all the 3-factor models give estimates of $\alpha_{0,j}$ in equation (7) that are close to and not significantly different from 0 when the models are estimated using a sample ending in Dec. 2007. The same generally also applies when the sample for estimating the models is extended to Dec. 2016, except for a few short maturities, where the intercepts in the SRM are close to but significantly different from 0.

Figure 5: Mincer-Zarnowitz Regression Slopes for Excess Returns Away from the ZLB

This figure reports slope coefficients from Mincer-Zarnowitz regressions of excess returns on an intercept and model-implied expected excess returns. Model-implied excess returns are computed by drawing 1,000 times from the conditional distributions of returns at each point in time. The point estimates and 95% confidence intervals for the 3-factor models are obtained using a block bootstrap procedure, with a block length of 189 months and 5,000 repetitions. The left column reports results when the model parameters are estimated using data from Jan. 1990–Dec. 2007. The right column reports results when the sample for estimating the model parameters is extended to Dec. 2016. The Mincer-Zarnowitz regressions are in both cases estimated using a sample from Jan. 1990–Dec. 2007. A well-specified model should return a slope coefficient equal to 1, which is highlighted using the heavy solid line.
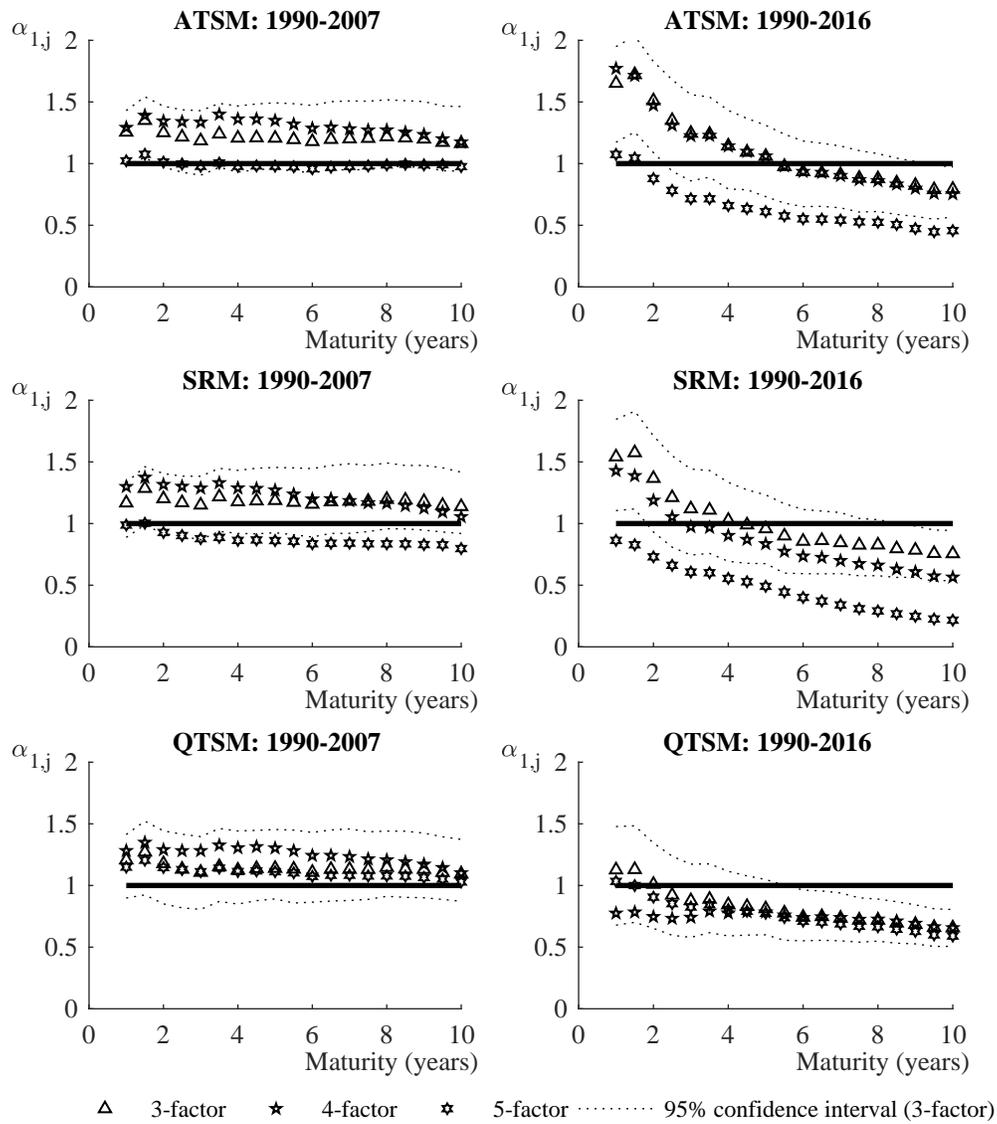
increasing the overall correlation between the slope of the yield curve and bond returns for all periods. This implies that excess returns from 1990–2007 become too volatile, and we therefore see values of $\alpha_{1,j}$ below 1 at long maturities.

Given that the 3-factor QTSM and SRM are unable to improve upon the ability of the ATSM to match the Campbell-Shiller loadings in Section IV.A, it is perhaps not surprising that these models also display values of $\alpha_{1,j}$ that are too low for long maturities when using the full sample for the model estimation. As we will show in Section V, the QTSM is marginally better than the SRM at matching the volatility compression of the short-term yields at the ZLB, which is why the QTSM generates values of $\alpha_{1,j}$ at short maturities that are closer to 1 than the SRM when using the full sample for model estimation.

