A new perspective on Gaussian shadow rate term structure models*

Adam Goliński Peter Spencer

December 29, 2018

Abstract

We propose a new approach to estimating shadow rate term structure models. This modifies the Joslin, Singleton, and Zhu (2011) factor rotation technique to allow for the zero lower bound using the Wu and Xia (2016) discrete-time closed-form approximation of the Black (1995) model. Compared with the standard approach based on the extended Kalman filter, our approach significantly improves convergence and greatly reduces the computation time. It has the added advantage of producing more robust estimates of the lower bound parameter and the path of the shadow rate. We apply the shadow rate model to recent U.S. data and show that expected inflation and unemployment gap are important unspanned macro factors that drive term premiums, as in Joslin, Priebsch, and Singleton (2014).

JEL classification: G12, C13.

Keywords: term structure, shadow rate.

*Goliński: University of York, Department of Economics and Related Studies Heslington, York, Y010 5DD, United Kingdom; E-mail: adam.golinski@york.ac.uk.

Spencer: University of York, Department of Economics and Related Studies, Heslington, York, Y010 5DD, United Kingdom; E-mail: peter.spencer@york.ac.uk.

1 Introduction

The Gaussian term structure model ($GTSM$) is routinely used to analyze the behavior of a wide range of financial markets, notably those for government and corporate bonds. Its popularity has been enhanced in recent years by a series of innovations that greatly reduce and, as Adrian, Crump, and Moench (2013) argue, ultimately remove the need for non-linear optimization methods. These innovations concentrate the likelihood function, helping to deal with multiple local optima and other difficult problems as well as greatly speeding up the estimation procedure.

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

The Gaussian version of the Black (1995) shadow rate term structure model, which represents the spot interest rate using the truncated normal distribution, has a great deal of appeal in this situation. Since this is viable only for models with one or two factors, rather than the three or more factors that are needed in practice, a popular compromise has been to work with a tractable pricing function that gives reasonable approximation to the 'true' Black prices (Krippner, 2013, Priebsch, 2013, Wu and Xia, 2016). However, the non-linearities in these models mean that researchers have resorted to the use of the extended Kalman filter (EKF) and other unsatisfactory methods that have been superseded by the new likelihood-concentration methods in the Gaussian setting.

This paper shows that these new methods can be applied to the shadow rate pricing model, allowing the researcher to estimate the shadow rate model almost as easily and quickly as the simple $GTSM$. Specifically, we show that the factor extraction approach of Joslin et al. (2011), henceforth JSZ, can be readily applied in this context with all of its well-known computational advantages. It has the added advantage of producing more robust estimates of the lower bound parameter and the path of the shadow rate than the EKF. In

the case of a well-known U.S. Treasury bond yield dataset, we typically achieve convergence in less than 1.5 minutes, compared with over 40 minutes or longer when using the EKF. Moreover, this apart, the fit and parameter estimates are economically indistinguishable from those obtained with the EKF. This technique opens the way to research with the shadow rate model that involves very large numbers of different model estimations, which we illustrate using a 2^{19} parameter combination search for an optimal macro-finance variant.

The paper is set out as follows. The next section describes the Gaussian shadow rate framework and shows how this rate is mapped into the policy rate using the approximation of Wu and Xia (2016). Section 3 shows how this is linearized by the EKF used in the existing shadow rate literature and how similar linearization methods allow the factor extraction methods of JSZ to be used in this context, outlining the many advantages of this methodology. Section 4.1 reports the results of using these various approaches to fit a three-factor shadow rate model to a well-known dataset for the US Treasury bond market. The macro-finance extension of this model, which introduces real activity and inflation variables as unspanned factor along the lines of JSZ is reported in section 4.2. The final section concludes and indicates other research problems that can now be tackled using this approach to estimating shadow rate models.

2 The pricing models

2.1 Black Gaussian shadow rate model

Assume that the observed short rate is specified as in Black (1995):

$$r_t \equiv \max(s_t, \underline{r}), \tag{1}$$

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

$$s_t = \delta_0 + \boldsymbol{\delta}_1' \mathbf{x}_t, \tag{2}$$

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

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

$$\mathbf{x}_t = \boldsymbol{\mu}^{\mathcal{Q}} + \boldsymbol{\Phi}^{\mathcal{Q}}\mathbf{x}_{t-1} + \mathbf{u}_t^{\mathcal{Q}}, \tag{4}$$

respectively, with $\mathbf{u}_t^{\mathcal{P}}, \mathbf{u}_t^{\mathcal{Q}} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ under their respective measure.

We adopt the parametrization scheme proposed by JSZ, i.e. $\boldsymbol{\Phi}^{\mathcal{Q}}$ is determined by K roots, which we can collect in a vector $\boldsymbol{\lambda}^{\mathcal{Q}}$, $\boldsymbol{\mu}^{\mathcal{Q}} = [\mu_\infty^{\mathcal{Q}}, 0, \ldots]'$, $\delta_0 = 0$ and $\boldsymbol{\delta}_1$ is a vector of ones. The physical dynamics, $\boldsymbol{\mu}^{\mathcal{P}}$, $\boldsymbol{\Phi}^{\mathcal{P}}$ and $\boldsymbol{\Sigma}$, are unrestricted. This parametrization plays a crucial role in our estimation scheme, as in the original JSZ Gaussian model, since it allows one to estimate the $\mathcal{P}-$parameters by OLS regression.

2.2 The shadow rate pricing approximation

If the short rate is well above the lower bound, it is reasonable to use the standard Gaussian model, which assumes: $r_t = s_t$. Let $P_{n,t}$ is the price of the $n-$period zero coupon bond at time t. The $1-$ period log forward rate at time t is defined as $f_{n,t} = \log(P_{n,t}/P_{n+1,t})$. The $n-$period zero coupon yield is

$$y_{n,t} = -\frac{1}{n}\log(P_{n,t}) = \frac{1}{n}\sum_{j=0}^{n-1} f_{j,t} \tag{5}$$

In the Gaussian framework, the log forward rates can be represented as an affine function $f_{n,t} = f_{n,t}^G$ of the factors:

$$f_{n,t}^G = a_{f,n} + \mathbf{b}_{f,n}'\mathbf{x}_t, \quad n = 0, 1, \ldots. \tag{6}$$

The coefficients $a_{f,n}$ and $\mathbf{b}_{f,n}$ follow well-known recursions, derived for example in Cochrane and Piazzesi (2009) and Wu and Xia (2016) :

$$\begin{aligned} \mathbf{b}'_{f,n} &= \boldsymbol{\delta}'_1 \left(\boldsymbol{\Phi}^{\mathcal{Q}}\right)^n, & (7) \\ a_{f,n} &= \delta_0 + \boldsymbol{\delta}'_1 \left(\sum_{j=0}^{n-1} \left(\boldsymbol{\Phi}^{\mathcal{Q}}\right)^j\right) \boldsymbol{\mu}^{\mathcal{Q}} - \frac{1}{2}\boldsymbol{\delta}'_1 \left(\sum_{j=0}^{n-1} \left(\boldsymbol{\Phi}^{\mathcal{Q}}\right)^j\right) \boldsymbol{\Sigma}\boldsymbol{\Sigma}' \left(\sum_{j=0}^{n-1} \left(\boldsymbol{\Phi}^{\mathcal{Q}}\right)^j\right)' \boldsymbol{\delta}_1. & (8) \end{aligned}$$

From (5) we obtain yields in the Gaussian system as $y^G_{n,t} = a_n + \mathbf{b}'_n \mathbf{x}_t$, where $a_n = (1/n)\sum_{j=0}^{n-1} a_{f,j}$ and $\mathbf{b}_n = (1/n)\sum_{j=0}^{n-1} \mathbf{b}_{f,j}$. Stacking the yields we can write this system in vector notation as:

$$\mathbf{y}^G_t = \mathbf{a} + \mathbf{B}\mathbf{x}_t. \tag{9}$$

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 or pricing fixed income derivatives. Shadow rate models deal with this inconvenience by treating the forward rates $f^G_{n,t}$ of the Gaussian model as 'shadow forward rates' and map them into the Black forward rates $f_{n,t}$. Unfortunately, (1) introduces a non-linearity into this system, making it difficult to estimate the model when there are more than 2 factors. Thus, a popular compromise has been to use a convenient approximation. Specifically, Wu and Xia (2016) propose the approximation:[1]

$$\begin{aligned} f_{n,t} &= f(\mathbf{x}_t, n; \boldsymbol{\Psi}) \\ &= \underline{r} + \left(a_{f,n} + \mathbf{b}'_{f,n}\mathbf{x}_t - \underline{r}\right) \Phi\left(\frac{a_{f,n} + \mathbf{b}'_{f,n}\mathbf{x}_t - \underline{r}}{\sigma^{\mathcal{Q}}_n}\right) + \sigma^{\mathcal{Q}}_n \phi\left(\frac{a_{f,n} + \mathbf{b}'_{f,n}\mathbf{x}_t - \underline{r}}{\sigma^{\mathcal{Q}}_n}\right), & (10) \end{aligned}$$

where $\Phi(.)$ and $\phi(.)$ are the cumulative distribution and probability density function of a standard normal distribution, respectively, $\sigma^Q_n = \sum_{i=0}^{n-1} \mathbf{b}'_{f,i}\boldsymbol{\Sigma}\mathbf{b}_{f,i}$, and $\boldsymbol{\Psi}$ is a vector of relevant parameters. The derivative of this function with respect to the latent state variables

[1] See Krippner (2013) for the counterpart in continuous time.

is available in a convenient closed form:

$$f'(\mathbf{x}_t, n; \Psi) = \frac{df(\mathbf{x}_t, n; \Psi)}{d\mathbf{x}_t} = \Phi\left(\frac{a_{f,n} + \mathbf{b}'_{f,n}\mathbf{x}_t - \underline{r}}{\sigma_n^{\mathcal{Q}}}\right)\mathbf{b}_{f,n}. \tag{11}$$

Similarly, we can write the $n-$period zero coupon bond yield in terms of the forwards by substituting (10) into (5) to get:

$$y_{n,t} = y(\mathbf{x}_t, n; \Psi) = \frac{1}{n}\sum_{j=0}^{n-1} f(\mathbf{x}_t, j; \Psi). \tag{12}$$

This approximation has been adopted by many researchers (e.g. Coroneo and Pastorello, 2017, Lemke and Vladu, 2017). Another method based on a second-order approximation was proposed by Priebsch (2013). However, to handle the the non-linearities in these models researchers have used Kalman-filter-based techniques. These have been superseded by likelihood-concentration methods in estimating the $GTSM$, which we now show can be used to estimate shadow rate models.

3 Factor identification and estimation strategies

This section sets out various schemes that mimic the JSZ factor identification and likelihood-concentration methods currently used in estimating the $GTSM$. The first employs a non-linear solution technique and the second employs variants of the linearization scheme used in the EKF.

