
Article:
Golinski, Adam orcid.org/0000-0001-8603-1171 and Spencer, Peter orcid.org/0000-0002-5595-5360 (2024) Unconventional monetary policies and the yield curve:Estimating non-affine term structure models with unspanned macro risk by factor extraction. Review of Asset Pricing Studies. pp. 119-152. ISSN: 2045-9920

https://doi.org/10.1093/rapstu/raad011

The Review of Asset Pricing Studies

Unconventional Monetary Policies and the Yield Curve: Estimating Non-Affine Term Structure Models with Unspanned Macro Risk by Factor Extraction

Adam Goliński
University of York, UK

Peter Spencer
University of York, UK

We show how the Joslin, Singleton, and Zhu (2011) factor extraction approach to estimating the Gaussian term structure model can be modified to handle the interest rate lower bound without the approximations used in other approaches. This drastically reduces the computation time and produces more robust estimates of the lower bound parameter and the shadow rate. It makes feasible the extensive specification search necessary to allow for unspanned factors as in Joslin, Priebsch, and Singleton (2014), allowing the term structure model to be used to better assess the effects of policy on the term premium and market expectations. (*JEL* G12, C13, E43)

Received June 28, 2022; editorial decision May 12, 2023 by Editor Hui Chen

The Gaussian term structure model (GTSM) is routinely used to analyze the behavior of a wide range of financial markets, including those for commodities as well as those for government and corporate bonds. Its popularity has been enhanced in recent years by a series of innovations that greatly reduce the need for numerical optimization methods. Instead of using the Kalman filter to estimate the latent variables underpinning the yield curve, these new methods restrict the measurement error covariance structure in a way that allows them to be extracted directly from yields or, perhaps, principal components of yields. This allows the likelihood function to be concentrated, helping to deal with multiple local optima and other difficult numerical problems as well as greatly speeding up the estimation procedure.

The Gaussian model, however, violates the zero lower bound constraint on interest rates. Although this may not be a problem at historical levels of the interest

Send correspondence to: Adam Goliński, adam.golinski@york.ac.uk.

The Review of Asset Pricing Studies 14 (2024) 119–152

https://doi.org/10.1093/rapstu/raad011 Advance Access publication 20 June 2023

rate, it is a serious problem at the near-zero interest rates seen in the developed economies since the onset of the financial crisis. This confronts the modeler with new numerical challenges.

In this situation, the Black (1995) shadow rate term structure model, which represents the spot interest rate using a truncated distribution, has a great deal of appeal. Unfortunately, this truncation introduces a sharp nonlinearity into this system, making the usual approach based on the partial differential equation solution infeasible when there are more than two factors. However, the formula proposed by Priebsch (2013) offers an exact closed-form solution for three or more factors when the underlying shadow rate is Gaussian. Simpler formulae based on close approximations to this Gaussian shadow rate term structure model (SRTSM) have been proposed by Krippner (2012) and Wu and Xia (2016). Other models that deal with the nonnegativity of the interest rate include the quadratic term structure model (QTSM), in which the short rate depends upon squared Gaussian factors.[1,2]

The nonlinearities in these various formulae mean that researchers have resorted to the use of the extended Kalman filter (EKF) and other methods that have been superseded by the new likelihood-concentration methods in the standard Gaussian setting. However, this paper shows how the factor extraction approach of Joslin, Singleton, and Zhu (2011), with all of its well-known computational advantages, can be readily applied to Gaussian shadow rate term structure models, allowing the researcher to estimate these without any approximation almost as easily and quickly as the simple GTSM. In the case of a well-known U.S. Treasury bond yield dataset, we find that the approximation error in using the commonly used Wu and Xia (2016) approximation to estimate the SRTSM is negligibly small, allowing the solution to be found in four to eight minutes. The Priebsch (2013) model naturally takes much longer, but we use it here as a benchmark to evaluate the error implied by the Wu and Xia (2016) approximation. Our approach is applicable to a broader class of Gaussian models, including the QTSM, which we also analyze in this paper.

These techniques open the way to research with shadow rate models that use daily and other large data sets or problems that require large numbers of different model estimations. Moreover, we find that, because they restrict the measurement error structure, these algorithms have the added advantage of producing more precise and robust estimates of the lower bound parameter and the path of the

[1] The origins of the quadratic model can be traced back to Longstaff (1989), Beaglehole and Tenney (1991, 1992), and Constantinides (1992). This model was fully developed by Ahn, Dittmar, and Gallant (2002) and Leippold and Wu (2002) in continuous time, and Realdon (2006) in discrete time.

[2] There are other approaches in the term structure literature that tackle the problem of nonnegativity of the interest rates that are based on non-Gaussian processes—e.g., Cox, Ingersoll, and Ross (1985) and Dai and Singleton (2000) proposed a class of models based on square-root processes in continuous time. Its discrete-time counterpart based on the autoregressive gamma process was developed by Gourieroux, Monfort, and Polimenis (2002) and Le, Singleton, and Dai (2010). Unfortunately, these processes make the lower bound a reflecting barrier and so do not allow the policy rate to remain at the bound for any length of time. However, Monfort et al. (2017) proposed a term structure model based on the autoregressive gamma process that overcomes this problem and allows for a point-mass at zero.

shadow rate than the EKF. This aside, the fit and parameter estimates are economically indistinguishable from those obtained with the EKF.

Our factor extraction approach is similar to the sequential regression procedure (SR) proposed by Andreasen and Christensen (2015) and implemented in the the context of a term structure model of interest rates by Andreasen and Meldrum (2019). They use nonlinear least squares to find the factors that optimize the fit of the cross-section of yields and then use these factors to fit the time-series dynamics using ordinary least squares (OLS). This method focuses on the cross-section without taking into account the time-series fit, and we find that the overall fit, reflected in the likelihood, is considerably worse than for our method.

To illustrate the ability of our factor extraction approach to handle problems that cannot feasibly be tackled using the EKF, we update the work of Joslin, Priebsch, and Singleton (2014) using a shadow rate version of their macrofinance model to analyze the way that macroeconomic shocks affect term premia. This model allows macroeconomic variables to influence the future evolution of the yield curve through the real-world dynamics even though they do not affect its current position and are in that sense "unspanned" (Duffee 2011). Unfortunately, introducing macro variables into the $\mathcal{P}$-dynamics greatly increases the number of parameters in the model. Moreover, as Joslin, Priebsch, and Singleton (2014) emphasize, many of these parameters, and hence the risk adjustments, are very poorly defined. Elimination of the insignificant price of risk parameters increases the precision of the $\mathcal{P}$-parameters by pushing them closer to the $\mathcal{Q}$-dynamics, which are defined with a very high degree of precision. However, as pointed out by Bauer (2018), imposing zero restrictions based on their individual or joint significance might result in the identification of a suboptimal model because it is generally path dependent. We follow Joslin, Priebsch, and Singleton (2014), who deal with this problem by searching over all possible combinations of exclusion restrictions to find the best combination and ensure that the model does not lead to misleading inferences. In our case, this involves searching over 2^{18} possible combinations, which would be infeasible using the EKF or similar techniques.

This analysis is of particular importance to monetary policy makers seeking to influence the economy through the yield curve. They routinely use term structure models to decompose yields into the contributions of market expectations and risk premia.[3] We show that the macrofinance model with the optimal set of price-of-risk restrictions (MFE^{opt}, described in detail in Section 4) views the behavior of the yield curve since the financial crisis in a very different light to the standard unrestricted yield-only model (FE). Figure 1 shows the way that these two models decompose the 10-year yield (continuous line) into the risk-neutral yield (panel A), which reflects market expectations, and the risk premium (panel B), which is the difference between the fitted and risk-neutral yields.

[3] See, e.g, https://www.newyorkfed.org/research/data_indicators/term-premia-tabs and https://www.federalreserve.gov/pubs/feds/2005/200533/200533abs.html

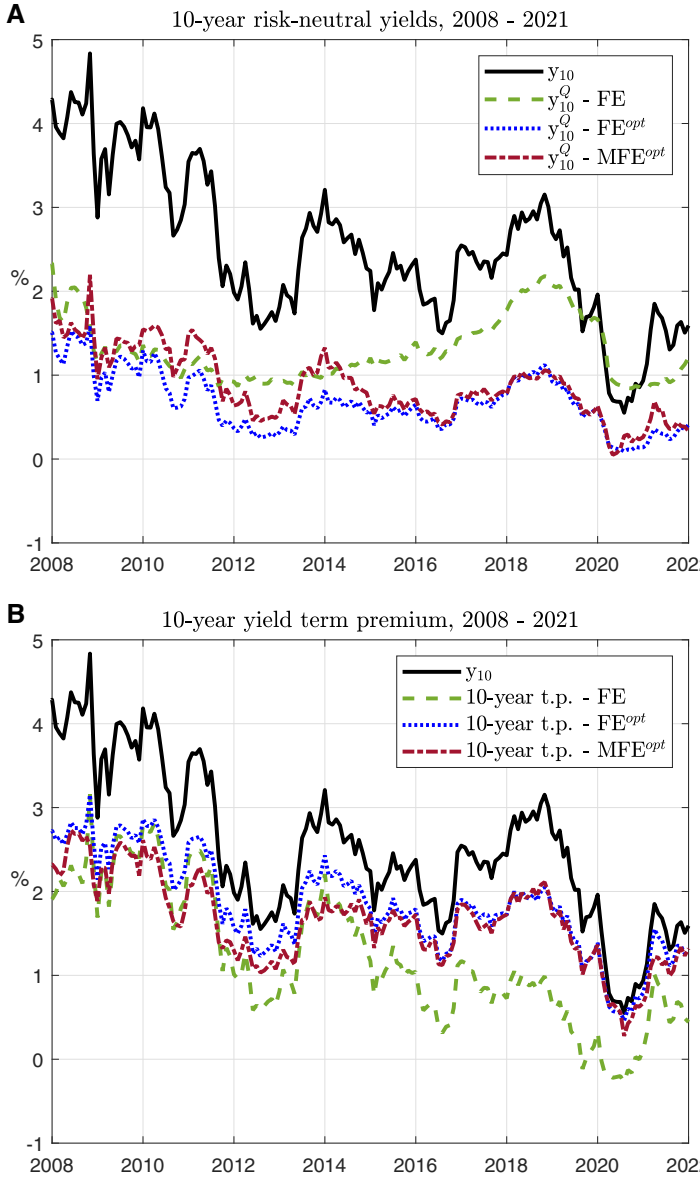

Figure 1
10-year risk-neutral yields and risk premia
The figure shows 10-year risk-neutral yields (panel A) and term premia (panel B) generated by different models. The FE model is a yield-only model and does not impose any restrictions, FE^{opt} is a yield-only model with optimal price-of-risk restrictions, and the MFE^{opt} is a macrofinance model with optimal price-of-risk restrictions. The 10-year yield is plotted as a reference.

FE suggests that the 10-year risk-neutral yields were relatively smooth over the period 2009–2021 and that the variations in observed yields were largely due to variations in the risk premium. This is consistent with the finding of Gagnon et al. (2011) using an affine term structure model that the QE policy announcements that followed the Lehman default in September 2008 affected the term premium rather than market expectations. However, we find that this conclusion is due to the presence of insignificant parameters in the time-series dynamics. When these are eliminated, this gives the optimal yield-only model FE^{opt} . The risk-neutral yields from this model are also shown in the figure. The two restricted models suggest that the variation in observed yields was much less due to the risk premium than implied by the standard model FE. We show that these variations still match announcements of changes in QE nicely.

Reflecting this, we show that the countercyclical nature of risk premiums found by Joslin, Priebsch, and Singleton (2014) in the pre-crisis period is also a feature of the post-crisis period. For example, the correlation between the 10-year term premium and expected inflation goes up from 4.4% in FE to 13.3% in FE^{opt} and to 20.05% in MFE^{opt}, while the correlation with economic growth falls from −4.6% to −8.3% in the FE and MFE^{opt} models, respectively. The latter estimate is consistent with the conventional view that risk premium tends to increase during an economic downturn. These findings support the argument of Joslin, Priebsch, and Singleton (2014) that it is important to eliminate insignificant time-series parameters and to allow for the effect of macroeconomic variables on risk premia in term structure models.

1. The Gaussian Term Structure Model

1.1 The model structure

Assume that we can write the n-period zero coupon bond yield as:

$$y_{n,t} = y(\mathbf{x}_t, n; \boldsymbol{\Psi}), \tag{1}$$

where $\boldsymbol{\Psi}$ is a vector of relevant parameters.

The dynamics of the state vector under the physical ($\mathcal{P}$) and risk-neutral measure ($\mathcal{Q}$) are:

$$\mathbf{x}_t = \boldsymbol{\mu}^{\mathcal{P}} + \boldsymbol{\Phi}^{\mathcal{P}}\mathbf{x}_{t-1} + \mathbf{u}_t^{\mathcal{P}}, \tag{2}$$

$$\mathbf{x}_t = \boldsymbol{\mu}^{\mathcal{Q}} + \boldsymbol{\Phi}^{\mathcal{Q}}\mathbf{x}_{t-1} + \mathbf{u}_t^{\mathcal{Q}}, \tag{3}$$

respectively, with $\mathbf{u}_t^{\mathcal{P}}, \mathbf{u}_t^{\mathcal{Q}} \sim N(0, \boldsymbol{\Sigma})$ under their respective measure. We adopt the parameterization scheme proposed by Joslin, Singleton, and Zhu (2011), that is, $\boldsymbol{\Phi}^{\mathcal{Q}}$ is determined by K roots, which we can collect in a vector $\boldsymbol{\lambda}^{\mathcal{Q}}$ and $\boldsymbol{\mu}^{\mathcal{Q}} = [\mu_\infty^{\mathcal{Q}}, 0, \ldots]'$. The physical dynamics, $\boldsymbol{\mu}^{\mathcal{P}}$, $\boldsymbol{\Phi}^{\mathcal{P}}$, and $\boldsymbol{\Sigma}$, are unrestricted. This parameterization plays a crucial role in our estimation scheme, as in the

original Joslin, Singleton, and Zhu (2011) Gaussian scheme, since it allows one to estimate the $\mathcal{P}$-parameters by OLS regression.