These results are broadly robust to the inclusion of additional pricing factors, although the picture is less clear-cut than for the LPY tests. The addition of a fourth factor generally worsens the performance of the models slightly, both close to and away from the ZLB. The 5-factor ATSM and SRM (but not the QTSM) generally perform a little better than the corresponding 3-factor models when estimated using pre-ZLB data. When including the ZLB period in the model estimation, the 5-factor ATSM and SRM imply slope coefficients that are closer to 1 at short maturities, but their performance at longer maturities is substantially worsened. However, the 5-factor QTSM implies slope coefficients that are below 1 at all maturities. In short, based on Mincer-Zarnowitz regressions, it is difficult to make a compelling case for increasing the number of pricing factors beyond the standard 3.

C.   Matching Survey Expectations

A further test of the models' ability to match conditional expectations of future yields is whether they can match expectations implied by surveys of professional forecasters. We therefore construct a survey-based measure of short-rate expectations using the mean of responses to Blue Chip Financial Forecasts surveys of federal funds rate expectations. Because survey respondents are asked to report their expectations of the average federal funds rate over specific calendar periods, we linearly interpolate to compute measures of expectations at

constant horizons of 6, 12, 24, and 36 months. These measures are available monthly at the 6- and 12-month horizons and semi-annually at the 24- and 36-month horizons.

In this section, we only report results for models estimated using the full sample from 1990–2016, but compare model-implied expectations with surveys separately for the pre-ZLB period from 1990–2007 and the ZLB period from 2009–2016. Table 2 reports the RMSEs between the survey-based measure and expected model-implied short rates. The 3-factor models provide an almost identical average fit during the pre-ZLB period. During the ZLB period, more substantial differences emerge between the models, as the SRM consistently outperforms the QTSM. However, the differences between squared forecast errors are not statistically significant according to unreported Diebold-Mariano tests, perhaps because the ZLB period is relatively short.

Table 2: Matching Short-Rate Expectations from Surveys

This table reports the root mean squared errors in annualized percentage points between the model-implied expected short rate and the expected federal funds rate derived from Blue Chip Financial Forecasts surveys. The survey-based measure of the expected short rate is computed as the mean response across all survey participants. These expectations are linearly interpolated to compute expectations at the reported constant horizons. The model-implied short rate expectations are derived from models estimated using data from Jan. 1990– Dec. 2016.

	3-Factor Models		4-Factor Models		5-Factor Models	
	1990–2007	2009–2016	1990–2007	2009–2016	1990–2007	2009–2016
(a) 6 months ahead						
ATSM	0.47	0.24	0.48	0.23	0.48	0.29
SRM	0.47	0.15	0.44	0.16	0.42	0.16
QTSM	0.46	0.18	0.50	0.14	0.47	0.24
(b) 12 months ahead						
ATSM	0.80	0.39	0.79	0.35	0.85	0.63
SRM	0.80	0.21	0.72	0.26	0.70	0.25
QTSM	0.79	0.29	0.80	0.21	0.79	0.40
(c) 24 months ahead						
ATSM	1.66	1.48	1.65	1.50	1.79	2.00
SRM	1.68	1.27	1.56	1.22	1.53	1.35
QTSM	1.70	1.46	1.66	1.33	1.80	1.56
(d) 36 months ahead						
ATSM	2.04	2.14	2.01	2.19	2.20	2.73
SRM	2.04	1.97	1.92	1.91	1.88	2.08
QTSM	2.06	2.25	2.02	2.04	2.25	2.31

Consistent with the results from the LPY tests, the gap between the performance of the models is narrowed by the addition of a fourth pricing factor to the QTSM (a fourth factor also generally improves the performance of the SRM, but by less). The case for adding a fifth factor to both models again appears weak, because the 5-factor models do less well at matching survey expectations during the ZLB period.

D. Out-of-Sample Forecasts

Finally, we evaluate the models' ability to match conditional expectations of bond yields by exploring how well they predict future yields in a recursive out-of-sample forecasting exercise. Specifically, we first estimate all the models using a sample from Jan. 1990–Jan. 2005 and produce forecasts of yields at horizons up to 12 months ahead. We then repeat the process recursively, adding 1 month of data to the estimation sample at each iteration. The final estimation sample ends in Dec. 2015 because we reserve the final 12 months of data for forecast evaluation. This forecast period seems challenging for the models, because it contains yields (i) far from zero, (ii) when hitting the ZLB, and (iii) a prolonged period at the ZLB.

The left column of Figure 6 shows the root mean squared prediction errors (RMSPEs) by maturity for 6- and 12-month forecast horizons. The 3-factor SRM and QTSM both out-perform the 3-factor ATSM, particularly at the shorter forecast horizon. However, there is very little difference between the average forecasts from the two ZLB-consistent models and unreported Diebold-Mariano tests show that the differences are not significant at a 95% confidence level. We also find that adding a fourth factor to the SRM and QTSM (shown using the gray markers) generally worsens the forecast performance of these models, particularly for medium- and long-term yields. Thus, in contrast to results elsewhere in this section, the inclusion of a fourth factor in the QTSM does not improve its forecasting performance. This result suggests that while a 4-factor model performs better in matching other measures of conditional expectations, its larger number of parameters harms its ability to forecast well out-of-sample. Unreported results show that 5-factor versions of the models perform extremely poorly in forecasting out-of-sample, with RMSPEs that are off the scales of the charts reported in Figure 6.

Overall, none of these benchmark DTSMs are particularly compelling when it comes to forecasting yields, as they all struggle to beat a random walk forecast, which is shown by the heavy black lines. This is a common finding for DTSMs with fully-flexible time-series dynamics, and previous studies have found that restricting those dynamics can substantially improve forecasts in ATSMs when yields are away from the ZLB. Diebold and Li (2006) and Christensen, Diebold, and Rudebusch (2011) obtain better forecasts when $\mathbf{h_x}$ is a diagonal matrix, while Duffee (2011a) finds a similar result when setting the first eigenvalue of $\mathbf{h_x}$ equal to 1.

