Shadow Rate Models and Monetary Policy

Ethan Struby
Carleton College

Michael F. Connolly
Colgate University and Boston College

August 2022

This paper represents the views of the author and does not necessarily reflect the opinion of Carleton College.

Recommended Citation
https://digitalcommons.carleton.edu/econ_repec/17
Shadow Rate Models and Monetary Policy

Michael F. Connolly and Ethan Struby *

August 1, 2022 †

Abstract

We examine the channels and efficacy of monetary policy at the zero lower bound (ZLB) through the lens of shadow rate models. We compare estimates across models with various factor structures and different assumptions about interest rate forecasts. We confirm that calendar-based forward guidance discretely shifted the implied duration of the ZLB and that large scale asset purchases (LSAPs) primarily lowered term premia. However, we find that the real effects of monetary policy are more muted relative to prior estimates: a 1 standard deviation fall in the shadow rate causes a peak decline in the unemployment rate of 0.003-0.01%.

*Connolly: Colgate University Department of Economics and Boston College Department of Economics. Struby: Carleton College Department of Economics. The authors would like to thank Leland Farmer for sharing his discretization filter code and Florian Huber for sharing his MSFVAR code. We also thank Dino Palazzo, Dongho Song, Dominique Brabant, Ryan Chahrour, and audiences at the Federal Reserve Bank of Kansas City, Macalester College, Colgate University, the Federal Reserve Bank of St. Louis Time Series Workshop and numerous conferences for helpful feedback. Alex Gallin provided excellent research assistance. The authors acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the research results reported within this paper. Any mistakes are the authors’ alone. We use data from the Blue Chip Financial Survey, which is a registered trademark of CCH Incorporated. This paper previously circulated under the title “Subjective Shadow Rate Beliefs at the Zero Lower Bound.”

†First version: April 2019
1 Introduction

Between 2008 and 2015, the Federal Reserve lowered its policy interest rate to the zero lower bound (ZLB) and employed new tools – large scale asset purchases (LSAPs) and forward guidance – to influence long-term interest rates. The consensus among policymakers is that these new tools were effective (Caldara et al. (2020)). However, the channels through which they operated are still not well understood. In particular, the nonlinearity resulting from the ZLB makes it difficult to separately identify interest-rate expectations from term premia in standard term-structure affine models. Despite the challenges of identifying how these new policies worked, the Federal Reserve employed similar strategies in response to the COVID-19 recession. Hence, understanding the effects and channels of monetary policy at the ZLB remains an important question.

Shadow rate models are a common tool for analyzing the effects of unconventional monetary policy. The shadow rate is the counterfactual one-period interest rate that would have obtained absent the ZLB (Black (1995)). Shadow rate models combine a time series process for the stochastic discount factor of financial market participants, the absence of arbitrage, and an effective lower bound on short-term interest rates. The path of the shadow rate reflects market expectations of the length of time that short-term interest rates would remain beneath its lower bound. These models have been used to summarize the stance of monetary policy and to evaluate its real effects (Wu and Xia (2016); Bauer and Rudebusch (2016)).

In this paper, we estimate a number of shadow rate models of U.S. Treasury forward rates in order to investigate the channels and efficacy of monetary policy at the ZLB. By comparing across numerous specifications of factor structure, data used in estimation, and structural assumptions about that data, we can determine which conclusions are robust and which are not. Our estimation procedure innovates in two dimensions relative to prior work. First, we use a fully non-linear estimator (the discretization filter developed in Farmer (2021)), which allows us to combine data before, during, and after the 2008-15 ZLB period in estimation.
Previous work (e.g., Wu and Xia (2016)) has relied on the extended Kalman filter, which locally linearizes to evaluate the likelihood of the data, or has estimated the parameters on pre-ZLB data alone (e.g. Bauer and Rudebusch (2016)). Using the discretization filter allows us to explore whether the choice of estimation method appreciably affects the properties of shadow rate estimates.

We extend Farmer’s filter to incorporate missing observations, which enables our second innovation: incorporating survey forecasts of average short-term interest rates into the estimated shadow rate. Although the use of survey forecasts is somewhat common in the affine term structure literature (see for example, Kim and Orphanides (2012)), to our knowledge we are the first to utilize this data in shadow rate estimation. The use of survey forecasts gives us information about the parameters governing the physical short rate process and allows the exploration of different models of forecast formation. In total, we estimate six different models: two- and three-factor models without forecasts, with “rational” forecasts, and “subjective” forecasts. Across our six models, no particular model is favored by (in- and out-of-sample) measures of fit.

With our estimates in hand, we revisit a number of important questions concerning the effects of monetary policy on financial markets and the macroeconomy during and after the Great Recession. First, we examine the shadow rates themselves and the model-implied real-time beliefs about the duration of the zero lower bound. Our models generally produce deeper estimates of the shadow rate in the ZLB period than Wu and Xia (2016) and the level of the shadow rate differs markedly across models. However, all of our estimates qualitatively agree that markets persistently expected a much shorter duration of the ZLB in the early stage of the Great Recession than occurred ex-post. Our results suggest that the introduction of calendar-based forward guidance in 2011 led to a discrete upward reassessment of ZLB duration (by 4-9 months depending on the model). Five of the six models imply the expected liftoff from the ZLB conformed to the FOMC’s calendar-based forward guidance. These findings broadly align with prior work on the expected liftoff from the ZLB (Swanson and
Williams (2014)) and suggest that calendar-based forward guidance was effective at shaping the beliefs (and behavior) of market participants.

Second, we examine the effect of monetary policy on the macroeconomy using our shadow rate estimates. A large literature has used the Wu and Xia (2016) estimates to replace the federal funds rate during the ZLB period. We estimate an identical Factor-Augmented VAR (FAVAR) as in Wu and Xia (2016) and find evidence of a structural break in either the effects of the lagged policy rate on macroeconomic variables or lagged macroeconomic variables on the policy rate for five of the models. This finding suggests caution in the use of shadow rates for applied work that spans this time period. Given the evidence of a structural break, we estimate a Markov-switching FAVAR (MSFAVAR) using our estimated shadow rates to account for potentially different regimes in the conduct and effects of monetary policy. We find that a 1 standard deviation surprise monetary easing lowered the unemployment rate by between 0.003% and 0.01% during the “unconventional” policy regime during the 2008-15 ZLB period. These estimated magnitudes are much smaller than in the existing literature (e.g. Wu and Xia (2016) and Corrado et al. (2021)) and are insignificant in some specifications.

Third, we use the estimated models to decompose the yield curve into the expectations hypothesis (EH) component of yields and term premia. We find that the first round of LSAPs largely affected long-term yields by reducing term premia, a result that is consistent with earlier findings (Gagnon et al. (2011)). The magnitude of the fall in term premia differs across models; however, the models that use survey data tend to attribute a larger portion of the change to reductions in the path of short rates. At a minimum, this result suggests a non-trivial role for unconventional monetary policy affecting yields via a signaling channel or path channel, as argued by Krishnamurthy and Vissing-Jorgensen (2011) and Bauer and Rudebusch (2016). Finally, we study the predicted effects of LSAPs on available Treasury supply and composition using the framework established in D’Amico et al. (2012). Like in that paper, we find effects largely attributable to changes in duration; however, we do not
find strong evidence supporting an economically important channel for “local scarcity” of medium-to-long term Treasuries.

The paper proceeds as follows. The next section reviews the relevant literature. Section 3 describes the shadow rate model, our information assumptions, and the estimation. Section 4 discusses the estimation results and the implied paths of the shadow rate; section 5 examines the macroeconomic and policy applications of our estimates, followed by the conclusion.

2 Literature Review

Our paper contributes to the growing literature on shadow interest rates, originated by Black (1995). Wu and Xia (2016) is the closest paper to ours; they use a non-linear approximation for forward rates (which we also adopt) for a three-variable latent factor model. They argue that their estimated shadow rate can be used to replace the effective federal funds rate in monetary VARs. Since their rate is widely used in applied work, we highlight comparisons between their results and ours throughout the rest of the paper. Bauer and Rudebusch (2016) estimate a term structure model with macroeconomic and (latent) financial factors using data from prior to the ZLB period associated with the Great Recession. They then use simulations over the ZLB period to find the modal forecast of the shadow rate. Their paper emphasizes the sensitivity of shadow rate level estimates to model specifications, a theme we explore in other dimensions. Gust et al. (2017) estimate a shadow rate in the context of a dynamic stochastic general equilibrium model. We differ from these papers by jointly explaining forecasts and forward rates, by generalizing the forecast formation process, and using a fully nonlinear estimation method.

Several papers in the affine term structure literature have also incorporated forecasts into estimation – for example, Kim and Wright (2005), Wright (2011), and Kim and Orphanides (2012). Piazzesi et al. (2015) show that risk premia constructed using survey forecasts have different time series properties than those typically calculated from market data alone. We detail the relationship of our paper to Piazzesi et al. (2015) in more detail in section 3.
Relative to these papers, we explicitly jointly account for survey forecasts and the ZLB in our framework.

A related literature connects forecast expectations to financial variables explicitly. Colacito et al. (2016) develop an equity pricing model that includes variance and skewness of professional forecasts, which they treat as exogenous. Barillas and Nimark (2018) and Struby (2018) estimate affine term structure models with dispersed information and survey forecasts. These papers do not incorporate the ZLB.