The observed short-rate is a function of K-dimensional state vector $\mathbf{x}_t$ and short-rate parameters collected in a vector $\mathbf{\Psi}_r$:

$$r_t = r(\mathbf{x}_t; \mathbf{\Psi}_r), \tag{4}$$

where $\mathbf{\Psi}_r$ is a subset of $\mathbf{\Psi}$. We now present some examples of Gaussian models.

1.1.1 The standard Gaussian model. The standard Gaussian term structure model (GTSM) was first proposed by Vasicek (1977) and since then, due to its analytical tractability, has become the most popular model in the literature. In the GTSM the short rate is linear in the state variables:

$$r_t = \delta_0 + \delta_1' \mathbf{x}_t \tag{5}$$

and, as such, the short-rate and other yields are unrestricted and can become negative. Following Joslin, Singleton, and Zhu (2011) we parameterize the model as $\mathbf{\Psi}_r = \{\delta_0 = 0, \delta_1 = 1_K\}$.

1.1.2 The shadow rate model. The shadow rate term structure model (SRTSM) was first proposed by Black (1995) with the observed short rate specified as:

$$r_t \equiv \max(s_t, \underline{r}), \tag{6}$$

where $\underline{r}$ is the lower bound and the shadow short rate s_t is driven by underlying the K dimensional process:

$$s_t = \delta_0 + \boldsymbol{\delta}_1' \mathbf{x}_t. \tag{7}$$

In his descriptive paper Black (1995) did not specify the nature of the process $\mathbf{x}_t$, but virtually all empirical implementations of the model assume Gaussian dynamics, as in Equation (2).[4] Our parameterization of the short rate is similar to the GTSM, in that we specify $\delta_0 = 0$ and $\delta_1 = 1_K$, with the additional lower bound parameter $\underline{r}$, which can be estimated or set to a prespecified value.

If the short rate is well above the lower bound, it is reasonable to use the GTSM, which assumes: $r_t = s_t$. However, as the short rate approaches the lower bound, the Gaussian model gives a significant probability mass to negative interest rates, which makes it impractical for many purposes, such as monetary policy analysis and pricing fixed income derivatives. Shadow rate models deal with this inconvenience by treating the forward rates of the Gaussian model as "shadow forward rates" and then map them into the Black forward rates. Unfortunately, Equation (6) introduces a sharp discontinuity into this mapping, making it difficult to estimate the model using the usual partial differential equation approach when there are

[4] See, e.g., Kim and Singleton (2012), Christensen and Rudebusch (2015), Bauer and Rudebusch (2016) among others.

more than two factors. However, Priebsch (2013) shows that a formula that uses the first two cumulants offers a tractable closed-form solution for the Gaussian shadow rate model. Simpler formulae based on close approximations have been proposed by Krippner (2012) and Wu and Xia (2016).

1.1.3 The quadratic model. The short-rate equation in the QTSM is given by:

$$r_t = \delta_0 + \delta_1' \mathbf{x}_t + \mathbf{x}_t' \Delta \mathbf{x}_t, \tag{8}$$

where Δ is a positive semi-definite parameter matrix.

The lower bound in the QTMS is given by $\underline{r} = \delta_0 - \frac{1}{4}\delta_1' \Delta \delta_1$,[5] which implies that we can reparameterize this as:

$$r_t = \underline{r} + \frac{1}{4}\delta_1' \Delta \delta_1 + \delta_1' \mathbf{x}_t + \mathbf{x}_t' \Delta \mathbf{x}_t. \tag{9}$$

The short rate parameters $\mathbf{\Psi}_r$ need to be restricted to achieve the model identification. We set $\delta_1 = 1_K$ and $\Delta = 1_{K,K}$. Although this parameterization means that the model is overidentified, the model has the same number of parameters as the shadow rate model and, if $\underline{r}$ is set to a prespecified value, the same as the GTSM.[6]

1.2 Fitting errors

If the yield model in Equation (1) is fitted without error, then, conditional on the model parameters, we could in principle invert any K of these relationships to identify the state vector $\mathbf{x}_t$. However, to allow for measurement and misspecification effects, we augment this specification with an additive error $v_{n,t}$ to get an empirical model of the observed yield $y_{n,t}^o$:

$$y_{n,t}^o = y(\mathbf{x}_t, n; \mathbf{\Psi}) + v_{n,t}, \tag{10}$$

where $v_{n,t} \sim N(0, \sigma_{v,n}^2)$. Stacking Equations (10) gives a system of J nonlinear yield equations:

$$\mathbf{y}_t^o = \mathbf{y}(\mathbf{x}_t; \mathbf{\Psi}) + \mathbf{v}_t, \tag{11}$$

where $\mathbf{v}_t \sim N(0, \mathbf{\Sigma}_v)$.

1.3 Current estimation strategies

Current estimation strategies are based on the EKF, which approximates the pricing function y in Equation (1) using a first-order Taylor expansion around

[5] See Realdon (2006, p. 282).

[6] Andreasen and Meldrum (2019) claim that a similarly overidentified model with $\delta_1 = 0_K$ and $\Delta = 1_{K,K}$ achieves a fit close to a maximally flexible quadratic model.

the expectation of the state under the physical measure $\mathcal{P}$, $\mathbf{x}_{t|t-1} = E^{\mathcal{P}}_{t-1}\mathbf{x}_t$, such that:

$$\begin{aligned} y_{n,t} &= y(\mathbf{x}_t, n; \boldsymbol{\Psi}) \\ &\approx y\left(\mathbf{x}_{t|t-1}, n; \boldsymbol{\Psi}\right) + y'\left(\mathbf{x}_{t|t-1}, n; \boldsymbol{\Psi}\right)\left(\mathbf{x}_t - \mathbf{x}_{t|t-1}\right) \end{aligned} \tag{12}$$

This introduces approximation errors that bias the model estimates. The iterated extended Kalman filter advocated by Krippner (2013) handles this problem, and it can be shown that, under the observability restriction (Equation (13)), this technique is equivalent to our factor extraction technique. However, both the EKF and iterated EKF require knowledge of the parameters $\boldsymbol{\mu}^{\mathcal{P}}$ and $\boldsymbol{\Phi}^{\mathcal{P}}$ to extract $\mathbf{x}_{t|t-1}$, which rules out the use of the likelihood concentration techniques discussed in the next section.[7]

2. Estimation by Factor Extraction

In this section we show how the factor identification and likelihood-concentration methods currently used in estimating the GTSM can be used to estimate an SRTSM in a general Gaussian framework.

2.1 Nonlinear factor extraction

The GTSM literature suggests ways of inverting this relationship to extract the latent factors $\mathbf{x}_t$ from the yields $\mathbf{y}^o_t$. Following Duffie and Kan (1996) and many others, we can assume that there are $K < J - 1$ yields or fixed combinations (or "portfolios") of yields, given by a fixed $J \times K$ weighting matrix $\mathbf{W}$, that are nevertheless fitted without error:

$$\mathbf{q}_t \equiv \mathbf{W}'\mathbf{y}^o_t = \mathbf{W}'\mathbf{y}_t \tag{13}$$

for all t. Substituting Equation (11) into Equation (13) shows that this is equivalent to assuming: $\mathbf{W}'\mathbf{v}_t = 0$. We will refer to this as the observability restriction (OR) and assume that the observable factors $\mathbf{q}_t$ are the first K principal components obtained from the covariance matrix of yields, $\mathrm{Cov}(\mathbf{y}^o_t)$. In this case, the weights $\mathbf{W}$ are given by the eigenvectors of this matrix. Denote by $\mathbf{q}(\mathbf{x}_t; \boldsymbol{\Psi})$ the vector-valued function in $\mathbb{R}^K$ that maps the latent state vector $\mathbf{x}_t$ to the observable principal components $\mathbf{q}_t$, and the inverse of this function by $\mathbf{q}^{-1}(\mathbf{q}_t; \boldsymbol{\Psi})$, that is:

$$\mathbf{q}_t = \mathbf{q}(\mathbf{x}_t; \boldsymbol{\Psi}) = \mathbf{W}'\mathbf{y}(\mathbf{x}_t; \boldsymbol{\Psi}) \Longleftrightarrow \mathbf{x}_t = \mathbf{q}^{-1}(\mathbf{q}_t; \boldsymbol{\Psi}) = \mathbf{x}(\mathbf{y}^o_t; \boldsymbol{\Psi}). \tag{14}$$

This is the basis of our extraction procedure, which we call the factor extraction estimator.[8] Either way, substituting $\mathbf{x}_t$ back into Equation (11) gives a nonlinear econometric model of the cross-section of J observed yields:

$$\mathbf{y}^{\mathbf{o}}_{\mathbf{t}} = \mathbf{y}(\mathbf{x}(\mathbf{y}^{\mathbf{o}}_{\mathbf{t}}; \boldsymbol{\Psi}); \boldsymbol{\Psi}) + \mathbf{v}_t. \tag{15}$$

[7] For more detail on the extended Kalman filter, see Durbin and Koopman (2001) .

[8] In practical application, the speed of this inversion can be improved by the use of the analytical gradient to the observability condition in Equation (13) with respect to $\mathbf{x}_t$.

Our estimation strategy is to use this nonlinear solution technique to recover $\mathbf{x}_t$, conditional on the risk-neutral parameters using Equation (14) and use it to fit the observed yields as in Equation (15). Further details regarding implementation are set out in Section 2.2.

The FE method is similar in spirit to the sequential regression (SR) estimator (see Andreasen and Christensen, 2015). As its name suggests, the sequential regression uses a sequence of regressions to fit the model. This method first finds the values of $\mathbf{x}_t$ and $\boldsymbol{\Psi}$ that minimize the sum of squares in Equation (11) without using the observability condition in Equation (13). Then, given the factors, we estimate the parameters of the physical dynamics, including $\boldsymbol{\Sigma}$, which also contributes to the cross-sectional fit of the model. Finally, given the state vector and the updated estimate of $\boldsymbol{\Sigma}$, the risk-neutral parameters are reestimated by minimizing the cross-sectional sum of squares.

2.2 The separability of the likelihood function

Let $\boldsymbol{\Theta} \equiv (\boldsymbol{\mu}^{\mathcal{P}}, \boldsymbol{\Phi}^{\mathcal{P}}, \mu_\infty^{\mathcal{Q}}, \boldsymbol{\lambda}^{\mathcal{Q}}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}_r, \boldsymbol{\Sigma}_v)$ denote the parameters to be estimated. The conditional log-likelihood function is:

$$\log \mathcal{L}(\boldsymbol{\Theta}) = \sum_{t=2}^{T} \ell(\mathbf{y}_t^o | \mathbf{y}_{t-1}^o; \boldsymbol{\Theta}). \tag{16}$$

The first advantage of the factor extraction schemes is that, in contrast to the EKF, they allow the parameters of the measurement errors $\boldsymbol{\Sigma}_v$ to be concentrated out of the likelihood function. Second, they allow for the separation of parameters in the conditional likelihood. In particular, the logarithm of the t-period conditional density can be decomposed into the log-likelihood of observing the yields given the noiseless yield portfolios, $\ell^{\mathcal{Q}}(\mathbf{y}_t^o | \mathbf{q}_t; \boldsymbol{\Theta})$, and the log-likelihood of observing the latter given the lagged fitted portfolios, $\ell^{\mathcal{P}}(\mathbf{q}_t | \mathbf{q}_{t-1}; \boldsymbol{\Theta})$:

$$\ell(\mathbf{y}_t^o, \mathbf{q}_t | \mathbf{q}_{t-1}; \boldsymbol{\Theta}) = \ell^{\mathcal{Q}}(\mathbf{y}_t^o | \mathbf{q}_t; \mu_\infty^{\mathcal{Q}}, \boldsymbol{\lambda}^{\mathcal{Q}}, \boldsymbol{\Sigma}, \underline{r}, \boldsymbol{\Sigma}_v) + \ell^{\mathcal{P}}(\mathbf{q}_t | \mathbf{q}_{t-1}; \boldsymbol{\mu}^{\mathcal{P}}, \boldsymbol{\Phi}^{\mathcal{P}}, \boldsymbol{\Sigma}), \tag{17}$$

allowing the latent state vector to be recovered independently of the $\mathcal{P}$-parameters. Conditional upon the risk-neutral parameters and the latent factors, the parameters of the physical dynamics that maximize the likelihood function can be obtained by the OLS estimates of Equation (2). They can be concentrated out of the likelihood function and recovered subsequently, as in the GTSM put forward by Joslin, Singleton, and Zhu (2011). This leaves us with only $K \times (K+1)/2 + 1$ parameters ($\mu_\infty^{\mathcal{Q}}$, $\boldsymbol{\lambda}^{\mathcal{Q}}$, and $\boldsymbol{\Sigma}$) that need to be found numerically.[9] The mapping between the observable principal components and the latent factors requires an adjustment to the likelihood function through the application of the change-of-variable technique. Further details on this are provided in the Appendix.