3.1 Fitting errors and factor extraction

If the yield model (12) 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 mis-specification effects, we augment this specification

with an additive error $u_{n,t}$ to get an empirical model of the observed yield $y^o_{n,t}$:

$$y^o_{n,t} = y(\mathbf{x}_t, n; \Psi) + u_{n,t}, \tag{13}$$

where $u_{n,t} \sim N(0, \sigma^2_{u,n})$. Stacking (12) gives a system of J nonlinear yield equations:

$$\mathbf{y}^o_t = \mathbf{y}(\mathbf{x}_t; \mathbf{\Psi}) + \mathbf{u}_t. \tag{14}$$

Now, to mimic JSZ, we assume that there are $K < J - 1$ combinations (or 'portfolios') of yields, given by a $J \times K$ weighting matrix $\mathbf{W}$, that are nevertheless fitted without error:

$$\mathbf{q}_t \equiv \mathbf{W}'\mathbf{y}^o_t = \mathbf{W}'\mathbf{y}_t \tag{15}$$

for all t. Substituting (14) into (15) shows that this is equivalent to assuming: $\mathbf{W}'\mathbf{u}_t = \mathbf{0}$, which we will refer to as the observability restriction. As in JSZ, we assume that the observable factors $\mathbf{q}_t$ are the first K principal components obtained from the covariance matrix of yields, $\text{Cov}(\mathbf{y}^o_t)$, so that the weights are given by the corresponding eigenvectors. Denote by $\mathbf{q}(\mathbf{x}_t; \mathbf{\Psi})$ the function that maps the latent state vector $\mathbf{x}_t$ to the observable principal components $\mathbf{q}_t$, and the inverse of this function by $\mathbf{q}^{-1}(\mathbf{x}_t; \mathbf{\Psi})$, i.e.:

$$\mathbf{q}_t = \mathbf{q}(\mathbf{x}_t; \mathbf{\Psi}) = \mathbf{W}'\mathbf{y}(\mathbf{x}_t; \mathbf{\Psi}) \Longleftrightarrow \mathbf{x}_t = \mathbf{q}^{-1}(\mathbf{q}_t; \mathbf{\Psi}) \tag{16}$$

Then, substituting $\mathbf{x}_t$ back into (14) gives our nonlinear econometric model of the cross-section of J observed yields:

$$\mathbf{y}^o_t = \mathbf{y}\left(\mathbf{q}^{-1}(\mathbf{q}_t; \mathbf{\Psi}); \mathbf{\Psi}\right) + \mathbf{u}_t. \tag{17}$$

Hence, our first estimation strategy is to use nonlinear solution technique to recover $\mathbf{x}_t$, conditional on the risk-neutral parameters, using (16) and use it to fit the observed yields as in (17). Further details regarding estimation are set out in Section 3.4 below. Note that

when the measurement errors are restricted so that the observability restriction (15) holds, this estimation technique is equivalent to the iterated extended Kalman filter advocated by Krippner (2013).

3.2 Approximating the pricing function

Current estimation strategies, in contrast, are based on linearization techniques. If the pricing function f is continuously differentiable, we can linearize it using a first order Taylor expansion around any point $\widetilde{\mathbf{x}}_{t-1}$ and use it to write the forward rate as:

$$\begin{aligned} f_{n,t} &= f(\mathbf{x}_t, n; \boldsymbol{\Psi}) \\ &\approx f(\widetilde{\mathbf{x}}_{t-1}, n; \boldsymbol{\Psi}) + f'(\widetilde{\mathbf{x}}_{t-1}, n; \boldsymbol{\Psi})(\mathbf{x}_t - \widetilde{\mathbf{x}}_{t-1}) \end{aligned} \tag{18}$$

where the derivative f' is defined in (11). Similarly the yields in (12) can be approximated as:

$$\begin{aligned} y_{n,t} &\approx y(\widetilde{\mathbf{x}}_{t-1}, n; \Psi) + y'(\widetilde{\mathbf{x}}_{t-1}, n; \Psi)(\mathbf{x}_t - \widetilde{\mathbf{x}}_{t-1}) \\ &= \underbrace{y(\widetilde{\mathbf{x}}_{t-1}, n; \Psi) - y'(\widetilde{\mathbf{x}}_{t-1}, n; \Psi)\widetilde{\mathbf{x}}_{t-1}}_{a_{n,t-1}} + \underbrace{y'(\widetilde{\mathbf{x}}_{t-1}, n; \Psi)}_{\mathbf{b}'_{n,t-1}}\mathbf{x}_t. \end{aligned} \tag{19}$$

When we use the Wu and Xia (2016) approximation to the forward rate, the approximation coefficients are:

$$\mathbf{b}'_{n,t-1} = \frac{1}{n}\left[\sum_{j=0}^{n-1} f'(\widetilde{\mathbf{x}}_{t-1}, j; \Psi)\right] = \frac{1}{n}\left[\sum_{j=0}^{n-1} \Phi\left(\frac{a_{f,j} + \mathbf{b}'_{f,j}\widetilde{\mathbf{x}}_{t-1} - \underline{r}}{\sigma_j^{\mathcal{Q}}}\right)\mathbf{b}'_{f,j}\right], \tag{20}$$

$$a_{n,t-1} = \frac{1}{n}\left[\sum_{j=0}^{n-1} (a_{f,j} - \underline{r})\,\Phi\left(\frac{a_{f,j} + \mathbf{b}'_{f,j}\widetilde{\mathbf{x}}_{t-1} - \underline{r}}{\sigma_j^{\mathcal{Q}}}\right) + \sigma_j^{\mathcal{Q}}\phi\left(\frac{a_{f,j} + \mathbf{b}'_{f,j}\widetilde{\mathbf{x}}_{t-1} - \underline{r}}{\sigma_j^{\mathcal{Q}}}\right)\right] \tag{21}$$

We give each yield coefficient a $t-1$ subscript because it is based on the information set available at that time. Our second estimation strategy exploits this lag, as we show in Section 3.4 below. In the limit when $\underline{r}$ takes a large negative value or when the spot rate is

very high, $\phi(.) = 0$ and $\Phi(.) = 1$, so this specializes to the Gaussian yield model.

We consider three candidates for the approximation points. The first approximation is taken with respect to the expectation of the state under the physical measure $\mathcal{P}$: $\widetilde{\mathbf{x}}_{t-1} = E_{t-1}^{\mathcal{P}}\mathbf{x}_t$. We call this the '$p$' expectation approach, which obviously requires knowledge of the parameters $\boldsymbol{\mu}^{\mathcal{P}}$ and $\boldsymbol{\Phi}^{\mathcal{P}}$. This approximation is implicit in the extended Kalman Filter and, as such, has been adopted by shadow rate modelers. The next approach exploits the high persistence embedded in interest rates and approximates the state vector using its realization at previous period, i.e. $\widetilde{\mathbf{x}}_{t-1} = \mathbf{x}_{t-1}$. We will call this the 'naive' approximation and denote it using an 'n'. Finally, we propose an approximation around the risk-neutral expectation of the state vector: $\widetilde{\mathbf{x}}_{t-1} = E_{t-1}^{\mathcal{Q}}\mathbf{x}_t$, which we call the '$q$' expectation approach. Since this only depends on the risk-neutral parameters, and not $\boldsymbol{\mu}^{\mathcal{P}}$ and $\boldsymbol{\Phi}^{\mathcal{P}}$, this (like the '$n$' approach), has the great advantage that it allows us to separate the likelihood function into cross-section and time-series components, as in JSZ.

3.3 Factor rotation

Stacking the yields in (19) gives a $J-$dimensional system:

$$\mathbf{y}_t = \mathbf{a}_{t-1} + \mathbf{B}_{t-1}\mathbf{x}_t. \tag{22}$$

If these yields were fitted without error, then, conditional on the model parameters and information available at time $t-1$, we could in principle invert this relationship to find the (estimate of the) state vector $\mathbf{x}_t$. Technically, the row rank of $\mathbf{B}_{t-1}$ needs to be K for this to be possible, which is a potential problem in a shadow rate framework because in this case $\mathbf{B}_{t-1}$ needs to be inverted in every time period. For example, it is often assumed (e.g. Hamilton and Wu, 2012) that specific rates are fitted without fitting error. The Wu and Xia (2016) model of forward rates reveals that it might be problematic. The time series of the cumulative distribution function $\Phi\left((a_{f,n} + \mathbf{b}'_{f,n}\mathbf{x}_t - \underline{r})/\sigma_n^{\mathcal{Q}}\right)$ that feature in (10), are presented in Figure (4). We can see that when the lower bound is estimated high enough

(about 16 b.p. as in the upper left panel), there were periods when these were zero or near-zero, especially for short maturities. If we were to select these rates as fitted without error, we would not be able to invert the system to recover the latent state $\mathbf{x}_t$ in such periods. In practice, this problem can occur even if the probabilities are positive but small, since rates for different maturities are strongly cross-correlated. This problem is greatly alleviated by our use of principal components, since they are uncorrelated and give a significant weight to the longer maturity rates. They also make an allowance for measurement error in individual rates, because they replace the assumption of noiseless rates with the assumption of noiseless portfolios of rates.

Augmenting (19) with additive errors gives in the matrix notation:

$$y^o_{n,t} = a_{n,t-1} + \mathbf{b}'_{n,t-1}\mathbf{x}_t + v_{n,t}, \tag{23}$$

where $v_{n,t} \sim N(0, \sigma^2_{v,n})$, or in matrix notation:

$$\mathbf{y}^o_t = \mathbf{a}_{t-1} + \mathbf{B}_{t-1}\mathbf{x}_t + \mathbf{v}_t. \tag{24}$$

Again, to mimic JSZ, we use the observability restriction (15). Substituting (24) into (15) shows that in this case, this is equivalent to assuming: $\mathbf{W}'\mathbf{v}_t = \mathbf{0}$. We then substitute (24) into $\mathbf{q}_t = \mathbf{W}'\mathbf{y}^o_t$ and, providing that the matrix $(\mathbf{W}'\mathbf{B}_{t-1})$ is full rank, we can recover the latent vector conditional on the information at time $(t-1)$:

$$\mathbf{x}_t = -(\mathbf{W}'\mathbf{B}_{t-1})^{-1}\mathbf{W}'\mathbf{a}_{t-1} + (\mathbf{W}'\mathbf{B}_{t-1})^{-1}\mathbf{q}_t. \tag{25}$$

Putting this back into (24) shows that the yields are *conditionally* affine in the observable vector:

$$\mathbf{y}^o_t = \mathbf{a}_{\mathbf{q},t-1} + \mathbf{B}_{\mathbf{q},t-1}\mathbf{q}_t + \mathbf{v}_{\mathbf{q},t}, \tag{26}$$

where $\mathbf{a}_{\mathbf{q},t-1} = \mathbf{a}_{t-1} - \mathbf{B}_{t-1}(\mathbf{W}'\mathbf{B}_{t-1})^{-1}\mathbf{W}'\mathbf{a}_{t-1}$ and $\mathbf{B}_{\mathbf{q},t-1} = \mathbf{B}_{t-1}(\mathbf{W}'\mathbf{B}_{t-1})^{-1}$.

The time-structure of these relationships means that, given the principal components,

reasonable estimates of $\mathbf{a}_0$ and $\mathbf{B}_0$ and an initial value $\mathbf{x}_1$, we can compute $\mathbf{a}_{\mathbf{q},1}$ and $\mathbf{B}_{\mathbf{q},1}$, hence $\mathbf{x}_2$, $\mathbf{a}_{\mathbf{q},2}$ and $\mathbf{B}_{\mathbf{q},2}$... $\mathbf{x}_T$, $\mathbf{a}_{\mathbf{q},T}$ and $\mathbf{B}_{\mathbf{q},T}$, and sequentially recover the time series of $\mathbf{x}_t$. Good estimates $\mathbf{a}_0$ and $\mathbf{B}_0$ are provided for example by Gaussian model estimates $\mathbf{a}$ and $\mathbf{B}$ in (9), so that $\mathbf{x}_1 = -(\mathbf{W}'\mathbf{B})^{-1}\mathbf{W}'\mathbf{a} + (\mathbf{W}'\mathbf{B})^{-1}\mathbf{q}_1$.

Since the 'n' and 'q' approximation schemes do not involve the parameters $\boldsymbol{\mu}^{\mathcal{P}}$ and $\boldsymbol{\Phi}^{\mathcal{P}}$, these can be concentrated out of the likelihood function using OLS regression formulae, leaving only $\mu_{\infty}^{\mathcal{Q}}$, $\boldsymbol{\lambda}^{\mathcal{Q}}$, $\boldsymbol{\Sigma}$ and (possibly) the lower bound parameter $\underline{r}$.to be estimated from the concentrated likelihood function, as in the original JSZ approach.

3.4 The separability of the likelihood function

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

$$\log \mathcal{L}(\boldsymbol{\Theta}) = \sum_{t=2}^{T} \log \ell(\mathbf{y}_t^o | \mathbf{y}_{t-1}^o; \boldsymbol{\Theta}). \tag{27}$$

The $t-$period conditional density can be decomposed into the likelihood of observing the yields given the fitted yield portfolios (g) and the likelihood of observing the latter given the lagged fitted portfolios (h):

$$\ell(\mathbf{y}_t^o | \mathbf{y}_{t-1}^o; \boldsymbol{\Theta}) = g(\mathbf{y}_t^o | \mathbf{q}_t; \boldsymbol{\Theta}) \times h(\mathbf{q}_t | \mathbf{q}_{t-1}; \boldsymbol{\Theta}). \tag{28}$$

The main advantage of the factor rotation and factor extraction schemes is that they allow for the separation of parameters in the conditional likelihood as:

$$\ell(\mathbf{y}_t^o, \mathbf{q}_t | \mathbf{q}_{t-1}; \boldsymbol{\Theta}) = g(\mathbf{y}_t^o | \mathbf{q}_t; \mu_{\infty}^{\mathcal{Q}}, \boldsymbol{\lambda}^{\mathcal{Q}}, \boldsymbol{\Sigma}, \underline{r}, \boldsymbol{\Sigma}_{\mathbf{v}}) \times h(\mathbf{q}_t | \mathbf{q}_{t-1}; \boldsymbol{\mu}^{\mathcal{P}}, \boldsymbol{\Phi}^{\mathcal{P}}, \boldsymbol{\Sigma}), \tag{29}$$

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 OLS estimates of (3).