We contribute to a large literature attempting to measure the effects of Federal Reserve policy at the zero lower bound, especially forward guidance and LSAPs. Many of these papers, such as Krishnamurthy and Vissing-Jorgensen (2011) and Nakamura and Steinsson (2018b) use event studies (in part) to measure the impact of policy announcements.¹ Gagnon et al. (2011) use an event study and a reduced-form model of the term premium to measure the effects of LSAPs. Wright (2012) estimates a VAR, identifying monetary policy shocks using heteroskedasticity, as well as an event study approach, and finds that monetary stimulus at the ZLB has a short-lived effect on longer-term Treasury yields and corporate yields. Hanson and Stein (2015) find large effects of FOMC announcements over a sample that includes the ZLB period. Bauer and Rudebusch (2014) estimate a suite of affine term structure models at a daily frequency to distinguish between the effect on term premia versus changes in the path of short rates (what they label the forward guidance effect). Their interest is on characterizing how model and parameter uncertainty affects the assessment of the two channels.

We differ from most of these papers by jointly estimating the dynamics of forecasts and forward rates in a structural model. The advantage of our approach is that it allows us to not just measure the raw effect of LSAPs, but understand the changes in expectations of short rates and risk premia, assess the perceived duration of the ZLB as it evolved over time, and examine the robustness of our results to different structural assumptions. D'Amico

¹Martin and Milas (2012) and Swanson (2018) survey this literature.
and King (2012) and Li and Wei (2013) examine the effects of changes in supply from LSAPs on term premia estimated using affine term structure models; we examine whether their interpretation of supply effects is robust to term premia estimated using shadow rate models.

Lastly, we contribute to a literature examining whether unconventional monetary policy stimulated real economic activity during the ZLB. Wu and Xia (2016) estimate a FAVAR using their shadow rate as the policy rate during the ZLB period associated with the Great Recession; they find monetary policy was effective at lowering the unemployment rate during this period. Corrado et al. (2021) find a similar result using the Wu and Xia (2016) and Krippner (2015) shadow rate estimates in the context of a Markov-switching FAVAR. We revisit these results using our suite of estimates based on alternative model specifications.

3 The shadow rate model and discretization filter

This section contains the details of the shadow rate model and estimation procedure. First, we outline the shadow rate model, following Wu and Xia (2016). Following that, we explain the alternative information assumptions we use to map the forecast data into the shadow rate model. Finally, we discuss estimation.

3.1 The Wu-Xia Shadow Rate Model

Following Wu and Xia (2016), the nominal short rate $r_t$ is given by

$$r_t = \max(r, s_t)$$

(1)

The shadow rate $s_t$ is affine in the state vector $X_t$:

$$s_t = \delta_0 + \delta_1 X_t$$

(2)

The stochastic discount factor $M_{t+1}$ is exponentially affine, and is related to the prices of risk $\lambda_t$ and innovations to the state $\varepsilon_{t+1}$:
\[
\ln M_{t+1} \equiv m_{t+1} = -r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\varepsilon_{t+1}
\]

(3)

The prices of risk are themselves a linear function of the state:

\[
\lambda_t = \lambda_0 + \lambda_1 X_t
\]

(4)

Using the superscript \(Q\) to indicate the risk-neutral probability measure, the law of motion for fundamental factors under the risk neutral measure is

\[
X_{t+1} = \mu^Q + \rho^Q X_t + \Sigma^Q X_{t+1} \varepsilon_{t+1} \sim^Q N(0, I)
\]

(5)

Under the physical measure, the law of motion is:

\[
X_{t+1} = \mu + \rho X_t + \Sigma X_{t+1} \varepsilon_{t+1} \sim N(0, I)
\]

(6)

The change of measure is related to the prices of risk (the \(\lambda\) terms) and the sizes of risks that bond traders face (\(\Sigma\)) in the following way:

\[
\mu - \mu^Q = \Sigma \lambda_0
\]

(7)

\[
\rho - \rho^Q = \Sigma \lambda_1
\]

(8)

Finally, we denote the forward rate from \(t + n\) to \(t + n + 1\) as

\[
f_{n,n+1,t} = (n + 1)y_{n+1,t} - ny_{n,t}
\]

(9)

where \(y_{n,t}\) is the log yield on a zero coupon bond that pays a dollar at time \(t + n\).

Given this setup,
\[ f_{n,n+1,t}^{SR} \approx r + \sigma_Q^2 g \left( \frac{a_n + b_n X_t - r}{\sigma_n^2} \right) \]  

where

\[ g(z) = z \Phi(z) + \phi(z) \]  

\( \Phi(\cdot) \) and \( \phi(\cdot) \) are a standard normal CDF and PDF, respectively (see Wu and Xia (2016) for details). \( a_n \) and \( b_n \) are nonlinear expressions of the prices of risk and parameters governing the state, and are defined explicitly in Appendix A.

### 3.2 Incorporating forecast survey data

We deviate from Wu and Xia (2016) and other earlier shadow rate estimates by incorporating forecasts data in the estimation. We utilize the Blue Chip Financial Forecasts survey, which is a monthly publication that collects macroeconomic and financial forecasts of market participants’ beliefs over subsequent quarters.\(^2\) We use the average forecasts of the 3-month constant-maturity Treasury bill yield to construct paths of average expected short-term rates over different horizons. We identify forecasts with the average short rate implied by the physical measure over the relevant horizon. Using analogous steps in the derivation of forward rates under the risk-neutral measure, one can show that expected short rates under the physical measure are

\[ E_t[r_{t+n}] \approx r + \sigma^n_P g \left( \frac{a^n_P + b^n_P X_t - r}{\sigma^n_P} \right) \]  

where

\[ a^n_P \equiv \delta_0 + \delta_1 \left[ \sum_{j=0}^{n-1} (\rho)^j \right] \mu \]  

\(^2\)In Appendix B we show that the average forecast from the survey is consistent with short-term asset prices.
Because the Blue Chip survey asks questions about quarterly averages, the horizon over which respondents are actually forecasting varies depending on the survey month. We account for this by extracting the purely forward-looking component from current-quarter forecasts (see Appendix C). Forecasts at horizons beyond the current quarter are raw average forecasts provided by the survey. Depending on the survey month, the horizon over which the survey reflects a forecast is changing, which we adjust for in our estimation method.

**Alternative assumptions on forecasts**  The affine term structure literature has incorporated short-term interest rate forecasts using a variety of assumptions. Kim and Orphanides (2012) propose estimating affine term structure models by incorporating forecasts to reduce small-sample problems and improve precision of estimates. They treat average forecasts as generated under the physical measure and full information rational expectations (FIRE), and observed with i.i.d. measurement error. We refer to estimates made using these assumptions as “KO” estimates for brevity, in contrast to models that do not incorporate survey forecasts, which we call yields-only models (“YO”). We assume this measurement error has a constant variance across forecast horizons.

A large literature has documented that there are aspects of forecast surveys that are inconsistent with FIRE. In Appendix B, we show that forecast errors are predictable in economically and statistically significant ways. Because we are interested in whether our results are sensitive to the choice of how to add forecasts into the model, we also estimate models using a different strategy following Piazzesi et al. (2015). They use short rate forecasts to construct “subjective” interest rate expectations and risk premia using quarterly data. They estimate a statistical model of yields and expected inflation, and then estimate parameters...
governing the risk-neutral measure and a subjective “distorted” measure in a second step by minimizing mean square error between the model-implied yields and forecasts. We modify their approach along several dimensions. In addition to estimating a model with the ZLB, we focus on forecasts of three-month Treasuries at horizons of 1-18 months ahead. We also estimate the dynamics of the physical, risk neutral, and distorted measure jointly in a single (quasi-) maximum likelihood step. Like Piazzesi et al. (2015), we assume that forecasts are formed under a “distorted” physical measure as in equation (12), but with $\rho$ replaced by

$$\rho - \Sigma k$$

where $k$ is a conformable matrix of parameters that govern the degree of distortion. We also allow for $i.i.d.$ measurement error in forecasts. We refer to this set of estimates as the “PSS” model. In some figures, we compare results based on our estimates to those of Wu and Xia (2016). We label their results as “WX” for brevity.

### 3.3 Estimation details

This section contains details about the factor structure, mapping of the model into a state-space representation, and the estimation details.

#### 3.3.1 Factor normalization and structure

We estimate the parameters governing the physical dynamics of our latent risk factors ((5) and (6)) and the prices of risk parameters $\lambda_0$ and $\lambda_1$. For three-factor models, we impose a similar normalization as Joslin et al. (2011) and Wu and Xia (2016) for the risk neutral factors:

$$\delta'_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

(16)
\[ \mu^Q = 0 \] (17)

We also assume \( \rho^Q \) is in real Jordan form with eigenvalues in descending order and \( \Sigma \) is lower triangular. Unlike Wu and Xia (2016), we do not impose a repeated eigenvalue in estimation. We also estimate two-factor versions of each model with analogous restrictions following Krippner (2015).

3.3.2 The nonlinear state-space representation

Throughout, we assume that the state equation is a VAR(1). The physical dynamics of the fundamental states are as in equation (6). The observed forward rate is the same as in equation (10) augmented with measurement error:

\[
f_{n,n+1,t} = r + \sigma_n^Q g \left( \frac{a_n + b_n'X_t - \bar{r}}{\sigma_n^Q} \right) + \omega e_{nt} \tag{18}\]

where \( \omega \) is a measurement error parameter common across horizons and \( e_{nt} \sim \mathcal{N}(0, 1) \). We allow this measurement error to have a different variance than the measurement error of forecasts. Generically, we collect observables by stacking them in an observation equation \( G_t(X_t) \).

3.3.3 Details of the estimation procedure

The data runs from 1987-2018 at a monthly frequency. Forward rates are constructed using the Gürkaynak et al. (2007) yield curve estimates and averaged over the month (to align the data with the Blue Chip question wording).