[9] As in Andreasen and Meldrum (2019), we set $\underline{r} = 0$.

In the GTSM the mapping between the observed and latent factors is linear given the OR in Equation (13). As Joslin, Singleton, and Zhu (2011) note, the no-arbitrage restrictions used to estimate the risk-neutral dynamics from the cross-section are irrelevant to the estimation of the physical dynamics of the observable factors $\mathbf{q}_t$ in this case. Moreover, as Joslin, Le, and Singleton (2013) show, this irrelevance proposition carries over to a macrofinance model in which the macro factors are unspanned (i.e., they do not have an immediate effect on the term structure, but drive the dynamics of the term structure factors). However, the mapping from the observed to the latent factors is nonlinear in models with a lower bound, which means that the irrelevance proposition does not hold. Indeed, we need to use the model of the cross-section of yields in order to extract the underlying latent factors from the yield observations and find their time-series dynamics.

2.3 Alternative factor normalization of the shadow rate model

The parameterization in terms of the dynamics of the Gaussian factors $\mathbf{x}_t$ is convenient from the point of view of econometric modeling, but unlike the classic level, slope, and curvature factors of conventional term structure models, they lack intuition and a clear economic interpretation.[10] To this end, since the factors can be arbitrarily rotated, we can transform them to interpretable quantities. As such, we follow Lemke and Vladu (2017) and rotate the latent factors $\mathbf{x}_t$ to Gaussian shadow principal components ($\mathbf{q}_t^s$):

$$\mathbf{q}_t^s = \mathbf{W}'\mathbf{y}_t^G, \tag{18}$$

where $\mathbf{W}$ are the loadings of the principal components of yields and $\mathbf{y}_t^G$ are shadow Gaussian yields—that is, hypothetical yields that would prevail if the lower bound did not exist:

$$\mathbf{y}_t^G = \mathbf{a} + \mathbf{B}\mathbf{x}_t, \tag{19}$$

where $\mathbf{a}$ and $\mathbf{B}$ are affine coefficients that follow well-known recursions (see, e.g., Wu and Xia, 2016) and are found from the risk-neutral parameters $\boldsymbol{\lambda}^{\mathcal{Q}}$, $\boldsymbol{\mu}^{\mathcal{Q}}$, and $\boldsymbol{\Sigma}$. The shadow principal components preserve the Gaussian dynamics:

$$\mathbf{q}_t^s = \boldsymbol{\mu}_s{}^{\mathcal{P}} + \boldsymbol{\Phi}_s^{\mathcal{P}}\mathbf{q}_{t-1}^s + \mathbf{u}_{s,t}^{\mathcal{P}}, \tag{20}$$

$$\mathbf{q}_t^s = \boldsymbol{\mu}_s{}^{\mathcal{Q}} + \boldsymbol{\Phi}_s^{\mathcal{Q}}\mathbf{q}_{t-1}^s + \mathbf{u}_{s,t}^{\mathcal{Q}}, \tag{21}$$

where $Var[\mathbf{u}_{s,t}^{\mathcal{P}}] = Var[\mathbf{u}_{s,t}^{\mathcal{Q}}] = \boldsymbol{\Sigma}_s$ and

$$\boldsymbol{\Phi}_s^{\mathcal{Q}} = (\mathbf{W}'\mathbf{B})\boldsymbol{\Phi}^{\mathcal{Q}}(\mathbf{W}'\mathbf{B})^{-1},$$

$$\boldsymbol{\mu}_s{}^{\mathcal{Q}} = (\mathbf{W}'\mathbf{B})\boldsymbol{\mu}^{\mathcal{Q}} + (\mathbf{I} - \boldsymbol{\Phi}_s^{\mathcal{Q}})\mathbf{W}'\mathbf{a},$$

[10] We are grateful to a referee of this journal for suggesting this.

and similar expressions for $\mathbf{\Phi}_s^{\mathcal{P}}$ and $\boldsymbol{\mu}_{\mathbf{s}}{}^{\mathcal{P}}$. The rotated system for the shadow Gaussian yields is:

$$\mathbf{y}_t^G = \mathbf{a}_s + \mathbf{B}_s \mathbf{q}_t^s, \tag{22}$$

where

$$\mathbf{a}_s = \left(\mathbf{I} - (\mathbf{W}'\mathbf{B})^{-1}\mathbf{W}'\right)\mathbf{a},$$
$$\mathbf{B}_s = \mathbf{B}(\mathbf{W}'\mathbf{B})^{-1}.$$

In particular, the shadow short rate is given by

$$s_t = \delta_{s,0} + \boldsymbol{\delta}'_{s,1}\mathbf{q}_t^s \tag{23}$$

with

$$\delta_{s,0} = \delta_0 - \boldsymbol{\delta}_1{}'(\mathbf{W}'\mathbf{B})^{-1}\mathbf{W}'\mathbf{a},$$
$$\boldsymbol{\delta}'_{s,1} = \boldsymbol{\delta}'_1(\mathbf{W}'\mathbf{B})^{-1}.$$

Note that the risk-neutral dynamics of the shadow principal components are still parameterized only in terms of $\boldsymbol{\lambda}^{\mathcal{Q}}$, $\boldsymbol{\mu}^{\mathcal{Q}}$, and $\mathbf{\Sigma}$. The parameters $\boldsymbol{\lambda}^{\mathcal{Q}}$ are the roots of the autoregressive matrix of the factors and, as such, determine factor persistence under the risk-neutral probability measure. Importantly, the roots are invariant to linear transformation. Moreover, following Zellner (1962), the estimates of $\mathbf{\Phi}_s^{\mathcal{P}}$ and $\boldsymbol{\mu}_s{}^{\mathcal{P}}$ are independent of the covariance matrix of $\mathbf{u}_{s,t}^{\mathcal{P}}$, which allows us to parameterize the model in terms of $\mathbf{\Sigma}_s$ rather than $\mathbf{\Sigma}$ since

$$\mathbf{\Sigma} = (\mathbf{W}'\mathbf{B})^{-1}\mathbf{\Sigma}_s(\mathbf{W}'\mathbf{B}). \tag{24}$$

Apart from an intuitive interpretation of the factors, the advantage of the model normalization in terms of shadow principal components is that, if the principal components and shadow principal components are highly correlated (which we find in practice), the estimate of the covariance matrix of shocks from the OLS regression of the dynamics of principal components can be used as a reliable starting value in the numerical search for $\mathbf{\Sigma}_s$.

In the case of a fully Gaussian model (i.e., without a lower bound) this normalization is identical to that proposed by Joslin, Singleton, and Zhu (2011). In their model, the parameters of the real-world dynamics $\mathbf{\Phi}_s^{\mathcal{P}}$ and $\boldsymbol{\mu}_s^{\mathcal{P}}$ can be estimated from the OLS regression of the VAR for principal components $\mathbf{q}_t$ independently of the numerical optimization of the likelihood function. For the shadow rate model, shadow principal components are conditional on the risk-neutral parameters and, as such, the estimation of the real-world dynamics has to be implemented inside the numerical optimization routine. Since they can be estimated by OLS, however, this adds only marginally to the computational burden of the optimization. Thus, the conditional log-likelihood function for the shadow rate model is:

$$\ell(\mathbf{y}_t^o, \mathbf{q}_t | \mathbf{q}_{t-1}; \mathbf{\Theta}) = \ell^{\mathcal{Q}}(\mathbf{y}_t^o | \mathbf{q}_t; \mu_\infty^{\mathcal{Q}}, \boldsymbol{\lambda}^{\mathcal{Q}}, \mathbf{\Sigma}_s, \underline{r}, \mathbf{\Sigma}_v) + \ell^{\mathcal{P}}(\mathbf{q}_t | \mathbf{q}_{t-1}; \boldsymbol{\mu}^{\mathcal{P}}, \mathbf{\Phi}^{\mathcal{P}}, \mathbf{\Sigma}_s). \tag{25}$$

3. Empirical Evaluation of the Factor Extraction Approach

3.1 Data and model specifications

We estimate these various models using Treasury yields constructed by Gurkaynak, Sack, and Wright (2007).[11] We use the same maturities as Wu and Xia (2016)—that is 3 and 6 months, and 1, 2, 5, 7, and 10 years—but we use a longer sample period, from January 1981 to December 2021. We follow Krippner (2013), Christensen and Rudebusch (2015), Coroneo and Pastorello (2020), and most of the literature on the term structure modeling within the Gaussian framework by estimating the model using yields rather than the forward rates fitted by Wu and Xia (2016).

Before we apply our factor extraction method to the analysis of monetary policy in the next section, we first examine its performance and compare its efficiency to other popular estimation methods. We test it on two nonnegative Gaussian models: the shadow rate model and the QTMS. These models are estimated both with a zero restriction on the lower bound parameter ($\underline{r} = 0$) and with $\underline{r}$ estimated as a free parameter ($\underline{r} = \widehat{r}$). We estimate the shadow rate model using the exact functional form derived by Priebsch (2013) and the approximation proposed by Wu and Xia (2016). The QTMS is described in detail in Realdon (2006), and in the implementation we adopt the same parameterization as in Andreasen and Meldrum (2019).

The first benchmark against which we compare our factor extraction method is the extended Kalman filter (henceforth EKF), arguably the most common in the literature. This estimation approach is based on the state-space system. For nonlinear systems, such as those that we consider here, it requires linearization of the measurement equation, which introduces an approximation error. For a direct comparison of the EKF with the factor extraction method, we first estimate the model with the OR imposed: $\mathbf{W}'\mathbf{v} = 0$.[12] The difference in the fit of the models estimated by the EF and this EKF method allows us to quantify the approximation errors due to the linearization in Equation (12) used in the EKF (see Section 3.4). We then estimate the shadow rate model by the EKF without imposing the OR, in order to examine how the restriction affects model properties. In the context of the Gaussian model, Joslin, Singleton, and Zhu (2011) show that relaxing this restriction has very little effect on the model, but this conclusion cannot generally be extended to nonlinear models and in particular the SRTSM, as we demonstrate in Section 3.5.

The second benchmark for comparison is given by the sequential regression procedure (SR) employed by Andreasen and Meldrum (2019). Recall that, this method sequentially fits the model of the cross-section (using nonlinear least

[11] The data is available from: http://www.federalreserve.gov/pubs/feds/2006/200628/feds200628.xls.

[12] We adopt the same assumptions as Joslin, Singleton, and Zhu (2011) (see their footnote 24) that the $J-K$ PCs not used to span the yield curve have an *i.i.d.* normal distribution. In particular, denote the base for these PCs (i.e., the remaining eigenvectors of the yield covariance matrix) by $\mathbf{W}_\perp$, which is orthonormal to $\mathbf{W}$. Then, we assume that $\mathbf{W}_\perp' \mathbf{y}_t^o \sim N(0, \sigma_v^2 \mathbf{I}_{J-K})$, or equivalently, the yield measurement errors are $\mathbf{v}_t \sim N(0, \sigma_v^2 \mathbf{W}_\perp' \mathbf{W}_\perp)$.

squares) and the time-series model (using OLS). To implement this procedure, we need to drop the OR. Importantly, unfortunately, this method is not likelihood-based, which makes it difficult to compare this method with likelihood-based methods.

3.2 Fitting a three-factor model

The Joslin, Singleton, and Zhu (2011) parameterization requires the roots of the model $\boldsymbol{\lambda}^{\mathcal{Q}}$ to be specified as real or complex, distinct or repeated. We find that with our data sample, the model specified in terms of real distinct roots achieves substantially higher likelihood values for all algorithms; therefore, we report only the estimates for this case.

Table 1 shows the fit of the three-factor model obtained using these various algorithms, reporting the likelihood of the model and the cross-sectional root-mean-square errors (RMSEs) for each maturity. Naturally, since it is designed to minimize the latter, the SR method does best in this respect. However, these errors are a tiny fraction of the overall prediction error, which takes into account the error in forecasting the factors one period ahead. It thus does worse than the other methods in terms of likelihood. The effect of the poor forecasting performance of this approach becomes clear when we split the log-likelihood into the cross-section ($\log \mathcal{L}^{\mathcal{Q}} = \sum_{t=2}^{T} \ell^{\mathcal{Q}}(\mathbf{y}_t^o|\mathbf{q}_t;\boldsymbol{\Theta})$) and time-series components ($\log \mathcal{L}^{P} = \sum_{t=2}^{T} \ell^{\mathcal{P}}(\mathbf{q}_t|\mathbf{q}_{t-1};\boldsymbol{\Theta})$) shown in the last three columns of the table.

The first row of Table 1 shows the fit of the Joslin, Singleton, and Zhu (2011) GTSM. As we would expect, it has a lower likelihood than models with the lower bound over this period, except the models estimated by SR. The RMSEs of the residuals in these models are generally in line with the bid-ask spreads in the Treasury market, although the RMSEs for the 5-year and 10-year maturities are higher than those for the other maturities.

In our sample, among maximum likelihood-based methods, we find that the FE method has a lower average RMSE than other estimators, while the EKF generally has the worst performance in this respect. The average RMSE is as much as 0.30 basis points below that for the shadow rate model (compare FE_{wx} versus EKF_{wx} with the estimated lower bound) and 0.32 basis points for the QTSM (see FE_{qd} versus EKF_{qd} with $\underline{r} = 0$). However, because the likelihood-based models optimize the cross-sectional maturity weights, rather than setting them to unity as in the SR approach, it is best to compare these models in terms of $\log \mathcal{L}^{\mathcal{Q}}$ rather than average RMSE. In this respect, the cross-sectional fit and indeed the overall fit of the EKF without the OR and lower bound restrictions is best. However, as we will see from the discussion of Section 3.5, this is due to overfitting.