To explore whether a similar result also holds for the SRM and QTSM, we next consider specifications where we restrict the $\mathbb{P}$ dynamics by letting $\mathbf{h_x}$ be a diagonal matrix with $h_{x,11} = 1$. The right column of Figure 6 shows the RMSPEs from these restricted models. The restricted 3-factor SRM and the restricted 3- and 4-factor QTSM now outperform a random walk in forecasting short-maturity yields, particularly at the 12-month horizon. The differences between the ZLB-consistent models remain fairly small, although the SRM does achieve slightly smaller RMSPEs at the 12-month horizon.[13]

E.   Summary

Despite the potential for non-linear terms to cause problems for the ZLB-consistent models (and particularly the QTSM) when it comes to matching conditional expectations of future yields, our results suggest that there is actually little to choose between the models in this respect. While a 3-factor SRM does offer some small advantages relative to a 3-factor QTSM in matching some of the tests considered in this section, the differences are not statistically significant and in most cases are eliminated with the addition of a fourth factor to the QTSM. In contrast, adding a fourth factor to the SRM or a fifth factor to either model results in a deteriotation in performance. In the remainder of this article, we therefore focus on the standard 3-factor models, although we also report results for the 4-factor QTSM when these are substantively different.

---

[13]Unreported results show that the ATSM and SRM with 4 and 5 factors do not benefit from letting $\mathbf{h_x}$ be diagonal. The same holds for the QTSM with 5 factors. This is because these additional factors increase the probability of getting near identical eigenvalues in $\mathbf{\Phi}$, and hence very similar loadings for some of the factors, which then generates strong factor cross-correlation under the $\mathbb{P}$ measure that is not captured by letting $\mathbf{h_x}$ be diagonal.

Figure 6: Out-Of-Sample Forecasting Performance

This figure reports root mean squared prediction errors between actual yields and model-implied forecasts. The forecasts are for Jan. 2006–Dec. 2015. They are computed by recursively estimating each of the models, starting with a sample ending in Dec. 2005. The forecasted yields in the SRM are computed by drawing 10,000 times from the conditional distributions of yields. The figure shows results for 3-factor versions of all the models, as well as the 4-factor QTSM (denoted QTSM(4)) and SRM (denoted SRM(4); only for the unrestricted case). Random walk forecasts are constructed by assuming that yields do not change from their values at the end of the relevant sample period.
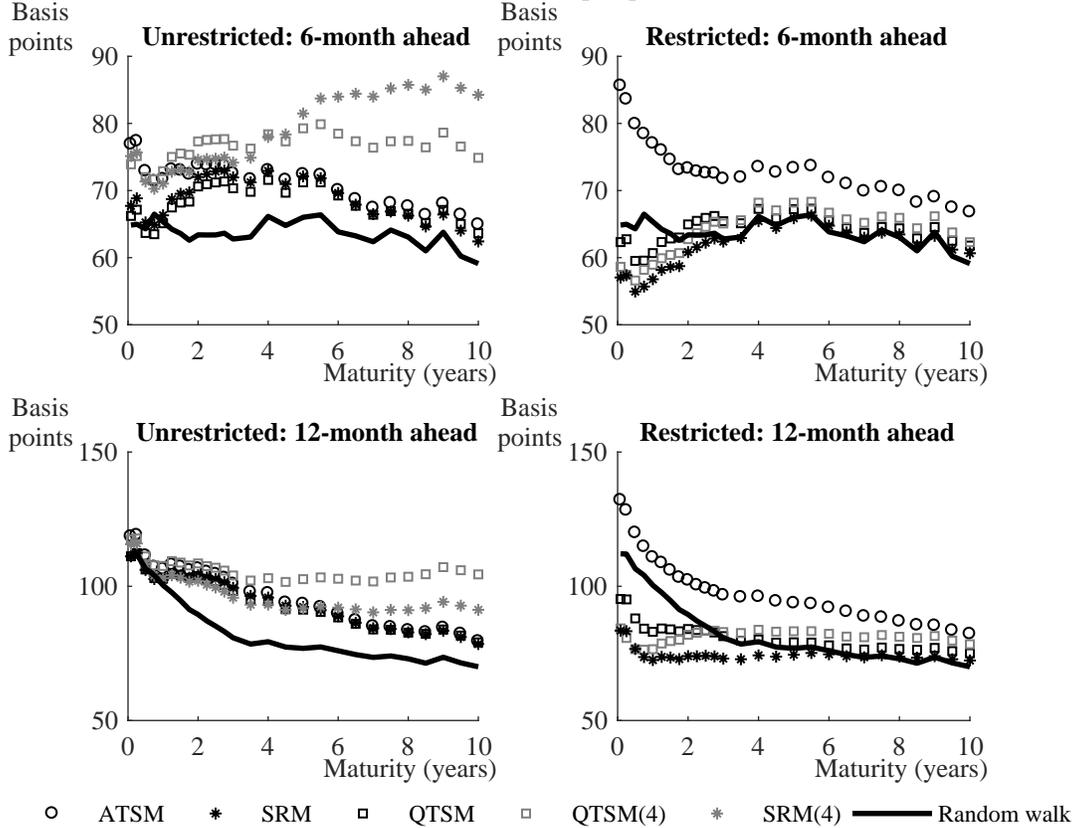
Perhaps more noteworthy than the small differences between the models are their common failings. In particular, when the SRM and QTSM are estimated on pre-ZLB data, both models replicate the satisfying ability of the ATSM to match the LPY tests and the desired coefficients from Mincer-Zarnowitz regressions. However, when the models are estimated on a sample that includes the ZLB period, both models are unable to match the LPY tests and the Mincer-Zarnowitz regressions when yields are away from the ZLB. Thus, we conclude that neither model fully captures the change in the dynamics of yields that occurred when the short rate reached the ZLB.