We estimate the nonlinear state space model using the discretization filter proposed by Farmer (2021). Farmer’s discretization filter approximates the state distribution on a discrete grid. We use the method outlined in Gospodinov and Lkhagvasuren (2014) to approximate the state distribution, choosing grid points to approximate the first two moments of the underlying Gaussian VAR. We then use \( G_t(X_t) \) to calculate predicted values of the observable.
forward rates and forecasts at each point on the grid. Treating each point as a regime for the data, the likelihood is estimated in a method similar to the Hamilton (1989) filter. Standard errors on parameter estimates are QMLE standard errors as in Hamilton (1989). We estimate smoothed states via an appropriately modified version of Kim’s smoother (Kim (1994)).

We adopt the discretization filter, rather than using the extended Kalman filter (EKF) as in Wu and Xia (2016). Both involve approximation: the discretization filter approximates the state space on a grid, while the EKF linearizes the observation and state equations locally. Farmer (2021) shows in the context of a different model that the discretization filter can achieve lower root mean square error and bias than the EKF. Although a detailed comparison of the two methods is beyond the scope of the paper, we present results from estimating a three-factor, yields-only version of the model in Appendix G. For that model, the within-sample fits for both the EKF and the discretization filter are similar. However, the out-of-sample performance of the discretization filter models is superior.\footnote{More practically, while the EKF is certainly faster than the discretization filter for the yields-only models, we found that it was much slower for models incorporating forecasts. This made it much more computationally burdensome to conduct the pseudo-out-of-sample exercise we used to compare across models.}

4 Estimation results

In this section we discuss parameter estimates and model fit, and present the estimated shadow rates and implied ZLB duration.

4.1 Parameter estimates and model fit

We report parameter estimates for all models in Appendix D. The estimated zero lower bound varies across models between 11 and 20 basis points. This is below the 25 basis points Wu and Xia (2016) assumed, but slightly above that used in Federal Reserve Board staff estimates in 2012 (around 10 basis points).

The models fit the yield curve well in sample. Appendix Figure D.1 shows the average yield curve predicted during the ZLB period versus the data. Each model appears to capture the shifts and changes in the shape of the yield curve during the ZLB period and the
nonlinearity at the short end.

In terms of comparisons across models, the three-factor models achieve higher likelihoods than the two-factor models, and the PSS model achieves a higher likelihood than the KO model conditional on the number of factors.\(^4\) However, as shown in panel A of Table 1, the three-factor YO model is favored by in-sample measures of root mean square error (RMSE) and mean absolute error (MAE) in fitting forward rates.\(^5\)

We also examine out-of-sample fit in a pseudo-real-time forecasting exercise for the ZLB period. Starting in 2007, we estimate the model parameters using only the data available as of December of that year. We then forecast forward rates at monthly horizons 1- to 12-months ahead, and we re-estimate adding the subsequent 12 months of data. Hence, for forward rates of each maturity, we have 120 (10 estimates of 12 horizons) sets of forecasts.\(^6\) Panel B of Table 1 displays the RMSE and MAE from this exercise. As shown in that panel, the two-factor KO and three-factor PSS models perform better out-of-sample for short-maturity bonds, while the three-factor YO and KO models perform better for long-maturity bonds (although the advantages over the PSS models are not large). Because there is no clear “winner” among different measures of in- and out-of-sample fit, we base our subsequent analysis on the entire suite of models to highlight where they display consensus and disagreement for a range of policy questions.

### 4.2 Shadow rate estimates and duration of the ZLB period

Figure 1 shows our shadow rate estimates alongside the Wu and Xia (2016) estimates during the ZLB period. We also indicate a selection of policy event dates during this period. The level of our estimates is generally lower than that estimated by Wu and Xia (2016), and the PSS level estimates are generally lower than the YO and KO estimates. Our estimates

\(^4\)The PSS model with three-factors is also overwhelmingly preferred by the Bayesian information criterion among the four models that use forecast data.

\(^5\)Note that unlike the KO and PSS models, the YO model is not trying to simultaneously match both prices and forecasts.

\(^6\)Because of the computational burden of estimating the model, re-estimating for each additional month of data is infeasible.
Table 1: Table reports model fits. Columns 1-6 report estimates for the mean absolute error (MAE) and Columns 7-12 report estimates for the root-mean-square error (RMSE) across models. Panel A contains estimates for the in-sample fit that uses all observations (384 months). Panel B contains estimates for the out-of-sample fit, which estimates the model each December from 2007-2018, and calculates forecasts for 1- to 12-months ahead. MAE and RMSE are reported across all horizons (10 sets of forecasts at 12 horizons each).

<table>
<thead>
<tr>
<th>Statistic</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>YO</td>
<td>KO</td>
</tr>
<tr>
<td>Factors</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3mo</td>
<td>0.20</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.11</td>
</tr>
<tr>
<td>6mo</td>
<td>0.13</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>0.16</td>
<td>0.08</td>
</tr>
<tr>
<td>12mo</td>
<td>0.14</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.18</td>
<td>0.10</td>
</tr>
<tr>
<td>24mo</td>
<td>0.19</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.09</td>
</tr>
<tr>
<td>60mo</td>
<td>0.15</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>0.19</td>
<td>0.13</td>
</tr>
<tr>
<td>84mo</td>
<td>0.11</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>0.14</td>
<td>0.11</td>
</tr>
<tr>
<td>120mo</td>
<td>0.18</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>0.22</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Panel B: Out-of-Sample Fit: 1-12 month-ahead forecasts (N=120)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>YO</td>
<td>KO</td>
</tr>
<tr>
<td>Factors</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3mo</td>
<td>0.19</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>0.26</td>
<td>0.30</td>
</tr>
<tr>
<td>6mo</td>
<td>0.24</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>0.34</td>
<td>0.38</td>
</tr>
<tr>
<td>12mo</td>
<td>0.29</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>0.48</td>
<td>0.52</td>
</tr>
<tr>
<td>24mo</td>
<td>0.41</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>0.64</td>
<td>0.67</td>
</tr>
<tr>
<td>60mo</td>
<td>0.55</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.77</td>
</tr>
<tr>
<td>84mo</td>
<td>0.60</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>0.77</td>
<td>0.79</td>
</tr>
<tr>
<td>120mo</td>
<td>0.69</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>0.86</td>
<td>0.82</td>
</tr>
</tbody>
</table>

also appear to react more strongly to the first two rounds of LSAPs and the introduction of calendar-based forward guidance/MEP than that of Wu and Xia. Conditional on the “type” of model – YO, KO, and PSS – the number of factors also appears to influence the level of the shadow rate.

The level of the shadow rate reflects the speed with which short-term interest rates are projected to revert to their mean. Arguably, when short-term interest rates rise above their lower bound is more informative. We translate the level of the shadow rate to the implied belief about how long short rates will remain at the ZLB. This duration is shown in Figure 2. All of our models suggest that market participants initially under-predicted the (ex-post) duration of the zero lower bound. In 2009 and 2010, the implied date of liftoff is one to two
Figure 1: Smoothed estimates of shadow rate during/post Great Recession, with event dates (three rounds of Large Scale Asset Purchases (LSAPs), the introduction of calendar-based forward guidance and the Maturity Extension Program (FG+MEP), Taper Announcement). FG and MEP were introduced in August and September 2011, respectively, but are shown in August 2011.

In August 2011, the Federal Open Market Committee (FOMC) introduced specific calendar-based forward guidance in their Statement of Economic Projections (SEP). In Figure 2, the black dashed lines indicate the range of dates for target rate liftoff implied by the SEP. These dates are shown as a range because they are only reported using end-of-quarter or end-of-year estimates for the federal funds rate. The introduction of calendar-based forward guidance corresponded with a discrete upward reassessment of the expected ZLB duration.
We display this effect in Figure 3, which uses the same estimates as those in the previous figure, but focuses on the period surrounding the introduction of the guidance. Averaging across models, the point estimates suggest that calendar-based forward guidance extended the market’s estimate of the ZLB duration by about seven months.

Although the models do not agree on the precise month of liftoff even after the introduction of calendar-based forward guidance, all of the models agree that liftoff would occur sometime after mid-2013. Five of the six models suggest market participants’ beliefs were consistent with the SEP’s guidance of “exceptionally low levels for the federal funds rate at least through mid-2013,” while the three-factor KO model suggests that market participants believed that rate increases would occur later than the FOMC was projecting at the time.

In short, our results confirm that markets initially under-estimated how long interest rates would remain at their lower bound in 2008 and 2009. Moreover, calendar-based forward guidance successfully extended the market perceived duration of the ZLB, although the exact degree to which that occurred differs across models.
Figure 3: Real-time implied mean duration of 2011 using the same estimates as those in Figure 2.

5 Monetary policy during and after the Great Recession

In the previous section, we showed how beliefs about the effective liftoff of monetary policy evolved over the course of the ZLB period. In this section, we explore how novel monetary policy tools affected asset prices and the macroeconomy during this period, viewed through the lens of our estimates.