A model that estimates the lower bound parameter does not necessarily fit the cross-section of yields better than one in which it is arbitrarily set to zero, since the gain in the time-series fit can offset the loss from the cross-sectional fit (as is illustrated by some of the results for the EKF). Nevertheless, setting the lower bound as a free parameter is more flexible than setting it to zero and must attain a

Table 1
Statistics of fit for different estimation methods

Estimation method	OR?	Root-mean-square error								Log-likelihood		
		3m	6m	1y	2y	5y	7y	10y	Av. RMSE	$\log \mathcal{L}^Q$	$\log \mathcal{L}^P$	$\log \mathcal{L}$
Gaussian model												
JSZ	Yes	4.80	4.06	5.00	2.91	6.24	2.59	5.76	4.48	11,733.24	6,269.25	18, 002.49
Shadow rate model, $\underline{r} = 0$												
FE_{wx}	Yes	4.62	3.99	4.67	2.79	5.91	2.51	5.48	4.28	11,824.03	6,389.62	18,213.65
FE_{pr}	Yes	4.63	3.99	4.70	2.81	5.93	2.53	5.52	4.30	11,815.66	6,387.73	18,203.39
EKF_{wx}	Yes	4.74	4.12	4.78	3.03	6.09	2.59	5.54	4.42	11,768.72	6,432.88	18,201.60
EKF_{pr}	Yes	4.75	4.15	4.80	3.04	6.12	2.61	5.58	4.44	11,760.34	6,430.50	18,190.84
EKF_{wx}	No	5.26	4.10	5.05	3.26	5.89	2.77	5.45	4.54	12,275.15	5,938.63	18,213.78
SR_{wx}	No	4.64	4.03	4.75	2.95	5.77	2.53	5.35	4.29	11,829.31	6,163.15	17,992.46
Shadow rate model, $\underline{r} = \widehat{\underline{r}}$												
FE_{wx}	Yes	4.59	3.98	4.63	2.81	5.84	2.48	5.42	4.25	11,840.31	6,393.85	18,234.16
FE_{pr}	Yes	4.58	3.98	4.64	2.83	5.83	2.50	5.43	4.25	11,839.30	6,391.12	18,230.42
EKF_{wx}	Yes	5.27	4.47	4.88	3.13	6.02	2.59	5.49	4.55	11,714.03	6,520.52	18,234.55
EKF_{pr}	Yes	5.27	4.49	4.91	3.14	6.04	2.61	5.53	4.57	11,709.09	6,514.20	18,223.29
EKF_{wx}	No	6.18	4.69	4.86	3.64	5.35	2.62	4.97	4.62	12,246.18	6.158.79	18,404.97
SR_{wx}	No	4.45	4.32	4.09	3.63	5.13	2.29	5.06	4.14	11,917.53	5,526.43	17,443.96
Quadratic model, $\underline{r} = 0$												
FE_{qd}	Yes	4.38	3.93	4.36	2.88	5.48	2.42	5.11	4.08	11,930.74	6,494.77	18,425.51
EKF_{qd}	Yes	5.10	4.05	4.45	3.32	5.87	2.83	5.17	4.40	11,794.57	6,635.19	18,429.76
SR_{qd}	No	4.38	3.89	4.37	3.08	5.31	2.42	4.97	4.06	11,948.54	5,935.61	17,884.15
Quadratic model, $\underline{r} = \widehat{\underline{r}}$												
FE_{qd}	Yes	4.38	3.93	4.37	2.88	5.47	2.42	5.11	4.08	11,930.90	6,494.62	18,425.52
EKF_{qd}	Yes	5.07	4.05	4.45	3.32	5.86	2.82	5.17	4.39	11,797.43	6,635.29	18,432.72
SR_{qd}	No	4.36	3.89	4.31	3.21	5.20	2.40	4.84	4.03	11,967.94	5,876.59	17,884.53

The table reports the root mean-square-error (RMSE) for each yield, the cross-sectional average RMSE, and the value of the log-likelihood function at the estimated maximum. The sample period is January 1981 to December 2021.

higher likelihood in the case of the likelihood-based methods (FE and EKF), though not necessarily hold for the SR estimator (compare the two SR results for the shadow rate model). Similarly, shadow rate models without the OR attain significantly higher likelihood values, but their cross-sectional fit is noticeably worse.

Finally, as noted in Section 2.3, we can normalize the shadow rate model in terms of shadow principal components. This offers the advantage of making the factors more readily interpretable and allows us to use the variance matrix of residuals from the VAR(1) system for the principal components as starting values for $\mathbf{\Sigma}_s$ in the numerical maximization of the likelihood function. To verify this conjecture, in Table 2 we report the Cholesky factor of the covariance matrix of VAR(1) residuals from principal components regression and the estimates of $\mathbf{\Sigma}_s^{1/2}$, where, with a slight abuse of notation, $\mathbf{\Sigma}_s^{1/2}$ denotes a lower triangular matrix, such that $\mathbf{\Sigma}_s^{1/2} \times \mathbf{\Sigma}_s^{1/2\prime} = \mathbf{\Sigma}_s$. The two sets of estimates are indeed close to each other. To complete this analysis, Figure 2 shows the principal components, shadow principal components and the underlying $\mathbf{x}_t$ factors from the shadow rate model estimated by FE_{wx} with $\underline{r} = 0$. The principal components have the classic interpretation as level, slope, and curvature factors. We can see that the two sets of

Table 2
Covariance matrix of VAR(1) residuals

Σ_q			Σ_s		
0.0082	0	0	0.0090	0	0
−0.0003	0.0035	0	−0.0005	0.0038	0
−0.0001	−0.0007	0.0014	−0.0002	−0.0008	0.0016

The table reports the covariance matrix of VAR(1) residuals from observed principal components (left panel) and shadow principal components from the shadow rate model estimated by factor extraction with $\underline{r} = 0$.

principal components (observed and shadow) indeed move closely together, except when the interest rates are at the lower bound, 2009-2015 and 2020-2021, which supports our conjecture in Section 2.3. The same, however, cannot be said about the latent factors $\mathbf{x}_t$; the first factor $x_{1,t}$ exhibits a simple down-trend, driven by the close-to-unit root behavior, while the other two exhibit more exaggerated cyclical and curvature patterns than the principal components do.

3.3 Estimation times

The main advantage of the factor extraction method is time efficiency. This results from its ability both to concentrate the parameters of the measurement error covariance matrix and the real-world dynamics from the likelihood and to extract the state vector using the analytical gradient of the mapping function $y(.)$. To illustrate this advantage, Table 3 reports the time needed to estimate these models with different algorithms using the Joslin, Singleton, and Zhu (2011) parameters as starting values. Naturally, the exact estimation time depends on many different factors, such as implementation, hardware, software, etc. Although we used our best efforts to make the comparison as fair as possible, the numerical optimization is strongly path-dependent and so our results should be treated as indicative.[13]

The table shows that there is little difference between the FE and SR methods in terms of estimation time. They both find the optimum within a few minutes. Estimating the shadow rate model with the Wu and Xia (2016) approximation takes less than 10 minutes with the FE, while it takes about 0.5–1 hour with the EKF. A much larger gain is obtained for the QTSM. The FE method allows us to estimate this in about 8–10 minutes, while it takes about 12–13 hours with the EKF. The sequential regression method is rapid but, as noted earlier, it does not maximize the likelihood. The Priebsch (2013) solution is more computationally demanding. The FE is handicapped by the difficulty of extracting the factors from a highly nonlinear system and the EKF by the large number of parameters. In both

[13] All computations are performed on a PC desktop Windows 10 Enterprise 64-bit operating system with the Intel^{R} Core^{TM} i5 3.20 GHz processor with 8 GB RAM using Matlab R2018a. As the numerical optimizers we use Matlab functions "fminsearch" and "fminunc" consecutively, in other words, we start another round of the optimization using "fminsearch" and "fminunc" with the results from the previous round as the starting values. Typically, the maximum value of the likelihood function found with this loop routine is much higher than a single-round optimization. We do not modify the default setting of the optimization routines, but to avoid stalling the routine with micro gains, we stop the estimation when the improvement of the log-likelihood between loops is smaller than 0.01, which is usually in the small proximity to the point in which no further progress can be made.

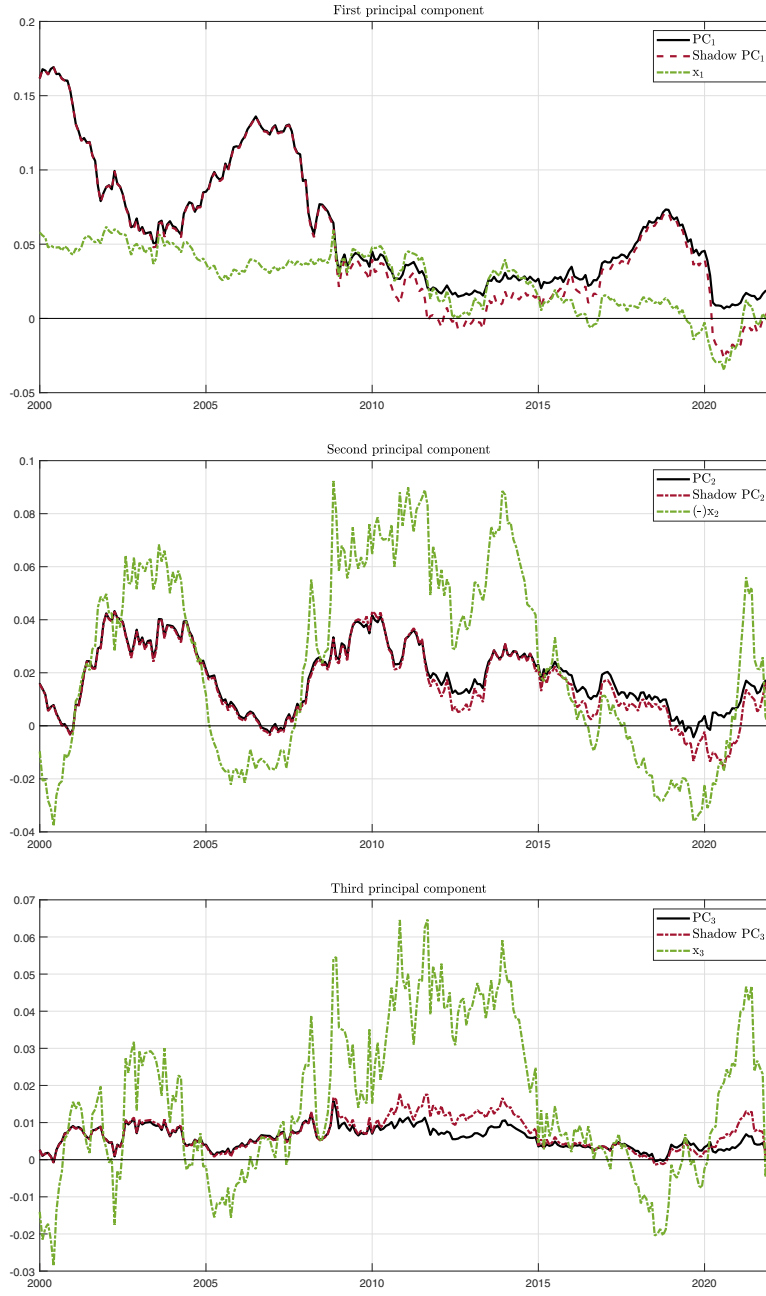

Figure 2
Principal components and shadow principal components
The figure shows principal components, shadow principal components, and the underlying $\mathbf{x}_t$ factors from the shadow rate model estimated by factor extraction with $\underline{r} = 0$.

Table 3
Estimation time for different estimation methods

Method	OR?	Time
		Model: Gaussian
JSZ	Yes	3s
		Shadow rate model, $\underline{r} = 0$
FE_{wx}	Yes	8m
FE_{pr}	Yes	13h:58m
EKF_{wx}	Yes	60m
EKF_{pr}	Yes	21h:38m
EKF_{wx}	No	2h:4m
SR_{wx}	No	2m
		Shadow rate model, $\underline{r} = \widehat{\underline{r}}$
FE_{wx}	Yes	4m
FE_{pr}	Yes	26h:38m
EKF_{wx}	Yes	35m
EKF_{pr}	Yes	16h:59m
EKF_{wx}	No	1h:9m
SR_{wx}	No	8m
		Quadratic model, $\underline{r} = 0$
EKF_{qd}	Yes	4h:21m
FE_{qd}	Yes	8m
SR_{qd}	No	4m
		Quadratic model, $\underline{r} = \widehat{\underline{r}}$
EKF_{qd}	Yes	17h:41m
FE_{qd}	Yes	10m
SR_{qd}	No	5m

All models in the table are specified with the observability restriction. The sample period is January 1981 to December 2021.

cases the estimation time can take between 14 to 26.5 hours. As noted, these were estimated using starting values from the Joslin, Singleton, and Zhu (2011) Gaussian model, but since the parameters of the Wu and Xia (2016) models are very close, considerable time can be saved using these as starting values.

In our preliminary analysis we found that the estimates of the Gaussian model provided reliable starting values for the shadow rate model, but this was generally not the case for the QTSM. To find the global optimum for the QTSM, we needed to implement the routine multiple times from different starting values, which compounds the estimation time considerably.

