Shadow Rate Models and Monetary Policy

Michael F. Connolly and Ethan Struby *

December 2024

*Connolly: Fannie Mae. Struby: Carleton College Department of Economics and Boston College Department of Economics. The views expressed are the authors' alone and not those of Fannie Mae or the Federal Housing Finance Agency. Michael F. Connolly's contributions to this article were conducted at Colgate University prior to his current employment at Fannie Mae. The authors would like to thank Leland Farmer for sharing his discretization filter code and Florian Huber for sharing his MSFAVAR 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."

Abstract

We examine the channels and efficacy of monetary policy at the zero lower bound (ZLB) through the lens of various shadow rate models. Our key methodological contribution is to extend the discretization filter to incorporate missing observations. This allows us to estimate shadow rate models that both incorporate survey forecasts and allow for departures from rational expectations. Although the models disagree about the level of the shadow rate and the duration of the ZLB in real time, they are consistent in attributing most of the effects of major Federal Reserve policy announcements to changes in term premia. We estimate small macroeconomic effects of shocks to the shadow rate relative to prior estimates, but this is due to differences in the sample, not the shadow rate estimation.

Keywords: Shadow rate, term structure models, term premia, forward guidance, nonlinear estimation

JEL: G12, E43, E44, E52, E58

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.

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 (2014)).

In this paper, we estimate different models using the discretization filter (Farmer (2021)): with two and three latent factors, with "rational" survey forecasts, and with "subjective" survey forecasts. Our set of models, while not exhaustive, encompasses a broad range of specifications commonly used in the shadow rate and affine term structure literature. We find that models estimated with the discretization filter tend to be competitive in forecasting forward rates relative to a yields-only model estimated with the extended Kalman filter as in Wu and Xia (2016), and outperform it during and after the COVID-19 recession. We also find that three factor models have superior ability to match the data (relative to two factor models) particularly after 2007, although they (surprisingly) have lower likelihoods for the whole sample. Given that our interest is understanding policy during the ZLB and afterwards, we subsequently mainly focus on the policy implications of our three-factor specifications.

Using our estimates, we re-evaluate a number of questions concerning the effects of monetary policy on financial markets and the macroeconomy during and after the Great Recession. First, we examine the model-implied real-time beliefs about the duration of the zero lower bound across our models. 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 small upward reassessment of ZLB duration, although there is some upward drift in duration expectations prior to the announcement. Models allowing for distorted forecasts imply market-expected durations that are longer than forward guidance suggested, while the other specifications usually align with the Fed's Summary of Economics Projections (SEP). These findings are consistent with prior work on the expected liftoff from the ZLB (Swanson and Williams (2014)) and suggest that calendar-based forward guidance was at least somewhat 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, replacing their shadow rate estimates with ours. We find similar impulse responses to monetary policy shocks during the Great Recession. However, we also find some 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 some specifications, contra Wu and Xia. We reject the null of no structural break for nearly all but one model when we extend the shadow rate estimation to 2023, even maintaining all the other aspects of the underlying FAVAR. We also find a substantial differences when we extend the macro dataset to just prior to the pandemic. Particularly, we find a 25 basis point decrease in the shadow rate decreases unemployment by about

0.03% at its peak, and the effects become insignificant in less than half a year. These estimated magnitudes are much smaller than those found in the existing literature (e.g. Wu and Xia (2016) and Corrado et al. (2021)). We argue that the differences in estimated impulse responses are mainly driven by differences in the macroeconomic sample, rather than the differences in shadow rate estimates.

Finally, we explore less structural applications of the estimated models for the assessment of policy. 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, but it explains at least 80% of the decline in yields. Finally, we study the predicted effects of LSAPs on available Treasury supply and composition following D'Amico et al. (2012). We find effects of LSAPs on long-term bond yields largely attributable to changes in duration and changes in "local scarcity" of long-term Treasuries. Although numerical details differ across models, they qualitatively agree that these two factors explain most of the supply-induced change in risk premia around the first two rounds of LSAPs.

Relative to existing work on shadow rates and the macroeconomy, we innovate, and synthesize, along a number of dimensions. Methodologically, we incorporate survey forecasts of average short-term interest rates into the data used in estimation. Survey forecasts of interest rates are commonly used to estimate affine term structure models to improve the precision of estimates of the short-rate process (Kim and Orphanides (2012)), which in turn materially impacts the properties of estimated term premia (Li et al. (2017)). In the context of shadow rate models, having precise estimates of the short rate process is especially important because the inferred dynamics of shadow rates depend on the estimated degree of mean reversion in short rates. Priebsch (2013, 2017) and Kim and Priebsch (2020) also incorporate surveys into shadow rates estimation as rational expectations forecasts of short rates. We go beyond this restriction by exploring an alternative model of forecast formation following Piazzesi et al. (2015). They estimate a term structure model which does not impose that survey forecasts are the same as the forecast of the marginal bond trader. We extend their work to incorporate the effective lower bound. We find in general that incorporating survey data improves the ability of the model to forecast yields, with Piazzesi et al. (2015) models having an advantage at the short end and rational survey forecasts having superior performance for bonds maturing in a year or more. Relative to the set of papers that incorporates "rational" forecasts (Priebsch (2013, 2017); Kim and Priebsch (2020)) we also explore the implications of short rates and term premia estimated using these models. Since much of the literature following Wu and Xia (2016) has focused on using their shadow rate series in other applications, it is particularly important to understand whether our assessment of monetary policy holds up to alternative data sets and structural assumptions.

We obtain our shadow rate estimates using the discretization filter (Farmer (2017, 2021)). Farmer (2017) applied the discretization filter to the estimation of a three factor yields-only shadow rate model and compared his estimated shadow rate series to Wu and Xia (2016). Relative to Farmer (2017), we consider a larger range of models and data (both after the liftoff from the ZLB and using forecasts) and explore the implications of our range of models to quantify the impact of monetary policy at the zero lower bound during the Great Recession and the COVID-19 period. Additionally, we extend the filter to incorporate missing data, in order to account for the different horizons of forecasts available in different periods.

The most widely used estimate of the shadow rate for applied work, Wu and Xia (2016), applies the extended Kalman filter (EKF). The EKF uses local linearization of the observation equation to evaluate the likelihood (Durbin and Koopman (2012).) Others (e.g. Bauer and Rudebusch (2016)) estimate parameters on pre-ZLB data alone. The discretization filter, by contrast, does not rely on linearization and its computational efficiency allows for comparison of a wide selection of models. We find that shadow rates estimated using this method tend to have competitive forecasting performance relative to those using the EKF, with mild costs of in-sample fit. Models estimated using the discretization filter and including forecasts also perform better during the COVID-19 recession and beyond.

Lastly, we examine whether the number of factors matters for our conclusions. Since at least Litterman (1991), the consensus has been that at least three factors are necessary to capture the statistical properties of the Treasury yield curve. Krippner (2015b) argues that two factor specifications are more suited for monitoring and summarizing the stance of monetary policy at the effective lower bound. Two-factor estimates are often used as an alternative benchmark in applied work (e.g., Carriero et al. (2021), Corrado et al. (2021), De Rezende and Ristiniemi (2023), King (2019), Johannsen and Mertens (2021), Nyholm (2021)). Finally, Krippner (2015b) advocates the use of two-factor shadow rate specifications rather than the three factors commonly used in the affine and shadow rate literature. The focus in that paper is mainly on monitoring and summarizing the stance of monetary policy. In our estimates, two factor models surprisingly achieve a higher likelihood, but three factor models outperform them in out-of-sample exercises and when fitting the *average* forward rate curve from 2007 onward. Given their superior performance in the main time period of interest, we focus mainly on the implications of three-factor models. Importantly, many of the qualitative and quantitative implications for policy are insensitive to the choice of the number of factors used to estimate the models.

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. We subsequently discuss results over the COVID-19 associated recession and its aftermath, and sensitivity to choice of yield curve estimates followed by the conclusion.

2 Literature Review

Our paper contributes to the growing literature on shadow interest rates in arbitrage-free models, originated by Black (1995). The most closely related papers are Wu and Xia (2016), Priebsch (2013, 2017), Kim and Priebsch (2020), and Farmer (2017). We adopt the Wu and Xia (2016) approximation for short rates at the zero lower bound, and like them we investigate the properties of estimated shadow rates in a FAVAR, but we find very different results on both the structural properties of these estimates and the estimated impacts of shadow rate innovations on macroeconomic variables. Farmer (2017) illustrated that the estimated parameters of a Wu and Xia (2016)-style three-factor model differ when obtained using the discretization filter. We extend the discretization filter to incorporate forecasts, estimate a wider range of models, and explore their policy implications. We also examine the out-of-sample performance of the different shadow rate models. Priebsch (2013, 2017) and Kim and Priebsch (2020) incorporate forecasts, but mainly focus on the implications for the yield curve. They also assume that forecasts are formed via rational expectations, an assumption that we relax in this paper. Our results complement theirs by showing that including forecasts can improve the out-of-sample forecasting performance of these models in a pseudo-real-time environment. We also contribute to prior work by investigating the implications of our estimates for macroeconomic variables, as well as for the assessment of LSAPs (D'Amico et al. (2013)) and structural VARs.

Our paper is related to two other papers that obtain shadow rate estimates. 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 and time to liftoff. 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 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

[1] Martin and Milas (2012) and Swanson (2018) survey this literature.