V. Matching Conditional Volatilities

We now turn to the second moment of yields. As discussed in Section III.B, the time-varying factor loadings for the SRM and QTSM imply that these models generate time-variation in the conditional volatilities of yields, in contrast to the ATSM. In Section V.A, we show that although the SRM and QTSM can generate a compression in the volatilities of short-term yields at the ZLB, similar to what we observe in the data, both models imply a link between volatilities and the level of yields that is too tight when yields are away from the ZLB. In Section V.B we consider whether a more flexible version of the QTSM can fit conditional volatilities more closely when yields are away from the ZLB, without implying a materially worse fit to the first moments of yields. Our results suggest that the QTSM is simply unable to match the first two conditional moments of yields simultaneously.[14]

A. The Tight Link Between the Level of Yields and Volatility

To estimate model-implied conditional volatilities, we use a local linearization of the relationship between the yields and pricing factors, i.e. we approximate the model-implied one-period-ahead conditional volatility of $y_{t+1,j}$ as

$$(8) \qquad \widehat{\sigma}_{t+1,j}\left(\hat{\mathbf{x}}_t; \boldsymbol{\theta}_1\right) \approx \sqrt{g_{j,\mathbf{x}}\left(\hat{\mathbf{x}}_t; \boldsymbol{\theta}_1\right)' \boldsymbol{\Sigma}\boldsymbol{\Sigma}' g_{j,\mathbf{x}}\left(\hat{\mathbf{x}}_t; \boldsymbol{\theta}_1\right)},$$

where $g_{j,\mathbf{x}}\left(\hat{\mathbf{x}}_t; \boldsymbol{\theta}_1\right)$ denotes the first derivative of $g_j\left(\mathbf{x}_t; \boldsymbol{\theta}_1\right)$ with respect to the pricing factors, evaluated at $\hat{\mathbf{x}}_t$.[15] Because we do not observe conditional volatilities in the data, we approximate them using the generalized autoregressive conditional heteroscedasticity (GARCH) model of Bollerslev (1986) applied separately to each bond yield.

The first row in Figure 7 shows plots of the fitted 1-year yields on the horizontal axes against their conditional volatilities from the 3-factor QTSM and SRM on the vertical axes (the

---

[14]We do not consider the SRM in this context because near-constant conditional volatilities away from the ZLB are an intrinsic feature of the SRM.

[15]For the ATSM and QTSM it is straightforward to compute analytical expressions for these derivatives. For the SRM, we evaluate the derivatives numerically.

black markers). We compare these results with the corresponding 1-year yields and conditional volatilities in the data (the gray markers). The bottom chart shows the same data points using a time-series plot.

Figure 7: Conditional Volatility and the Level of One-Year Bond Yields

The top row of charts plots one-year yields $y_{t+1,12}$ on the horizontal axes against their conditional volatilities $\sigma_t(y_{t+1,12})$ on the vertical axes, both in the data and implied by the 3-factor models estimated using data from Jan. 1990–Dec. 2016. Yields and conditional volatilities are expressed as annualized percentages. The model-implied conditional volatilities are computed using a local linearization of the relationship between yields and the pricing factors, evaluated at the estimated factor values. The conditional volatilities in the data are estimated using a univariate GARCH(1,1) model applied to the changes in the yield. The bottom chart plots the conditional volatilities from the QTSM and SRM over time, together with the GARCH(1,1) estimates.
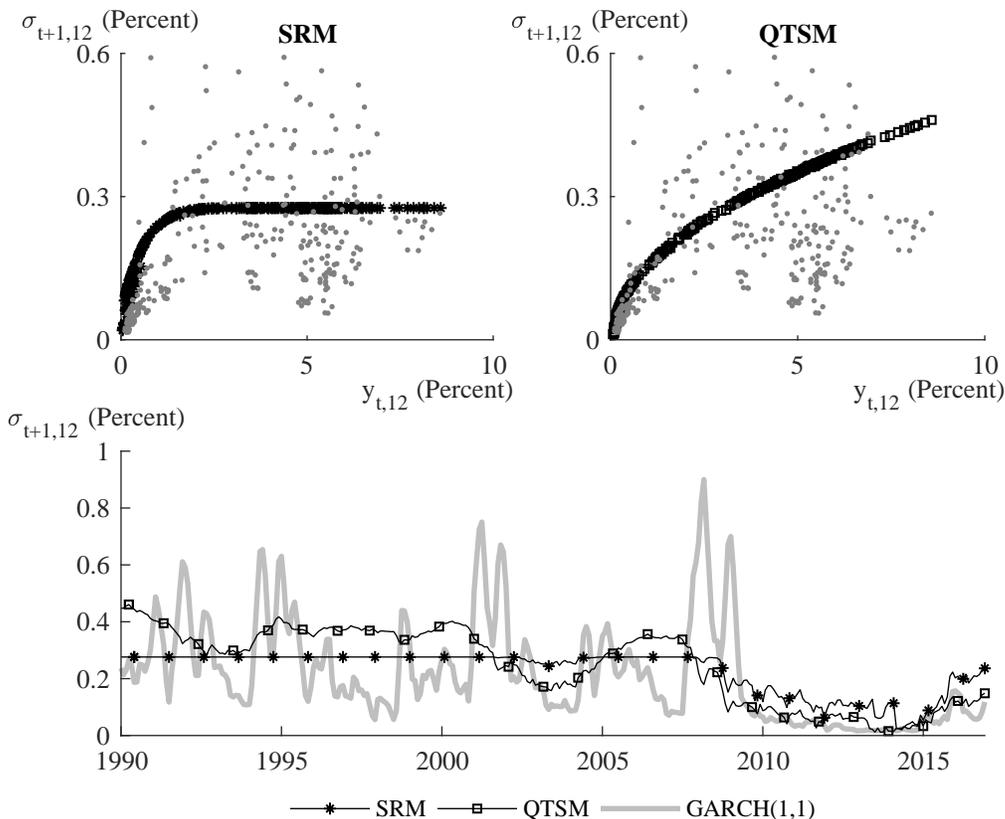
As suggested by the time-varying factor loadings reported in Section III.B, both the SRM and QTSM imply a compression in conditional volatilities as yields approach the ZLB, with the QTSM achieving a slightly better fit to the GARCH estimates. Thus, close to the ZLB, the models have realistic implications for conditional volatilities, consistent with the results of Kim and Singleton (2012) for Japanese yields.