3.4 Approximation errors

In this section, we analyze two possible types of approximation error: (a) the linearization in Equation (12), which is necessary in the EKF filtering procedure, and (b) the WX approximation of the mapping function in Equation (6). This issue is important from the perspective of using the model for economic inference, such as measuring term premia or policy expectations.

First, when assessing the effect of the linearization at (a), we take the estimates of the model at the optimum as reported in Table 1 and examine the difference in the fitted values between the EKF (with the OR) and the FE estimator, which does

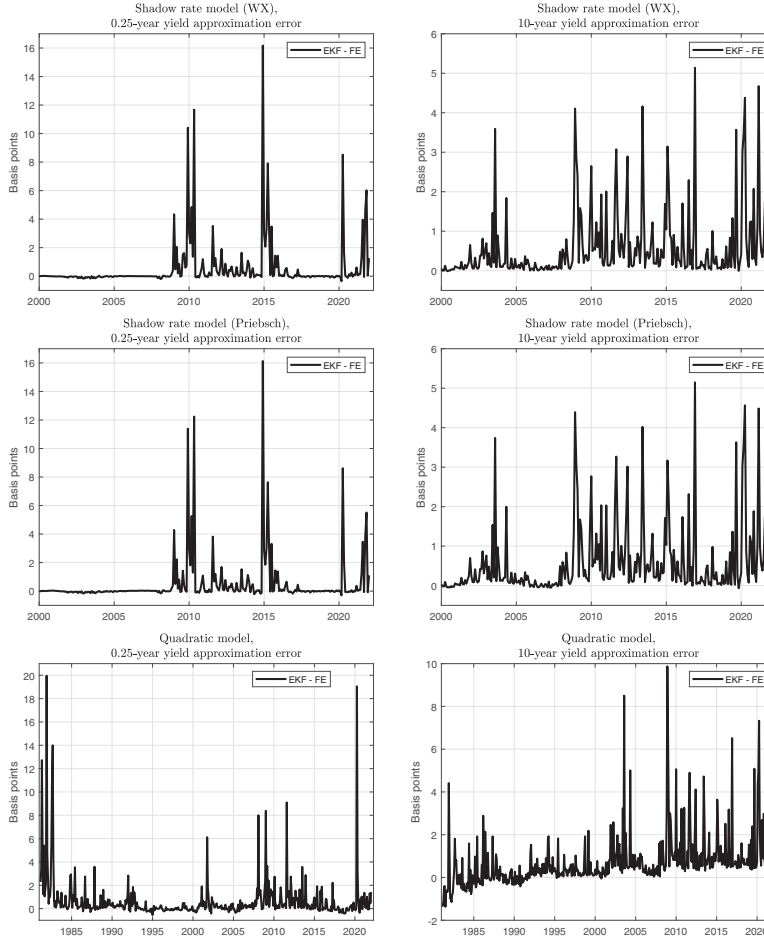

Figure 3
Linearization errors
Linearization errors in the extended Kalman filter for different models. The estimation sample is January 1981 to December 2021.

not involve any linearization. Figure 3 shows these differences for the shadow rate model with the Wu and Xia (2016) approximation (top row) and the Priebsch (2013) solution (middle row), and the QTSM (bottom row). We plot the errors for the 3-month yield (left column) and the 10-year yield, as the errors for other yields take the intermediate values between these two yields.[14] The plots for the shadow rate model start from January 2000, because before that the interest rates were far from the lower bound and for that period the linearization error is virtually zero.

[14] To save space, all plots are for models with the lower bound parameter set to zero ($\underline{r} = 0$), but the results for models with the estimated lower bound are similar.

It is interesting to note that this particular linearization induces a positive bias in the fitted values of yields. In the SRTSM, the errors in the short rate basically occur only when the interest rates were at the lower bound, that is, in the period 2009–2016 and since 2020. For the long rate, however, the linearization errors are also present around 2003–2004 and then throughout the whole period since 2008. This indicates that the shadow rate model gains traction when interest rates are low but not necessarily at the lower bound. On the other hand, the linearization errors in the long rate are less prone to taking large values and rarely exceed 5 basis points. The errors in the QTSM shown in the bottom panel seem to be present over the whole sample. They also tend to be bigger in magnitude.

We evaluate the distortion (b) introduced by the approximation in the mapping function proposed by Wu and Xia (2016) for the shadow rate model by comparing it with the estimates obtained by using the Priebsch (2013) formula, which is exact for the SRTSM. To simulate the practical application, both formulae are evaluated at the respective maximum likelihood. Figure 4 shows the difference in fitted values of the 3-month (upper panel) and 10-year (lower panel) yields between the two schemes (estimated with the lower bound parameter set to zero). The figure makes it evident that the Wu and Xia approximation does a very good job. In our sample the approximation error is always smaller than 1 basis point, which makes it negligible for virtually all practical purposes. This is important in light of the substantial computational burden of the estimation of the model with the Priebsch mapping function noted in Section 3.3.

Finally, the impact of linearization errors in the context of a shadow rate model can also be illustrated by its impact on the estimates of the shadow short rate. The first two panels of Figure 5 show the shadow rate estimated by the EKF and the FE method with Wu and Xia approximation.[15] Although the estimates of the shadow rate by the EKF and FE methods with the OR are similar for most of the time, a sudden divergence emerges from time to time. This is more visible when the lower bound parameter is estimated (middle panel). In general, however, the estimates of the shadow rate with the estimated lower bound parameter are more likely to exhibit sudden jumps, which could indicate model overfitting. This observation makes us favor the more parsimonious model with $\underline{r}$ set to zero.

3.5 The estimates of the shadow rate and shadow yields

Given its relevance to the analysis of monetary policy, this section focuses on the way that the choice of estimation method affects the properties of the shadow rate model. Figure 5 shows the federal funds policy rate alongside the shadow short rates (SSRs) implied by different estimation methods: the standard EKF, the EKF with the OR restriction enforced, and the FE method, which is squarely based on the OR restriction. The top panel presents the SSRs estimated with the lower bound parameter set to zero, while the middle panel shows these when this parameter is

[15] Note that the linearization errors shown in the first two panels of Figure 3 for the Wu and Xia (2016) and Priebsch (2013) estimators are visually identical. We do not consider the QTMS because it does not have a comparable shadow rate.

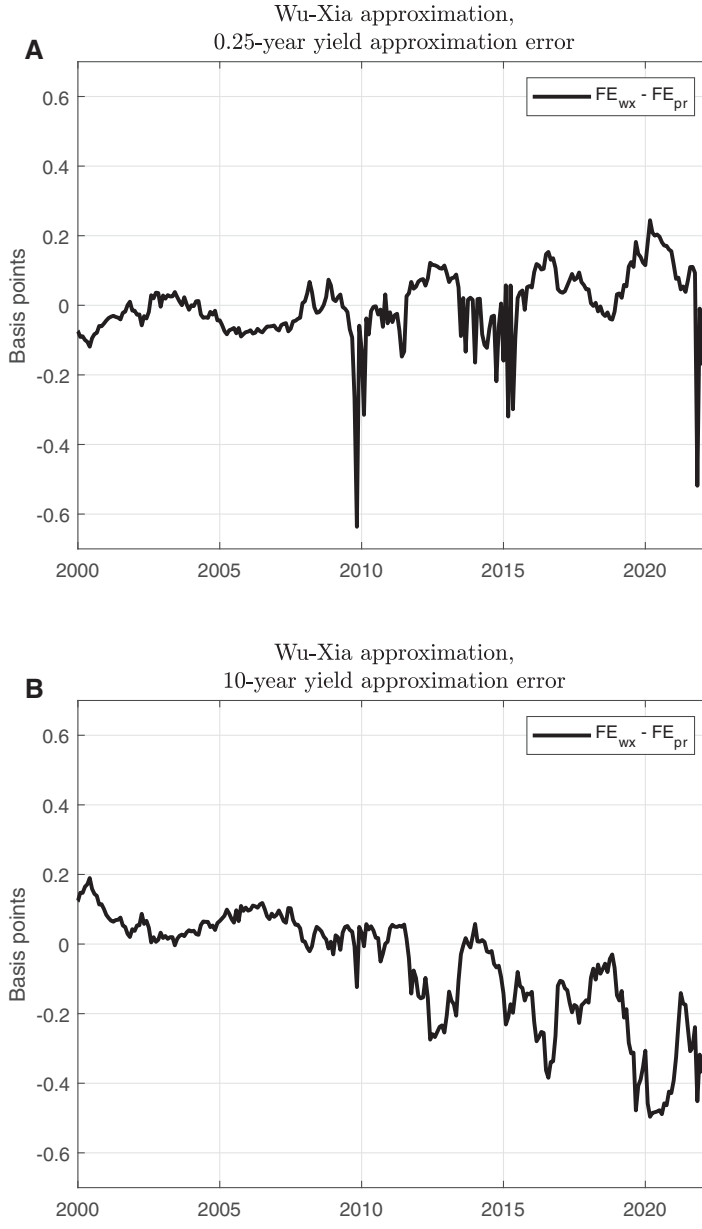

Figure 4
Wu-Xia approximation errors
The figure shows the difference between the fitted 0.25-year (panel A) and 10-year (panel B) yields obtained from the exact Priebsch (2013) shadow rate formula and its approximation proposed by Wu and Xia (2016).

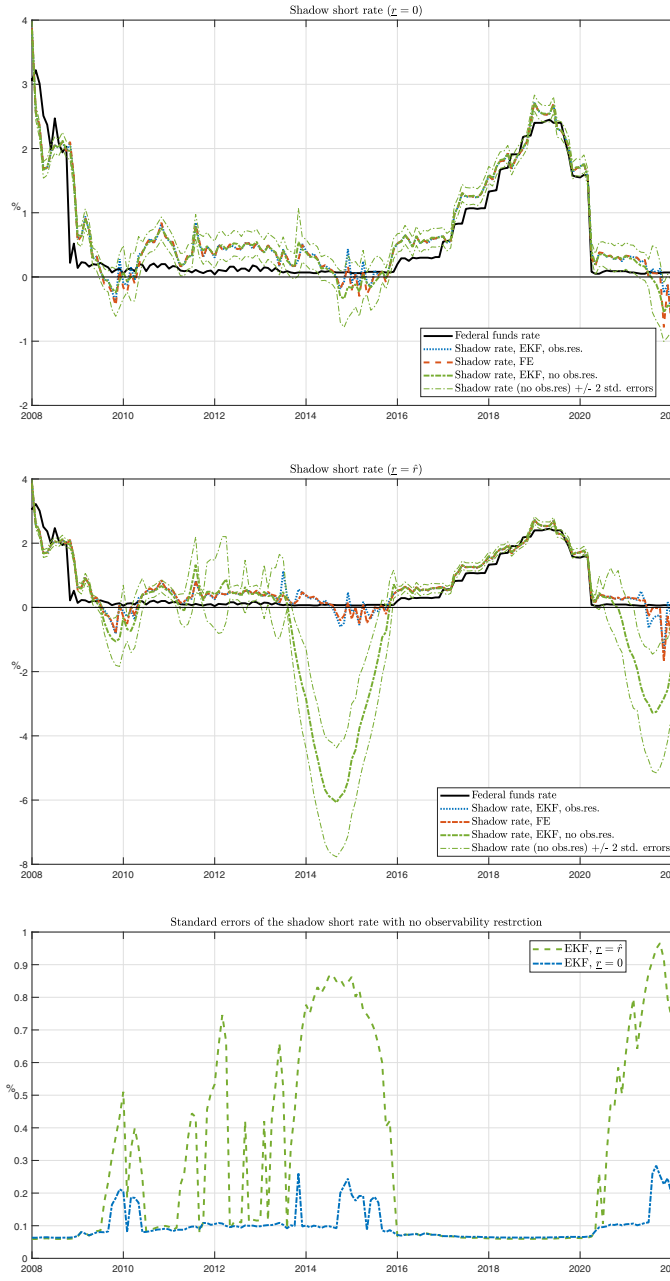

Figure 5
Short rate and shadow short rate implied by different estimation methods
The top panel presents the shadow short rates with the imposed restriction $\underline{r} = 0$, while in the middle panel the lower bound is estimated. The bottom panel shows standard errors of the shadow short rate estimated without the observability restriction. The federal funds rate is plotted for comparison.

freely estimated. The middle panel immediately reveals that unless the OR or lower bound restrictions are applied, the EKF can generate wildly negative values for the SSR when the federal funds rate is close to the lower bound and that these have large error bounds.

This effect was first noted by Krippner (2013), who argued that it was largely spurious, due to the flexibility with which the standard three-factor EKF can estimate the SRTSM. It occurs because the weights that map the short shadow rates into the Black estimates are close to zero at the lower bound. Raising the lower bound parameter effectively reduces these weights to zero. Since the short rates can be approximated by the lower bound parameter instead of term structure factors, this leaves the algorithm with more freedom to fit the long end, resulting in overfitting. This is clear from Table 1, which shows that if the standard EKF (i.e., with no OR) is free to estimate the lower bound, it trades off a small deterioration in the RMSEs of the short maturities to get a marked improvement in the long maturities. The log-likelihood components reported in Table 1 show that this allows it to return a relatively large value for the likelihood of the cross-section, with smaller deterioration in the time-series component.