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 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 (2015b) 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

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

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

$$s_t = \delta_0 + \delta_1 X_t \tag{2}$$

The stochastic discount factor M_{t+1} is exponentially affine, and is related to the prices of risk λ_t and innovations to the state ε_{t+1}:

$$\ln M_{t+1} \equiv m_{t+1} = -r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\varepsilon_{t+1} \tag{3}$$

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

$$\lambda_t = \lambda_0 + \lambda_1 X_t \tag{4}$$

Using the superscript $\mathbb{Q}$ to indicate the risk-neutral probability measure, the law of motion for fundamental factors under the risk neutral measure is

$$X_{t+1} = \mu^{\mathbb{Q}} + \rho^{\mathbb{Q}} X_t + \Sigma \varepsilon_{t+1}^{\mathbb{Q}} \quad \varepsilon_{t+1}^{\mathbb{Q}} \sim^{\mathbb{Q}} N(0, I) \tag{5}$$

Under the physical measure, the law of motion is:

$$X_{t+1} = \mu + \rho X_t + \Sigma \varepsilon_{t+1} \quad \varepsilon_{t+1} \sim N(0, I) \tag{6}$$

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

$$\mu - \mu^{\mathbb{Q}} = \Sigma \lambda_0 \tag{7}$$

$$\rho - \rho^{\mathbb{Q}} = \Sigma \lambda_1 \tag{8}$$

Finally, we denote the forward rate from $t + n$ to $t + n + 1$ as

$$f_{n,n+1,t} = (n + 1) y_{n+1,t} - n y_{n,t} \tag{9}$$

where $y_{n,t}$ is the log yield on a zero coupon bond bond that pays a dollar at time $t + n$. Given this setup,

$$f^{SR}_{n,n+1,t} \approx \underline{\mathrm{r}} + \sigma^{\mathbb{Q}}_n g\left(\frac{a_n + b_n X_t - \underline{\mathrm{r}}}{\sigma^{\mathbb{Q}}_n}\right) \tag{10}$$

where

$$g(z) = z\Phi(z) + \phi(z) \tag{11}$$

$\Phi(\cdot)$ 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

[2]In Appendix B we show that the average forecast from the survey is consistent with short-term yields.

are

$$E_t\left[r_{t+n}\right] \approx \underline{\mathrm{r}} + \sigma_n^{\mathbb{P}} g\left(\frac{a_n^{\mathbb{P}} + b_n^{\mathbb{P}} X_t - \underline{\mathrm{r}}}{\sigma_n^{\mathbb{P}}}\right) \tag{12}$$

where

$$a_n^{\mathbb{P}} \equiv \delta_0 + \delta_1\left[\sum_{j=0}^{n-1}(\rho)^j\right]\mu \tag{13}$$

$$b_n^{\mathbb{P}} \equiv \delta_1(\rho)^n \tag{14}$$

$$(\sigma_t^{\mathbb{P}}) \equiv \mathrm{Var}_t^{\mathbb{P}} = \sum_{j=0}^{n-1}\delta_1\left[(\rho)^j\Sigma\Sigma'(\rho')^j\right]\delta_1' \tag{15}$$

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 as detailed in section 3.3.3.

The addition of surveys to the data set requires making an assumption about their data generating process. In an affine context, Kim and Orphanides (2012) 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.

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 ρ replaced by

$$\rho - \Sigma k$$

where k is a conformable matrix of parameters that govern the degree of distortion. We refer to this set of estimates as the "PSS" model. In both the KO and PSS cases, we assume the *i.i.d.* measurement error has a constant variance across forecast horizons. Finally, we also estimate models without including forecasts, which we refer to as yields-only models ("YO"). The three-factor YO model (YO3) is most analogous to Wu and Xia (2016)'s estimates, with the main difference being the underlying yields data and the use of the discretization filter rather than the EKF. In some figures, we compare results based on our estimates to those of Wu and Xia (2016). We label their results as "WX."

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 (equations (5) and (6)) and the prices of risk parameters λ_0 and λ_1. For three-factor models, we impose a similar normalization to Joslin et al. (2011) and Wu and Xia (2016) for the risk neutral factors:

$$\delta_1' = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \tag{16}$$

$$\mu^{\mathbb{Q}} = \mathbf{0} \tag{17}$$

We also assume $\rho^{\mathbb{Q}}$ is in real Jordan form with eigenvalues in descending order and Σ is lower triangular. Joslin et al. (2011) show that this is the maximally flexible specification for latent dynamics that is econometrically identified. 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 (2015b).

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:

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

where ω is a measurement error parameter common across horizons and $e_{nt} \sim \mathbb{N}(0, 1)$. 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 1987-2019 at a monthly frequency. Forward rates are constructed using the Liu and Wu (2021) 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 adjust the Farmer (2021) procedure slightly to account for the fact that we do not observe each forecast horizon every month. Our treatment of missing values is analogous to the treatment of missing data in other state space models (e.g. Durbin and Koopman (2012)). $G_t(X_t)$ has a time-varying dimension to capture the fact that forecasts of particular horizons are periodically unobservable. Because which forecasts are missing is a deterministic function of time and unrelated to the (hypothetical) values of those forecasts, the missing observations are "ignorable" in the sense of Rubin (1976). Hence, they do not affect inference of the parameters governing the hidden Markov model (see Speekenbrink and Visser (2021)). The process for which observations are missing can be factored out of the likelihood function, the resulting likelihood is proportional to the likelihood function without missing data, and the convergence proofs in Douc et al. (2004), Douc et al. (2011) and Farmer (2021) are unaffected.

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. We present results from estimating a three-factor, yields-only version of the model with the EKF in Appendix G. For that model, the within-sample fits for both the EKF and the discretization filter are similar. Out of sample, the EKF tends to do similarly or slightly worse at short horizons and better at ten year horizons. Its performance during the COVID-19 pandemic and afterwards, however, is much worse at short horizons, particularly compared to the models that incorporate forecasts. In practice, we also found that the EKF was much more computationally expensive in models that incorporate forecasts.[3]

[3] The EKF is much quicker than the discretization filter for yields-only models. However, at each time t, the EKF requires linearization of the observation equation G_t to evaluate the likelihood, which adds a considerable computational burden when the dimen-

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. In the estimation, we constrain the lower bound parameter to fall below 25 basis points, which is the value assumed by Wu and Xia (2016). For the 3-factor yields only model, that constraint binds, but the remaining models fall between 12 and 21 basis points, slightly above that used in Federal Reserve Board staff estimates in 2012 (around 10 basis points).

Model-implied shadow rate expectations depend on the set of parameters governing the physical dynamics. The inference about the values of these parameters differs across the models, in general. This is mainly driven by differences by factors after the first. Comparing the three-factor models, the largest eigenvalue for each system is both highly persistent and roughly similar across the models (although slightly larger for the models using forecast data). The remaining eigenvalues for the KO model are complex (with modulus of about .92) while the YO and PSS models have three distinct eigenvalues. The difference is both what information is used to inform the estimates (e.g., forecasts are an additional set of observations that depend on the parameters governing risk-neutral dynamics) and identifying assumptions

sion of G_t varies over time, as in the KO and PSS models.

(e.g., whether we restrict the forecasts to be governed by exactly the same persistence parameters as the yields). One implication of these differences is that the predicted long-run short rate differs substantially across models; just under 2.0% for the three-factor KO model, 3.4% for the PSS model and 4.2% for the YO model. In other words, the models disagree about the "long run" nominal short-run interest rate.

The PSS models also allow for differences in the physical dynamics and the time-series process perceived by forecasters. Interestingly, the PSS two- and three-factor models disagree on the characteristics of that distortion. For the two-factor model, the eigenvalues for the distorted state-space process are smaller than for the physical process. This result would imply that forecasters thought short rates were *less* persistent than they actually are. For the three-factor model, the differences are somewhat more subtle. The two largest eigenvalues under the distorted measure are larger but the third is smaller. Impulse responses to s_t under each measure (shown in figure 1) are qualitatively different for the first and third factors as a result. Shocks to the third factor are forecast to have larger effects at medium-to-long horizons under the physical measure than the distorted measure, while shocks to the first factor result in expected shadow rates that are higher at long horizons in the distorted measure. The differences in the perceived evolution of short rates will also imply differences in perceived duration of the ZLB across the two measures, as we will see below.

In- and out-of-sample fit Having estimated the complete suite of models, we now investigate how well they can explain the data.

In terms of comparisons across models, different criteria point to different models as providing the best fit. For instance, the two factor models (surprisingly) achieve higher values for the log-likelihood over the complete sample than the analogous three factor models; the two-factor KO model has the highest likelihood of models that use forecasts, but the PSS three-factor model outperforms the KO three-factor model.

Arguably, we should focus on the ability of the model to capture the data in the post-2007 period – i.e., how well do shadow rate models capture the data at the zero lower bound and in the post-ZLB regime? 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. Panel A of table 1 summarizes in-sample fit (mean absolute error and root mean square error), on average, for each model. The three-factor YO model appears to have the best performance (although it does not have to simultaneously explain forecasts and yields). The fit of the three-factor YO model is also similar to the average fit for the analogous model estimated with the EKF (see appendix table G.1) .