Away from the ZLB, however, neither the SRM nor the QTSM can provide a realistic

description of volatilities in the data. In the SRM, conditional volatilities are by construction approximately constant when yields are sufficiently far from the ZLB. In the QTSM there is a stronger positive relationship between conditional volatilities and yields even far from the ZLB. In contrast, in the data volatilities vary over time (unlike in the SRM) but are only weakly correlated with the level of yields (unlike in the QTSM). Finally, unreported results show that there are no substantive differences to the results discussed in this section when we add additional pricing factors to either model. This failure to match conditional volatilities when yields are away from the ZLB is consistent with the findings of some previous studies. In particular, Kim (2007) finds that a 2-factor QTSM with only positive eigenvalues in $\mathbf{\Delta}_{xx}$ (to enforce the ZLB, as discussed in Section II.A.2) does not produce a good fit to conditional volatilities in the data.

B.   Improving the Fit to Conditional Volatilities

As discussed above, it is no surprise that the SRM generates essentially constant conditional volatilities when yields are away from the ZLB. To understand why our benchmark QTSM implies an extremely tight relationship between the level of yields and volatilities, recall that we impose that the short rate is the square of a linear combination of the pricing factors. As discussed in Section II.A.2, this gives the model the flexibility to match variation in long-term yields when the short rate is at the ZLB. However, this feature comes at the cost of generating an extremely tight link between the level of the short rate and its volatility, because $\text{var}_t\left[r_{t+1}\right] \approx 4\boldsymbol{\beta}'\mathbf{\Sigma}\mathbf{\Sigma}\boldsymbol{\beta} \times r_t$, i.e. the conditional volatility of the short rate is approximately proportional to the short rate itself. The results reported in Section V.A suggest that there is a similar tight link between volatilities and the level of yields at longer maturities.

To explore whether this tight link can be relaxed without materially harming the model fit to the first moments of yields, we reestimate the 3-factor QTSM but with two differences relative to the QTSM discussed elsewhere in this article. First, to give the model the best possible chance of matching yields and volatilities simultaneously, we allow the off-diagonal elements of $\mathbf{\Delta}_{xx}$ in equation (3) to be free parameters (subject to $\mathbf{\Delta}_{xx}$ remaining symmetric and

positive semi-definite to ensure identification and to enforce the ZLB, respectively, as explained in Section II.A.2). Second, we include the GARCH estimates of 1-month-ahead conditional volatilities of the 1-, 5- and 10-year yields as "observables", along with the same set of yields used to estimate the benchmark model. Specifically, at Step 1 of the SR approach, we modify equation (4) such that the estimated parameters are given by

$$(9) \quad \hat{\boldsymbol{\theta}}_1^{step1} = \arg\min_{\boldsymbol{\theta}_1 \in \Theta_1} \frac{1}{2N} \sum_{t=1}^{T} \left[\sum_{j=1}^{n_{y,t}} \left(y_{t,m_j} - g_{m_j}\left(\mathbf{x}_t; \boldsymbol{\theta}_1\right) \right)^2 + \sum_{j=1}^{n_{v,t}} \left(\sigma_{t+1,j}^{GARCH} - \hat{\sigma}_{t+1,j}\left(\mathbf{x}_t; \boldsymbol{\theta}_1\right) \right)^2 \right],$$

where $\hat{\sigma}_{t+1,j}\left(\mathbf{x}_t; \boldsymbol{\theta}_1\right)$ is the approximate model-implied volatility computed according to equation (8), $\sigma_{t+1,j}^{GARCH}$ is the corresponding GARCH estimate, $n_{v,t}$ is the number of observed volatilities in period t, and $N \equiv \sum_{t=1}^{T}\left(n_{y,t} + n_{v,t}\right)$.[16] Below we report results for fitted yields and volatilities from Step 1 of the SR approach, although in principle we could also make a similar modification to Step 3 in the SR approach.

Figure 8 shows that this modified version of the QTSM does indeed achieve a closer fit to the conditional volatility of the 1-year yield when compared to the benchmark QTSM. However, the modified QTSM has three important deficiencies. First, the conditional volatilities of the 5- and 10-year yields are still not anywhere close to the GARCH-based estimates. Second, the fit to yields deteriorates, e.g. the RMSEs for the 1- and 10-year yields increase to 13 and 16 basis points, respectively, compared with just 8 basis points at both maturities in the QTSM estimated using only data on yields. Third, the extracted factors imply estimates of $\mathbf{h}_x$ that induce explosive factor dynamics under the $\mathbb{P}$ measure. Unreported results show that adding a fourth factor to the QTSM does not substantively change these results. Thus, we conclude that the ZLB-consistent QTSM is simply unable to match the first and second moments of yields simultaneously.

---

[16]This version of the SR approach is closely related to the estimator of Andersen, Fusari, and Todorov (2015).