They can effectively be concentrated out of the likelihood function and recovered subsequently, as in JSZ. As usual, the cross-sectional variance parameters, $\mathbf{\Sigma_v}$, can also be concentrated out of the likelihood function. This leaves us with only $K \times (K+1)/2 + 2$ parameters in $\mathbf{\Sigma}$, $\mu_\infty^{\mathcal{Q}}$ and $\underline{r}$ that need to be found numerically, just one more ($\underline{r}$) than in the Gaussian term structure model. 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.

In a maximally-flexible Gaussian $DTSM$, (i.e. one in which there are no restrictions across the two sets of dynamics) the mapping between the observed and latent factors is linear given the observability restriction (15). As JSZ 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 macro-finance 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 the shadow rate model, which means that the irrelevance proposition does not hold. Indeed, we need to use the model of the yields cross-section of yields in order to extract the underlying latent factors from the principal components of yields and find their time series dynamics.

3.5 Summary

We can summarize our estimation procedure for 'q' and 'n' approximation schemes as:

1. Estimate the observable factors $\mathbf{q}_t$ and the weights $\mathbf{W}$ from the first K principal components of the yields;

2. Conditional on the risk-neutral parameters Ψ, obtain initial estimates of $\mathbf{a}$, $\mathbf{B}$ and the state vector: $\mathbf{x}_1 = -(\mathbf{W}'\mathbf{B})^{-1}\mathbf{W}'\mathbf{a} + (\mathbf{W}'\mathbf{B})^{-1}\mathbf{q}_1$;

3. Use $\mathbf{x}_1$ to find $\mathbf{a}_1$ and $\mathbf{B}_1$ as in (20) and (21);

4. Find $\mathbf{x}_2$ using (25): $\mathbf{x}_2 = -(\mathbf{W}'\mathbf{B}_1)^{-1}\mathbf{W}'\mathbf{a}_1 + (\mathbf{W}'\mathbf{B}_1)^{-1}\mathbf{q}_2$;

5. Repeat steps (3)-(4) for $t = 2, \ldots, T$;

6. Find the parameters of the $\mathcal{P}-$dynamics by OLS estimation using the time series of the latent state vector;

7. This gives the value of the concentrated likelihood (28) conditional on the risk-neutral parameters Ψ. Quasi-ML estimates of the parameters follow by optimizing this numerically over Ψ.

4 Empirical evaluation of the factor extraction approach

4.1 The yield-only model

4.1.1 Data and model specifications

We estimate the model using Treasury yields constructed by Gurkaynak, Sack, and Wright (2007).[2] We use the same maturities as Wu and Xia (2016), that is 3 and 6 months, 1, 2, 5, 7 and 10 years, but we use longer sample period, from January 1981 to December 2016. We follow Krippner (2013), Christensen and Rudebusch (2015), Coroneo and Pastorello (2017) and most of the literature on the term structure modelling within the Gaussian framework by estimating the model using yields rather than the forward rates fitted by Wu and Xia (2016).

We first estimate three benchmark models. The first (which we denote JSZ) is the Gaussian model (with $\mathbf{a}_t = \mathbf{a}$ and $\mathbf{B}_t = \mathbf{B}$) estimated by the PC factor rotation method proposed by Joslin et al. (2011). The second and third benchmarks are provided by the

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

shadow rate models of yields with the Wu and Xia forward rate approximation estimated by the EKF, both unrestricted and with the observability restriction $\mathbf{W}'\mathbf{v} = \mathbf{0}$ on the fitting errors, which we denote KF_u and KF_r, respectively.

We then estimate the shadow rate model by factor rotation, starting with the nonlinear factor extraction method (17) outlined in Section 3.1, which we denote by FE. Since the model estimated with factor rotation using the 'p' approximation coincides with the KF_r model estimated by the EKF,we do not consider it separately. Instead, we consider only the 'n' and 'q' approximations, denoted respectively by FR_n and FR_q. Although it is common to initiate the Kalman filter from the unconditional mean of the latent factors, we estimate all models conditional on the first observation assumed to be generated by the Gaussian model, i.e.:

$$\mathbf{y}_1^o = \mathbf{a} + \mathbf{B}\mathbf{x}_1 + \mathbf{v}_1. \tag{30}$$

We do this because the unconditional mean is not available at the beginning of the iterative factor rotation procedure for the FR_n and FR_q methods. This allows us to compare directly all models not only in terms of the cross-sectional fit, but also the value of the likelihood function.

The summary of different methods is given in Table 1. We consider 3−factor models. All models, except JSZ, are estimated both with imposed restriction on the zero lower bound parameter ($\underline{r} = 0$) and with $\underline{r}$ estimated as a free parameter ($\underline{r} = \widehat{r}$). The JSZ parametrization 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 value for all methods, therefore we report only the estimates for this case.

4.1.2 Fitting a three factor model

Table 2 shows the fit of the 3 factor model obtained using these various methods of estimation, reporting the likelihood of the model and the root-mean-square error ($RMSE$)

for each maturity. The upper panel shows the results with the lower bound set at zero ($\underline{r} = 0$), while the lower panel treats this as a parameter to be estimated ($\underline{r} = \widehat{\underline{r}}$). The first row shows the fit of the JSZ Gaussian model, which, as we would expect, is not as good as it is for the shadow rate models over this period. 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. Recall that model KF_u allows all principal components to be measured with an error, while KF_r uses the observability restriction (15). Imposing this restriction improves the cross-sectional fit of the model (as indicated by the average $RMSE$) when the lower bound is set at zero, but with the estimated $\underline{r}$ parameter, the average $RMSE$ is broadly the same.

Looking at the cross-sectional fit, it is apparent that when the model is estimated using the Kalman filter without the observability restriction (15) in KF_u, there is a marked improvement in the fit of the $10-$year yield, which comes at the price of a deterioration at the short end. This is particularly pronounced when the lower bound is a free parameter. Although this has the effect of boosting the likelihood, this reflects a well-known problem with the lower bound parameter in the shadow rate model (Christensen and Rudebusch, 2015, Bauer and Rudebusch, 2016 and Krippner, 2015). As Krippner (2015), puts it: "shorter-maturity data more below the lower bound parameter will result in greater effective downweighting of the influence of that data relative to longer maturity data (...) to better fit the steeper slope in the longer-maturity data."[3] In other words, when the lower bound is a free parameter, KF_u, boosts the likelihood by setting a high value of the lower bound parameter: 16 b.p. in our data set.[4]

In order to explore the 'downweighting' effect we need to look at the shadow rates, shadow forwards and the probability weights $\phi(z)$ and $\Phi(z)$ that the model gives uses in (10). But

[3]Krippner (2015), p.18.

[4]Since this is well above the minimum federal funds rate of $4bp$ seen over this period, it certainly cannot be regarded as a lower bound. In our sample, the yields are as low as $3bp$. These models attribute the difference between sub-bound observations and the lower bound to a fitting error. The minimum observed forward rate is actually negative ($-0.5bp$ reached by the $6-$month forward rate on July 2014). Indeed, the recent European experience suggests that the lower bound could be potentially much lower still, with rates of $-40bp$ persisting in the Euro area and even more negative rates seen at times in Switzerland and Sweden.

first, we would note that, as Table 2 shows, estimating the model with the FE and 'q' approximation methods gives almost identical results to the KF_r method, both in terms of the cross-sectional fit and the likelihood value. The 'n' approximation also works reasonably well, although the goodness of fit statistics are slightly inferior to those of other methods. All in all, there is little to choose between these approaches in terms of fit, so the choice can be made in terms of speed of computation, which is an order of magnitude faster for our rotation approaches, as Section 4.1.5 reveals.

In Figures 1 and 2 we plot the fitting errors, defined as the difference between the observed yields and fitted yields, $\mathbf{y}_t^o - \widehat{\mathbf{y}}_t$, with the lower bound set to zero and estimated as a free parameter, respectively. Since the fit of all models was very similar before 2008, the charts focus on the period from January 2008 to December 2016, i.e. the period of ultra low interest rates. First of all, consistent with our results from Table 2, we note that the fitting errors of the shadow rate models with restricted errors are almost identical, so that the lines of the KF_r, FE, FR_n and FR_q fitting errors are almost perfectly overlapping. Second, the fit of the Gaussian JSZ model, although at times quite different from the shadow rate model, is surprisingly, not much worse during the lower bound period. Third, Figure 2 supports our view about overfitting the KF_u model. If we do not impose the observability restriction (15), the KF_u fitting errors for the 3 and 6 month yield are higher and more volatile than those obtained from the methods that restrict the measurement errors. Finally, we can see some persistence in the fitting errors. Admittedly, although it would be desirable to address this statistical feature in the estimation, we do not deal with this issue as is common in the literature.[5]

4.1.3 Shadow rates and the probability of 'lift-off'

Figure 3 plots the shadow short rates implied by the different approaches for the low interest rate period $2008-2016$. (In the preceding period, when the interest rates are higher, the estimates of the short rate implied by all models are virtually identical to each other

[5]For methods to deal with the persistent fitting errors, see Adrian et al. (2013), Goliński and Spencer (2017) or Hamilton and Wu (2014).

and close to the actual short rate). For comparison, this also shows the actual 1−month rate and the estimate from the JSZ Gaussian model. Interestingly, the short rate from this model does not violate the zero lower bound restriction at any point in time. This is largely because our sample includes the 3−month rate, which is close to the 1−month rate and which the model tries to fit closely. Nevertheless, the Gaussian model has the awkward feature that given a low starting value, the probability of a negative short rate in the near future is non-negligible.

With $\underline{r} = 0$, the shadow short rates implied by different shadow rate estimates are all very similar to each other, irrespective of whether fitting errors are restricted. When the lower bound parameter is freely estimated but the observability restriction (15) is applied, the shadow rate specifications give similar paths to each other for the shadow rate, which at times (for example in 2009-2010 and 2014-2015) is more negative than when the lower bound is set to zero.

As noted, however, when the fitting errors are unrestricted, KF_u chooses a much higher value of $\underline{r}$ than the other methods. Figure 3 shows that the model then implies much lower values of the shadow spot rate (and the short forward rates), particularly during 2014 and 2015, meaning that the yields at the short end of the curve are 'downweighted' and lose traction. This is due to the effect of the probability density and distribution functions $\phi(z_t)$ and $\Phi(z_t)$ used in the Wu and Xia (2016) forward rate approximation (10). When the shadow forward rate for any maturity is one or more standard deviations lower than $\underline{r}$, these two functions are small. The fitted forward rate converges upon $\underline{r}$ which centres in the middle of the distribution of short maturities. This leads to a small deterioration in their fit but allows the optimizer to capture the upward slope in the longer maturities, as noted in the quote that we cited earlier from Krippner (2015).

Figure 4 plots the CDFs, $\Phi(z_t)$, for the various estimation methods and maturity horizons, which can be interpreted as the (risk-neutral) probability of rates being above $\underline{r}$, i.e. that 'lift-off' has occurred at that horizon. We focus on the period 2000-2016, since interest rates were high enough to keep the CDFs very close to unity for all maturities in earlier years. This

figure shows that when using KF_u with a freely estimated lower bound parameter (Panel (b)), the CDFs for 3−month, 6−month and even 1−year rates are effectively zero at times post-crisis. This is not a problem when we assume that some portfolios of yields are fitted without error (Panel (d)) or when we set the lower bound parameter to zero (Panels (a) and (c)).