We conduct four exercises. First, we test whether there was a structural break during this period; we find evidence of a break in five of six models. This cautions against using the level of the shadow rate as a measure of the monetary policy stance during the ZLB period without accounting for structural or regime changes. In light of this finding, we proceed to estimate a Markov-switching FAVAR to assess the effects of policy shocks on macroeconomic variables.
Across all models, we find that the magnitude of the impulse responses of the inflation and unemployment rates to monetary shocks are much smaller than those estimated using the Wu and Xia (2016) shadow rate. We also generally find that the effects of monetary policy shocks are larger in the “unconventional” regime (which approximately corresponds with the ZLB period). Third, we examine the effects of policy announcements on term premia. We find substantial reductions in yields and term premia around the first two rounds of LSAPs and the introduction of calendar-based forward guidance, but not for the third round of LSAPs. While these results are broadly consistent across models, the PSS model finds more muted effects (and hence stronger effects on interest rate expectations). Fourth, we relate our term premia estimates to measures of Treasury supply and duration in the pre-ZLB period and decompose the first round of LSAPs into different supply channels. Consistent with prior studies, we find that term premia fell as aggregate duration was removed from the market. We do not find large effects from the “local supply” of Treasuries after conditioning on duration.

5.1 Structural break test

Following Wu and Xia (2016), we estimate a Factor-Augmented Vector Autoregression (FAVAR) as in Bernanke et al. (2005), where we substitute the shadow rate for the effective federal funds rate when policy rates are constrained. We use the same data and specification as Wu and Xia (2016) to estimate the FAVAR. The unrestricted model is:

\[
\begin{bmatrix}
F_t \\
S_t
\end{bmatrix} = 1_{t \leq \text{Dec 2007}} B_1(L) \begin{bmatrix}
F_{t-1} \\
S_{t-1}
\end{bmatrix}
+ 1_{t > \text{Dec 2007} \& < \text{July 2009}} B_2(L) \begin{bmatrix}
F_{t-1} \\
S_{t-1}
\end{bmatrix}
+ 1_{t \geq \text{July 2009}} B_3(L) \begin{bmatrix}
F_{t-1} \\
S_{t-1}
\end{bmatrix} + v_t
\]
Table 2: First column: p-values for test of structural break in effect of lagged shadow rate on macroeconomic factors. Second column: p-values for test of structural break in effect of lagged macroeconomic factors on shadow rate. Null is no structural break.

By construction the macroeconomic factors ($F_t$) have been purged of effects from the policy/shadow rate, and the variance-covariance matrix of the structural shocks $v_t$ is lower triangular.

The null hypothesis of the structural break test is that sub-elements of $B_1(L)$ and $B_3(L)$ are equal. The first column of Table 2 reports the p-value for the coefficients of lagged shadow rates vis-a-vis the macroeconomic factors. We reject this null at the 5% level for the KO and PSS three-factor models, and at the 10% level for the KO two-factor model. The second column reports the p-value for the coefficients of lagged macroeconomic factors on shadow rates. We reject the null for both YO models. In short, for five of the six models, we find at least some of the VAR coefficients significantly differed after July 2009, relative to before the crisis. This result cautions against assuming continuity in macro-monetary policy relationships before and after the Great Recession.\(^7\)

5.2 MSFAVAR results

Because we find evidence of a structural break in the conventional FAVAR, we use a Markov-switching FAVAR (MSFAVAR) model to examine the effectiveness of policy shocks. Our particular implementation follows Huber and Fischer (2018).\(^8\) Consistent with Huber and

---

\(^7\)This result is in contrast to Wu and Xia (2016)'s suggestion that “the continuity of our shadow rate allows researchers to update their favorite VAR during and post the ZLB period.” While it is true that their shadow rate does not appear to have a structural break during and after the ZLB period associated with the Great Recession, this appears to be more specific to their estimates.

\(^8\)Corrado et al. (2021) similarly estimate a MSFAVAR using the Wu and Xia (2016) and Krippner (2015) shadow rate estimates to study the effectiveness of monetary policy at the ZLB; we compare the results using
Fischer (2018), we allow for transition probabilities between states to be time-varying and depending on a set of observable variables. Define $Y_t$ to be a vector of macroeconomic variables. A subset of these data include all variables that are assumed to be measured without error ($Y_t^o$) (in our application, the annual CPI inflation rate, the civilian unemployment rate, and the federal funds/shadow rate). In our setting, we estimate the MSFAVAR separately using the shadow rate from each of our six models, as well as from the Wu and Xia (2016) rate. Define $F_t$ as a vector of unobserved macroeconomic variables extracted from $Y_t$ using principal component analysis. The observation equation is given by:

$$Y_t = \Lambda^F F_t + \Lambda^o Y_t^o + \epsilon_t$$  \hspace{1cm} (19)

where the upper $K \times K$ block of $\Lambda^F$ is an identity matrix, the upper $K \times N$ block of $\Lambda^o$ equals zero, and $\epsilon_t$ is an $N \times 1$ vector normally distributed errors with diagonal variance-covariance matrix $\Sigma_{\epsilon}$. Similar to Huber and Fischer (2018), we allow $Z_t = [F_t', (Y_t^o)']'$ to follow a regime-switching VAR:

$$Z_t = a_{S_t} + \sum_{i=1}^{Q} A_{i,S_t} z_{t-i} + \epsilon_t$$  \hspace{1cm} (20)

In equation 20, $S_t$ is an indicator variable equal to 0 in the “conventional” regime and 1 in the “unconventional” regime. $\epsilon_t$ is a normal distributed error term with mean zero and variance-covariance matrix $\Sigma_{\epsilon,S_t}$. The constant term $a_{S_t}$ and coefficient matrices $A_{i,S_t}$ are allowed to vary across states. We identify regimes by assuming that the constant term corresponding to the federal funds/shadow rate ($j$-th position of the vector $A_{S_t}$) is larger in the conventional relative to the unconventional state:

$$a_{j,S_t=0} > a_{j,S_t=1}.$$  \hspace{1cm} (21)

We estimate the MSFAVAR using the shadow rate from our six models from 1987-2018.
For macroeconomic data, we use a set of 98 variables from McCracken and Ng (2015), transformed to be stationary. We choose $K = 3$ macro factors and include $Q = 13$ months of lags.\footnote{For additional detail on implementing the MSFAVAR, we refer the reader to Huber and Fischer (2018) and Corrado et al. (2021).} We also estimate a version using the updated WX estimates for the 1990-2018 period. We simulate 70,000 draws, with the first 35,000 discarded as burn-in. Figure 4 displays the time-varying transition probabilities of the likelihood of switching to the unconventional state ($S_t = 1$) conditional on being in the conventional state ($S_t = 0$). We find that these probabilities increased substantially at the beginning of the ZLB in 2008 and remained elevated until approximately 2015.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{transition_probabilities.png}
\caption{MSFAVAR Transition Probabilities: This figure displays estimates of the mean probabilities of transitioning to the unconventional regime, conditional on being in the conventional regime.}
\end{figure}

Figure 5 reports impulse responses from the MSFAVAR to an unexpected one-standard deviation easing in the federal funds/shadow rate. Estimates calculated for the conventional (unconventional) regime are shown in blue (red). We highlight two primary results from this analysis. First, using the WX estimates, we find that an unexpected decrease in the shadow rate leads to higher inflation and lower unemployment rates on impact in both regimes. Second, across all of the six models that we consider, the magnitudes of the estimated elasticities are substantially smaller (or their posterior confidence intervals include zero). To provide
context, consider an unexpected one-standard deviation decline in the fed funds/shadow rate in the unconventional regime. Using the WX estimates, we find that this shock has large and persistent effects on the unemployment rate, peaking at -0.037% after 20 months. This is similar to the magnitude of the effect estimated in Wu and Xia (2016), albeit eight months after their peak. In contrast, 20 months after the initial shock, the response of the unemployment rate was significantly different from zero in only the 3-factor models. The magnitudes of these estimates were much smaller than the WX responses: -0.011% (YO), -0.003% (KO), and -0.009% (PSS). \(^{10}\) According to our estimates, while monetary policy does seem to be

\(^{10}\)In Appendix G, we show that the impulse responses obtained from a version of the YO3 model estimated
We find dramatically reduced magnitudes of these effects relative to previous studies.

5.3 Decomposition of yields around policy events

Changes to the stance of monetary policy can simultaneously affect both expected future short-term rates and expected future risk premia. We use our estimates of shadow rate paths to decompose changes in the monthly 10-year Treasury yield following select Federal Reserve unconventional policy event dates during the ZLB period. The effect of unconventional policy on term premia is debated in the literature. For example, Gagnon et al. (2011) argue that early asset purchases were consistent with a portfolio rebalancing channel through which the reduction in supply of long-duration assets reduced term premia and hence long-term yields. Swanson (2018) finds that LSAPs affected long term yields, while forward guidance affected short-term yields. But Krishnamurthy and Vissing-Jorgensen (2011) attribute some change in yields to changes in the expectations hypothesis effects and Bauer and Rudebusch (2014) attribute roughly 40-50% of the reduction in 10-year yields around LSAP events to changes in policy rate expectations.

Using our collection of models, we decompose rates on 10-year Treasuries into their expectations hypothesis and term premium components, in order to understand whether our results suggest these events primarily operated through term premia or not. Figure 6 displays the 10-year U.S. Treasury yield in black along with the EH component of yields calculated from each model. The decomposition shows that the 10-year yield experienced steep declines in the months of major early Fed policy announcements. For example, between

---

11Hanson and Stein (2015) find significant effects of changes in the two-year U.S. Treasury yield on long-term real rates in a two-day window of FOMC announcements. In contrast, Nakamura and Steinsson (2018a) provide high-frequency evidence that term premia are virtually unaffected by monetary policy shocks. Kuttner (2018) surveys the evidence on the effectiveness of unconventional policy.

12The June 2020 assessment of the Fed’s monetary-policy framework (Caldara et al. (2020)) cites effects on the term premium from LSAP1 event dates of Gagnon et al. (2011).