This problem can be mitigated by imposing the lower bound constraint, or as Krippner has suggested, reducing the number of factors. It is clear that the OR also helps to mitigate the problem. To see the reason for this, recall that the SSR is the simple sum of the latent $\mathbf{x}_t$ factors in Equation (7). If the OR is imposed, the factors and therefore the SSR are effectively observed, since they are determined by rotating the PCs, which are observed without error. This rotation depends upon the risk-neutral parameters, which are estimated extremely precisely. Given the factors, this pins down the short end of the shadow yield curve conditional on the values of the risk-neutral parameters. Moreover, the OR restriction reduces the dimension space of the measurement errors, making it more difficult for the algorithm to generate negative values without a marked deterioration in the fit of the longer maturities.

The bottom panel of Figure 5 demonstrates the effect of relaxing the OR constraint. It shows the standard deviations of the standard EKF model with and without the lower bound restricted to zero. The small cross-sectional errors mean that the standard EKF determines the factors with a high degree of precision in an affine structure like the Gaussian model. This property holds in the SRTSM for the period before 2008 since the model is essentially Gaussian until then. However, these properties do not carry over to the zero lower bound (ZLB) period because, as noted, the weights that map the shadow short rates into the estimates of $\mathbf{x}_t$ are close to zero, weakening the links upon which the EKF depends to estimate the factors, reducing the precision with which they and the SSRs are estimated. Figure 5C clearly illustrates this effect. The shadow rates are precisely determined until the mid-December 2008 FOMC meeting, which cut the target fed funds range to 0–25 basis points. Then the shadow rates turn negative and the variance increases, especially if the lower bound restriction is not imposed.

The zero lower bound affects not only the short rate but all other yields through future interest rates expectations. Figure 6 shows the fitted 1-year, 5-year, and 10-year yields and their shadow counterparts, as defined in Equation (19), over the period 2008 to 2021. The top panel shows that the shadow 1-year yield was negative for most of the time between 2010 and 2015. The shadow 5-year rate fell to zero in 2012, putting it about 1% lower than the fitted yield, but it recovered to 1% in 2013. The figure shows a 1% spread between the 10-year yield and its shadow, but its shadow never reached zero during this episode. The coronavirus crisis looks particularly interesting in this light. The shadow short rate and 1-year yield were about –1% at the time. However, both the 5-year and 10-year shadow yields fell to lower values (below –1.5%) by mid-2020. This suggests that the market expected a prolonged recession, with shadow short rates remaining negative for longer than one year.

4. The Macrofinance Shadow Rate Model

Having validated the new factor extraction method, in this section we explore the implications of the macrofinance model pioneered by Ang and Piazzesi (2003). For this purpose we focus on the shadow rate model, since it lends itself to a natural interpretation for monetary policy purposes.

Term structure models are based on the observation that the cross-section has a low-dimension factor structure and that the risk-neutral dynamics thus have a similarly low dimension, as in Equations (3) and (21). However, the real-world dynamics can be much richer, typically involving macroeconomic variables that can influence the future evolution of the yield curve even though they do not affect its current position. In that sense they are "unspanned" (Duffee, 2011). In other words we need to re-specify the $\mathcal{P}$-dynamics in Equation (20), while keeping Equation (21) as it is. In Joslin, Priebsch, and Singleton (2014) the $\mathcal{P}$-dynamics are specified in terms of the PCs ($\mathbf{q}_t$), which span the term structure, while the physical dynamics involve $\mathbf{q}_t$, as well as a vector of macroeconomic or other (unspanned) state variables ($\mathbf{m}_t$) that do not. Consequently, the macro variables do not affect the term structure contemporaneously, but they can influence interest rate expectations under the $\mathcal{P}$-measure and hence the term premium. Joslin, Priebsch, and Singleton (2014), assume that under the $\mathcal{P}$-measure the spanned and unspanned factors can be specified jointly as a Gaussian VAR(1).

Joslin, Priebsch, and Singleton (2014) note that because their VAR includes only observable state variables ($\mathbf{q}_t$ and $\mathbf{m}_t$), these VAR parameters can be estimated by OLS independently of the $\mathcal{Q}$-dynamics, as in Joslin, Singleton, and Zhu (2011): the irrelevance proposition. However, in the shadow rate framework, the mapping from the principal components ($\mathbf{q}_t$) to the shadow principal components ($\mathbf{q}_t^s$) is nonlinear when interest rates are low. Reflecting this, $\mathbf{q}_t^s$ is Gaussian, while the observed variables $\mathbf{y}_t^o$ and hence $\mathbf{q}_t$ need not be. Unfortunately, as noted at the beginning of this paper, allowing $\mathbf{m}_t$ to affect the $\mathcal{P}$-dynamics increases the number of parameters in the VAR, many of which are insignificant in practice.

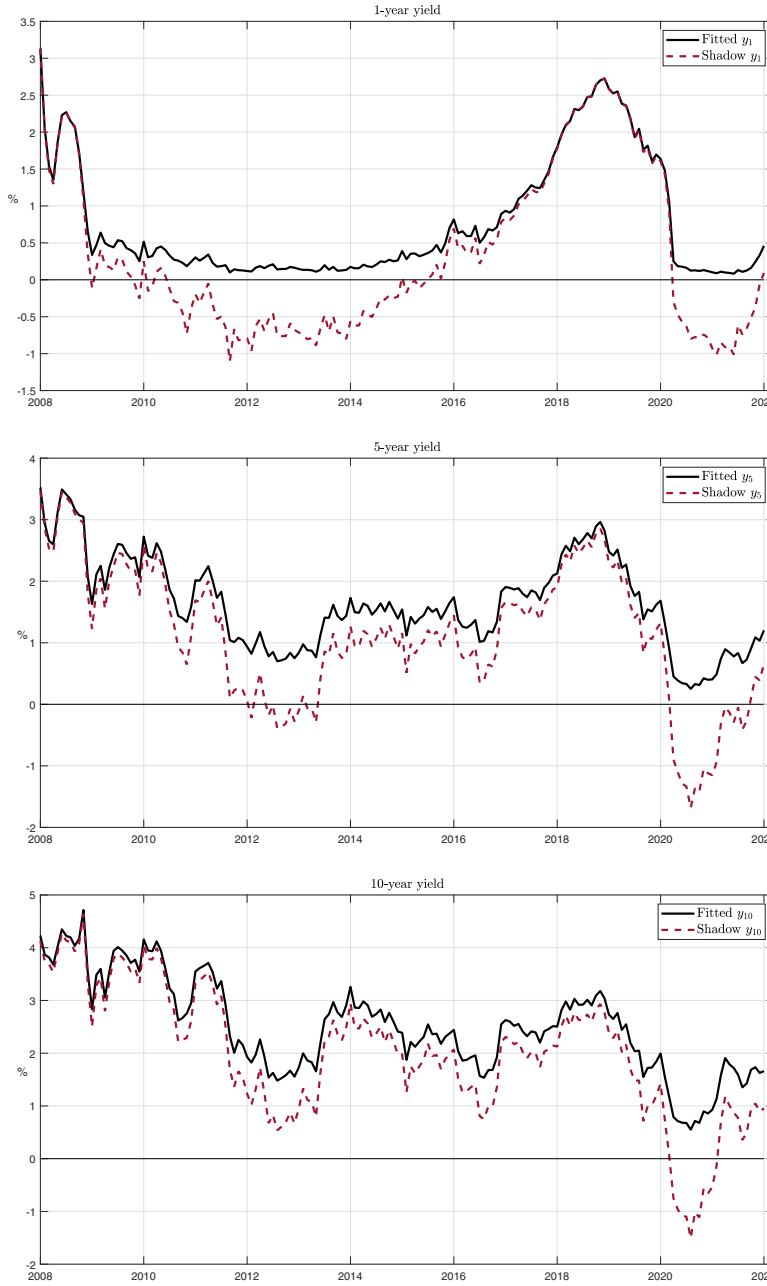

Figure 6
Fitted yields and shadow yields
The figure shows fitted yields and shadow yields with maturities 1-year, 5-years, and 10-years from the shadow rate model estimated by factor extraction with $\underline{r} = 0$.

Joslin, Priebsch, and Singleton (2014) deal with this problem by using exclusion restrictions found using a specification search procedure described in the next section. While it is not strictly necessary to do this, it is important to be confident about model properties and in particular the implications for the effects of policy on the term premium.

Because the shadow principal components are specific to the Black model of the $\mathcal{Q}$-dynamics, the parameters of this VAR cannot be estimated separately, as they can in Joslin, Priebsch, and Singleton (2014)—their irrelevance proposition does not hold. Following Joslin, Priebsch, and Singleton (2014), we assume that under the $\mathcal{P}$-measure the dynamics can be specified jointly as a Gaussian VAR(1):

$$\begin{bmatrix} \mathbf{q}_t^s \\ \mathbf{m}_t \end{bmatrix} = \begin{bmatrix} \mu_q^{\mathcal{P}} \\ \mu_m^{\mathcal{P}} \end{bmatrix} + \begin{bmatrix} \Phi_{qq}^{\mathcal{P}} & \Phi_{qm}^{\mathcal{P}} \\ \Phi_{mq}^{\mathcal{P}} & \Phi_{mm}^{\mathcal{P}} \end{bmatrix} \begin{bmatrix} \mathbf{q}_{t-1}^s \\ \mathbf{m}_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_{q,t}^{\mathcal{P}} \\ \mathbf{u}_{m,t}^{\mathcal{P}} \end{bmatrix}. \tag{26}$$

Assuming that $\mathbf{m}_t$ are unspanned, the $\mathcal{Q}$-dynamics in Equation (21) can also be written in this format as:

$$\mathbf{q}_t^s = \boldsymbol{\mu}_q^{\mathcal{Q}} + \begin{bmatrix} \Phi_{qq}^{\mathcal{Q}} & 0_{3,2} \end{bmatrix} \begin{bmatrix} \mathbf{q}_{t-1}^s \\ \mathbf{m}_{t-1} \end{bmatrix} + \mathbf{u}_{q,t}^{\mathcal{Q}}. \tag{27}$$

To remain close to the Gaussian benchmark in Joslin, Priebsch, and Singleton (2014), we employ similar macroeconomic variables, to account for economic growth and expected inflation. We use a 3-month moving average of the Chicago Fed National Activity Index as the growth measure and the Federal Reserve Bank of Cleveland estimates of the 1-year expected inflation.[16] For interest rates, we use the same set of yields as in the previous section. Our sample period is January 1982 (the beginning of the Cleveland Fed's expected inflation series) to December 2021.

Modeling the dynamics of the macro variables jointly with the shadow principal components allows this specification to be compared directly with those of Joslin, Priebsch, and Singleton (2014) for the Gaussian model. It is also consistent with the argument of Wu and Xia (2016) and Wu and Zhang (2016) that estimates of the shadow rate s_t rather than the constrained policy rate r_t indicate the policy thrust that the authorities desire and indeed try to achieve using unconventional monetary policies (UMPs) at the ZLB. This rate is the sum of the latent factors in our framework, but we use all three shadow factors separately in the FAVAR, on the argument of Ang et al. (2011) that monetary policy is implemented using the whole yield curve. They were concerned with the pre-ZLB period, but their argument holds a fortiori at the ZLB. Once the lower bound is reached, UMPs were used to lower longer-term yields through forward interest rate guidance, reflected in the expectations component of the term structure, and open market operations, reflected in the risk premium (Gagnon et al., 2011).

[16] The data are available on the website of the Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org.

A term structure model is able to use the physical factor dynamics in Equation (26) to decompose a forward rate (or the value implied by the risk-neutral dynamics) into an interest rate expectation and a residual, which is a measure of the risk premium, throwing light on the way that UMPs affect the yield curve. However, the existing literature uses the standard Gaussian model to make this decomposition, without respecting the ZLB. Our shadow rate model can allow for the effect of this constraint on the decomposition, as we show in Section 4.2.

4.1 The 2^{12} and 2^{18} parameter restriction search

In contrast to the risk-neutral dynamic system in Equation (3), which, with $\underline{r} = 0$, has four parameters (three roots and the level parameter $\mu_\infty^{\mathcal{Q}}$) in addition to those of $\boldsymbol{\Sigma}$, the system of the physical dynamics for the yield-only model in Equation (2) has 12 time-series parameters, and 30 parameters with macro variables in Equation (26). Moreover, as Cochrane and Piazzesi (2009) note, the risk-neutral dynamics are estimated with much greater precision than the physical dynamics, since the cross-sectional errors are "tiny" compared to the forecasting errors in Equation (26). Indeed, we find that most of the parameters in the latter are insignificant. Bauer (2018) argues that this problem of weak identification can be resolved by writing the parameters of the $\mathcal{P}$-dynamics in terms of those of the $\mathcal{Q}$-dynamics using:

$$\boldsymbol{\mu}_q^{\mathcal{P}} = \boldsymbol{\mu}_q^{\mathcal{Q}} + \mathbf{l}_0, \quad \boldsymbol{\Phi}_{qq}^{\mathcal{P}} = \boldsymbol{\Phi}_{qq}^{\mathcal{Q}} + \mathbf{L}_1, \tag{28}$$

and testing zero restrictions on the risk-adjustment parameters $\mathbf{l}_0$ and $\mathbf{L}_1$, as well as $\boldsymbol{\Phi}_{qm}^{\mathcal{P}}$, for which the counterpart under the risk-neutral measure is a matrix of zeros, $\boldsymbol{\Phi}_{qm}^{\mathcal{Q}} = 0_{3,2}$, as shown in Equation (27).