Figure 8: Conditional Volatility of Bond Yields in a QTSM with Observed Volatilities
This figure plots conditional volatilities $(\sigma_t(y_{t+1,j}))$ of the 1-, 5- and 10-year yields in the data and implied by the 3-factor models estimated using data from Jan. 1990–Dec. 2016. Conditional volatilities are expressed as annualized percentages. The model-implied conditional volatilities are computed using a first-order linearization of the relationship between bond yields and the pricing factors, evaluated at the estimated factor values. The conditional volatilities in the data are estimated using a univariate GARCH(1,1) model applied to the change in a given yield.
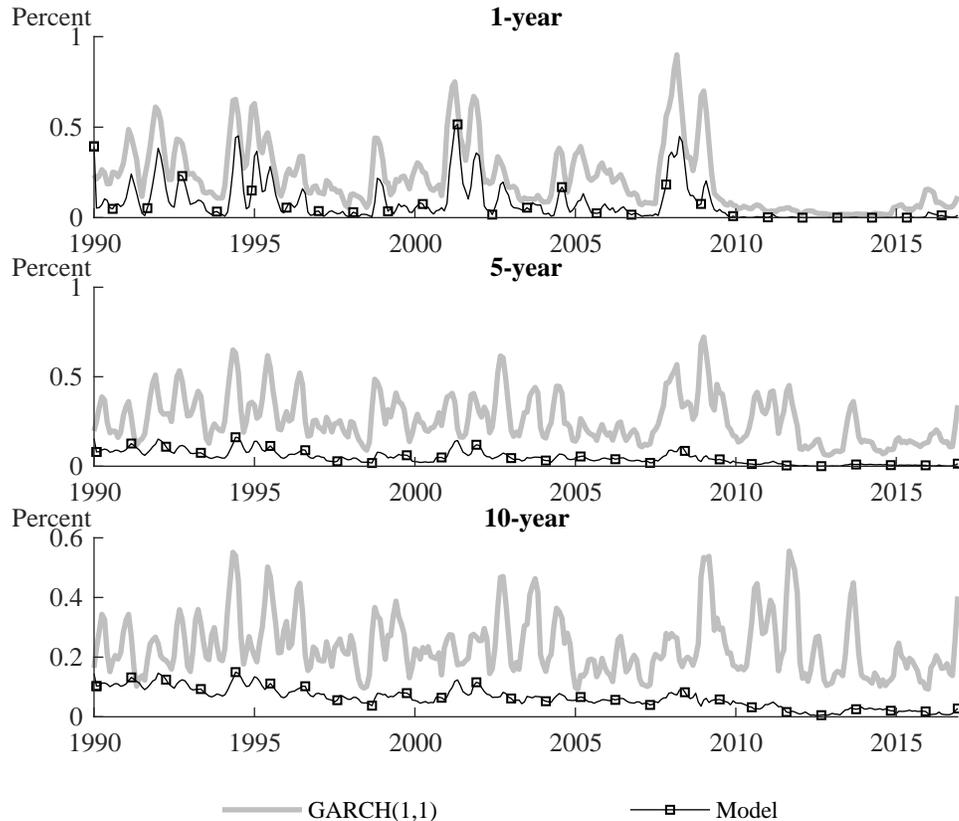
VI.   Sharpe Ratios and Return Predictability

The fact that the SRM and QTSM have different implications for conditional volatilities means that they are also likely to have different implications for other important aspects of bond yields. In this section, we therefore consider whether the SRM and QTSM have plausible implications for model-implied Sharpe ratios (i.e. the ratio of the first conditional moment of excess returns to the second) and return predictability (i.e. the ratio of the variance of model-implied expected excess returns to the variance of observed excess returns). Specifically, we consider whether the models can replicate the three "robust properties" reported by Duffee (2010) for ATSMs estimated using pre-ZLB data. We largely confirm that these robust properties

also hold for the SRM and QTSM when yields are away from the ZLB but that the models have different implications for model-implied Sharpe ratios as the short rate approaches the ZLB.

A. Average Conditional Sharpe Ratios

Duffee's first robust property of ATSMs is an inverse relationship between a bond's average Sharpe ratio and its maturity. The conditional Sharpe ratio on a j-period bond is defined as

$$(10) \qquad S_{t,j} = \frac{\mathrm{E}_t\left[R_{t+1,j} - R_{t+1,1}\right]}{\sqrt{\mathrm{var}_t\left[R_{t+1,j} - R_{t+1,1}\right]}},$$

where $R_{t+1,j} \equiv P_{t+1,j-1}/P_{t,j}$. Following Duffee (2010), we report average population conditional Sharpe ratios, although we additionally condition on whether the short rate is above or below 1%.

To estimate the Sharpe ratios, we simulate a single time series from each of the models until we have at least 1,000 periods in which the short rate is below 1% and 1,000 periods in which it is 1% or higher. For each simulated observation we make 1,000 draws from the model-implied conditional return distributions and compute the conditional Sharpe ratios according to equation (10).

Table 3 reports the means of these conditional Sharpe ratios for the 3-factor models when the short rate is above or below 1%. In periods in which the short rate is above 1%, all three models generate an inverse relationship between the average Sharpe ratio and maturity, confirming Duffee's first robust property of ATSMs also holds in the SRM and QTSM when yields are away from the ZLB.

However, the patterns change when the short rate is below 1%. To understand why, consider first the ATSM. Recall that the conditional variance of returns is constant, so conditional Sharpe ratios only depend on the level of the short rate through its effect on expected excess returns. On the one hand, the fact that the model is stationary means that the further yields fall below their unconditional mean, the more bond prices tend to be expected to

fall, and the lower the expected return tends to be, lowers Sharpe ratios. On the other hand, if the short rate is low, the cost of financing a long-term bond position is also very low, which increases Sharpe ratios. At short maturities, the former effect dominates and the Sharpe ratio of a 1-year bond is loewr when the short rate is below 1% than when it is above 1%. However, because the yield curve tends to be steeply upward sloping when the short rate is low, long-term bonds offer abnormally higher expected excess returns over the 1-month rate, and their Sharpe ratio rises slightly. Thus, when the short rate is low we observe a flatter average relationship between the Sharpe ratio and maturity in the ATSM.

The QTSM and SRM introduce an additional complication, because the conditional variance of returns (i.e. the denominator in equation (10)) also falls as the short rate approaches the ZLB. At long maturities, this compression is on average relatively unimportant, such that the average Sharpe ratio of a 10-year bond is higher when the short rate is below 1% than when it is above 1%, as in the ATSM. At shorter maturities, the effect on Sharpe ratios depends on the relative speed with which the first and second conditional moments of returns approach zero. In the QTSM, the volatility compression dominates and the average Sharpe ratio of a 1-year bond is higher when the short rate is below 1%, while the reverse is true in the SRM. Finally, adding a fourth factor to the QTSM raises the conditional Sharpe ratios for short-maturity bonds somewhat, to about 0.3 when the short rate is above 1% and to 0.8 when it is below 1%.