13The decomposition of the yields across maturities for the entire sample is available upon request.
November and December 2008 (LSAP1), the 10-year yield fell by 119 basis points. The YO and KO models, shown in the blue and red lines, attribute little of this decline to the EH component (16-30 basis points across models). In contrast, results from the PSS model, shown in the green lines, suggest that 41-46% of the fall in yields was due to changes in expected short-term rates. We find a similar ordering of results during the months of subsequent announcements (LSAP2 and FG). The third LSAP announcement had negligible effects on yields. Finally, the EH component in yields increased markedly following the announcement that the Fed would begin tapering its asset purchases, with most pronounced effects for the PSS model. Narrowing in on three major event dates, Figure 7 plots the cumulative change in the 10-year yield (black line) and EH components across models from the month before LSAPs 1 and 2 and the introduction of calendar-based forward guidance (FG). The bulk of the change in the 10-year yield following LSAP1 was due to changes in term premia. In contrast, changes in the EH component were relatively more important following LSAP2 and FG (which included the MEP announcement in September 2011). In particular, the PSS models attribute a large fraction of the change in yields to changes in expectations of future short-term rates, whereas the YO models attribute almost all of the change to term premia.\textsuperscript{14}

The PSS estimates imply, generally, a larger role for the EH component of yields (e.g., emphasizing the expected future short rates/forward guidance interpretation), similar to the results of Bauer and Rudebusch (2014). The other models imply a modest (KO) or nearly-nonexistent (YO) effect on expectations. In summary, the interpretation of how major policy announcements affected yields appears to be sensitive to both the inclusion of forecast data and to the structural assumptions about those forecasts. Given this sensitivity,\textsuperscript{14} this result is consistent with the argument in Kim and Orphanides (2012), that (in an affine term structure context) the short sample problem tends to lead to upward bias in the degree of mean reversion of short rates. This would imply that they would be relatively less important for explaining long-term yields. While a natural concern is that including forecasts “mechanically” increases the contribution of the EH component of prices, the Monte Carlo evidence in Kim and Orphanides (2012) and the variance decomposition reported in Appendix E suggest that estimates including forecasts generally imply a larger role for the EH component in general because of the information they give about parameter estimates rather than the Kalman smoother attempting to mechanically fit the forecasts.
it is unsurprising that previous efforts have found mixed evidence on the precise effects of policy announcements. However, the range of disagreement appears to be about whether term premia explain the entire cumulative change in yields versus slightly more than half.

### 5.4 Supply effects and term premia

Previous studies have found that the composition of medium- to long-term Treasury securities in the Fed’s portfolio can have sizable effects on yields.\(^{15}\) To the extent that LSAPs change that composition, the effects could operate either through changes in the EH component of yields or through the term premium. The former channel would suggest that LSAPs signal changes in expectations of the path of future short-term rates (Krishnamurthy and Vissing-Jorgensen (2011)). Alternatively, the Fed’s purchases could coincide with lower interest-rate risk through the removal of aggregate duration of Treasury securities (Gagnon et al. (2011)) or changes in the scarcity of assets with similar maturities (D’Amico et al. (2012)). The results in the previous section suggest that the majority of the change in yields during the first round of LSAPs was due to changes in term premia. Hence, in this section, we examine the channels for those changes in term premia.

To construct measures of Treasury supply, we first merge the CUSIP identifiers of all outstanding U.S. Treasury securities from the Center for Research in Securities Prices with the Fed’s weekly System Open Market Account (SOMA) holdings and Treasury buyback operations. Following D’Amico et al. (2012), we proxy for local scarcity using privately held nominal Treasuries (PHNT), the share of Treasury securities held by the private sector - outside the Federal Reserve and U.S. government. We focus on the holdings of securities with maturities ranging from 2 to 10 years as a share of total Treasury debt outstanding, due to the Fed’s concentration in purchases of these assets in 2008. To proxy for duration risk, we calculate the duration gap (DG), the difference between aggregate duration risk in the 2-10 year maturity bucket and the duration of the on-the-run 10-year Treasury bond.

\(^{15}\)D’Amico et al. (2012) and Huther et al. (2017) provide thorough historical descriptions of the Fed’s balance sheet policies.
Aggregate duration risk is the sum of modified duration weighted by PHNT for each CUSIP. In addition, we control for the slope of the term structure, proxied by the difference between the 10-year and 2-year nominal Treasury yields. The regression equation is:

\[
TP(10\text{yr})_t = \beta_0 + \beta_1 PHNT(m: 2 - 10)_t + \beta_2 DG_t + \beta_3 \text{Slope}(10-2\text{yr})_{t-1} + \epsilon_t
\]  (22)

Table 3 displays results from these regressions. In the first column, we regress 10-year U.S. Treasury yields (rather than term premia) against our local scarcity, duration, and slope proxies. The adjusted \( R^2 \) for this regression is about 55%. We then use the model-implied monthly 10-year term premium as the dependent variable in regression (22) to examine whether the impacts of policy differ across models. We find robust evidence that the duration gap and slope significant explain 10 year yields and term premia. However, the local scarcity measure is only significant in the regression models including term premia for the two KO models and the YO3 model. D’Amico et al. (2012), using weekly data over the same sample period, found point estimates of 4.34 and 123.47 for local scarcity and duration, respectively. In both cases, these variables were robustly significant in explaining 10-year term premia. Our estimated point elasticities on local supply are about half as large (for all but the KO3 specification), while the elasticity of term premia with respect to the duration gap is larger in the YO and KO models, but not for PSS. The difference in significance between our results and those in D’Amico et al. (2012) is possibly attributable to the use of weekly versus monthly data; our point estimates (while certainly in the same neighborhood as in their paper) do not point to as large effects of local supply. As emphasized by Wu and Xia (2016) and Bauer and Rudebusch (2016), the behavior of short-term rates (and hence expected rates and term premia) is quite different for affine versus nonlinear models which likely affects the results. Our suite of models suggests that the significance of local scarcity is more sensitive to the underlying estimate of term premia than had been noted previously.

Setting aside statistical significance, we use the estimated point elasticities from the
pre-2008 sample to predict the effects of changes in supply on yields and term premia. D’Amico et al. (2013) document that the first (second) round of LSAPs decreased privately held nominal treasuries by about 4.69% (6.98%) and decreased the average duration gap by about 0.12 (0.10) years. Using our estimated results from Table 3, we calculate the predicted change in 10-year Treasury yields and term premia, with results reported in Table 4. Based on these estimates, we would have predicted yields to decrease by about 37.4 basis points overall as a result of LSAP1 and 38.8 basis points as a result of LSAP2. We interpret these numbers as the predicted change in yields attributable to the supply factors in the reduced-form model. Term premia are predicted to fall between 23 and 33 basis points, depending on the model. Given that the (predicted) change in yields must be attributed to either term premia or expectations, we interpret the residual as the variation in short-rate expectations that was induced by the supply changes from the LSAP programs. These effects are largest for the PSS and YO models, but much smaller for the KO model.

We are cautious to not draw a causal interpretation from these regressions. We simply note that the extent of disagreement about the size (and sign) of short-rate expectations effects reported in Table 4 demonstrates the difficulty of precisely estimating these effects using reduced-form methods. The robust finding across our six specifications is that the change in duration is associated with economically and statistically significant changes in term premia for 10-year Treasuries, while the association with changes in local supply are only significant in some specifications; the estimated magnitude of duration effects is between about 18% larger (KO3) to more than three times as large (YO2). The lack of robust association and imprecision of the estimated magnitude of this channel of LSAPs is somewhat at odds with some predictions of preferred habitats models of bond prices discussed by D’Amico et al. (2012).
6 Conclusion

Using data on forecasts and financial prices, we estimate shadow rates, interest rate expectations, and term premia for US Treasury markets during the zero lower bound period associated with the Great Recession. We extend on previous work by fully estimating a nonlinear state space model, incorporating interest rate forecasts alongside forward rates, and allowing for deviations from full information rational expectations. While there is qualitative agreement between most of our models about the effects of forward guidance, the quantitative differences in level of the shadow rate and expected duration of the ZLB are sizeable. We test for (and find) evidence of a structural break in the impact of policy on the macroeconomy before and after the Great Recession for all but one of our models, and find that all of our estimates imply relatively small effects of changes in the measured shadow rate on unemployment and inflation relative to previous estimates. We robustly find that the majority of the effect of LSAPs was on term premia, and that term premia were mainly affected by changes in duration during the first two rounds of LSAPs.