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. Although our main sample ends in 2019, we conduct the pseudo-forecasting exercise all the way through the end of 2023. Hence, for forward rates of each maturity, we have 216 (18 estimates of 12 horizons) sets of forecasts.[4] Panel B of Table 1 displays the RMSE and MAE from this exercise, and panels C and D break the out-of-sample forecasting results between the pre-COVID (2007-2019) and post-COVID period (2020-2023). Barring one set of forecasts, three factor models have lower error than their two factor counterparts.[5] Taken together, these results seem to point towards the superiority of three factor models for actually capturing the phenomena of interest, relative to their two-factor counterparts. In our subsequent discussion, we focus mainly on the implications of the three-factor models. The results are less defininitve on which specification – YO, KO, or PSS – is most useful for applied work. The PSS models appear to do somewhat better at forecasting at shorter horizons and in-sample than the KO models, while the KO models do better at the medium and longer end. The YO model has the best overall fit of the actual yields data during the period of interest. Hence, we generally focus on areas of consensus and disagreement across the three specifications.

[4]Because of the computational burden of estimating the model, re-estimating for each additional month of data is infeasible.

[5]Particularly, the two-factor PSS model has the lowest error for ten year forecasts during the 2020-23 period. However, the MAE is only one basis point lower than the three-factor KO model.

4.2 Shadow rate estimates and the forecasted length of the ZLB period

Figure 2 shows our shadow rate estimates alongside the Wu and Xia (2016) estimates during the ZLB period for the three-factor models. 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), at least through the taper announcement. Our estimated rates are somewhat smoother than the EKF estimates, and tighten more quickly following the taper announcement.[6]

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 3. All of our models suggest that market participants initially under-predicted the (ex-post) duration of the zero lower bound.

In August 2011, the Federal Open Market Committee (FOMC) introduced specific calendar-based forward guidance in their Summary of Economic Projections (SEP). In Figure 3, 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

[6]Two factor estimates generally seem to have a lower level than the three factor estimates, as shown in appendix figure F.1.

estimates for the federal funds rate. The introduction of calendar-based forward guidance corresponded with an upward reassessment of the expected ZLB duration. We display this effect in Figure 4, which focuses on the period surrounding the introduction of the guidance. The point estimates suggest that calendar-based forward guidance extended the market's estimate of the ZLB duration by between 2 and 4 months. There is also some upward drift in the months prior to the announcement, although that could simply be markets reacting to the same information that contributed to the FOMC's new policy.

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. The yields only model suggests market participants' beliefs were consistent with the SEP's guidance of "exceptionally low levels for the federal funds rate at least through mid-2013." The KO model suggests the central tendency of beliefs was consistent with the SEP up through 2014, when the median path diverges (albeit with very wide confidence bands that include the SEP's corridor). The beliefs implied by the physical and distorted dynamics diverge, with the physical dynamics implying a longer duration of the ZLB throughout (and much longer than the SEP's guidance). The distorted dynamics, by contrast, imply beliefs that are more-or-less consistent with the SEP's forward guidance. The KO and PSS results are informative about each other. The PSS model essentially relaxes a requirement imposed by the KO model that the market forecast and the

Blue Chip forecast must have the same data generating process. The results imply that Blue Chip forecasters' beliefs were essentially consistent with the forward guidance provided by the FOMC. By contrast, the "market" forecast (implied by the physical measure) implied more persistence in the underlying factors that drive short rates in order to match the actual behavior of forward rates. In other words, the estimated physical dynamics implied that once rates were low, they implied they would stay low for longer. The KO model, in attempting to reconcile a possible change in the beliefs of participants with what it sees in yields, appears to split the difference relative to the PSS measures.[7] The results in table 1 imply that the three-factor KO and PSS models have similar ability to *forecast* forward rates which makes their starkly different predictions about the implied duration of the ZLB all the more surprising. However, the ability to flexibly reconcile the divergence between physical and subjective beliefs is probably why the PSS model appears to have superior fit in-sample and forecasting over short horizons in particular.

In short, our results confirm that markets initially under-estimated how long interest rates would remain at their lower bound in 2008 and 2009. There is evidence that calendar-based forward guidance shifted the market's perceived duration of the ZLB in the desired direction, although the exact degree to which that occurred differs across models, as does whether the

[7] Interestingly, relatively more persistence in physical than subjective short rates is the opposite of the estimated result in Piazzesi et al. (2015).

models imply beliefs consistent with the SEP's calendar.

5 The effects of 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 some ways, the features of these beliefs reflect very different beliefs about monetary policy. 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 three exercises. First, we estimate a structural FAVAR as in Bernanke et al. (2005); Wu and Xia (2016). We test whether there was a structural break during this period; we find mixed evidence for a break. 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. Using an updated macroeconomic sample from 1987-2019, we find shocks to policy rates that are less persistent and have much smaller effects than reported Wu and Xia (2016). But, we also show that differences in estimated impulse responses to a shock to the shadow rate are driven mainly by differences in the underlying macroeconomic dataset rather than differences in the covariance between shadow rates and other macroeconomic indicators. Second, we examine the effects of policy announcements on term premia. All of our models suggest there were 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. Third, 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 and that there is evidence of "local scarcity" effects on term premia.

5.1 Estimation of FAVAR and the effects of monetary policy shocks

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. To isolate the difference driven by shadow rate estimates alone, we initially se the same data specification and code as Wu and Xia (2016) to estimate the FAVAR. The unrestricted model is:

$$
\begin{aligned}
\begin{bmatrix} F_t \\ s_t \end{bmatrix} =& \mathbf{1}_{t<\text{Dec 2007}} \mathbf{B_1(L)} \begin{bmatrix} F_{t-1} \\ s_{t-1} \end{bmatrix} \\
&+ \mathbf{1}_{t>\text{Dec 2007 \& } < \text{ July 2009}} \mathbf{B_2(L)} \begin{bmatrix} F_{t-1} \\ s_{t-1} \end{bmatrix} \\
&+ \mathbf{1}_{t\geq \text{ July 2009}} \ \mathbf{B_3(L)} \begin{bmatrix} F_{t-1} \\ s_{t-1} \end{bmatrix} + v_t
\end{aligned}
$$

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 $\mathbf{B_1(L)}$ and $\mathbf{B_3(L)}$ are equal. Like in Wu and Xia (2016), we examine whether this is the case using a likelihood ratio test adjusting for small-sample bias (Sims (1980)). The first column of table 2, panel (a) reports the p-value for the coefficients of lagged shadow rates vis-a-vis the macroeconomic factors. Focusing on the three-factor specifications, we fail to reject the null for the YO and PSS models, but do reject it for the KO model at the 95% level. The second column reports the p-value for the coefficients of lagged macroeconomic factors on shadow rates. We do not reject the null for any of the three-factor models in this case.[8]

Because our shadow rate estimates are based on smoothed factors, they cover the full sample of yields. Since the contribution of Wu and Xia (2016), we have an additional ZLB period that may be informative about the parameters governing the shadow rate. As an experiment, we substitute shadow rates based on our yield and parameter estimates from 1987-2019 to those using data from 1987-2023 (e.g., the COVID-19 pandemic period and its aftermath; we discuss these estimates more in section 6). We similarly estimate the FAVAR using the same data and specification as Wu and Xia, covering the period 1960-2013, so any differences are driven completely by new information about the shadow rate alone and subsequent macroeconomic developments. The results are shown in panel (b) of table 2. We reject the null of no structural break for all of our models except the YO three-factor model. As we discuss later, this model breaks down during the COVID period. While not definitive, these result cautions against assuming continuity in macro-monetary policy relationships before and after the Great

[8]While we focus mainly on the three-factor specifications, we note that we *do* reject the null for the two-factor specifications. As mentioned in the introduction, a number of authors continue to use shadow rates estimated with two factors in applied work, at least as a robustness check; our results suggest they should be very cautions about the presence of a structural break.

Recession.[9]

Is it the shadow rate or the macroeconomic data? Up to this point, we have focused on whether swapping our shadow rate estimates in for the Wu and Xia (2016) estimate materially impacts the FAVAR structural break test. One way of seeing if the *substantive* conclusions differ across our rates is to focus on the "post-crisis" period. Figure 5 compares the impulse response for the unemployment rate to 25 basis point decreases in the policy rate for an estimated FAVAR(1) using the same macroeconomic data as Wu and Xia, but swapping our shadow rate estimates. From the figure, we see that the median impulse responses across models are relatively similar; the impulse responses using our shadow rate estimates appear to have shorter half-lives than the Wu and Xia estimate, but the peak effects are similar. One might conclude that monetary policy has slightly less long-lived real effects than previous expected.

Subsequently, we estimate the FAVAR on a different macroeconomic dataset, the "FRED-MD" dataset from McCracken and Ng (2015), from 1987 to 2019.[10] Other than changing the dataset, we maintain the same FAVAR(13) specification for the complete sample and FAVAR(1) for the

[9]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 models estimated with their shadow rate to not appear to have a structural break at the Great Recession, this is not necessarily true of any given shadow rate estimate. We note that estimating using the Liu and Wu (2021) yields and our extended sample, but using the EKF as in Wu and Xia (2016), we do not reject the null of no structural break before and after the Great Recession.

[10]Particularly, we use the vintage as of November 2024. We drop data series with missing values and use McCracken and Ng's suggestions for stationarizing the underlying series.