In summary, we conclude that Duffee's first robust property of ATSMs also holds in the SRM and QTSM when the short rate is away from the ZLB, although the patterns do change when the short rate approaches the ZLB and with the addition of a fourth factor to the QTSM.

B.   Return Predictability

Duffee's second robust property of ATSMs is that they imply that 15–20% of annual excess bond returns are predictable. The fourth and fifth columns of Table 3 therefore report the population ratio of the variance of expected excess returns implied by the 3-factor models to the variance of actual excess returns in our simulated samples, again conditioning on whether the short rate is above or below 1%.

Table 3: Sharpe Ratios and the Predictability of Excess Returns

This table reports mean conditional Sharpe ratios, the predictable proportion of excess returns at different horizons, and the fraction of expected returns at different horizons explained by the first principal component of monthly expected returns at different maturities. All results are for models estimated using data from Jan. 1990–Dec. 2016. The results are reported separately for periods in which the short rate is below 1% and in which it is 1% or higher. To obtain these model-implied properties, we first simulate from each of the models until we have at least 1,000 periods in which the short rate is below 1% and 1,000 periods in which it is 1% or higher. For each simulated sample period we then obtain 1,000 draws from the 1-month- and 1-year-ahead conditional excess return distributions to compute various model-implied moments. The first three columns show the resulting estimates of mean conditional monthly Sharpe ratios. The fourth and fifth columns report the predictable portion of excess returns at a monthly and annual horizon, defined as the ratio of the variance of model-implied expected excess returns on a 10-year bond to the variance of subsequent excess returns. The final two columns show the R^2 from unrestricted linear regressions of monthly and annual expected excess returns on the first principal component of expected 1-month excess returns on bonds with maturities of 2, 3, 4, 5, 6, 7, 8, 9, and 10 years. All models have 3 factors except QTSM(4), which denotes the 4-factor QTSM.

		Mean Conditional Sharpe Ratio (Maturity in Years)			Predictable Proportion of Excess Returns on a 10-year Bond		Fraction of Expected Returns Explained by a "Monthly Return Factor"	
		1	5	10	Monthly	Annual	Monthly	Annual
ATSM	$r_t \geq 1\%$	0.31	0.20	0.16	0.02	0.18	1.00	0.94
	$r_t < 1\%$	0.19	0.21	0.21	0.02	0.24	1.00	0.94
SRM	$r_t \geq 1\%$	0.25	0.19	0.15	0.02	0.22	1.00	0.92
	$r_t < 1\%$	0.15	0.23	0.20	0.02	0.27	0.99	0.92
QTSM	$r_t \geq 1\%$	0.26	0.14	0.12	0.03	0.24	0.97	0.88
	$r_t < 1\%$	0.33	0.24	0.21	0.02	0.25	0.97	0.91
QTSM(4)	$r_t \geq 1\%$	0.32	0.16	0.11	0.03	0.22	0.97	0.81
	$r_t < 1\%$	0.76	0.21	0.14	0.03	0.29	0.97	0.89

Focusing first on periods in which the short rate is above 1%, we find that 2–3% of monthly returns on a 10-year bond are predictable in all the models, only slightly lower than the 5% reported by Duffee (2010) for a 3-factor ATSM estimated using pre-ZLB data. For annual excess returns, this proportion increases to 18% in the ATSM, close to the 19% reported by Duffee (2010). The proportions are slightly higher in the QTSM and SRM (but not materially so) at 24% and 22% respectively.

When the short rate is below 1%, return predictability rises slightly at an annual horizon in all three models (although the differences are still not particularly large), whereas return predictability is little changed at a monthly horizon. Finally, adding a fourth factor to the QTSM increases return predictability at an annual horizon a little further, although it remains below 30% irrespective of whether the short rate is above or below 1%. In summary, we interpret these results as evidence that the second robust property of Duffee (2010) also holds

approximately in the QTSM and SRM.

Finally, Duffee's third robust property is that the variation in excess returns on different maturity bonds over a holding period of 1 month are almost exclusively explained by a single factor, but the same factor explains less of the variation in excess results for an annual holding period. Following his approach, we construct a "monthly return factor" by taking the first principal component of model-implied 1-month expected excess returns on bonds with maturities from 2 to 10 years at 1-year intervals during our simulated sample, again conditioning on whether the short rate is above or below 1%.

The results reported in the final two columns of Table 3 show that we can also confirm the third robust property for all of the models. The monthly return factor explains almost all of the monthly expected excess returns on the 10-year bond, irrespective of whether the short rate is above or below 1%. However, if we regress expected annual excess returns on this factor, then the R^2 is somewhat lower, again irrespective of whether the short rate is above or below 1%.

VII.   Conclusion

This article examines the performance of the SRM and QTSM when estimated using U.S. Treasury yields. While the SRM has received more attention in recent empirical work, the two models actually perform remarkably similarly against a number of criteria. While the SRM achieves marginally superior performance in some respects, the small and generally statistically-insignificant differences between the models do not appear sufficiently large to support the clear preference that has recently emerged in the literature in favor of the SRM, particularly if we add a fourth factor to the QTSM and take into account its greater computational tractability.

Perhaps more noteworthy than the modest differences between the models are their common failings. Although both the SRM and the QTSM outperform the ATSM when yields are close to the ZLB, neither of the two ZLB-consistent models appear to fully capture the change in the yield dynamics at the ZLB. The problem seems to be that the time-series dynamics of yields during the ZLB period change, and simply modifying the functional form of

the short rate does not fully capture this change. A potential solution to this problem might be to modify the factor dynamics of the $\mathbb{P}$ measure as the short rate approaches the ZLB. In addition, neither of the models has the flexibility to provide a good description of conditional volatilities of yields when yields are away from the ZLB, suggesting that it may be necessary to incorporate some form of unspanned stochastic volatility into the models. We leave these and other extensions of the models to future research.