We have not closely investigated whether innovations to the shadow rate were driven by particular factor innovations that can be linked to macroeconomic or financial developments. But our estimated factors are correlated with macroeconomic indicators such as labor market variables; like interest rate forecasts (Caldara et al. (2020)), forecasts of labor market variables were also subject to substantial over-optimism and revision during the recovery from the Great Recession. Accounting for variation in the yield curve using a macro-factor structure, would be a natural next step.
Figure 6: Decomposition of the 10-year U.S. Treasury yield (black line) during the ZLB. The Expectations Hypothesis component of the 10-year yield is shown for the YO (blue line), KO (red line), and PSS (green line) models. Results for the two-(three-)factor model are in the upper (lower) panel.
Figure 7: Decomposition of the change in the 10-year U.S. Treasury yield (black line) during specific ZLB dates. All measures are shown related to the month preceding the following events: LSAP1 (November 2008), LSAP2 (August 2010), the introduction of calendar-based forward guidance (FG) (July 2011). The change in the Expectations Hypothesis component of the 10-year yield is shown for the three-factor YO (blue line), KO (red line), and PSS (green line) models. Results for the two-(three-)factor model are given by dashed (solid) lines.
<table>
<thead>
<tr>
<th>PHNT(m:2-10)</th>
<th>Duration Gap</th>
<th>Slope (1mo lag)</th>
<th>Adjusted R²</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.51 1.37 2.09* 2.65** 3.87* 2.30 2.28</td>
<td>213.26*** 168.23*** 144.91*** 166.87*** 149.55*** 121.95*** 104.92***</td>
<td>-0.09* 0.37*** 0.51*** 0.30** 0.94*** 0.49*** 0.52***</td>
<td>0.55 0.79 0.86 0.60 0.66 0.92 0.81</td>
<td>71 71 71 71 71 71 71</td>
</tr>
<tr>
<td>(1.58) (1.16) (1.15) (1.28) (1.79) (1.46) (1.42)</td>
<td>(26.66) (18.93) (19.54) (20.70) (43.55) (19.35) (21.65)</td>
<td>(0.05) (0.04) (0.04) (0.04) (0.07) (0.04) (0.05)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PHNT(3:10-15)</th>
<th>Duration Gap</th>
<th>Slope (1mo lag)</th>
<th>Adjusted R²</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.09 2.65** 3.87* 2.30 2.28</td>
<td>121.95*** 104.92***</td>
<td>0.52***</td>
<td>0.83</td>
<td>71</td>
</tr>
<tr>
<td>(1.28) (1.79) (1.46) (1.42)</td>
<td>(19.35) (21.65)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PHNT(3:15-20)</th>
<th>Duration Gap</th>
<th>Slope (1mo lag)</th>
<th>Adjusted R²</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.28</td>
<td>104.92***</td>
<td>0.83</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>(1.42)</td>
<td>(21.65)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Supply regressions: First column: coefficients from regression of 10-year U.S. Treasury zero-coupon yield on supply factors and yield curve slope. Columns two through six: coefficients from regression of term premia on supply factors and yield curve slope. ***, **, * indicate significance at the 1, 5, and 10 percent levels, respectively calculated using the Newey-West correction for standard errors.
<table>
<thead>
<tr>
<th></th>
<th>10yr Yield</th>
<th>Term Premia on 10-year Bond</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 fac</td>
<td>3 fac</td>
</tr>
<tr>
<td></td>
<td>2 fac</td>
<td>3 fac</td>
</tr>
<tr>
<td></td>
<td>2 fac</td>
<td>3 fac</td>
</tr>
<tr>
<td></td>
<td>2 fac</td>
<td>3 fac</td>
</tr>
<tr>
<td></td>
<td>2 fac</td>
<td>3 fac</td>
</tr>
<tr>
<td></td>
<td>2 fac</td>
<td>3 fac</td>
</tr>
</tbody>
</table>

Sample: December 2002 - October 2008

<table>
<thead>
<tr>
<th>LSAP 1</th>
<th>Predicted effect from scarcity</th>
<th>-11.76</th>
<th>-6.44</th>
<th>-9.81*</th>
<th>-12.41**</th>
<th>-15.79*</th>
<th>-10.78</th>
<th>-10.68</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predicted effect from duration</td>
<td>-25.59***</td>
<td>-20.19***</td>
<td>-17.39***</td>
<td>-20.26***</td>
<td>-17.95***</td>
<td>-14.63***</td>
<td>-12.59***</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>-37.35***</td>
<td>-26.63***</td>
<td>-27.20***</td>
<td>-32.68***</td>
<td>-33.74***</td>
<td>-25.42***</td>
<td>-23.27***</td>
</tr>
<tr>
<td></td>
<td>Residual (expectations effects)</td>
<td>-10.73</td>
<td>-10.15</td>
<td>-4.67</td>
<td>-3.62</td>
<td>-11.93</td>
<td>-14.08*</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LSAP 2</th>
<th>Predicted effect from scarcity</th>
<th>-17.51</th>
<th>-9.59</th>
<th>-14.60*</th>
<th>-18.48**</th>
<th>-23.50*</th>
<th>-16.05</th>
<th>-15.89</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predicted effect from duration</td>
<td>-21.33***</td>
<td>-16.82***</td>
<td>-14.49***</td>
<td>-16.89***</td>
<td>-14.95***</td>
<td>-12.19***</td>
<td>-10.49***</td>
</tr>
<tr>
<td></td>
<td>Residual (expectations effects)</td>
<td>-12.42</td>
<td>-9.74</td>
<td>-3.47</td>
<td>-0.38</td>
<td>-10.59</td>
<td>-12.45</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Predicted effects from supply regressions: First column: predicted change (in basis points) of 10-year U.S. Treasury zero-coupon yield due to duration and scarcity effects. Columns two through six: predicted change (in basis points) of term premia due to duration and scarcity effects. ***, **, * indicate significance at the 1, 5, and 10 percent levels, respectively calculated using the Newey-West correction for standard errors.
References


A  Explicit expressions from the Wu-Xia shadow rate model

We include the complete expression for the recursions in Wu and Xia (2016). Interested readers should refer to their paper for a complete derivation.

\[ \bar{a}_n = \delta_0 + \delta_1 \sum_{k=0}^{n-1} (\rho^Q)_k \mu^Q \]  

(23)

\[ a_n = \bar{a}_n - \frac{1}{2} \sum_{j=0}^{n-1} \delta_1 \left[ (\rho^Q)^j \Sigma \Sigma^\prime ((\rho^Q)^\prime)^j \right] \delta_1^\prime \]  

(24)

\[ b_n = \delta_1 (\rho^Q)^n \]  

(25)

And

\[ E_t(s_{t+n}) = \bar{a}_n + b_n X_t \]

B  Examining beliefs during the ZLB period and the usefulness of forecast data

Hamilton (2018) argues that event-study estimates of the impact of monetary policy actions make it difficult to separately identify the pure effects of LSAPs from informational effects. For example, figure B.1 shows the yield curve on US Treasuries at the end of day on March 17, 2009 and March 19, 2009. On March 18, 2009, the FOMC announced it would be maintaining a target for the Federal Funds rate at 0-25 basis points for an “extended period” and expanded the scale of LSAPs.\(^{16}\) The shift in the long end of the yield curve conflates

---

\(^{16}\) The March 18 2009 FOMC statement included the following language: “The Committee will maintain the target range for the federal funds rate at 0 to 1/4 percent and anticipates that economic conditions are likely to warrant exceptionally low levels of the federal funds rate for an extended period. To provide
this news about short term interest rates and economic conditions which affect risk premia.

Figure B.1: US Treasury yield curve on March 17 (dashed) and March 19 (circles), 2009. Data from Gürkaynak et al. (2007).

In principle, shadow rate models allow for the separation of these forces by identifying the pure EH component of yields separately from risk premia, even when short term interest rates are stuck at or near the zero lower bound. Forecasts are potentially an additional source of information about expectations. Because the decision to use forecast data is not innocuous (see Li et al. (2017)), it is worth briefly rationalizing our approach.

First, the Blue Chip panelists are primarily private sector forecasters, and policymakers frequently make use of the Blue Chip surveys as an indicator of market expectations that
are free of effects from priced risk premia, both in public speeches (see, for example, Clarida (2019)) and internally as a benchmark (D’Amico et al. (2013), Cieslak (2018)). This is consistent with their use in other shadow rate studies; for example, Bauer and Rudebusch (2016) verify their model-based forecasts are sensible by comparing them to surveys.

Second, graphical evidence suggests that forecasts for short-term bond yields – which one might expect have relatively small, if any, risk premia – are reasonably close to what would be implied by prices. For instance, figure B.2 compares the yield on a 12-month zero coupon Treasury bond to the average expected short-term interest rate over the next 12 months. In general, the forecasts are consistent with prevailing prices.

![Figure B.2: Yields on 12 month Treasuries and average expected short rates of 12 month Treasuries from the Blue Chip Financial Survey.](image)

Third, and most importantly, our approach in this paper is neither to ignore forecasts or assume they are the same as market expectations. We estimate several models that allow for surveys to be identified with traders directly (a-la Kim and Orphanides (2012)), as well as allowing for interest rate forecasts from surveys to be related to those implied by yields but
possibly distorted (as in Piazzesi et al. (2015)). Allowing for distortion may be important given a large literature (e.g. Coibion and Gorodnichenko (2015)) which has demonstrated professional forecasts often significantly deviate from the FIRE benchmark.

To test for the existence of distorted beliefs in Blue Chip short-rate forecasts, we regress future forecast errors on revisions of the same forecast (as in Coibion and Gorodnichenko (2015)) in each month. Define $E_t[\bar{r}_{t+n-2,t+n}]$ to be the consensus forecast made in month $t$ of the average level of short rates between months $t + n - 2$ and $t + n$. Because Additionally, call $FE_t(\bar{r}_{t+n}) = \bar{r}_{t+n} - E_t[\bar{r}_{t+n-2,t+n}]$ the forecast error from month $t$ to month $t + n$ and $FR_t(\bar{r}_{t+n}) = E_t[\bar{r}_{t+n-2,t+n}] - E_{t-1}[\bar{r}_{t+n-2,t+n}]$ the forecast revision between months $t - 1$ and $t$.

We regress forecast errors across horizons $n$ on forecast revisions:

$$FE_t(\bar{r}_{t+n}) = \alpha(n) + \beta(n)FR_t(\bar{r}_{t+n}) + \epsilon_{t+n} \tag{26}$$

Under the null hypothesis of FIRE, rational expectations errors would be unpredictable ($\beta(n) = 0$) as would be efficiently incorporating all information available at $t$. However, as the results in figure B.3 suggest, such errors can be predicted using revisions to forecasts from time $t - 1$ to $t$. This effect is nearly always significant at the 95% level. The results imply that knowing forecasts for the next quarter had been revised upward by 25 basis points between the first and second month of the current quarter implies a likely underestmate of the actual average 3-month rate by around 25 basis points (despite the upward revision). This economically and statistically significant result is inconsistent with FIRE.