Krippner (2015) argues that when conventionally estimated (using KF_u), the three factor shadow rate model is too flexible to provide a reliable indicator of the shadow rate. As such, this shadow rate should not be regarded as a policy indicator, contrary to the claims of Wu and Xia (2016). He goes on to develop a two factor model that fits the data less-well, but is more robust in this respect. However, restricting the flexibility of the Kalman filter using the observability restriction (15), in addition to its well-known advantages, factor extraction also has the effect of stopping the filter from exploiting this flexibility to 'downweight' the short maturities and focus on the fit of the long maturities.

4.1.4 Parameter estimates

While practitioners are likely to be interested in the ability of a method to fit the cross-section of bond yields, a researcher is more likely to be interested in drawing statistical inferences about the risk-neutral and the physical parameters. Table 3 shows the $\mathcal{Q}$ parameter estimates from the JSZ Gaussian model and the shadow rate model estimated by the Kalman filter-based and factor extraction approaches. The reported standard errors are calculated by using the quasi-maximum likelihood estimator proposed by White (1982), which is robust to mis-specification.[6,7] The table shows that the effect of imposing the observability restriction (15) on the estimates of the risk-neutral parameters is of negligible economic significance, although in statistical terms these differences can as usual be large relative to

[6]See Hamilton (1994),p. 145. We found that its standard errors are more conservative in comparison to those calculated using the inverse of the Hessian matrix or the Jacobian outer product.

[7]Alternatively, the standard errors could be calculated in a bootstrap exercise. Using the traditional Kalman filter method to estimate the standard errors for the shadow rate model using simulation methods would be highly impractical due to the long estimation time, but with the factor rotation method this can be obtained relatively quickly.

the minuscule standard errors implied by the 'tiny' fitting errors in the cross-section. The only difference of any note is in the large estimate of $\underline{r}$ given by KF_u, when this parameter is freely estimated, but as noted in the previous section, this is highly problematic. Importantly for our analysis, the estimates of the parameters of the model obtained by the factor rotation method (FR_n and FR_q) with $\underline{r} = 0$, are virtually identical to those obtained by the Kalman filter (KF_r), with small differences at the fourth decimal place.

The estimation method has relatively little effect on the estimates of the physical dynamics. In Section 4.1.6 below, using simulations we document robustness of the $\mathcal{P}$ estimates with respect to different estimation methods. Thus, to conserve the space, in Table 4 we report only the estimates obtained by the FE method with the estimated lower bound parameter. The White (1982) standard errors are reported in small font. As commonly found in the literature, under the physical measure the factor dynamics are slightly less persistent than under the risk-neutral measure, with the eigenvalues of the $\mathbf{\Phi}^{\mathcal{P}}$ matrix equal to 0.9895, 0.9660 and 0.8062. These estimates, however, can potentially suffer from the small sample autoregressive bias. In Section 4.2.1 we show how to effectively deal with it within the shadow rate framework by imposing zero restrictions on the price of risk parameters.

4.1.5 Estimation times

To illustrate the estimation time advantage, Table 5 reports the time (in seconds) needed to estimate the model with different methods using the JSZ parameters as starting values. We tried different starting values but we found that these are reliably provided by the estimates of the Gaussian model, and thus we use them in each estimation. We use zero as the starting value for the lower bound parameter.[8] The estimation time for the FR_n and

[8] All computations are performed on a PC desktop Windows 10 Enterprise $64-$bit operating system with the Intel(R) Core(TM) $i5$ $3.20GHz$processor with 8 GB RAM using Matlab R2017a. Using the built-in Matlab Coder, we speed up parts of the computing routine (in particular the function calculating the normal cumulative density, which is the most time consuming part of the routine) by compiling the program in Mex files using the code in C programming language.As the numerical optimizers we use Matlab functions 'fminsearch' and 'fminunc' consecutively. We do not modify their setting, but to ensure that we found the local maximum we loop the estimation routine until the optimizers do not make further progress, i.e. we start another round of the optimization using 'fminsearch' and 'fminunc' with the results from the previous

FR_q factor rotation methods is less than 2 minutes, irrespective of whether the the lower bound parameter is estimated or imposed. However, the estimation time using the Kalman filter is much longer. Estimating the the model with unrestricted fitting errors takes about 40 minutes. Restricting the fitting errors reduces the estimation time considerably to about $10 - 13$ minutes, but this is still much longer than for the factor rotation method.

4.1.6 Estimating the approximation errors

In this section we investigate the effect of using 'q' and 'n' approximations to estimate the latent state variables from the yield data, relative to (a) nonlinear extraction and (b) the 'p' approximation used in the EKF. To focus on the effect of the Taylor approximation, we assume that the PCs are observable. Since the $\mathcal{P}-$dynamics are crucial in determining a model's implications for the term premium and the like, we then investigate the extent to which different approximation schemes affect the $\mathcal{P}-$parameters.

To this end, we perform a Monte Carlo simulation. Using the parameters of the $\mathcal{P}-$dynamics obtained using the $KF_r(\underline{r} = \widehat{\underline{r}})$ method as reported in Tables 3 and 4, we simulate $150,000$ paths of the state variables $\mathbf{x}_t$, each with $10,000$ observations. We use this very long sample to eliminate any possible autoregressive bias in subsequent estimates of the $\mathcal{P}-$dynamics. This allows us to focus on distortions of the dynamics due to the Taylor approximation.

We consider each sample path s of the state vector $\mathbf{x}_t^{(s)}$ to be its 'true' value and use the model $\mathcal{Q}-$parameters to construct the model implied cross-section of forward rates and yields $\mathbf{y}_t^{(s)}$ using the WX formula (10). Importantly, since there is no measurement or other cross-sectional error, the observability restriction holds and a researcher could infer the 'true' value of the latent factors $\mathbf{x}_t^{(s)}$ in any simulation from any 3 of these yields or portfolios of yields as suggested in (16) using the nonlinear function inverse. So, $\mathbf{x}_t^{(s)}$ is the natural benchmark to use in this exercise.

Next, from each panel of fitted WX yields $\mathbf{y}_t^{(s)}$ we recover the 'estimated' state vector

round as the starting values. We repeat this process until there is no progress in the value of the likelihood function. We found that the maximum value of the likelihood function found with this loop routine is typically much higher than a single round optimization.

using the sequential factor rotation technique summarized in Section 3.5, using the 'p', 'q' and 'n' approximation schemes. We use the same $\mathcal{Q}-$parameters and the identification matrix $\mathbf{W}$ in all simulations. Thus, we obtain $150,000$ paths of the estimated state vectors $\mathbf{x}_{p,t}^{(s)}$, $\mathbf{x}_{q,t}^{(s)}$ and $\mathbf{x}_{n,t}^{(s)}$. These would all coincide with $\mathbf{x}_t^{(s)}$ in an affine model under the observability assumption, so the differences between estimated and true values reflect the degree of nonlinearity and the ability of the various linearization schemes to handle it.

First, we want to quantify the size of distortions in the state variables introduced by the different Taylor approximation schemes. As such, in Table 6 we report the distribution statistics of the bias and $RMSE$ calculated for each approximation scheme as

$$bias_i^{(s)} = \frac{1}{10,000} \sum_{t=1}^{10,000} \left(\mathbf{x}_{i,t}^{(s)} - \mathbf{x}_t^{(s)} \right) \tag{31}$$

and

$$RMSE_i^{(s)} = \left(\frac{1}{10,000} \sum_{t=1}^{10,000} \left(\mathbf{x}_{i,t}^{(s)} - \mathbf{x}_t^{(s)} \right) \cdot \left(\mathbf{x}_{i,t}^{(s)} - \mathbf{x}_t^{(s)} \right) \right)^{1/2}, \tag{32}$$

where in the second equation the dot symbol $(\cdot)$ denotes element by element multiplication and $i \in \{p, q, n\}$ denotes the approximation scheme. In particular, we report the mean and median of the distortion bias and $RMSE$ as well as the first and $99th$ percentile of the distribution, which give us the idea about possible extreme distortions. Since the factors have different magnitude, to make the statistics comparable, before calculating the bias and $RMSE$, we normalize the recovered factors by dividing by the standard deviations of the state variable $\mathbf{x}_t^{(s)}$ for each simulation.

As is any Monte Carlo simulation, the presence of outliers in the simulated latent factor shocks can lead to large distortions in their subsequent estimates. The table suggests that the distortions are small and similar for all three linearization schemes. Generally, as should be expected, the 'p' approximation scheme generates smaller deviations than the other approximation schemes. However, surprisingly, despite the goodness of fit statistics discussed earlier, the 'n' approximation scheme does very well in this exercise, almost as well as the 'p'

approximation scheme of the EKF. The most noticeable distortions are for the third latent factor, which is the smallest and least persistent and thus the least precisely estimated. Still, the median bias for the third factor is only about 2%. The variation in the estimates of the third factor as measured by the $RMSE$, however, can be quite substantial. The median of the $RMSE$ varies from 15% for the 'p' approximation to 26% for the 'q' approximation.

Next, we compare the estimates of the $\mathcal{P}-$dynamics obtained from different approximation methods to the estimates obtained from the 'true' state vector.

That is, for each simulation and each approximation scheme, we calculate the OLS estimates of $\widehat{\mu}_i^{\mathcal{P}(s)}$ and $\widehat{\Phi}_i^{\mathcal{P}(s)}$ from the $\mathcal{P}-$dynamics calculated from the state vectors

$$\mathbf{x}_{i,t}^{(s)} = \mu_i^{\mathcal{P}} + \mathbf{\Phi}_i^{\mathcal{P}} \mathbf{x}_{i,t-1}^{(s)} + \mathbf{u}_{i,t}^{(s)\mathcal{P}}. \tag{33}$$

and see how they deviate from the baseline values obtained from from the true vectors:[9]

$$\mathbf{x}_{t}^{(s)} = \mu^{\mathcal{P}} + \mathbf{\Phi}^{\mathcal{P}} \mathbf{x}_{t-1}^{(s)} + \mathbf{u}_{t}^{(s)\mathcal{P}}. \tag{34}$$

Table 7 reports the mean and standard deviation of the differences of the parameters obtained using the estimated series $\mathbf{x}_{i,t}^{(s)}$ and the true series $\mathbf{x}_t^{(s)}$. Again, the 'n' approximation scheme does almost as well as the 'p' approximation scheme. Nevertheless, these distortions are all small compared to the parameter standard errors and there is little to choose between them. However, the $\mathcal{P}-$dynamics are unrestricted and as such likely to be over-fitted. This issue is investigated in the next section.

Finally, we examine the total effect of approximation distortions on the yield term premium, which is often the object of most interest to an econometrician modelling the term structure of interest rates. As such, we calculate the bias and $RMSE$ of the deviations of the 1, 5 and 10 year term premium calculated based on the factors obtained from different

[9] Alternatively, we could compare the deviations of the parameters estimates obtained by different approximation schemes from the 'true' parameters used in the simulations. We believe, however, that for practical purposes more relevant is the comparison to the sample estimates obtained by using the original simulated factors. The alternative comparison does not change the relative classification of the approximations and is available from the authors upon request.

approximations and the resulting estimates of the $\mathcal{P}-$dynamics from the term premium calculated using the original (simulated) factors and the corresponding $\mathcal{P}-$parameters for each simulation. The results are presented in Table 8. It is evident that all three approximation methods perform well, with the 'p' and 'n' approximations performing almost equally and the 'q' approximations slightly worse. The median $RMSEs$ for the deviations in the $10-$year term premium for the 'p', 'n' and 'q' approximations are 6, 8 and 12 basis points, respectively. In practice, the lines of the four term premia overlap almost exactly.

4.2 The macro-finance model

Encouraged by these results, we decided to explore the implications of the shadow rate model for monetary policy, developing a macro-finance model, by adding observable macro target variables to the model. To stay as close as possible to the Gaussian benchmark in Joslin et al. (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 survey of consumers conducted by the University of Michigan as the expected inflation measure.[10] For interest rates, we use the same set of yields as in the previous section.

Following Joslin et al. (2014), we assume that these macro variables are unspanned, only affecting the term structure with a lag through the $\mathcal{P}-$dynamics. This means the structure of the risk-neutral dynamics (4) remains the same, specified in terms of the K latent variables, while the physical dynamics involve the K latent variables ($\mathbf{x}_t$) that span the term structure as well as a vector of macroeconomic or other 'unspanned' variables ($\mathbf{m}_t$) that do not. Although the macro variables do not affect the term structure contemporaneously, they can influence interest rates expectations under the $\mathcal{P}-$measure and hence the term premium. Following Joslin et al. (2014), we assume that the spanned and unspanned factors can be specified jointly as a $VAR(1)$.