"post-crisis" (July 2009-onward) sample as Wu and Xia (2016). The resulting impulse responses are shown in figures 6 and 7, respectively. The first column of each figure, labeled WX (2016), replicates the existing estimates from Wu and Xia. The second uses the most recent (as of December 2024) vintage of Wu and Xia shadow rate estimates from the Federal Reserve Bank of Atlanta, combined with the FRED-MD data. Subsequent columns report impulse responses using our shadow rate estimates and the FRED-MD dataset. We find that impulse responses with the FRED-MD dataset and our more limited time frame mean revert much more quickly and do not differ substantially across estimation method; the peak decrease in unemployment is less than two-third as large as in Wu and Xia, and the effects become insignificant within six months. For the post-crisis sample (incorporating an additional six years of data), the effects of monetary shocks are minuscule. One some level, this is unsurprising; Ramey (2016) points out that the effects of monetary policy shocks have been more difficult to identify since the 1980s because monetary policy is more systematic. On the other hand, it is somewhat surprising that even the extraordinary policy during the Great Recession was not particularly informative about the effects of monetary shocks, or that those policies did not have sizable real effects.

It is worth emphasizing that the presence of a structural break, and the differences in impulse responses, do not necessarily imply the shadow rates are not a useful replacement for the federal funds rate in monetary VARs. Our argument is that care is needed; the choice of sample appears to be more material than the method of estimating the shadow rate in this particular application.

5.2 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.[11] 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. Term premia are calculated by using the estimated parameters to calculate the expected path of nominal *actual* short rates out to ten years – that is, the expectations hypothesis component reflects the effective lower bound on interest rates.

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.[12] 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

[11] Hanson 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.

[12] The 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).

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.[13] Figure 8 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 Federal Reserve policy announcements. For example, between November and December 2008 (LSAP1), the 10-year yield fell by about 112 basis points.

All of our models attribute little of this decline to the EH component of yields (ranging from 2 basis points for the 3-factor KO model model to 18 basis points for the 3 factor YO model). This implies the share of the yield decrease from the LSAP announcement attributable to changes in the EH component is between 1.7% and 16.5%. Narrowing in on three major event dates, Figure 9 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 and the forward guidance

[13] The decomposition of the yields across maturities for the entire sample is available upon request.

announcement were due to changes in term premia. In contrast, changes in the EH component were relatively more important following LSAP2.

In summary, the interpretation of how major policy announcements affected yields appears to be sensitive to a number of features of the underlying estimates of risk premia. Given this sensitivity, it is unsurprising that previous efforts have found mixed evidence on the precise effects of policy announcements. However, the results here imply that the range of disagreement appears to be about whether term premia explain the entire cumulative change in yields versus about four-fifths.[14]

5.3 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 previ-

[14]This particular result is quantitatively sensitive to the choice of underlying yield curve data, a point we discuss more in section 7.

[15]D'Amico et al. (2012) and Huther et al. (2017) provide thorough historical descriptions of the Fed's balance sheet policies.

ous 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 and whether conclusions about those channels differs based on the method of estimating 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. 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:

$$\text{TP}(10\text{yr})_t = \beta_0 + \beta_1 PHNT(m:2-10)_t + \beta_2 DG_t + \beta_3 \text{Slope}(10\text{-}2\text{yr})_{t-1} + \epsilon_t \tag{19}$$

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 56%. We then use the model-implied monthly 10-year term premium as the dependent variable in regression (19) to examine whether the impacts of policy differ across models. We find robust evidence that local scarcity and the duration gap both significantly explain term premia; the slope is also generally significant. D'Amico et al. (2012), using weekly data over the same sample period and a term premium estimated using an affine term structure model, 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 scarcity are about 30% smaller (except in the PSS 3 factor case), while the elasticity of term premia with respect to the duration gap ranges from about 35% smaller (PSS3) to nearly 40% larger (YO2). 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. But our suite of models suggests that the local scarcity channel's statistical and economic significance is not sensitive to the term structure model used to estimate term premia.[16]

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. Their estimates imply that the first LSAP program decreased term premia by about 42 basis points overall. 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 39.6 basis points overall as a result of LSAP1 and 43.7 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 28 and 35 basis points for LSAP1 and 34 to 44 basis points for LSAP2, 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 induced by the supply changes from the LSAP programs. These effects are only marginally significant for the 3-factor YO model for LSAP 1.

We are cautious to not draw a causal interpretation from these regres-

[16]However, it may be sensitive to the underlying data used to estimate the term premia as we note in section 6.

sions. Moreover, these results in isolation do not imply that there were no signaling effects of the LSAP programs, since they are isolated to the effects predicted by duration and local scarcity effects. The robust finding across our specifications is that the change in duration and local scarcity associated with economically and statistically significant changes in term premia for 10-year Treasuries. However, the precise magnitudes and *relative* importance of each channel varies across the specifications used to estimated term premia. Moreover, our estimated term premium changes are about 15-30% smaller than the elasticities reported in D'Amico et al. (2012), and we generally estimate relatively smaller contributions from scarcity and larger from duration relative to that paper.

6 What do we learn from the COVID-19 recession?

Our focus in this paper has been on the ZLB period following the 2008 financial crisis and the effects of monetary policy at that time. Accordingly, we ended our main sample in 2019, in order to avoid influencing our estimates with the dramatic, but short-lived, recession associated with the COVID-19 pandemic. But, of course, the Federal Reserve lowered rates to their lower bound during 2020 and kept them there for two years, so it is worth examining how our models handle this period. We plot the yields from this period in figure 11. The decline in yields was quite sharp at the beginning of both crises, but the length of time at the lower bound was notably shorter for the pandemic recession, and the increase in rates at the short end of the yield curve was much more rapid once the Federal Reserve began tightening.

We have already noted in section 5 that yields data from this period inform the estimates of shadow rates for earlier periods. This affects conclusions about applications back to the Great Recession period. We report parameter estimates for the extended sample in online appendix H, as well as a measure of average fit. It is clear from the measures of fit that some specifications of the model struggle to fit the yield curve with the additional observations from 2020-2023. In particular, the three-factor YO model displays bizarre average predictions.[17] The three-factor KO and PSS models, however, have good fit, capturing the inverted average yield curve from 2022 and 2023.

We also examine the implied durations of the ZLB during this period, compared to the guidance from the Fed's SEP. The median forecast for the short rate at the start of the pandemic was zero for the entire forecast horizon, before narrowing in late 2021. The three-factor models with forecast information seem to be broadly consistent with these paths, although we once again see that there is a disconnect between the estimated physical and distorted paths of short rates – in particular, the median Blue Chip forecast appears to imply tightening would occur earlier than the SEP indicated.

[17] The two-factor models also suffer a degradation of performance, further affirming our focus on three factor models for our policy applications.

Taking these results alongside those from section 5, it seems that three-factor shadow rate models that incorporate forecast information are capable of bridging both recent ZLB episodes. We are less sanguine about yields-only models (and models with fewer factors). Concretely, it appears that markets and forecasters broadly understood the path of policy, especially as the Federal Reserve approached normalization.

7 Sensitivity to yield curve data

In an earlier working paper version of this paper, we used the Gürkaynak et al. (2007) (GSW) zero-coupon yield curve estimates, following the choice made by Wu and Xia (2016). Subsequent to our first completed draft, Liu and Wu (2021) (LW) published new estimates of the zero-coupon yield curve, which we have used in this iteration.[18] The Liu and Wu (2021) yield curve estimation methodology differs from Gürkaynak et al. (2007) in two important respects: Liu and Wu's estimates include data on Treasury bills and securities with less than three months to maturity, and they use nonparametric methods to estimate the constant-maturity zero coupon curve. Liu and Wu (2021) emphasize that their estimates capture the local variation in maturities and have smaller pricing errors, especially at the short and long end of the yield curve. We plot the annualized difference in forward rate estimates in figure 13. They are often substantial, especially around the financial crisis period.

[18]We thank an anonymous referee for encouraging us to use the Liu and Wu data.

Some of the qualitative and quantitative conclusions of our estimation differ between the two datasets. This is unsurprising, as the underlying yield curve data are quite different, and this difference is somewhat magnified for forward rates, particularly around the start of the Great Recession. Liu and Wu find a similar result in the context of Cochrane and Piazzesi (2005) regressions.

Rather than double the reported estimates, we summarize which conclusions are sensitive to the choice of zero-coupon yield curve data in this section.

- **Model fit and forecasting performance**: Unsurprisingly, the non-parametrically estimated yield curve is harder to match than the relatively smooth GSW estimates. Using the GSW data, we found that the EKF did a marginally better job fitting in-sample, but had worse forecasting performance. The picture is more nuanced with the LW data, which can be observed by comparing table 1 to appendix table G.1. Using the LW data set, the performance for the EKF is similar to the best discretization-fllter estimates except at the 7- and 10-year horizon, where the RMSE is 7 and 22 bp lower for the EKF (panel A in the respective tables). Similarly, the average out-of-sample forecasting performance is somewhat better for the discretization filter estimates at the short end but not for the long end (panel B in each table). The benefit of the short-end forecasts is especially apparent during the 2020-2024 period (panel D). Some of the degradation of fit and forecasting performance is likely due to the greater variability of the yield curve. On average, the 3 factor models estimated with the discretization filter appear to perform well up through 2019 regardless of the data set.

- **Structural break tests and monetary policy VARs**: Using GSW data, we rejected the null of no structural break for most, but not all, of our models (particularly, we failed to reject the null at for the PSS 2-factor model, and rejected it marginally for the KO 2-factor model). Using the LW data, we get a similarly mixed pattern. We also find that the shadow rates estimated with the EKF on the LW data do not reject the null of no structural break.