While we believe surveys are a source of information about the beliefs of traders, we are cognizant that there is a possible tension in (1) treating them as FIRE and (2) identifying them with traders’ beliefs. Since the literature has not reached a consensus, we examine whether our results are robust to assuming forecasts are FIRE or perhaps generated by a distorted belief about the underlying state.

17Further details of the construction of Blue Chip forecasts are provided in appendix C.
Figure B.3: Each line represents confidence intervals for coefficient estimates of forecast error on forecast revision as in equation (26), by forecast horizon and month within the quarter.

C Incorporating the Blue Chip Financial Forecasts Survey into the structural estimates

The Blue Chip Financial Forecasts survey has been conducted at a monthly frequency since 1982. Survey participants are asked for their quarterly average forecasts of a range of financial-market variables at horizons of 1- to 5-quarters ahead (6-quarters ahead beginning
Appendix – for online publication

in 1997).\textsuperscript{18} The analysis in this paper utilizes forecasts of 3-month Treasury bill constant-maturity yields, which proxies for the risk-free short-term interest rate.

The Blue Chip survey is generally published on the first day of each month. However, forecasters complete the survey over a two-day period in the prior week. We follow Cieslak (2018) and choose the “survey date” to be the earliest business day in the range of the 23rd-27th of the month for January through November and the 17th-20th for December. Yields used in estimation are selected on those dates to correspond with the forecasters’ true information set.

Current-quarter forecasts published in the second and third months of a quarter already contain past realizations of yields. To address this issue, we adjust forecasts for prior yields within a given quarter.\textsuperscript{19} Consider the case of Q1 forecasts published in February. These forecasts reflect interest rates that already occurred in January. We calculate a forward-looking forecast by subtracting the average of 3-month interest rates (taken from the Fed’s H.15 release) over the first three weeks of January. The two-month ahead forecast then equals \( E_t[\bar{r}_{t+1,t+2}] = (3 \times E_t[\bar{r}_{t,t+2}] - \bar{r}_t)/2 \). Now consider the case of Q1 forecasts published in March, which are made in February. These forecasts reflect interest rates that already occurred in January and the first three weeks of February. The one-month ahead forecast subtracts the monthly average of yields in January and the average of the first three weeks of February: \( E_t[\bar{r}_{t+1}] = 3 \times E_t[\bar{r}_{t-1,t+1}] - \bar{r}_t - \bar{r}_{t-1} \). In both cases, the average of the first three weeks of the month is assumed to be approximately equal to the monthly average.

\section{D Parameter estimates and model fit}

This appendix first presents tables of the parameter estimates for each model. Following that, figure D.1 plots the average fit for the yield curve during the zero lower bound period. The top figure shows results for two-factor models, while the bottom shows results for three-

\textsuperscript{18}The Blue Chip also publishes long-horizon forecasts semi-annually, which we do not utilize due to the sparse time series.

\textsuperscript{19}This procedure is identical to Xu (2019), except that we use a slightly different forecast horizon convention.
<table>
<thead>
<tr>
<th></th>
<th>1200µ</th>
<th>1200ρ</th>
<th>1200Σ</th>
<th>1200r</th>
<th>1200δ₀</th>
<th>1200 (yield meas. err)</th>
<th>Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.1720</td>
<td>-0.1515</td>
<td>-0.0000</td>
<td>0.1718</td>
<td>13.5222</td>
<td>0.1276</td>
<td>19798.5126</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0005)</td>
<td>(0.0002)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0001)</td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>0.9748</td>
<td>-0.0077</td>
<td>0.4629</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.0123</td>
<td>0.9525</td>
<td>0.5926</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.0123</td>
<td>0.9525</td>
<td>0.5926</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9989</td>
<td>0.9441</td>
<td>0.9754</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td></td>
</tr>
</tbody>
</table>

Table D.1: Estimated parameters for 3 factor model without forecasts (YO model). QMLE standard errors in parentheses

Figure models. The average yield curve implied by the model is shown by the individual markers, while the actual yield curve is plotted as a solid line.
Table D.2: Estimated parameters for 2 factor model without forecasts (YO model). QMLE standard errors in parentheses.
Table D.3: Estimated parameters for 3 factor model including forecasts (KO model). QMLE standard errors in parentheses

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate 1</th>
<th>Estimate 2</th>
<th>Estimate 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1200μ</td>
<td>-0.1310</td>
<td>-0.0550</td>
<td>-0.0159</td>
</tr>
<tr>
<td></td>
<td>(0.0135)</td>
<td>(0.0012)</td>
<td>(0.0059)</td>
</tr>
<tr>
<td>ρ</td>
<td>0.9731</td>
<td>0.0248</td>
<td>-0.2807</td>
</tr>
<tr>
<td></td>
<td>(0.0015)</td>
<td>(0.0384)</td>
<td>(0.2537)</td>
</tr>
<tr>
<td></td>
<td>-0.0055</td>
<td>0.9762</td>
<td>0.1227</td>
</tr>
<tr>
<td></td>
<td>(0.0017)</td>
<td>(0.0055)</td>
<td>(0.0402)</td>
</tr>
<tr>
<td></td>
<td>-0.0022</td>
<td>0.0026</td>
<td>0.8548</td>
</tr>
<tr>
<td></td>
<td>(0.0018)</td>
<td>(0.0086)</td>
<td>(0.1239)</td>
</tr>
<tr>
<td>diag(ρ^Q)</td>
<td>0.9981</td>
<td>0.9568</td>
<td>0.9185</td>
</tr>
<tr>
<td></td>
<td>(0.0009)</td>
<td>(0.0006)</td>
<td>(0.0256)</td>
</tr>
<tr>
<td>1200Σ</td>
<td>0.2615</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0355)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.0999</td>
<td>0.2555</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0225)</td>
<td>(0.0719)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.0126</td>
<td>0.0065</td>
<td>0.0269</td>
</tr>
<tr>
<td></td>
<td>(0.0136)</td>
<td>(0.0076)</td>
<td>(0.0088)</td>
</tr>
<tr>
<td>1200 Σ</td>
<td>0.1132</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0170)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1200 δ₀</td>
<td>10.4698</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.4299)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1200 (yield meas. err)</td>
<td>0.4017</td>
<td>(0.0246)</td>
<td></td>
</tr>
<tr>
<td>1200 (fcst meas. err)</td>
<td>0.0923</td>
<td>(0.0155)</td>
<td></td>
</tr>
</tbody>
</table>

Log Likelihood: 34368.0697
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate 1</th>
<th>Estimate 2</th>
<th>(Standard Error 1)</th>
<th>(Standard Error 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1200μ</td>
<td>-0.2503</td>
<td>-0.0925</td>
<td>(0.0003)</td>
<td>(0.0031)</td>
</tr>
<tr>
<td>ρ</td>
<td>0.9597</td>
<td>0.0077</td>
<td>(0.0006)</td>
<td>(0.0013)</td>
</tr>
<tr>
<td></td>
<td>0.0050</td>
<td>0.9576</td>
<td>(0.0005)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>diag(ρ^2)</td>
<td>0.9986</td>
<td>0.9808</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>1200Σ</td>
<td>0.4023</td>
<td></td>
<td>(0.0078)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.2303</td>
<td>0.4366</td>
<td>(0.0051)</td>
<td>(0.0007)</td>
</tr>
<tr>
<td>1200 r</td>
<td>0.1235</td>
<td></td>
<td>(0.0036)</td>
<td></td>
</tr>
<tr>
<td>1200 δ₀</td>
<td>12.6821</td>
<td></td>
<td>(0.0686)</td>
<td></td>
</tr>
<tr>
<td>1200 (yield meas. err)</td>
<td>0.3470</td>
<td></td>
<td>(0.0185)</td>
<td></td>
</tr>
<tr>
<td>1200 (fcst meas. err)</td>
<td>0.2225</td>
<td></td>
<td>(0.0033)</td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>33221.2481</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table D.4: Estimated parameters for 2 factor model including forecasts (KO model). QMLE standard errors in parentheses.
Table D.5: Estimated parameters for 3 factor model with distorted forecaster dynamics (PSS model). QMLE standard errors in parentheses

\[
\begin{array}{ccc}
1200 \mu & -0.2376 & -0.0967 & 0.0196 \\
& (0.0026) & (0.0024) & (0.0005) \\
\rho & 0.9721 & -0.0050 & 0.2241 \\
& (0.0002) & (0.0169) & (0.1309) \\
& -0.0131 & 0.9453 & 1.0963 \\
& (0.0040) & (0.0020) & (0.0190) \\
& 0.0021 & 0.0034 & 0.8501 \\
& (0.0004) & (0.0011) & (0.0176) \\
\text{diag}(\rho^Q) & 0.9976 & 0.9565 & 0.9153 \\
& (0.0001) & (0.0078) & (0.0191) \\
1200 \Sigma & 0.3809 \\
& (0.0163) \\
& 0.1763 & 0.3851 \\
& (0.0095) & (0.1358) \\
& -0.0066 & 0.0037 & 0.0337 \\
& (0.0001) & (0.0009) & (0.0029) \\
1200 r & 0.1206 \\
& (0.0025) \\
1200 \delta_0 & 13.8656 \\
& (0.1626) \\
k & -12.1696 & -1.0860 & -99.9713 \\
& (1.8598) & (1.7633) & (0.3307) \\
& -17.5792 & -9.0648 & 60.0199 \\
& (2.9041) & (1.4347) & (0.0059) \\
& 39.5478 & 99.9749 & 60.0199 \\
& (3.8120) & (0.2874) & (0.0059) \\
\end{array}
\]