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

In order to allow a direct comparison with the results in JPS for the Gaussian model, we follow Lemke and Vladu (2017) and rotate the latent factors $\mathbf{x}_t$ to get 'shadow principal components', $\mathbf{q}_t^s$:

$$\mathbf{q}_t^s = \mathbf{W}'\mathbf{y}_t^s = \mathbf{W}'(\mathbf{a}_y + \mathbf{B}_y\mathbf{x}_t), \tag{35}$$

where $\mathbf{y}_t^s$ are 'shadow yields' (i.e. yields that would prevail in the absence of the zero lower bound), and have a Gaussian distribution. Importantly, unlike the *observed* principal components of interest rates in the Gaussian model of JSZ, the *shadow* principal components are specific to the model of the $\mathcal{Q}-$dynamics, meaning that the $\mathcal{P}-$dynamics cannot be estimated separately as they can in JSZ.

The joint system of the $\mathcal{P}-$dynamics can then be written as:

$$\begin{bmatrix} \mathbf{q}_t^s \\ \mathbf{m}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_q^{\mathcal{P}} \\ \boldsymbol{\mu}_m^{\mathcal{P}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Phi}_{qq}^{\mathcal{P}} & \boldsymbol{\Phi}_{qm}^{\mathcal{P}} \\ \boldsymbol{\Phi}_{mq}^{\mathcal{P}} & \boldsymbol{\Phi}_{mm}^{\mathcal{P}} \end{bmatrix} \begin{bmatrix} \mathbf{q}_{t-1}^s \\ \mathbf{m}_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_{q,t}^{\mathcal{P}} \\ \mathbf{u}_{m,t}^{\mathcal{P}} \end{bmatrix}. \tag{36}$$

The $\mathcal{Q}-$dynamics (4) can also be written in this format as:

$$\begin{bmatrix} \mathbf{q}_t^s \\ \mathbf{m}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_q^{\mathcal{Q}} \\ \mathbf{0}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Phi}_{qq}^{\mathcal{Q}} & \mathbf{0}_{3,2} \\ \mathbf{0}_{2,3} & \mathbf{0}_{2,2} \end{bmatrix} \begin{bmatrix} \mathbf{q}_{t-1}^s \\ \mathbf{m}_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_{q,t}^{\mathcal{Q}} \\ \mathbf{0}_2 \end{bmatrix}, \tag{37}$$

where: $\boldsymbol{\Phi}_{qq}^{\mathcal{Q}} = \mathbf{W}'\mathbf{B}\boldsymbol{\Phi}^{\mathcal{Q}}(\mathbf{B}'\mathbf{W})^{-1}$ and $\boldsymbol{\mu}_q^{\mathcal{Q}} = \mathbf{W}'(\mathbf{I}-\mathbf{B}\boldsymbol{\Phi}_{qq}^{\mathcal{Q}}(\mathbf{B}'\mathbf{W})^{-1}\mathbf{W}')\mathbf{a}_y+\mathbf{W}'\mathbf{B}\boldsymbol{\mu}^{\mathcal{Q}}$. Although the theoretical structure of the risk-neutral dynamics is the same as in (4), the introduction of extra observable factors into (36) can shift the empirical covariance matrix $\boldsymbol{\Sigma}$, which is common to both risk-neutral and physical dynamics:

$$\boldsymbol{\Sigma}_q = \mathtt{Var}(\mathbf{u}_{q,t}^{\mathcal{P}}) = \mathtt{Var}(\mathbf{u}_{q,t}^{\mathcal{Q}}) = \mathbf{W}'\mathbf{B}\boldsymbol{\Sigma}(\mathbf{B}'\mathbf{W})^{-1}, \tag{38}$$

Modelling the dynamics of the macro variables jointly with the shadow principal components allows the results to be compared directly with those of Joslin et al. (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 at the ZLB. These authors neglect the constraint (38) and use the shadow rate in their factor augmented vector autoregression model ($FAVAR$). This rate is the sum of the latent factors in our framework, but we use all three shadow factors in the $FAVAR$, on the argument that the whole yield curve becomes a monetary policy instrument at the ZLB. Once the lower bound is reached, unconventional monetary policies 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, Raskin, Remache, and Sack, 2011).

A term structure model is able to use the physical factor dynamics (36) 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 unconventional monetary policies affected the yield curve. The existing literature uses the Gaussian $DTSM$ to make this decomposition, without respecting the ZLB, but our model can allow for the effect of this constraint on the decomposition, as we show in Section 4.2.2.

4.2.1 The 2^{19} model selection search

In contrast to the risk-neutral dynamic system (4), which has 4 parameters excluding Σ (3 roots and the level parameter $\mu_\infty^{\mathcal{Q}}$), the system of the physical dynamics (36) has 30 time-series parameters. 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 (36). 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:

$$\boldsymbol{\mu}_q^{\mathcal{P}} = \boldsymbol{\mu}_q^{\mathcal{Q}} + \mathbf{l}_0, \quad \boldsymbol{\Phi}_{qq}^{\mathcal{P}} = \boldsymbol{\Phi}_{qq}^{\mathcal{Q}} + \mathbf{L}_1, \tag{39}$$

and testing zero restrictions on the risk-adjustment parameters $\mathbf{l}_0$ and $\mathbf{L}_1$, as well as $\boldsymbol{\Phi}_{qm}^{\mathcal{P}}$ which is a matrix of zeros under the risk-neutral measure.

Exploiting the overwhelming time advantage of the factor rotation method, we follow Joslin et al. (2014) and search for the best macro finance model in terms of standard model selection criteria. Following Joslin et al. (2014), we estimate all possible combinations of zero restrictions on these parameters. Additionally, as in Joslin et al. (2014), we also consider restricting the largest eigenvalue of the $FAVAR$ response matrix under the $\mathcal{P}-$measure to be equal the largest root under the risk-neutral measure. This restriction is designed to correct the small-sample bias in autoregressive models.[11] Thus, in total we estimate 2^{19} models with different combinations of restrictions (18 risk-adjustment parameters and the eigenvalue restriction).

Although in our sample the 'q' approximation generally gives higher likelihood value, Taking advantage of the speed in estimation with the factor rotation approach, we estimate the macro-finance model using both the n and 'q' approximation schemes. Although the factor rotation method brings a dramatic improvement in speed, it would not be feasible to estimate this number of models on a single personal computer and we perform this task using parallel computing.[12] Both factor rotation methods produced nearly identical results.

The model selection led to the same set of zero restrictions in the risk premium and very similar parameter estimates for both factor rotation methods. In particular, the BIC and $HQIC$ information criteria lead to the selection of the same model, with 9 zero risk-adjustment restrictions and the eigenvalue restriction. As is well-known, the AIC generally leads to a much less parsimonious model; in our case it suggests 5 risk-adjustment restrictions

[11] Alternative approaches to correcting the autoregressive bias have been proposed by Bauer, Rudebusch, and Wu (2012), Jardet, Monfort, and Pegoraro (2013), and Joslin et al. (2011).

[12] The estimation of the macro-finance models were performed using Data Analysis Cluster at the Alcuin Research Resource Centre, University of York.

and a marginal rejection of the maximum eigenvalue restriction. Since the AIC, as opposed to BIC and $HQIC$, is not a consistent model selection criterion, we rely on the model suggested by the other two information criteria,

To eliminate any distortions due to the linearization, using selected risk-adjustment restrictions the inputs from the 'q' and 'n' results, we re-estimate the model using the nonlinear factor extraction technique, which we consider to be optimal and call MFE^{opt}. The estimates of the risk-adjustment parameters are reported in Table 9. The price of level risk, which is reflected in the risk-adjustment in the first row of this table, is influenced by all factors except expected inflation. The macro factors are statistically significant influences upon the price of slope risk in the second row (often interpreted as reflecting the stance of conventional monetary policy), while the second and third shadow principal components are not. Also, we find that the shadow level factor is significant in the price of curvature risk in the third row. The positive sign of the GRO coefficient in the first row implies that price of level risk is cyclical. On the other hand, the price of slope risk behaves in a countercyclical fashion, since it declines with both growth and inflation expectations.

Our results for the shadow rate model differ from the Gaussian model results reported in Table 2 of Joslin et al. (2014) in several ways. In particular, in Joslin et al. (2014) expected inflation has a positive effect on the price of level risk and no effect on the price of slope risk, while in our Table 9 it has no effect on the price of level risk but a negative effect on the price of slope risk. This difference appears to be due to the effect of the longer data sample rather than the use of a shadow rate model, because when we run the model selection exercise using the Gaussian model on our longer data sample we find a similar sign pattern. Indeed, the only difference between the two models for this sample is that the curvature factor is significant in pricing curvature risk in the Gaussian model. We conclude that the difference between our results and those reported in Joslin et al. (2014) is largely due to the difference in the data sample.

4.2.2 The term premium

Table 10 reports the estimates of the parameters of the $\mathcal{P}-$dynamics for our optimal macro-finance model (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 9. The unrestricted coefficients in the final two columns are all statistically significant, which indicate that the macro variables contain incremental information for predicting the term structure factors. On the other hand, the first and third principal components are statistically significant in predicting expected inflation, while the second principal component has a significant effect on the growth factor.

Given any estimates of the parameters of the $\mathcal{P}-$dynamics, such as those shown in Table 10, we can decompose the $10-$year yield into components representing expectations and risk premia. The expectations component represents 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 et al. (2012) we first find the risk-neutral shadow forward rates $f_{n,t}^{G}$ by using these parameters to compute the coefficients a_f and $\mathbf{b}_f$ using (6) and hence the risk-neutral yields using (20 - 21). The risk premium for any model and maturity follows by subtracting the risk-neutral yield from the fitted yield.

To the best of our knowledge, we are the first to estimate the term premium for a shadow rate macro-finance rate model. The estimate of the $10-$year yield term premium implied by our optimal macro-finance model is plotted as MFE^{opt} in Figure 5, alongside the the $10-$year yield, the vertical distance being the risk-neutral yield. For comparison, MFE shows the term premium implied by unrestricted shadow rate macro-finance model and FE shows the premium from the shadow rate yield-only model, with the parameters reported in the last row of Table 3. Finally, JPS^{opt} shows the premium from a Gaussian macro-finance model with the same zero restrictions that are selected for the shadow rate model. Although the model selection procedure for the Gaussian model suggested one restriction less than for model MFE^{opt} we show this to emphasize the difference implied by the affine framework, *ceteris paribus*.

Panel (a) of Figure 5 shows these various estimates of the term premium for the whole sample period. There is a marked difference between the restricted and unrestricted model specifications over the first part of this sample. This was a time of relatively high interest rates, which is why the term premiums implied by the restricted Gaussian and shadow rate models JPS^{opt} and MFE^{opt} are very similar. However, in the 1980s and early 1990s and some other times, they are well below the term premiums computed from the unrestricted shadow rate models MFE and FE. The differences become more complex as interest rates fall towards the ZLB. For this reason, Panel (b) focuses the picture on the period January 2008 to December 2016. Taken literally, the Gaussian model JPS^{opt} implies that the average expected short rate is negative over the 10−year horizon, even in 2016 after the federal funds rate 'lifts off' from the range $0 - 25$ basis point. This means that the Gaussian term premium is unrealistically high, often exceeding the 10−year rate. In this low interest rate environment, the term premium implied by the unrestricted shadow rate yield-only model FE is generally more volatile than other estimates of the term premium and it is frequently the lowest, reaching zero in the middle of 2016. Adding the macro variables (in MFE) helps to stabilize the term premium somewhat, although this remains very low, especially in 2012 and 2016. However, excluding insignificant risk adjustment parameters to get MFE^{opt} gives a more realistic term premium. It is about 4% at the beginning of 2009 when the federal funds rate reached its lower bound and gradually fell to about 1.5% in the middle of 2010. It moved up to 2.5% late in 2010, as the second phase of quantitative easing began, and since easing back in late 2011 it has remained relatively stable, ranging between 0.8% and 2%.