- **Event analysis**: In general, the estimates using LW attribute a greater fraction of variation around the announcement of LSAPs and forward guidance to risk premia, relative to those using GSW. In particular, the change in expected duration of the ZLB around the forward guidance announcement is greatly attenuated using the LW estimates. Figure 4, for example, implies that that the largest jump in anticipated duration of the ZLB around the forward guidance announcement was about 4 months; using the GSW data, the largest duration was about 9 months. Similarly, we found a greater explanatory power for the expectations hypothesis component of yields in the analysis of LSAPs using the GSW data. The PSS 3-factor model implied more than 40% of the drop in yields was due to changes in short rates, while using the LW data, that fraction is less than 20%. However, it is worth remembering that the GSW short end is a parametric extrapolation, while the LW estimates use near-to-maturity securities. This seems to make LW preferable for the event analysis

- **Mean duration of ZLB** Using the GSW yield curve estimates, we found that real-time mean duration of the ZLB was similar to that implied by the forward guidance from the SEP, except for one model (KO3), whose duration estimates were generally longer. Using the LW data, we get qualitatively different results. The three factor KO model has wide confidence bands, but the median path diverges from the SEP's guidance. The PSS model over-estimates yields, but the implied path of the "distorted" forecast path (which draws mainly on the forecasts) is consistent with the SEP.

  Effects of Treasury supply changes: Using GSW's yield curve estimates, we find qualitatively and quantitatively different evidence on the channels of asset purchases. Our estimates based on GSW imply that local scarcity is only significant in half of the term premium regressions. By contrast, the results in tables 3 and 4, which calculate term premia and the yield curve's slope using the Liu and Wu dataset, suggest changes in local scarcity and duration are both statistically associated with changes in the term premium observed during the first two LSAP rounds, in a way that is economically significant.

8 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, allowing for deviations from full information rational expectations in those forecasts, and incorporating new yield curve estimates that incorporate more information from the short end of the maturity spectrum. We find that forecasting performance is often better using the discretization filter. Our recommendation for applied work, particularly work that focuses on in-sample properties of interest rate expectations or term premia, would be to compare robustness across a variety of both model assumptions and estimation methods.

Our goal in this paper has been to identify robust effects of monetary policy at the zero lower bound and, additionally, to provide some guidance for applied researchers who are interested in producing or using estimated shadow rates. To summarize our conclusions:

1. The Farmer (2021) discretization filter appears to be an accurate and computationally competitive alternative to the extended Kalman filter for estimating shadow rate models that use the Wu and Xia (2016) approximation for the ZLB.

2. Three factor models that incorporate financial forecast data appear to

do the best job of actually capturing the relevant features of the yield curve since 2007, relative to their two-factor or yields-only alternatives.

3. The method of estimating shadow rates and the underlying data set for yields, forecasts, and structural assumptions about forecasts will materially affect the conclusions about the level of the shadow rate and the expected duration of the ZLB, in practice. This is true despite the fact that average fit or forecasting ability may be relatively similar across different models. This may affect conclusions about whether forward guidance was successful at shaping market forecasts or the extent to which "unconventional" policy announcements affected yields through expectations or term premia.

4. Despite differences in underlying shadow rate estimates, many conclusions about policy are qualitatively similar regardless of the shadow rate model, or depend more on data choices than choices about estimation. In particular:

    (a) Conclusions about the effects of monetary policy in FAVARs depend highly on the macroeconomic data, rather than the shadow rate.

    (b) The majority of the effect of the LSAP1 and LSAP2 announcements was a change in term premia, regardless of how those term premia are estimated.

(c) Reduced-form tests about the effects of duration versus local scarcity for LSAPs depend on the yield data used to estimate them (and underlying term premia). For estimates using Liu and Wu (2021) yields, the effects of duration and local scarcity are both statistically and economically significant regardless of term structure model. This is not true for estimates using Gürkaynak et al. (2007) yields.

In this paper, we focused mainly on the period prior to the COVID-19 pandemic. Our extension to that period revealed that models incorporating forecast data performed better at predicting the data out-of-sample, and that two-factor models with forecasts have trouble fitting both the financial crisis and pandemic ZLB episodes. For applied work, we recommend authors consider using estimated shadow rates informed by interest rate forecasts.

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 and respecting the zero lower bound would be a natural extension.

9 Tables

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 (396 months). Panel B contains estimates for the out-of-sample fit, which estimates the model each December from 2007-2023, and calculates forecasts for 1- to 12-months ahead. Panels C and D report subcomponents of the out-of-sample forecasts splitting the sample before and after the COVID-19 pandemic. MAE and RMSE are reported across all horizons (10 sets of forecasts at 12 horizons each).

Statistic	MAE						RMSE					
Model	YO		KO		PS		YO		KO		PS	
Factors	2	3	2	3	2	3	2	3	2	3	2	3
Panel A: In-Sample Fit (N=396)												
3mo	0.18	**0.11**	0.21	0.33	0.21	0.18	0.23	**0.14**	0.27	0.40	0.26	0.23
6mo	0.11	**0.10**	0.23	0.21	0.18	0.21	0.15	**0.14**	0.34	0.31	0.25	0.29
12mo	0.18	**0.13**	0.32	0.31	0.23	0.23	0.22	**0.16**	0.44	0.43	0.29	0.30
24mo	0.22	**0.16**	0.31	0.32	0.25	0.31	0.30	**0.21**	0.41	0.43	0.32	0.40
60mo	0.25	**0.23**	0.27	0.32	0.27	0.27	0.34	**0.31**	0.36	0.41	0.36	0.37
84mo	0.22	**0.21**	0.28	0.26	0.24	0.23	0.31	**0.28**	0.38	0.35	0.32	0.30
120mo	0.34	**0.26**	0.40	0.37	0.37	0.43	0.50	**0.36**	0.57	0.51	0.50	0.58
Panel B: Out-of-Sample Fit: 1-12 month-ahead forecasts (N=216)												
3mo	0.56	0.48	0.50	0.40	0.47	**0.34**	0.92	0.97	0.95	**0.68**	0.82	0.73
6mo	0.55	0.51	0.57	0.33	0.45	**0.29**	0.89	0.92	0.97	**0.63**	0.79	0.65
12mo	0.56	0.44	0.62	0.37	0.47	**0.34**	0.80	0.67	0.92	0.60	0.69	**0.59**
24mo	0.56	0.85	0.57	**0.42**	0.58	0.46	0.72	1.45	0.79	**0.54**	0.83	0.59
60mo	0.66	1.62	0.57	**0.52**	0.97	0.61	0.83	2.51	0.74	**0.64**	1.23	0.77
84mo	0.81	1.98	0.97	**0.54**	1.37	0.62	1.02	3.00	1.11	**0.69**	1.64	0.84
120mo	1.01	2.01	1.30	**0.72**	1.42	0.77	1.24	3.02	1.52	**0.86**	1.69	0.90
Panel C: Out-of-Sample Fit, 2007-2019, 1-12 month-ahead forecasts (N=132)												
3mo	0.38	0.22	0.31	0.29	0.30	**0.21**	0.57	0.35	0.47	0.36	0.41	**0.32**
6mo	0.37	0.20	0.39	0.23	0.32	**0.18**	0.54	0.33	0.56	0.30	0.44	**0.30**
12mo	0.42	0.25	0.47	**0.23**	0.34	0.23	0.57	0.36	0.65	**0.31**	0.42	0.31
24mo	0.53	0.33	0.52	**0.32**	0.38	0.41	0.72	0.49	0.76	**0.39**	0.54	0.48
60mo	0.74	0.53	0.56	**0.50**	0.70	0.63	0.90	0.67	0.73	**0.63**	0.87	0.82
84mo	0.98	0.66	0.90	**0.63**	0.99	0.67	1.17	0.82	1.06	**0.78**	1.18	0.93
120mo	1.17	**0.71**	1.15	0.75	1.07	0.77	1.39	**0.86**	1.38	0.90	1.30	0.93
Panel D: Out-of-Sample Fit, 2020-2023,-1-12 month-ahead forecasts (N=48)												
3mo	0.19	0.12	0.27	0.32	0.15	**0.05**	0.24	0.14	0.45	0.36	0.21	**0.06**
6mo	0.24	0.11	0.39	0.13	0.12	**0.09**	0.31	0.15	0.60	0.17	0.19	**0.11**
12mo	0.34	0.20	0.60	**0.14**	0.19	0.18	0.45	0.29	0.85	**0.18**	0.26	0.22
24mo	0.52	0.44	0.93	**0.30**	0.41	0.33	0.70	0.63	1.18	**0.39**	0.55	0.44
60mo	0.77	0.61	0.66	**0.54**	0.67	0.59	0.88	0.77	0.89	**0.69**	0.80	0.82
84mo	0.63	0.58	0.65	**0.53**	0.60	0.64	0.78	0.74	0.88	**0.73**	0.73	0.87
120mo	0.75	0.59	0.78	0.59	**0.58**	0.67	0.86	0.79	1.04	0.78	**0.73**	0.85

(a) Shadow rate estimated from 1987-2019

Wu and Xia 2016	0.289	1.000
YO	1.000	0.175
YO 2 factor	1.000	0.022
KO	0.047	1.000
KO 2 factor	0.017	1.000
PSS	0.139	1.000
PSS 2 factor	0.004	1.000
YO3 factor using EKF, LW data	0.999	0.538

(b) Discretization-filter based shadow rates estimated from 1987-2023

YO	1.000	0.166
YO 2 factor	1.000	0.007
KO	0.096	1.000
KO 2 factor	0.017	1.000
PSS	0.000	1.000
PSS 2 factor	0.006	0.566

Table 2: p-values for tests of a structural break in the FAVAR estimated in Wu and Xia (2016), substituting different estimates of the shadow rate. In each panel, the first column contains the p-values for the test of structural break in effect of lagged shadow rate on macroeconomic factors, and the second column contains p-values for test of structural break in effect of lagged macroeconomic factors on shadow rate. The test is a likelihood ratio test adjusting for small-sample bias as in Sims (1980), with a null is that there are no structural breaks. In panel (a), the shadow rates estimated using the discretization filter are estimated using yields data from 2019, while in panel (b) the data goes to 2023. In both cases, the FAVAR itself is estimated from 1960-2013 using identical data as Wu and Xia (2016) other than the shadow rate during the ZLB period.