Exploiting the overwhelming speed advantage of the factor extraction method, we follow Joslin, Priebsch, and Singleton (2014) and search for the best combination among all possible zero restrictions on the risk-adjustment parameters. In order to distinguish between the effects of imposing parameter restrictions and adding macro variables, we estimate the restricted versions of both the yield-only and macrofinance models. Thus, in total we estimate 2^{12} and 2^{18} different combinations of restrictions, respectively. We conduct the specification search with the factor extraction method using parallel computing.

Eliminating insignificant parameters using the BIC leads to the selection of a specification with seven restrictions for the yield-only model and 11 restrictions for the macrofinance model.[17] We label our optimal yield-only model and macrofinance model FE^{opt} and MFE^{opt}, respectively. The estimates of the risk-adjustment parameters for the MFE^{opt} are reported in Table 4 (the results for the yields-only models are available upon request). For the yield-only model the price-of-risk restrictions related to the shadow PCs are exactly the same as for the

[17] As is well known, the AIC generally leads to a less parsimonious model; in our case it suggests eight risk-adjustment restrictions in the macrofinance model. Since the AIC, as opposed to BIC, is not a consistent model selection criterion, we rely on the more parsimonious model suggested by the BIC.

Table 4
Estimates of risk adjustment parameters

	$\mathbf{l}_0$	$\mathbf{L}_1$				
	const	PC_1^s	PC_2^s	PC_3^s	GRO	INF
PC_1^s	0	0	0	0.3514	0.0020	−0.0014
	–	–	–	0.2812	0.0005	0.0005
PC_2^s	−0.0013	0.0117	0	0	0	0
	0.0009	0.0051	–	–	–	–
PC_3^s	0.0006	−0.0047	0	0	0	0
	0.0002	0.0013	–	–	–	–

Maximum likelihood estimates of the risk adjustments ($\mathbf{l}_0$ and $\mathbf{L}_1$ in Equation (28)) for our preferred model with unspanned macro risk, MFE^{opt}. Standard errors are reported in small font. The sample period is January 1982 to December 2021.

Table 5
Estimates of the $\mathcal{P}$-dynamics

	$\boldsymbol{\mu}^{\mathcal{P}}$	$\boldsymbol{\Phi}^{\mathcal{P}}$				
	const	PC_1^s	PC_2^s	PC_3^s	GRO	INF
PC_1^s	0.0003	1.0085	0.0700	0.0603	0.0020	−0.0014
	0.0019	0.0054	0.0656	0.0838	0.0005	0.0005
PC_2^s	−0.0015	0.0027	0.9755	0.2799	0	0
	0.0013	0.0018	0.0250	0.1379	–	–
PC_3^s	0.0008	−0.0008	−0.0009	0.8731	0	0
	0.0003	0.0009	0.0036	0.0302	–	–
GRO	0.0508	0.9627	1.6785	−6.0865	0.7298	−0.0596
	0.0894	0.7616	1.7000	6.8097	0.0322	0.0548
INF	0.4935	5.2689	2.0755	−1.6099	0.0639	0.5885
	0.0575	0.4986	1.0463	4.6012	0.210	0.0358

Maximum likelihood estimates of the $\mathcal{P}$-dynamics for our preferred model with unspanned macro risk, MFE^{opt}. Standard errors are reported in smaller font. The sample period is January 1982 to December 2021.

macrofinance model (results available upon request). The price of level risk, but not slope or curvature risk, is influenced by the third factor and the macro variables. However, the first shadow principal component does influence the price of slope risk (often interpreted as reflecting the stance of conventional monetary policy) and curvature risk. The positive sign of the GRO coefficient in the first row implies that price of level risk is cyclical.

Table 5 reports the estimates of the parameters of the $\mathcal{P}$-dynamics for MFE^{opt}. By construction, the coefficients in the top right 3×2 block, showing the effect of the macro variables on the principal components, are the same as the risk adjustments shown in Table 4. On the other hand, the first and second principal components are statistically significant in explaining expected inflation, while none of the principal components has a significant effect on the growth factor. In the following section we present a detailed analysis of the yield decomposition.

4.2 The term premium

Given any estimates of the parameters of the $\mathcal{P}$-dynamics, we can decompose the 10-year yield into components that represent market expectations and the risk

premium. The expectations component is the so-called risk-neutral yield that would obtain in a risk-neutral world governed by the physical rather than the risk-neutral dynamics. Specifically, following Bauer, Rudebusch, and Wu (2012) we first use the VAR parameters to compute the expected short rate over the bond maturity and hence, by taking the average, obtain the risk-neutral yields using Equation (1).

The risk premium for any model and maturity follows by subtracting the risk-neutral yield from the fitted yield. The 10-year risk-neutral yields for FE, FE^{opt}, and MFE^{opt} models are shown alongside the 10-year yield in Figure 1A, and the 10-year yield term premium is shown in Figure 1B. The difference between the restricted and unrestricted model specifications become more pronounced after 1995 as interest rates begin to fall toward the lower bound. For this reason, the figure focuses on the period from January 2008 to December 2021. As noted in the introduction, FE suggests that the 10-year risk-neutral yields were relatively smooth over this period and that the variations in observed yields were largely due to variations in the risk premium. However, when the insignificant parameters in the time-series dynamics are eliminated, the two restricted models suggest that much of the variation in observed yields was due to variations in the risk premium. Table 6, left panel, shows how these models decompose the 10-year yield. According to the unrestricted yield-only model, the term premium constitutes about 90% of the total variation of the 10-year yield. Imposing zero restrictions on the insignificant price-of-risk parameters reduces the term premium contribution in the yield variance to 45% and adding the macro variables to the constrained model decreases it further to 34%, although we should note a substantial correlation between the risk-neutral yield and the term premium in the restricted models.

Figure 1 makes it clear that most of the difference between the standard and macro models is due to the elimination of these insignificant parameters rather than the introduction of macro variables. Nevertheless, ignoring the macro variables can lead to mismeasurement of the risk premium by as much as 50 bp. Also, importantly, the model with macro variables substantially increases business cycle variation in the term premium, making it more countercyclical. In the right panel of Table 6 we report the correlation of the 10-year yield with real economic growth and expected inflation for the period from January 2009 to December 2021. The correlation of the term premium with real growth for FE^{opt} is –3.63%, while that for the MFE^{opt} is -8.34%. Similarly, the correlation with expected inflation for FE^{opt} is 13.30%, and it increases to 20.05% for MFE^{opt} . These findings support the argument of Joslin, Priebsch, and Singleton (2014) that it is important to eliminate insignificant time-series parameters and to allow for the effect of macroeconomic variables on risk premia in term structure models.

4.3 What does the macrofinance model tell us about QE?

In this section, we use MFE^{opt} to analyze the effects of the various episodes of quantitative easing (QE) that were used by the Federal Reserve to help bring down

Table 6
Variance decomposition and term premium variability

	Variance decomposition			Business cycle correlation	
	$\frac{Var[y_{10}^{Q}]}{Var[y_{10}]}$	$\frac{Var[TP_{10}]}{Var[y_{10}]}$	$\frac{2\times Cov[y_{10}^{Q},TP_{10}]}{Var[y_{10}]}$	$Corr(TP_{10},GRO)$	$Corr(TP_{10},INF)$
FE	0.1870	0.9039	–0.0909	–0.0465	0.0438
FE^{opt}	0.1279	0.4513	0.4208	–0.0363	0.1330
MFE^{opt}	0.1921	0.3418	0.4661	–0.0834	0.2005

Variance decomposition of the 10-year yield and correlation of the the 10-year term premium with macro variables for three different models over the period from January 2009 to December 2021.

longer-term yields during the financial crisis. We analyze the effect of these policies on two types of indicators, term premia and cumulative density functions (CDFs), used in the Wu and Xia (2016) approximation.

Term premia are informative because, as Figure 1B shows, they account for most of the variation in long-term yields. Moreover, Gagnon et al. (2011) maintain that changes in asset supplies should affect term premia rather than market interest rate expectations, and so, as they argue, a sharp fall in the term premium upon the announcement of an asset purchase program indicates that this had the effect of reducing yields via the scarcity effect. To measure the term premium, however, Gagnon et al. (2011) use a GTSM (Kim and Wright, 2005). Term premia are also important because they tend to increase when policy interest rates fall, blunting the effect on long yields (the so-called conundrum effect; see Hanson, Lucca, and Wright 2018). The Fed's QE policies can be seen as a way of mitigating this effect and allowing long rates to fall too. On the other hand, it could be argued that the risk-neutral yields, which reflect market expectations, show the effect of the Fed's forward guidance on interest rates.

The CDFs from our shadow rate model provide another type of monetary policy indicator. Recall that one minus the CDF is the weight that the yield model gives to the effect of the lower bound. Thus, a sharp fall in the CDF upon a policy announcement indicates that the policy has pushed the yield closer to zero, because either the long-term rate expectation or the term premium have fallen, or perhaps both. Figure 7 shows the periods covered by the various policy episodes alongside the term premium and the CDF for the 2-year (top panel) and 10-year (bottom panel) yields, both from our optimal MFE^{opt} model.

Figure 7 shows an increase in the term premium as policy rates were cut back following the Lehman bankruptcy in September 2008, which is largely reversed following the announcement of the Fed's first Long Term Asset Purchase program (QE1) announced in November 2008. This is visible particularly for the indicators based on the 2-year yield (top panel). The fall in the 2-year term premium from 0.54% in October to virtually zero in November is accompanied by a a fall in the CDF from 0.9 to 0.7, pushing the 2-year rate closer to zero. The announcement of the expansion of the QE program in March 2009 also coincided with dips in the term premium and CDFs. The 10-year term premium fell from 2.28% in February

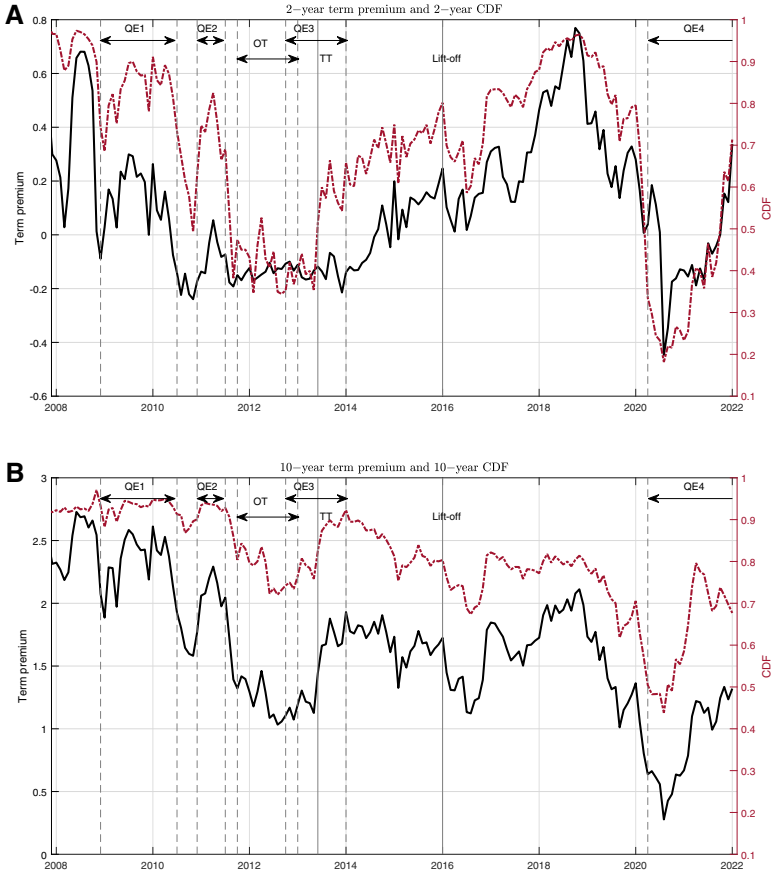

Figure 7
Quantitative easing and the macrofinance model
The continuous graph shows the term premium (as in Figure 1B) and the dashed graph the CDF from MFE^{opt} for the 2-year (panel A) and 10-year (panel B) yields. The dashed vertical lines show the announcement and end dates for each episode of quantitative easing (QE) as well as Operation Twist (OT) on an end-month basis. The continuous vertical lines indicate the announcement of major contractionary policy changes. TT marks the beginning of the Taper Tantrum following Ben Bernanke's testimony to Congress in May 2013. Lift-off marks the first increase in the fed funds rate in December 2015.

2009 to 1.97% in March, while the 10 year CDF fell from 0.82 to 0.75. The 2-year term premium became negative at the end of the first phase of QE, in May 2010, and stayed below zero until February 2015. The indicators for the 10-year yield were moving slower, reaching a minimum at the beginning of 2013. On May 22, 2013 Ben Bernanke's testimony to Congress triggered the so-called Taper Tantrum, upon which the 2-year CDF moved up from 0.35 in April to 0.52 in May and 0.60 in June, while the 10-year term premium moved up from 1.13% to 1.44% to 1.66%. This speech marked the beginning of a sustained increased in the

CDF. The fed funds rate was finally increased in December 2015. Our results, which extend those of Gagnon et al. (2011), suggest that new policy initiatives and reversals of policy do appear to have immediate effects upon announcement, but extensions of existing programs seem to have little immediate effect.

It is clear from the charts that the onset of the Covid pandemic pushed the risk premia and CDFs to historic lows right along the curve, with a sharp reversal occurring as the economy recovers in 2021.