5 Conclusion

This paper shows how the factor rotation method of Joslin et al. (2011) can be applied to the shadow rate pricing model, allowing the researcher to estimate the shadow rate term structure model almost as easily and quickly as the simple Gaussian term structure model. We find that restricting the flexibility of the standard Kalman shadow rate estimator, by

adopting the JSZ assumption that the principle components are measured without error, has the has the added advantage of producing more plausible estimates of the lower bound parameter and the path of the shadow rate. This arguably provides a more reliable indicator of the thrust of monetary policy.

This technique opens the way to research with the shadow rate model that require large numbers of different datasets or models to be analyzed. We illustrate this by estimating the macro-finance variant that involves searching over large numbers of different parameter combinations. The traditional filtering approach makes such exercises infeasible. 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), which involve potentially many different combinations of common and idiosyncratic factors.

Appendix: The likelihood function

We maximize the joint log likelihood conditional in (27). The $t-$period conditional density can be written as in (28). As explained in Section 3.4, under the 'q' and 'n' approximation schemes but also when using the nonlinear factor extraction technique, with the exception of the $\mathbf{\Sigma}$ matrix, the g and h parts of the 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 likelihood function with the risk-neutral parameters g has the standard form:

$$g(\mathbf{y}_t^o|\mathbf{q}_t) = -\frac{J-K}{2}\left(1+\log(2\pi)\right) - \frac{J-K}{2}\log\sigma_v^2, \tag{A-1}$$

where σ_v^2 is the (homoscedastic) variance of measurement errors $v_{n,t}$ in (23), 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, h, the part of the 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 4.1 we estimate with respect to the $\mathcal{P}-$dynamics for the latent state vector as in the JSZ parametrization. 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:[13]

$$h(\mathbf{q}_t|\mathbf{q}_{t-1}) = h_x(\mathbf{x}_t|\mathbf{x}_{t-1}) \times \left|\det\left(\mathbf{J}_t\right)\right|^{-1}, \tag{A-3}$$

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

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

$$\mathbf{J}_t = \left[\frac{\partial \mathbf{q}_t}{\partial x_{1,t}}, \cdots, \frac{\partial \mathbf{q}_t}{\partial x_{K,t}}\right] = \mathbf{W}'\mathbf{B}_{t-1}. \tag{A-4}$$

The logarithm of the $P-$likelihood is then:

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

In our application of macro-finance model in Section 4.2, 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 (35). Applying the change-of-variable technique once more, we can write the log likelihood in terms of $\mathbf{q}_t^s$:

$$\log h(\mathbf{q}_t|\mathbf{q}_{t-1}) = \log h_{q^s}(\mathbf{q}_t^s|\mathbf{q}_{t-1}^s) - \log\left|\det\left(\mathbf{J}_t\right)\right| + \log\left|\det\left(\mathbf{J}\right)\right|, \tag{A-7}$$

where

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

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

$$\log h(\mathbf{q}_t, \mathbf{m}_t|\mathbf{q}_{t-1}, \mathbf{m}_{t-1}) = \log h_z(\mathbf{z}_t|\mathbf{z}_{t-1}) - \log\left|\det\left(\mathbf{J}_t\right)\right| + \log\left|\det\left(\mathbf{J}\right)\right|, \tag{A-9}$$

where the dynamic system for $\mathbf{z}_t$ is given in (36).

References

Adrian, T., Crump, R. K., Moench, E., 2013. Pricing the term structure with linear regressions. Journal of Financial Economics 10, 110–138.

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

Bauer, M. D., Rudebusch, G. D., 2014. The signaling channel for Federal Reserve bond purchases. International Journal of Central Banking 10, 233–289.

Bauer, M. D., Rudebusch, G. D., 2016. Monetary policy expectations at the zero lower bound. Journal of Money, Credit and Banking 48, 1439–1465.

Bauer, M. D., Rudebusch, G. D., Wu, J. C., 2012. Correcting estimation bias in dynamic term structure models. Journal of Business & Economic Statistics 30, 454–467.

Black, F., 1995. Interest rates as options. Journal of Finance 50, 1371–1376.

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

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

Coroneo, L., Pastorello, S., 2017. European spreads at the interest rate lower bound. Discussion Papers 17/10, Department of Economics, University of York.

Egorov, A. V., Li, H., Ng, D., 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., Raskin, M., Remache, J., Sack, B., 2011. The financial market effects of the Federal Reserve's large-scale asset purchases. International Journal of Central Banking 7, 3–43.

Goliński, A., Spencer, P., 2017. The advantages of using excess returns to model the term structure. Journal of Financial Economics 125, 163–181.

Greene, W. H., 2011. Econometric Analysis. 7th Edition, Pearson Education.

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

Hamilton, J., 1994. Time series analysis. Princeton University Press.

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

Hamilton, J. D., Wu, J. C., 2014. Testable implications of affine term structure models. Journal of Econometrics 178, 231–242.

Jardet, C., Monfort, A., Pegoraro, F., 2013. No-arbitrage near-cointegrated VAR(p) term structure models, term premia and GDP growth. Journal of Banking & Finance 37, 389–402.

Joslin, S., Le, A., Singleton, K. J., 2013. Why Gaussian macro-finance term structure models are (nearly) unconstrained factor-VARs. Journal of Financial Economics 109, 604–622.

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

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

Krippner, L., 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, The Australian National University.

Krippner, L., 2015. A comment on Wu and Xia (2015), and the case for two-factor shadow short rates. CAMA Working Papers 2015-48, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.

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

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.).

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1–25.

Wu, J. C., Xia, F. D., 2016. Measuring the macroeconomic impact of monetary policy at the zero lower bound. Journal of Money, Credit and Banking 48, 253–291.

Wu, J. C., Zhang, J., 2016. A shadow rate new Keynesian model. NBER Working Papers 22856, National Bureau of Economic Research, Inc.

Tables

Model	Specification
JSZ	Gaussian model estimated using observable principal components
KF_u	Shadow rate model estimated by extended Kalman filter, unrestricted errors
KF_r	Shadow rate model estimated by extended Kalman filter, restricted errors
FE	Shadow rate model estimated by nonlinear factor extraction
FR_n	Shadow rate model estimated by factor rotation using 'n' approximation
FR_q	Shadow rate model estimated by factor rotation using 'q' approximation

Table 1. Summary of the model specification.

Yields	Root-mean-square error, $K = 3$								
Model	3m	6m	1y	2y	5y	7y	10y	Av.RMSE	Log-lik.
JSZ	4.69	4.26	4.88	3.09	6.31	2.48	5.97	4.53	15,725.55
					$\underline{r} = 0$				
KF_u	4.85	4.18	4.83	3.14	5.89	2.52	5.56	4.42	15,932.07
KF_r	4.48	4.14	4.55	2.93	5.97	2.37	5.69	4.31	15,920.08
FE	4.48	4.15	4.53	2.94	5.95	2.37	5.67	4.30	15,924.60
FR_n	4.50	4.16	4.56	2.94	5.98	2.37	5.69	4.31	15,910.64
FR_q	4.47	4.14	4.54	2.93	5.98	2.37	5.69	4.30	15,926.94
					$\underline{r} = \widehat{\underline{r}}$				
KF_u	4.99	4.31	4.52	3.32	5.25	2.32	4.97	4.24	16,077.33
KF_r	4.44	4.12	4.50	2.95	5.88	2.34	5.59	4.26	15,953.00
FE	4.43	4.14	4.47	2.96	5.83	2.33	5.56	4.25	15,959.15
FR_n	4.47	4.15	4.52	2.96	5.91	2.34	5.62	4.28	15,929.50
FR_q	4.44	4.13	4.51	2.95	5.90	2.35	5.61	4.27	15,948.02

Table 2. Statistics of fit for different estimation methods. 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 2016.

Yields	Parameter estimates				
Model	λ_1^Q	λ_2^Q	λ_2^Q	$\mu_\infty^Q \times 100$	$\underline{r} \times 10,000$
JSZ	0.9973 0.00010	0.9663 0.0039	0.8741 0.0061	0.0379 0.0004	-
			$\underline{r} = 0$		
KF_u	0.9978 0.0003	0.9652 0.0020	0.8607 0.0253	0.0325 0.0023	0 –
KF_r	0.9977 0.0009	0.9654 0.0064	0.8574 0.0442	0.0328 0.0056	0 –
FE	0.9978 0.0001	0.9652 0.0004	0.8578 0.0016	0.0326 0.0005	0 –
FR_n	0.9978 0.0001	0.9653 0.0004	0.8589 0.0016	0.0326 0.0005	0 –
FR_q	0.9978 0.0001	0.9653 0.0004	0.8567 0.0016	0.0329 0.0005	0 –
			$\underline{r} = \widehat{\underline{r}}$		
KF_u	0.9979 0.0003	0.9642 0.0021	0.8667 0.0249	0.0314 0.0019	16.02 0.69
KF_r	0.9978 0.0003	0.9649 0.0019	0.8562 0.0223	0.0317 0.0020	9.03 0.91
FE	0.9978 0.0001	0.9648 0.0004	0.8585 0.0018	0.0317 0.0005	7.54 0.30
FR_n	0.9978 0.0001	0.9650 0.0005	0.8591 0.0020	0.0318 0.0005	6.40 0.31
FR_q	0.9978 0.0001	0.9652 0.0004	0.8569 0.0018	0.0318 0.0005	6.81 0.30

Table 3. Maximum likelihood estimates of the risk-neutral dynamics for the model estimated by different estimation methods. Standard errors are reported in small font. The sample period is January 1981 to December 2016.

$\boldsymbol{\mu}^{\mathcal{P}}$	$\boldsymbol{\Phi}^{\mathcal{P}}$			$\boldsymbol{\Sigma}$		
0.0005 0.0005	0.9933 0.0094	0.0257 0.0109	0.0265 0.0216	0.0044 0.0002		
−0.0023 0.0011	0.0347 0.0208	1.0041 0.0267	0.0963 0.0519	−0.0036 0.0006	0.0074 0.0010	
0.0034 0.0011	−0.0696 0.0226	−0.0771 0.0313	0.7643 0.0507	0.0004 0.0006	−0.0071 0.0012	0.0049 0.0006

Table 4. Maximum likelihood estimates of the $\mathcal{P}-$dynamics for the model estimated by the nonliner factor extraction and estimated lower bound parameter, $FE(\underline{r} = \widehat{\underline{r}})$. Standard errors are reported in small font. The sample period is January 1981 to December 2016.

model	time (in seconds)
	$\underline{r}=0$
KF_u	2,061
KF_r	635
FE	351
FR_n	91
FR_q	98
	$\widehat{\underline{r}}=\widehat{\underline{r}}$
KF_u	2,832
KF_r	787
FE	802
FR_n	71
FR_q	106

Table 5. Estimation time for different methods in seconds. The sample period is January 1981 to December 2016.

	1%			99%			Median			Mean		
	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
'p' approximation												
Bias	0.0007	0.0028	0.0043	0.0012	0.0079	0.0334	0.0009	0.0050	0.0149	0.0009	0.0050	0.0156
RMSE	0.0055	0.0267	0.0808	0.0110	0.0653	0.3803	0.0075	0.0398	0.1491	0.0077	0.0408	0.1608
'q' approximation												
Bias	0.0009	0.0021	0.0087	0.0016	0.0066	0.0476	0.0012	0.0039	0.0238	0.0012	0.0040	0.0245
RMSE	0.0069	0.0326	0.1295	0.0158	0.0927	0.8003	0.0103	0.0532	0.2575	0.0104	0.0548	0.2854
'n' approximation												
Bias	0.0008	0.0024	0.0061	0.0015	0.0072	0.0399	0.0011	0.0044	0.0186	0.0011	0.0045	0.0194
RMSE	0.0063	0.0303	0.0997	0.0126	0.0730	0.4000	0.0088	0.0450	0.1780	0.0089	0.0461	0.1883

Table 6. Statistics of the distortions of the factors due to different approximations. Using the parameters obtained from the FE model, we simulate $150,000$ samples of the state vector $\mathbf{x}_t$, each path consisting of $10,000$ time series observations. Based on these vectors we find the yields as given by the WX formula. Then, we apply different linearization schemes to recover the state vector from those yields by the factor rotation. For each simulation we calculate the bias and $RMSE$ of the estimated normalized factors. The table reports the statistics of these biases and $RMSEs$ obtained across $100,000$ simulations.