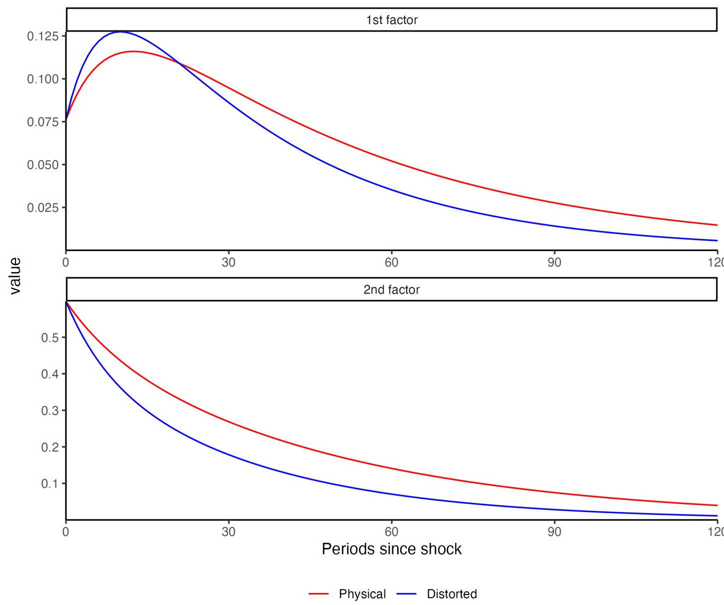
	Sample: December 2002 - October 2008						
	10yr	Term Premia on 10-year Bond					
	Yield	YO	YO	KO	KO	PSS	PSS
		2 fac	3 fac	2 fac	3 fac	2 fac	3 fac
PHNT(m:2-10)	3.48**	3.24***	3.46***	2.93***	3.77***	3.14***	5.18***
	(1.44)	(1.07)	(0.97)	(1.07)	(0.96)	(1.07)	(1.35)
Duration Gap	194.41***	171.52***	101.28***	141.35***	124.15***	129.92***	79.04***
	(27.22)	(19.73)	(16.93)	(22.91)	(23.74)	(17.06)	(22.83)
Slope (1mo lag)	-0.08*	0.28***	0.45***	0.06	0.15***	0.51***	0.53***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Adjusted R^2	0.56	0.63	0.81	0.29	0.27	0.85	0.80
N	71	71	71	71	71	71	71

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.

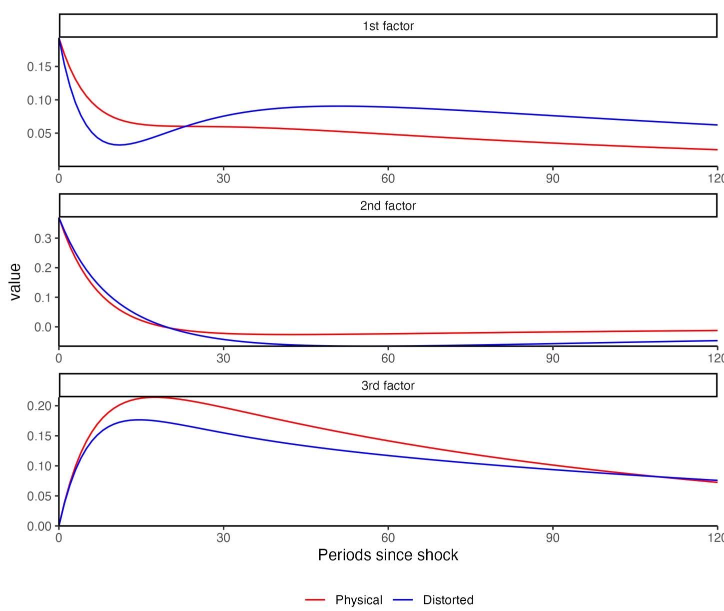
	Sample: December 2002 - October 2008						
	10yr	Term Premia on 10-year Bond					
	Yield	YO	YO	KO	KO	PSS	PSS
		2 fac	3 fac	2 fac	3 fac	2 fac	3 fac
LSAP 1							
Predicted effect from scarcity	-16.29**	-15.19***	-16.20***	-13.73***	-17.69***	-14.71***	-24.30***
Predicted effect from duration	-23.33***	-20.58***	-12.15***	-16.96***	-14.90***	-15.59***	-9.48***
Total	-39.62***	-35.77***	-28.36***	-30.69***	-32.58***	-30.30***	-33.79***
Residual (expectations effects)		-3.84	-11.26*	-8.93	-7.04	-9.32	-5.83
LSAP 2							
Predicted effect from scarcity	-24.24**	-22.61***	-24.11***	-20.43***	-26.32***	-21.89***	-36.17***
Predicted effect from duration	-19.44***	-17.15***	-10.13***	-14.14***	-12.42***	-12.99***	-7.90***
Total	-43.69***	-39.76***	-34.24***	-34.57***	-38.74***	-34.88***	-44.08***
Residual (expectations effects)		-3.92	-9.44	-9.12	-4.95	-8.81	0.39

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 signfiicance at the 1, 5, and 10 percent levels, respectively calculated using the Newey-West correction for standard errors.

10 Figures



(a) 2 factor IRF



(b) 3 factor IRF

Figure 1: Estimated impulse response of the shadow rate s_t to a 1-standard deviation shock to each latent factor under physical and subjective (distorted) dynamics for the estimated PSS models. The red line shows the evolution of the shadow rate using the physical estimated physical dynamics, while the blue line shows the evolution using the subjective dynamics. Panel (a) shows the results for the two-factor PSS model; Panel (b) shows the three-factor PSS model.

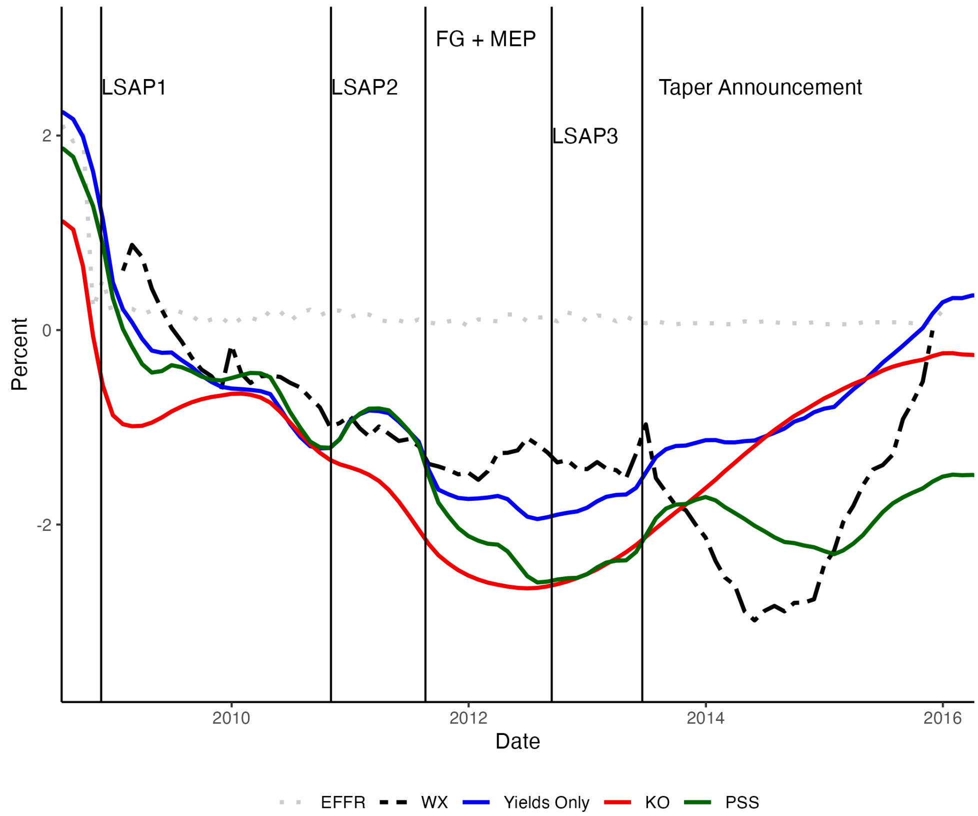


Figure 2: Smoothed estimates of shadow rate during/post Great Recession, with event dates (three rounds of Large Scale Asset Purchases (LSAPs), the introduction of calendar-based forward guidance and the Maturity Extension Program (FG+MEP), Taper Announcement). FG and MEP were introduced in August and September 2011, respectively, but are shown in August 2011.