5. Conclusion

This paper shows how the factor extraction and likelihood concentration techniques developed for the Gaussian term structure model can be applied to non-affine pricing models. This allows the researcher to estimate these models almost as easily and quickly as the simple Gaussian model, while avoiding the various approximations used by previous approaches, arguably providing a more reliable assessment of the thrust of monetary policy.

We illustrate these advantages by estimating two models that eliminate insignificant time-series parameters. This involves searching over large numbers of different parameter combinations. The traditional filtering approach makes such exercises infeasible. Exploiting the speed and consistency of our factor extraction algorithm makes it feasible to conduct such exercises.

The first of these models is a standard yield-only model, and the second introduces two unspanned macro variables. This macrofinance model views the behavior of the yield curve since the financial crisis in a very different light to the standard model: its term premium explains less of the total variation in yields and is more countercyclical. The macrofinance model with optimal price-of-risk restrictions suggests that after the financial crisis of 2008–2009 the risk premium fluctuated around 2% per annum until about 2019, while the standard model implies a steady decline of the term premium to below 1%.

Other applications that involve the estimation of large numbers of alternatives include policy simulations such as those of Bauer and Rudebusch (2014) and international bond market models such as those of Egorov, Li, and Ng (2011) and Bauer and de los Rios (2012), which potentially involve many different combinations of common and idiosyncratic factors. Central bank and other practitioners, who need models of the yield curve fitted to daily data for analysis and pricing and still rely on the standard Gaussian term structure model to provide these, can now use these methods to deal effectively with the lower bound.

Appendix. The Likelihood Function

We maximize the joint log-likelihood conditional in Equation (16). As explained in Section 2.2, when using the nonlinear factor extraction technique, with the exception of the $\mathbf{\Sigma}$ matrix, the $\ell^{\mathcal{Q}}$ and $\ell^{\mathcal{P}}$ parts of the log-likelihood will have disjoint sets of parameters. In the following, to keep the notation simple, we suppress the conditioning set of parameters in the likelihood function.

The part of the log-likelihood function with the risk-neutral parameters $\ell^{\mathcal{Q}}$ has the standard form:

$$\ell^{\mathcal{Q}}(\mathbf{y}_t^o|\mathbf{q}_t) = -\frac{J-K}{2}\Big(1+\log(2\pi)\Big) - \frac{J-K}{2}\log\sigma_v^2, \tag{A-1}$$

where σ_v^2 is the (homoscedastic) variance of measurement errors $v_{n,t}$ in Equation (10) calculated as:

$$\sigma_v^2 = \frac{1}{(T-1)\times(J-K)}\sum_{t=2}^{T}\sum_{j=1}^{J}\widehat{v}_{t,j}. \tag{A-2}$$

The dynamics of $\mathbf{q}_t$ are generally nonlinear, and thus $\mathbf{q}_t$ needs to be rotated to the latent state vector $\mathbf{x}_t$ or a linear transformation of it. Hence, $\ell^{\mathcal{P}}$, the part of the log-likelihood that describes the time series of the latent factor dynamics, needs to be adjusted accordingly. Recall the relation between $\mathbf{q}_t$ and $\mathbf{x}_t$ from (26). The yields-only model examined in Section 3 we estimate with respect to the $\mathcal{P}$-dynamics for the latent state vector as in the Joslin, Singleton, and Zhu (2011) parameterization. To write down the $\mathcal{P}$-likelihood in terms of the probability density function of $\mathbf{x}_t$, we apply the change-of-variable technique:[18]

$$\ell^{\mathcal{P}}(\mathbf{q}_t|\mathbf{q}_{t-1}) = \ell^{\mathcal{P}}(\mathbf{x}_t|\mathbf{x}_{t-1}) - \log|\det(\mathbf{J}_t)|, \tag{A-3}$$

where $\mathbf{J}_t$ is the Jacobian term resulting from the change of variables:

$$\mathbf{J}_t \equiv \left[\frac{\partial\mathbf{q}_t}{\partial x_{1,t}},\ldots,\frac{\partial\mathbf{q}_t}{\partial x_{K,t}}\right]. \tag{A-4}$$

The logarithm of the $\mathcal{P}$-likelihood is then:

$$\begin{aligned}\ell^{\mathcal{P}}(\mathbf{q}_t|\mathbf{q}_{t-1}) &= \ell^{\mathcal{P}}(\mathbf{x}_t|\mathbf{x}_{t-1}) - \log|\det(\mathbf{J}_t)| \\ &= -\frac{K}{2}\log(2\pi) - \frac{1}{2}\log\left(\det(\boldsymbol{\Sigma}\boldsymbol{\Sigma}')\right) \qquad \text{(A-5)} \\ &\quad -\frac{1}{2}\left(\mathbf{x}_t - \boldsymbol{\mu}^{\mathcal{P}} - \boldsymbol{\Phi}^{\mathcal{P}}\mathbf{x}_{t-1}\right)'\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\right)^{-1}\left(\mathbf{x}_t - \boldsymbol{\mu}^{\mathcal{P}} - \boldsymbol{\Phi}^{\mathcal{P}}\mathbf{x}_{t-1}\right) \\ &\quad -\log|\det(\mathbf{J}_t)|. \qquad \text{(A-6)}\end{aligned}$$

In our application of macrofinance model in Section 4, where we apply further a linear rotation of the state vector $\mathbf{x}_t$ to the shadow principal components $\mathbf{q}_t^s$ as in Equation (18). Applying the change-of-variable technique once more, we can write the log-likelihood in terms of $\mathbf{q}_t^s$:

$$\ell^{\mathcal{P}}(\mathbf{q}_t|\mathbf{q}_{t-1}) = \ell^{\mathcal{P}}(\mathbf{q}_t^s|\mathbf{q}_{t-1}^s) - \log|\det(\mathbf{J}_t)| + \log|\det(\mathbf{J})|, \tag{A-7}$$

where

$$\mathbf{J} = \left[\frac{\partial\mathbf{q}_t^s}{\partial x_{1,t}},\ldots,\frac{\partial\mathbf{q}_t^s}{\partial x_{K,t}}\right] = \mathbf{W}'\mathbf{B}. \tag{A-8}$$

Adding macro variables does not add any further complication. In particular, define $\mathbf{z}_t = \left[\mathbf{q}_t^{s\prime}, \mathbf{m}_t'\right]'$. Then, the part of the log-likelihood that describes the time series of the latent factor dynamics is:

$$\ell^{\mathcal{P}}(\mathbf{q}_t, \mathbf{m}_t|\mathbf{q}_{t-1}, \mathbf{m}_{t-1}) = \ell^{\mathcal{P}}(\mathbf{z}_t|\mathbf{z}_{t-1}) - \log|\det(\mathbf{J}_t)| + \log|\det(\mathbf{J})|, \tag{A-9}$$

where the dynamic system for $\mathbf{z}_t$ is given in Equation (26).

Declarations of interest: none.

[18] See Greene (2011), Appendix B.

References

Ahn, D., R. Dittmar and R. Gallant. 2002. Quadratic term structure models: Theory and evidence. *Review of Financial Studies* 15:243–88.

Andreasen, M. M., and B. J. Christensen. 2015. The SR approach: A new estimation procedure for non-linear and non-Gaussian dynamic term structure models. *Journal of Econometrics* 184:420–51.

Andreasen, M. M., and A. Meldrum. 2019. A shadow rate or a quadratic policy rule? The best way to enforce the zero lower bound in the United States. *Journal of Financial and Quantitative Analysis* 54:2261—92.

Ang, A., J. Boivin, S. Dong, and R. Loo-Kung. 2011. Monetary policy shifts and the term structure. *Review of Economic Studies* 78:429–57.

Ang, A., and M. Piazzesi. 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics* 43:75–87.

Bauer, G., and A. D. de los Rios. 2012. An international dynamic term structure model with economic restrictions and unspanned risks. Staff Working Papers 12-5, Bank of Canada.

Bauer, M. D. 2018. Restrictions on risk prices in dynamic term structure models. *Journal of Business & Economic Statistics* 36:196–211.

Bauer, M. D., and G. D. Rudebusch. 2014. The signaling channel for Federal Reserve bond purchases. *International Journal of Central Banking* 10:233–89.

———. 2016. Monetary policy expectations at the zero lower bound. *Journal of Money, Credit and Banking* 48:1439–65.

Bauer, M. D., G. D. Rudebusch, and J. C. Wu. 2012. Correcting estimation bias in dynamic term structure models. *Journal of Business & Economic Statistics* 30:454–67.

Beaglehole, D., and M. Tenney. 1992. Corrections and additions to "a nonlinear equilibrium model of the term structure of interest rates." *Journal of Financial Economics* 32:345–53.

———. 1991. General solutions of some interest rate-contingent claim pricing equations. *Journal of Fixed Income* 1:69–83.

Black, F., 1995. Interest rates as options. *Journal of Finance* 50:1371–76.

Christensen, J. H. E., and G. D. Rudebusch. 2015. Estimating shadow-rate term structure models with near-zero yields. *Journal of Financial Econometrics* 13:226–59.

Cochrane, J. H., and M. Piazzesi. 2009. Decomposing the yield curve. 2009 Meeting Papers 18, Society for Economic Dynamics.

Constantinides, G. M. 1992. A theory of the nominal term structure of interest rates. *Review of Financial Studies* 5:531–52.

Coroneo, L., and S. Pastorello. 2020. European spreads at the interest rate lower bound. *Journal of Economic Dynamics and Control* 119:1–21.

Cox, J., J. Ingersoll, and S. Ross. 1985. A theory of the term structure of interest rates. *Econometrica* 52:385–407.

Dai, Q., and K. Singleton. 2000. Specification analysis of affine term structure models. *Journal of Finance* 55:415–41.

Duffee, G. R. 2011. Information in (and not in) the term structure. *Review of Financial Studies* 24:2895–2934.

Duffie, D., and R. Kan. 1996. A yield-factor model of interest rates. *Mathematical Finance* 6:379–406.

Durbin, J., and S. J. Koopman. 2001. *Time series analysis by state space methods*. Oxford: Oxford University Press.

Egorov, A. V., H. Li, and D. Ng, 2011. A tale of two yield curves: Modeling the joint term structure of dollar and euro interest rates. *Journal of Econometrics* 162:55–70.

Gagnon, J., M. Raskin, J. Remache, and B. Sack. 2011. The financial market effects of the Federal Reserve's large-scale asset purchases. *International Journal of Central Banking* 7:3–43.

Gourieroux, C., A. Monfort, and V. Polimenis. 2002. Affine term structure models. Working Papers 2002-49, Center for Research in Economics and Statistics.

Greene, W. H. 2011. *Econometric analysis*, 7th ed. Upper Saddle River, NJ: Pearson.

Gurkaynak, R. S., B. Sack, and J. H. Wright. 2007. The U.S. Treasury yield curve: 1961 to the present. *Journal of Monetary Economics* 54(8):2291–2304.

Hanson, S., D. O. Lucca, and J. H. Wright. 2018. Interest rate conundrums in the twenty-first century. Staff Reports 810, Federal Reserve Bank of New York.

Joslin, S., A. Le, and K. J. Singleton. 2013. Why Gaussian macro-finance term structure models are (nearly) unconstrained factor-VARs. *Journal of Financial Economics* 109:604–22.

Joslin, S., M. Priebsch, and K. J. Singleton. 2014. Risk premiums in dynamic term structure models with unspanned macro risks. *Journal of Finance* 69:1197–1233.

Joslin, S., K. J. Singleton, and H. Zhu. 2011. A new perspective on Gaussian dynamic term structure models. *Review of Financial Studies* 24:926–70.

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

Kim, D. H., and J. H. Wright. 2005. An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates. Finance and Economics Discussion Series 2005-33, Board of Governors of the Federal Reserve System (U.S.).

Krippner, L. 2012. Modifying gaussian term structure models when interest rates are near the zero lower bound. Reserve Bank of New Zealand Discussion Paper Series DP2012/02, Reserve Bank of New Zealand.

——— 2013. A tractable framework for zero-lower-bound Gaussian term structure models. CAMA Working Papers, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, Australian National University.

Le, A., K. J. Singleton, and Q. Dai. 2010. Discrete-time affine$^{\mathbb{Q}}$ term structure models with generalized market prices of risk. *Review of Financial Studies* 23:2184–2227.

Leippold, M., and L. Wu. 2002. Asset pricing under the quadratic class. *Journal of Financial and Quantitative Analysis* 37:271–95.

Lemke, W., and A. L. Vladu. 2017. Below the zero lower bound: A shadow-rate term structure model for the euro area. Working Paper Series 1991, European Central Bank.

Longstaff, F. A. 1989. A nonlinear general equilibrium model of the term structure of interest rates. *Journal of Financial Economics* 23:195–224.

Monfort, A., F. Pegoraro, J.-P. Renne, and G. Roussellet. 2017. Staying at zero with affine processes: An application to term structure modelling. *Journal of Econometrics* 201:348–66.

Priebsch, M. A. 2013. Computing arbitrage-free yields in multi-factor Gaussian shadow-rate term structure models. Finance and Economics Discussion Series 2013-63, Board of Governors of the Federal Reserve System (U.S.).

Realdon, M. 2006. Quadratic term structure models in discrete time. *Finance Research Letters* 3:277–89.

Vasicek, O. 1977. An equilibrium characterisation of the term structure. *Journal of Financial Economics* 5:177–88.

Wu, J. C., and F. D. Xia. 2016. Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking* 48:253–91.

Wu, J. C., and J. Zhang. 2016. A shadow rate new Keynesian model. NBER Working Papers 22856.

Zellner, A. 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association* 57:348–68.