	μ_1	μ_2	μ_3	$\phi_{1,1}$	$\phi_{1,2}$	$\phi_{1,3}$	$\phi_{2,1}$	$\phi_{2,2}$	$\phi_{2,3}$	$\phi_{3,1}$	$\phi_{3,2}$	$\phi_{3,3}$
'p' approximation												
Mean	0.0000	-0.0002	0.0008	-0.0005	0.0058	-0.0214	-0.0012	0.0066	-0.0275	-0.0037	0.0117	-0.0469
Std.dev.	0.0000	0.0001	0.0006	0.0011	0.0032	0.0167	0.0017	0.0043	0.0237	0.0035	0.0086	0.0457
'q' approximation												
Mean	0.0000	-0.0006	0.0023	-0.0035	0.0155	-0.0599	-0.0056	0.0194	-0.0807	-0.0130	0.0369	-0.1492
Std.dev.	0.0001	0.0002	0.0013	0.0022	0.0064	0.0356	0.0034	0.0083	0.0496	0.0070	0.0164	0.0948
'n' approximation												
Mean	0.0000	-0.0004	0.0011	-0.0012	0.0086	-0.0291	-0.0021	0.0102	-0.0379	-0.0059	0.0199	-0.0671
Std.dev.	0.0000	0.0001	0.0006	0.0012	0.0036	0.0173	0.0019	0.0047	0.0243	0.0039	0.0095	0.0466

Table 7. Statistics of the distortions of the $\mathcal{P}-$dynamics due to different approximations. For each simulation (please see notes to Table 6) we calculate the $VAR(1)$ dynamics obtained from the original (simulated) factors and from those obtained by factor rotation. The table reports the statistics of these mean and standard deviations of the differences between the estimates of the parameters obtained from the factors estimated by factor rotation and the parameters obtained from the estimation of the original (simulated) factors.

	1%			99%			Median			Mean		
	$1y$	$5y$	$10y$	$1y$	$5y$	$10y$	$1y$	$5y$	$10y$	$1y$	$5y$	$10y$
'p' approximation												
Bias	1	1	2	2	3	3	2	2	3	2	2	3
RMSE	3	3	3	22	20	17	6	7	6	7	7	7
'q' approximation												
Bias	2	2	2	8	5	6	4	3	4	4	3	4
RMSE	5	6	5	31	28	23	13	13	12	14	14	12
'n' approximation												
Bias	0	1	2	4	3	4	2	2	3	2	2	3
RMSE	4	4	4	23	21	18	8	9	8	9	9	8

Table 8. Statistics of the distortions of the term premium due to different approximations. For each simulation (please see notes to Table 6), based on the factors recovered from different approximation schemes and corresponding estimated dynamics, we calculate the bias and $RMSE$ between the 1, 5 and 10−year yield term premium implied by different approximation schemes and the term premium implied by the 'true' parameters and original (simulated) factors. The reported values are in basis points.

	$const$	PC_1^s	PC_2^s	PC_3^s	GRO	INF
PC_1^s	0	-0.0112 0.0024	-0.1057 0.0287	0.3330 0.0647	0.0022 0.0006	0
PC_2^s	0	0.0175 0.0006	0	0	-0.0007 0.0002	-0.0007 0.0001
PC_3^s	0.0008 0.0001	-0.0054 0.0001	0	0	0	0

Table 9. Maximum likelihood estimates of the risk adjustments ($\mathbf{l}_0$ and $\mathbf{L}_1$ in (39)) for our preferred model with unspanned macro-risk, MFE^{opt}. Standard errors are reported in small font. The sample period is January 1981 to December 2016.

	$\boldsymbol{\mu}^{\mathcal{P}}$	$\boldsymbol{\Phi}^{\mathcal{P}}$				
	$const$	PC_1^s	PC_2^s	PC_3^s	GRO	INF
PC_1^s	0.0006 0.0000	0.9939 0.0024	-0.0096 0.0287	-0.0094 0.0647	0.0022 0.0006	0
PC_2^s	-0.0004 0.0000	0.0102 0.0006	0.9549 0.0001	0.3195 0.0003	-0.0007 0.0002	-0.0007 0.0001
PC_3^s	0.0012 0.0001	-0.0021 0.0000	0.0058 0.0001	0.8584 0.0000	0	0
GRO	0.0962 0.0500	0.1699 0.1173	2.1723 0.8824	-0.3880 3.5024	0.9349 0.0156	-0.0541 0.0170
INF	0.4626 0.0757	0.7732 0.1771	-1.5913 1.3232	9.9862 5.2880	0.0252 0.0234	0.8078 0.0258

Table 10. Maximum likelihood estimates of the $\mathcal{P}-$dynamics for our preferred model with unspanned macro-risk, MFE^{opt}. Standard errors are reported in small font. The sample period is January 1981 to December 2016.

Figures

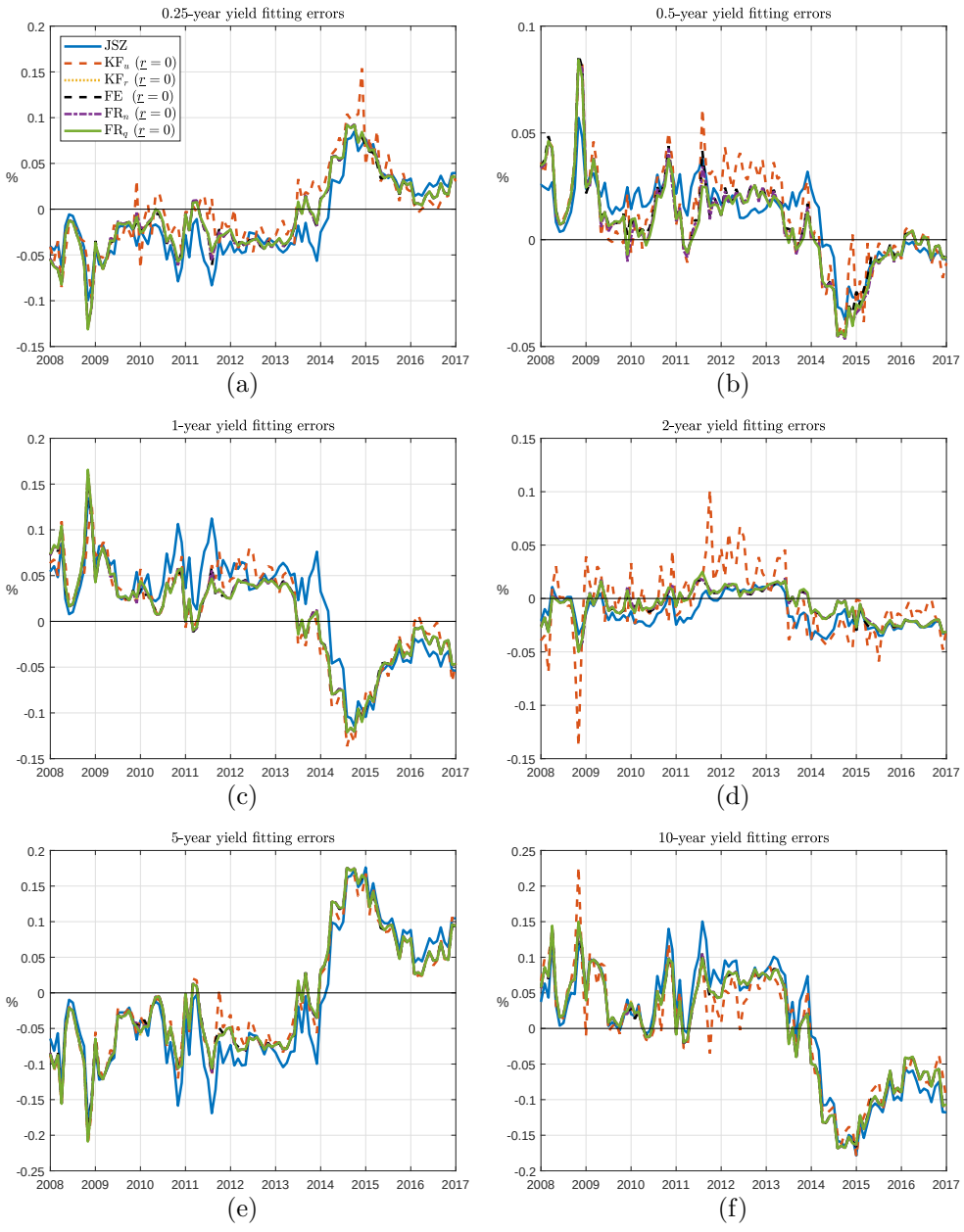


Figure 1. Fitting errors for models with $\underline{r} = 0$. The figure shows the difference between the observed 0.25, 0.5, 1, 2, 5 and 10−year yields and their fitted counterparts for particular models. The shadow rate models are estimated with the imposed restriction $\underline{r} = 0$. The estimation sample is January 1981 to December 2016.