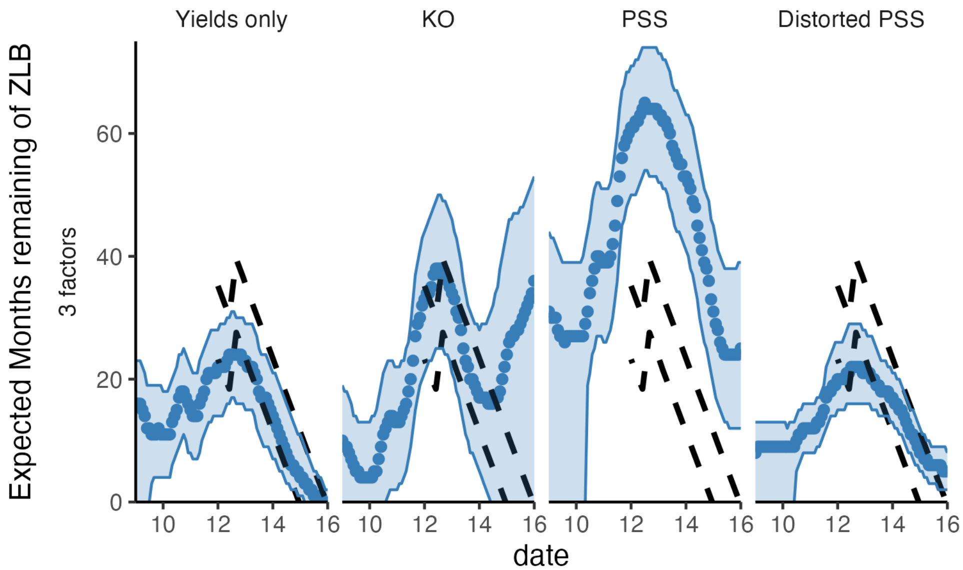


Figure 3: Real-time implied mean duration of ZLB period. Bands indicate 99th percentile of liftoff dates.

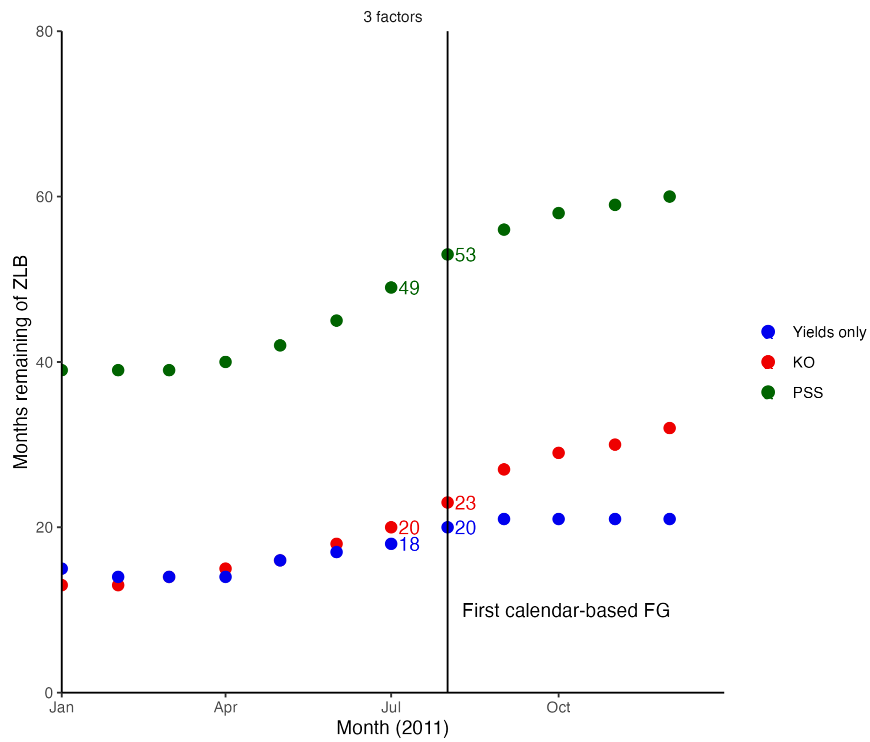


Figure 4: Real-time implied mean duration of 2011 using the same estimates as those in Figure 3.

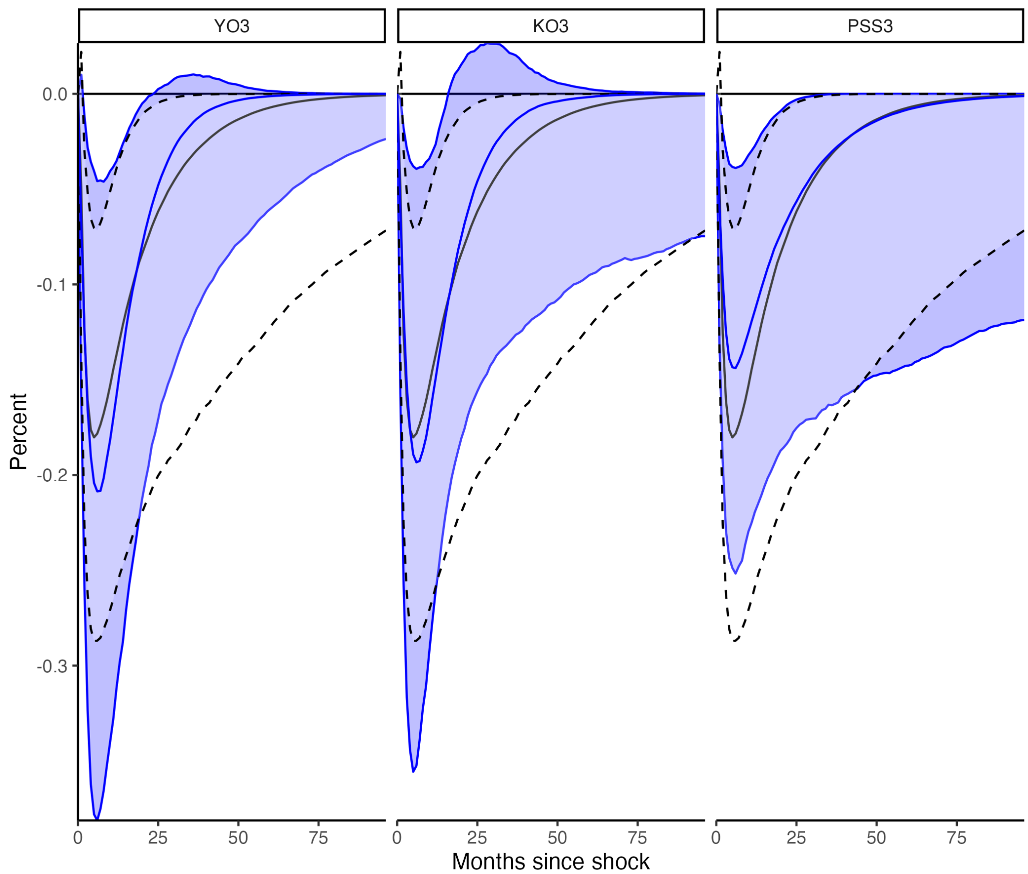


Figure 5: Estimated impulse response of unemployment to a 25 basis point decrease in the policy rate (Federal funds rate/shadow rate). Solid lines indicate median response and bands indicate 90% confidence intervals. Black is the impulse response reported in Wu and Xia (2016) and blue indicates estimates with shadow rates estimated with the discretization filter. The model is estimated as a FAVAR(1) using the “post-crisis” specification and data in Wu and Xia.