1200 (yield meas. err) 0.2416 \\
& (0.0146) \\
1200 (fcast meas. err) 0.1236 \\
& (0.0195) \\
Log Likelihood 35301.4976

Table D.6: Subjective physical dynamics, 3 factor PSS model

\[
eig(\rho - \Sigma k) \approx 7.6162 + 9.3590i \\
eig(\rho - \Sigma k) \approx 7.6162 - 9.3590i \\
eig(\rho - \Sigma k) \approx -7.0499
\]

Table D.6: Subjective physical dynamics, 3 factor PSS model
Table D.7: Estimated parameters for 2 factor model with distorted forecaster dynamics (PSS model). QMLE standard errors in parentheses

Table D.8: Subjective physical dynamics, 2 factor PSS model
Figure D.1: Average fit of 2-figure (top) and 3-figure (bottom) model across years, using smoothed state estimates. Line indicates average yield curve in the indicated year.
E  Variance decomposition of yields

In this appendix, we report the variance decomposition of yields across models and horizons. All of the models imply that the variation in short-maturity yields is due to expectations of future short-term rates. However, the YO and two-factor KO models attribute relatively more of the variation in medium- and long-term yields to the term premium component than the other models. The PSS models and three-factor KO model cannot distinguish between these two components for long-maturity bonds. The difference in results across models emphasizes that these decompositions are sensitive to the underlying structural model and the presence of survey forecasts in the estimation.
Figure E.1: Decomposition of unconditional variance of the level of yields into Term Premium and Expectations Hypothesis components for yields only, KO, and PSS models for the full sample (1987-2018). Newey-West 99% confidence bands are shown for point estimates. Maturity is reported in months.
Figure E.2: Decomposition of unconditional variance of the change in yields into Risk Premium and Expectations Hypothesis components for yields only, KO, and PSS models for the ZLB sample (1987-2018). Newey-West 99% confidence bands are shown for point estimates. Maturity is reported in months.
F Additional ZLB duration figures

This appendix displays two additional figures on the expected duration of the ZLB after the Great Recession. Figure F.1 focuses on the post-2011 period and groups together the real-time expected mean duration across models, separated by the number of factors. Figure F.2 compares the three-factor model point estimates of expected duration to those implied by Wu and Xia (2016).
Figure F.1: Real-time implied mean duration of ZLB period. Dashed lines indicate the duration implied by the SEP liftoff dates.
Figure F.2: Real-time implied mean duration of ZLB period for the three-factor model. Dashed lines indicate the duration implied by the SEP liftoff dates. White squares are the Wu and Xia estimates.
In this section, we include comparisons between our results and ones estimated with the extended Kalman filter. We focus on a three-factor, yields-only model. Unlike the main results in the paper, but like Wu and Xia (2016), we use annualized yields data for the EKF exercises (as opposed to estimating on un-annualized data and then annualizing after the fact).

We compare four different sets of results. First, we employ the same procedure as is used for the discretization filter – global minimization without constraining the lower bound of the short rate $r$. We then restrict $r = 0.25$ as in Wu and Xia (2016), but use global search and the complete dataset as in the previous case. We then use the same local search and the same data as Wu and Xia (2016), but using smoothed state estimates. Finally, we include the results obtained from Wu and Xia (2016)’s code. Average fits for each model during the ZLB period (within the sample used to estimate) are shown in figure G.1. Within the period 2009-2012 (when all four sets of models are on equal footing), no model is clearly superior.

The estimated shadow rates (along with the Wu and Xia (2016) shadow rate) are shown in figure G.2. Here, global search with an unrestricted ZLB yields nonsensical estimates for the shadow rate (that are positive when the rate was constrained). Calibrating the lower bound as in Wu and Xia (2016) gives an estimated path more similar to theirs. The differences between the green and black shadow rates are attributable to differences between smoothed and filtered estimates and the computing environment. Turning to the duration plot (figure G.3), we see that both the sets of global search results give extremely long implied horizons for the ZLB, while the local search is slightly more consistent with Wu and Xia (2016). Notably, neither set of 'local' results is completely consistent with the calendar-based forward guidance provided by the FOMC, although the WX results are closer.

For the remainder of the comparison, we focus on comparing our main results with EKF results estimated using global methods and a fixed lower bound at $r = .25$. We view
Figure G.1: Predicted and actual forward curves for 3 factor YO model estimated with extended Kalman filter. Blue line: Gürkaynak et al. (2007) forward rate curves. Square: estimated with global search and without fixing $\bar{r}$. Triangle: Global search with $\bar{r} = 0.25$. Circle: local search with $\bar{r} = 0.25$. Plus sign: Results from Wu and Xia (2016).
Figure G.2: Estimated shadow rates from 3 factor YO model estimated with extended Kalman filter. Blue: estimated with global search and without fixing $r$. Red: Global search with $r = 0.25$. Green: local search with $r = 0.25$. Black: Results from Wu and Xia (2016).
Figure G.3: Implied duration of ZLB from 3 factor YO model estimated with extended Kalman filter. Blue circle: estimated with global search and without fixing $r$. Red triangle: Global search with $r = 0.25$. Green square: local search with $r = 0.25$. Black plus sign: Results from Wu and Xia (2016) based on simulation. The black dashed corridor is the implied range of liftoff dates based on the FOMC SEP as described in the main text.
Table G.1: Table reports model fits for models estimated using the Extended Kalman Filter as described in the text. The first column reports mean absolute error (MAE) while the second column reports RMSE. Panel A contains estimates for the in-sample fit that uses all observations (384 months). Panel B contains estimates for the out-of-sample fit, which estimates the model each December from 2007-2018, and calculates forecasts for 1- to 12-months ahead. MAE and RMSE are reported across all horizons (10 sets of forecasts at 12 horizons each).

<table>
<thead>
<tr>
<th>Statistic</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 month</td>
<td>0.064</td>
<td>0.079</td>
</tr>
<tr>
<td>6 month</td>
<td>0.051</td>
<td>0.071</td>
</tr>
<tr>
<td>12 month</td>
<td>0.059</td>
<td>0.074</td>
</tr>
<tr>
<td>24 month</td>
<td>0.040</td>
<td>0.054</td>
</tr>
<tr>
<td>60 month</td>
<td>0.084</td>
<td>0.101</td>
</tr>
<tr>
<td>84 month</td>
<td>0.055</td>
<td>0.070</td>
</tr>
<tr>
<td>120 month</td>
<td>0.085</td>
<td>0.104</td>
</tr>
</tbody>
</table>

Panel A: In-Sample Fit (N=384)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 month</td>
<td>0.507</td>
<td>0.899</td>
</tr>
<tr>
<td>6 month</td>
<td>0.557</td>
<td>0.936</td>
</tr>
<tr>
<td>12 month</td>
<td>0.671</td>
<td>1.018</td>
</tr>
<tr>
<td>24 month</td>
<td>0.951</td>
<td>1.228</td>
</tr>
<tr>
<td>60 month</td>
<td>1.213</td>
<td>1.492</td>
</tr>
<tr>
<td>84 month</td>
<td>1.149</td>
<td>1.385</td>
</tr>
<tr>
<td>120 month</td>
<td>0.929</td>
<td>1.128</td>
</tr>
</tbody>
</table>

Panel B: Out-of-Sample Fit: 1-12 month-ahead forecasts (N=120)

This as being a more reasonable comparison to our main results because it does not imply counterfactual short rates (unlike the case where $r$ is estimated). We report the in-sample fitting error in panel A of table G.1. Compared to the results in panel A of table 1, the fitting errors are marginally smaller (within two or three basis points except for 10 years, which have a 5 bp error). This suggests that the in-sample fit of the DF does not suffer much due to approximating on a grid.

Next we turn to the pseudo-out-of-sample fit. Here, the EKF model does significantly worse than the YO model estimated with the discretization filter, with forecast errors about twice as large on average. The relatively poor performance is somewhat puzzling. Two possibilities are that the EKF is over-fitting in the rolling sample, or that the (approximate)
likelihood surface is flat (so that parameters are not well identified, and the estimation procedure happens to land on a set of parameters with particularly bad forecasting performance).

Finally, for the sake of comparison, we report the MSFAVAR estimates for the restricted YO3 model. Figure G.4 reports impulse responses for both models with estimates from the conventional (unconventional) regime shown in blue (red). For both models an unexpected decrease in the fed funds/shadow rate leads to higher inflation and lower unemployment rates on impact. However, in the unconventional state the inflation rate declines after two years. In general, the magnitude of the impulse responses for the YO model are larger in the unconventional state than those of the WX model, although both sets of responses are within the other model’s confidence bands.
Figure G.4: MSFAVAR IRFs: This figure displays impulse responses of the federal funds/shadow rate (upper panel), the annual CPI inflation rate (middle panel), and the unemployment rate (lower panel) to a one-standard deviation monetary shock. Estimates calculated for the conventional (unconventional) regime are shown in blue (red). The units of the horizontal axis are the number of months from the initial monetary shock. For each model, the solid line displays the median estimate and the shaded bands indicate 68% confidence intervals. Estimates for the WX and YO3 (Extended Kalman Filter) models are shown in columns 1 and 2, respectively.