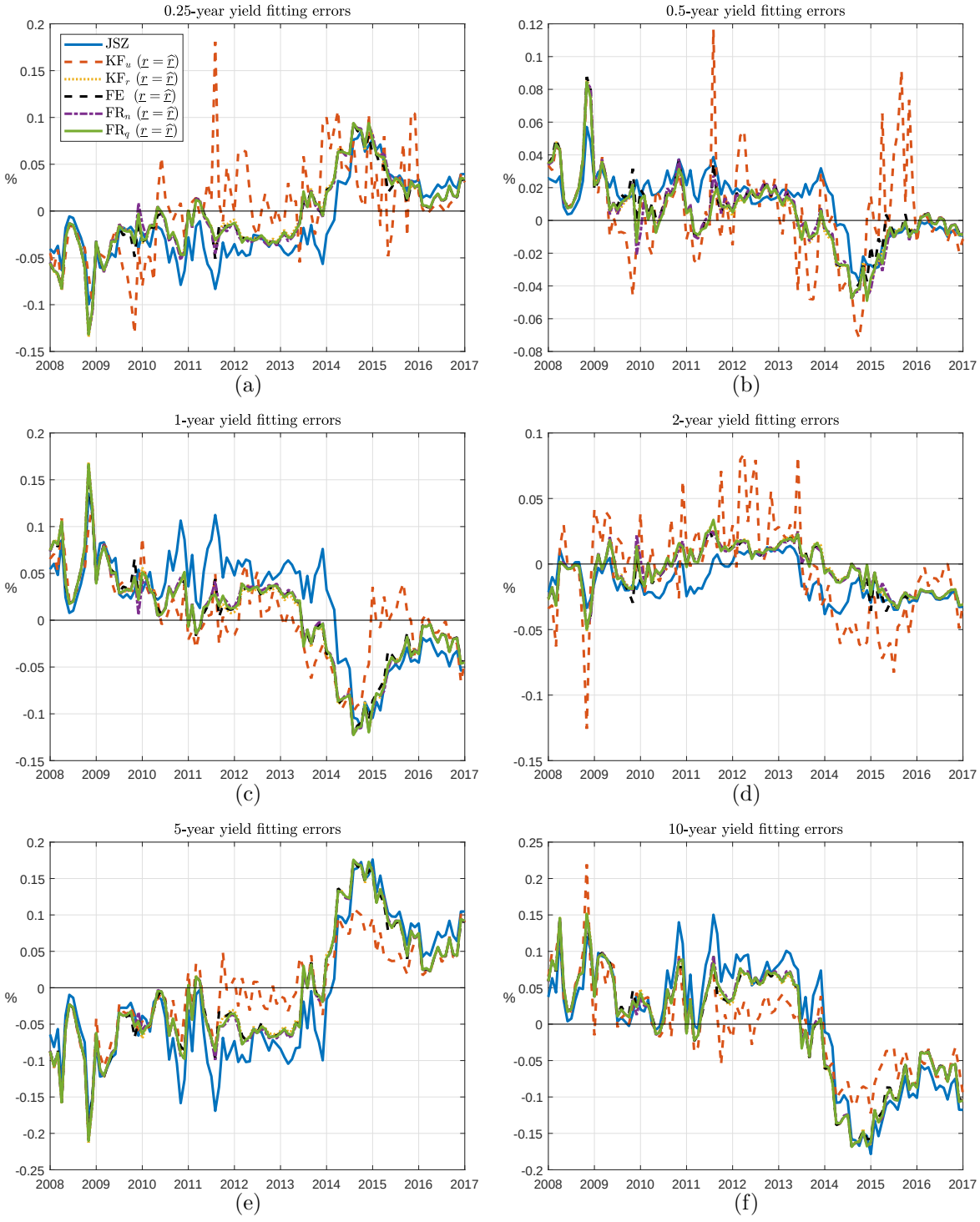
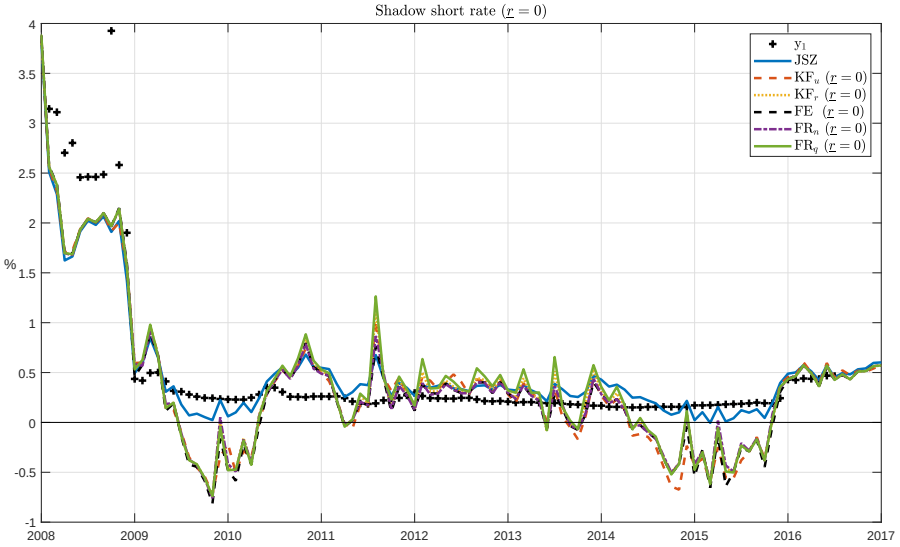
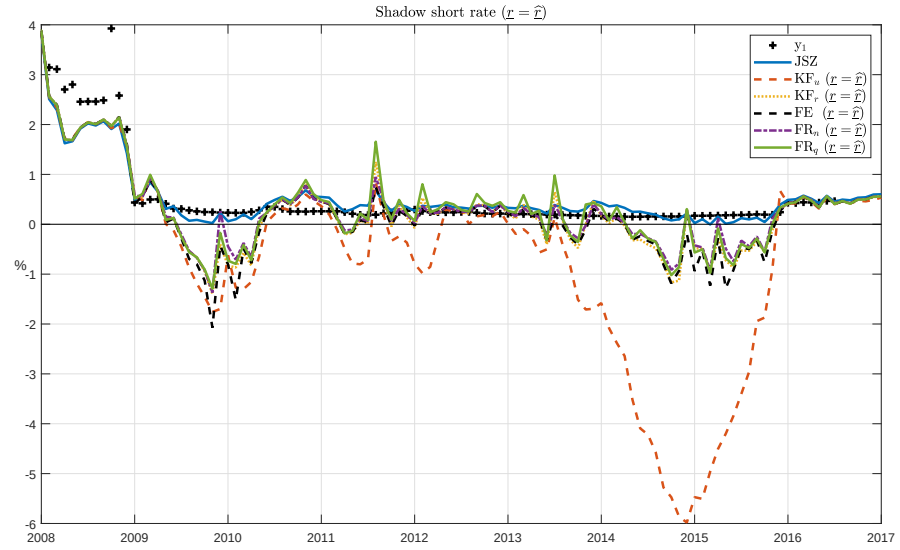


Figure 2. Fitting errors for models with $\underline{r} = \widehat{r}$. The figure shows the difference between the observed 0.25, 0.5, 1, 2, 5 and 10−year yields and their fitted counterparts for particular models. The lower bound parameter in the shadow rate models is freely estimated. The estimation sample is January 1981 to December 2016.



(a)



(b)

Figure 3. Short rate and shadow short rate implied by different models. Panel (a) presents the shadow short rates with the imposed restriction $\underline{r} = 0$, while in Panel (b) the lower bound is estimated by the shadow rate models. The observed 1−month rate and the fitted short rate from the Joslin et al. (2011) models are plotted for comparison.

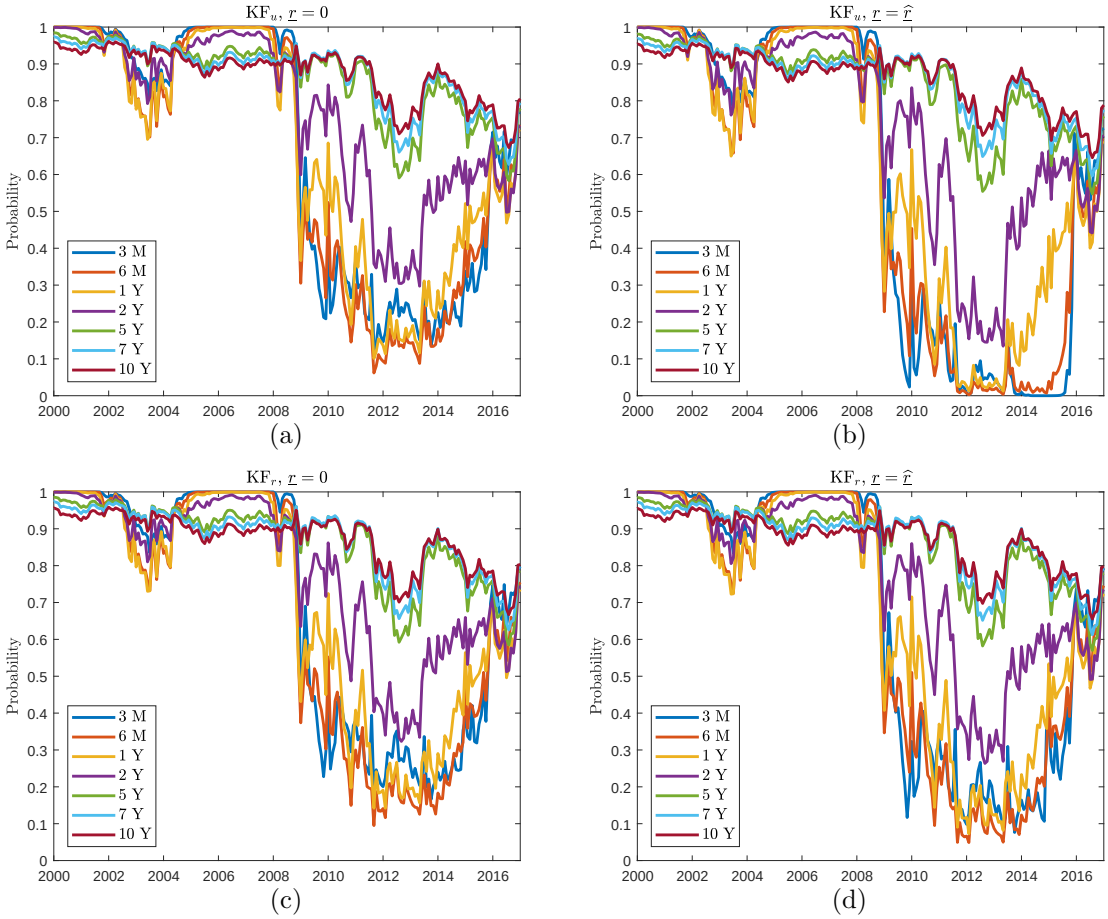


Figure 4. Cumulative density functions $\Phi(z_t)$ used in the Wu and Xia (2016) forward rate approximation for different maturities. The CDFs for the model with unrestricted measurement errors with the lower bound parameter set to zero and estimated freely are presented in Panels (a) and (b), respectively, while the CDFs for the model with imposed observability restriction and with the lower bound parameter set to zero and estimated freely are presented in Panels (c) and (d), respectively.

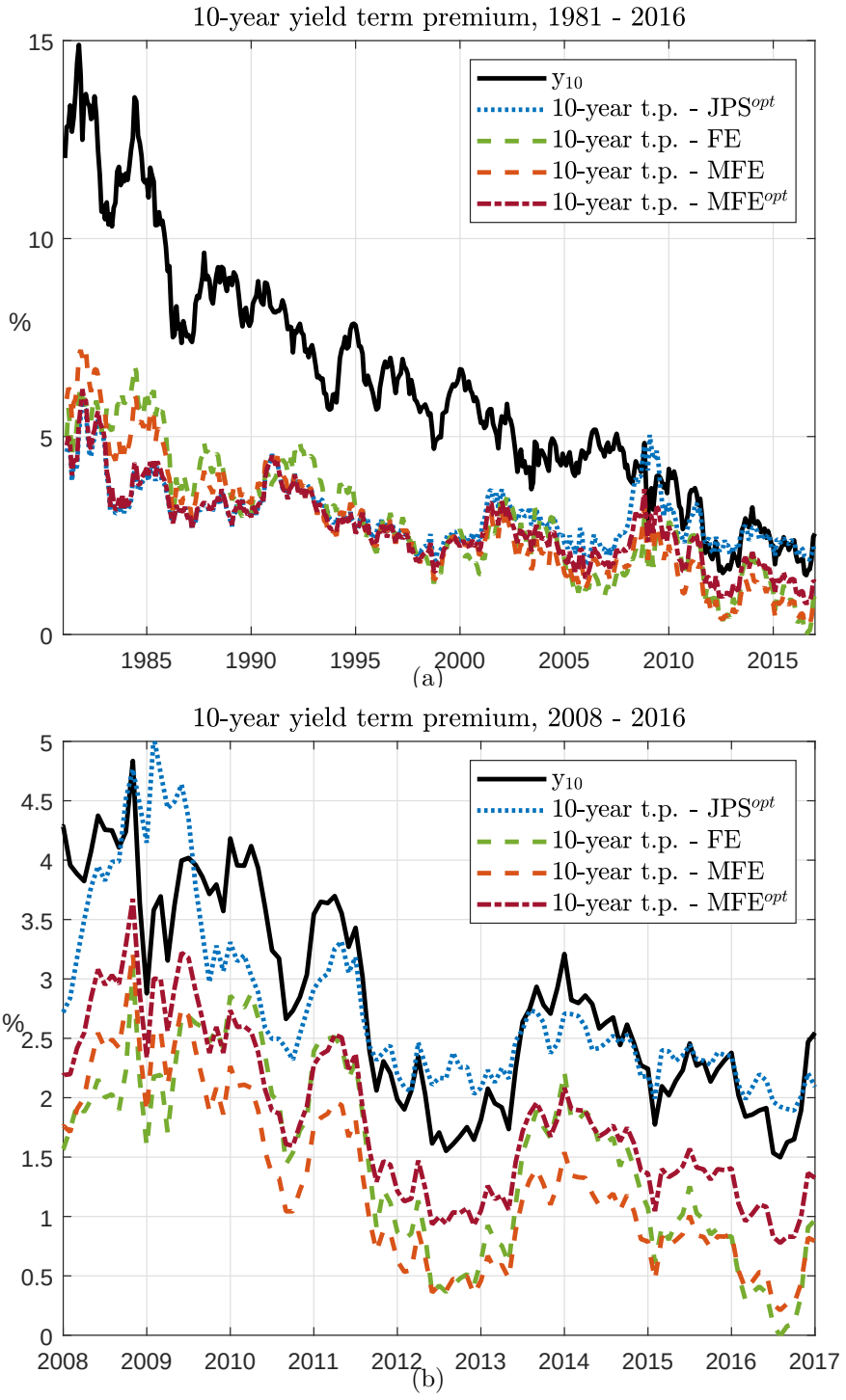


Figure 5. 10−year term premium generated by different models. The JPS^{opt} model and MFE^{opt} models are macro-finance models and they use the same set of restrictions on the price of risk and the maximum eigenvalue restriction. The FE model is a yield-only model and does not impose any restrictions. Similarly, the MFE is a shadow rate macro-finance model estimated without any restrictions. The 10−year yield is plotted as a reference. Panel (a) shows the term premium estimates for the whole sample period, January 1981 to December 2016, while Panel (b) focuses on the period from January 2008 onwards.