References

Adrian, T., R. K. Crump, and E. Moench. "Pricing the Term Structure with Linear Regressions." *Journal of Financial Economics*, 110 (2013), 110–138.

Ahn, D.-H., R. F. Dittmar, and A. R. Gallant. "Quadratic Term Structure Models: Theory and Evidence." *Review of Financial Studies*, 15 (2002), 243–288.

Andersen, T. G., N. Fusari, and V. Todorov. "Parametric Inference and Dynamic State Recovery from Option Panels." *Econometrica*, 83 (2015), 1081–1145.

Andreasen, M. M. and B. J. Christensen. "The SR Approach: A New Estimation Procedure for Non-Linear and Non-Gaussian Dynamic Term Structure Models." *Journal of Econometrics*, 184 (2015), 420–451.

Bauer, M. D. and G. D. Rudebusch. "Monetary Policy Expectations at the Zero Lower Bound." *Journal of Money, Credit and Banking*, 48 (2016), 1440–1465.

Bauer, M. D., G. D. Rudebusch, and J. C. Wu. "Correcting Estimation Bias in Dynamic Term Structure Models." *Journal of Business and Economic Statistics*, 30 (2012), 454–467.

Black, F. "Interest Rates as Options." *Journal of Finance*, 50 (1995), 1371–1376.

Bollerslev, T. "Generalized autoregressive conditional heteroskedasticity." *Journal of Econometrics*, 31 (1986), 307–327.

Campbell, J. Y. and R. J. Shiller. "Yield Spread and Interest Rate Movements: A Bird's Eye View." *Review of Economic Studies*, 58 (1991), 495–514.

Christensen, J. H. E., F. X. Diebold, and G. D. Rudebusch. "The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models." *Journal of Econometrics*, 164 (2011), 4–20.

Christensen, J. H. E. and G. D. Rudebusch. "Estimating Shadow-Rate Term Structure Models with Near-Zero Yields." *Journal of Financial Econometrics*, 13 (2015), 226–259.

Cochrane, J. H. and M. Piazzesi. "Decomposing the yield curve." Working Paper, University of Chicago (2008).

Cox, J. C., J. E. Ingersoll, and S. A. Ross. "A Theory of the Term Structure of Interest Rates." *Econometrica*, 53 (1985), 385–407.

Dai, Q. and K. J. Singleton. "Specification Analysis of Affine Term Structure Models." *Journal of Finance*, 55 (2000), 1946–1978.

Dai, Q. and K. J. Singleton. "Expectation Puzzles, Time-Varying Risk Premia and Affine Models of the Term Structure." *Journal of Financial Economics*, 63 (2002), 415–441.

Diebold, F. X. and C. Li. "Forecasting the Term Structure of Government Bond Yields." *Journal of Econometrics*, 130 (2006), 337–364.

Duffee, G. R. "Term Premia and Interest Rate Forecasts in Affine Models." *Journal of Finance*, 57 (2002), 405–443.

Duffee, G. R. "Sharpe Ratios in Term Structure Models." Working Paper, Johns Hopkins University (2010).

Duffee, G. R. "Forecasting with the Term Structure: The Role of No-Arbitrage Restrictions." Working Paper, Johns Hopkins University (2011a).

Duffee, G. R. "Information in (and not in) the Term Structure." *Review of Financial Studies*, 24 (2011b), 2895–2934.

Fama, E. F. and R. R. Bliss. "The Information in Long-Maturity Forward Rates." *American Economic Review*, 77 (1987), 680–692.

Feunou, B., J.-S. Fontaine, A. Le, and C. Lundblad. "Tractable Term-Structure Models and the Zero Lower Bound." Working Paper, Bank of Canada (2015).

Filipovic, D., M. Larsson, and A. B. Trolle. "Linear-Rational Term Structure Models." *Journal of Finance*, 72 (2017), 655–704.

Joslin, S., K. J. Singleton, and H. Zhu. "A New Perspective on Gaussian Dynamic Term Structure Models." *Review of Financial Studies*, 24 (2011), 926–970.

Kim, D. H. "Spanned stochastic volatility in bond markets: A reexamination of the relative pricing between bonds and bond options." Working Paper, BIS (2007).

Kim, D. H. and K. J. Singleton. "Term Structure Models and the Zero Bound: An Empirical Investigation of Japanese Yields." *Journal of Econometrics*, 170 (2012), 32–49.

Leippold, M. and L. Wu. "Asset Pricing under the Quadratic Class." *Journal of Financial and Quantitative Analysis*, 37 (2002), 271–295.

Mincer, J. A. and V. Zarnowitz. *The Evaluation of Economic Forecasts*, NBER, chap. 1, 3–46 (1969).

Monfort, A., F. Pegoraro, J.-P. Renne, and G. Roussellet. "Staying at Zero with Affine Processes: A New Dynamic Term Structure Model." *Journal of Econometrics*, 201 (2017), 348–366.

Priebsch, M. A. "Computing Arbitrage-Free Yields in Multi-factor Gaussian Shadow-Rate Term Structure Models." Finance and Economics Discussion Series, Federal Reserve Board (2013).

Realdon, M. "Quadratic Term Structure Models in Discrete Time." *Finance Research Letters*, 3 (2006), 277–289.

Realdon, M. "Gaussian Models for Euro High Grade Government Yields." *European Journal of Finance*, 23 (2016), 1–44.

Rudebusch, G. D. and T. Wu. "Accounting for a Shift in Term Structure Behavior with No-Arbitrage and Macro-Finance Models." *Journal of Money, Credit and Banking*, 39 (2007), 395–422.

Wu, J. C. and F. D. Xia. "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound." *Journal of Money Credit and Banking*, 48 (2016), 253–291.