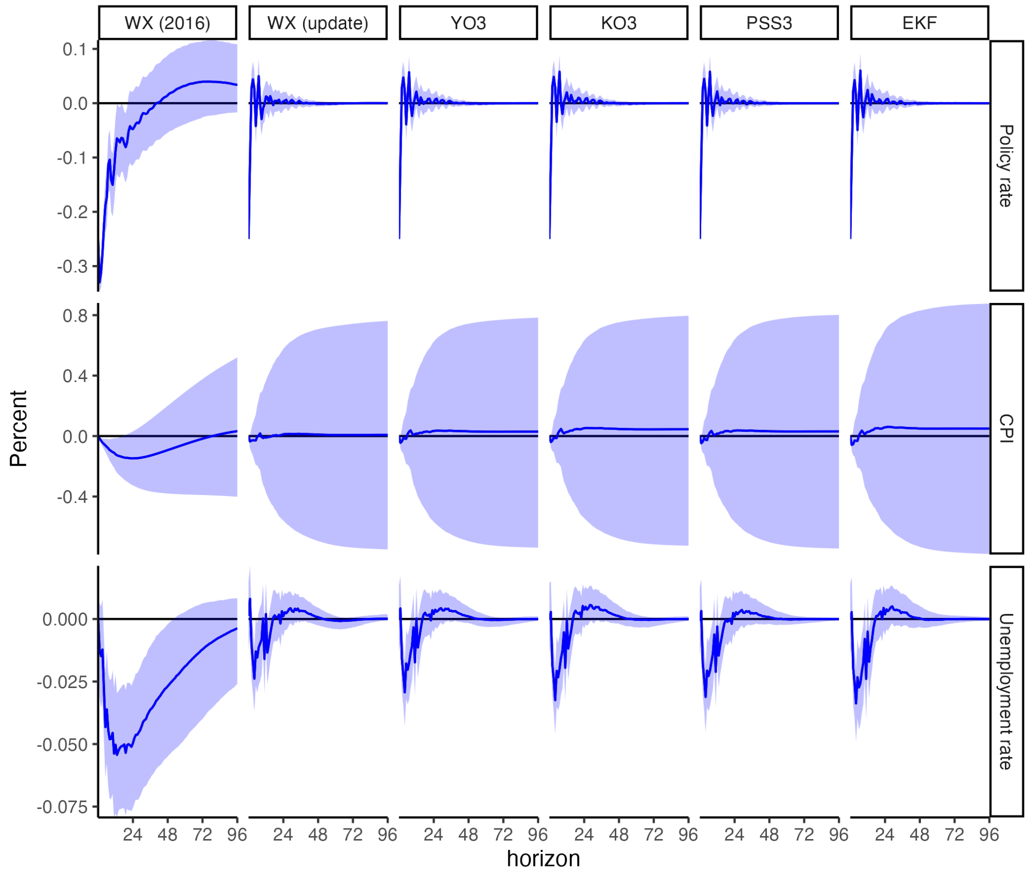


Figure 6: Estimated impulse response of unemployment and inflation to a 25 basis point decrease in the policy rate (Federal funds rate/shadow rate). Solid lines indicate median response and bands indicate 90% confidence intervals. Impulse responses are estimated using a FAVAR(13) estimated using data from 1987-2019. The difference across columns is the underlying estimate of the shadow rate. "WX (2016)" indicates the original impulse response reported in Wu and Xia (2016); "WX (updated)" uses the most recent vintage of data as of December 2024. "EKF" is the YO3 factor model estimated using Liu and Wu (2021) yields and the extended Kalman filter.

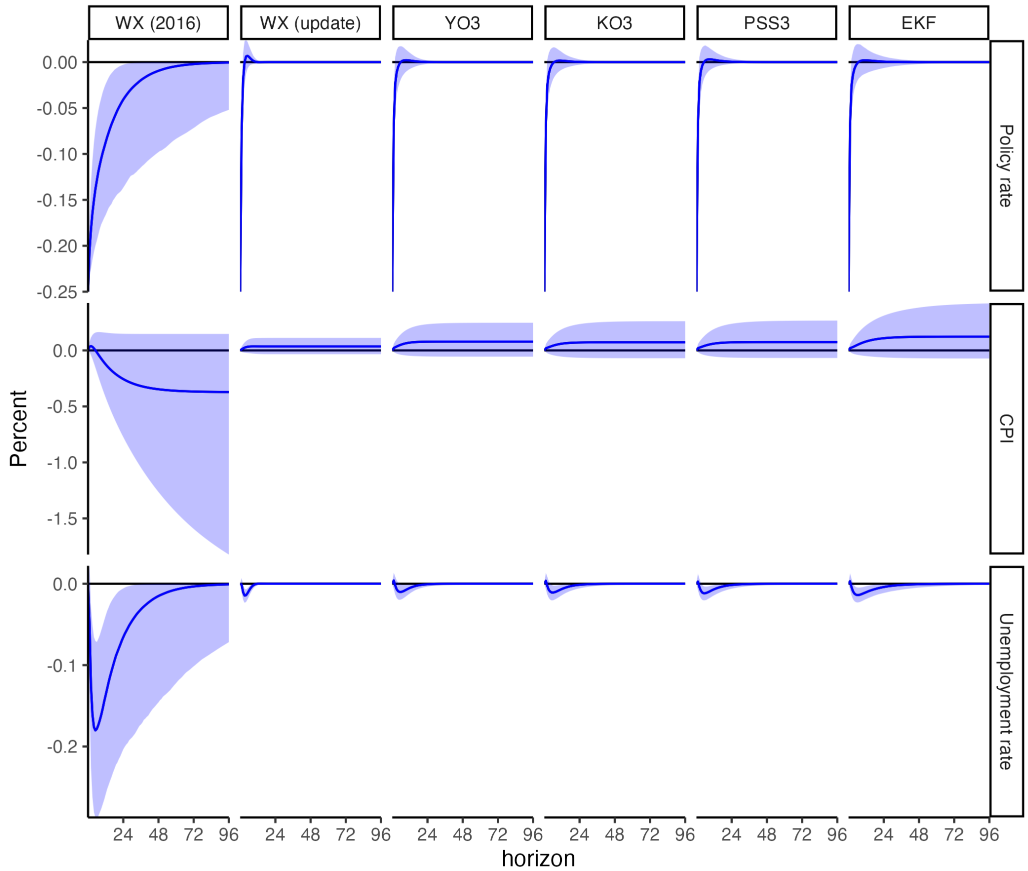
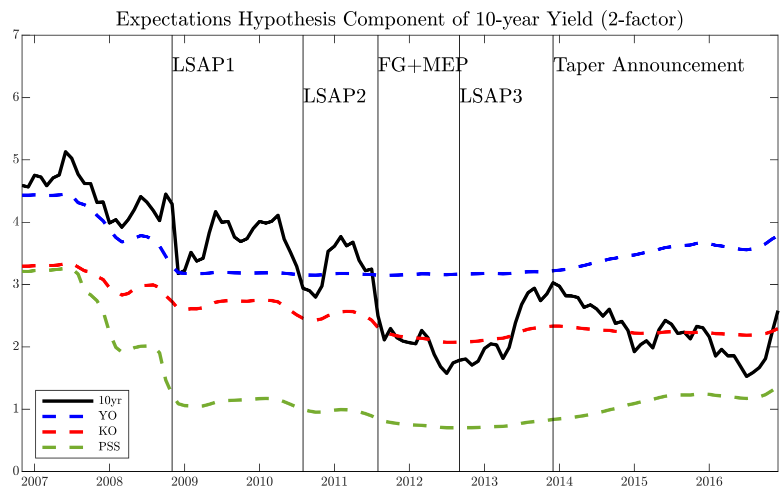


Figure 7: Estimated impulse response of unemployment and inflation to a 25 basis point decrease in the policy rate (Federal funds rate/shadow rate). Solid lines indicate median response and bands indicate 90% confidence intervals. Impulse responses are estimated using a FAVAR(1) estimated using data from July 2009-December 2019. The difference across columns is the underlying estimate of the shadow rate. “WX (2016)” indicates the original impulse response reported in Wu and Xia (2016); “WX (updated)” uses the most recent vintage of data as of December 2024. “EKF” is the YO3 factor model estimated using Liu and Wu (2021) yields and the extended Kalman filter.



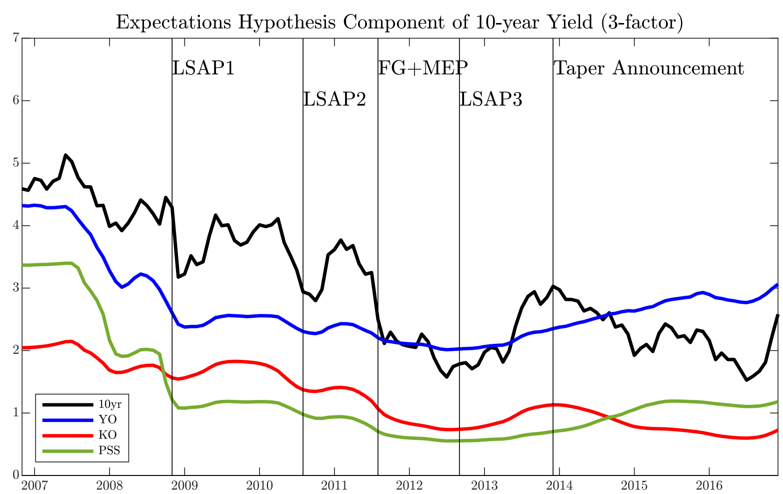
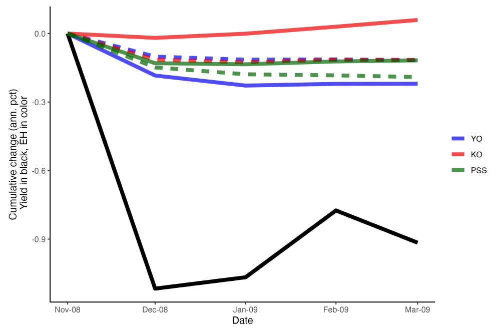
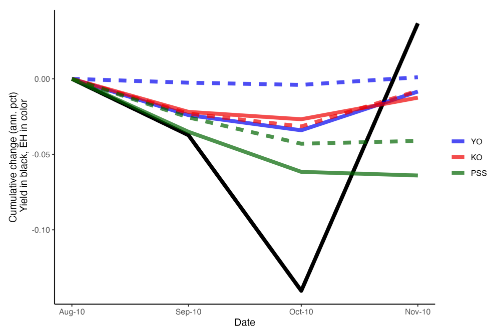


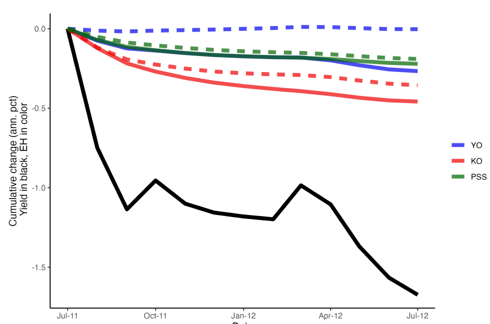
Figure 8: Decomposition of the 10-year U.S. Treasury yield (black line) during the ZLB. The Expectations Hypothesis component of the 10-year yield is shown for the YO (blue line), KO (red line), and PSS (green line) models. Results for the two-(three-)factor model are in the upper (lower) panel.



(a) LSAP1



(b) LSAP2



(c) FG

Figure 9: 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.

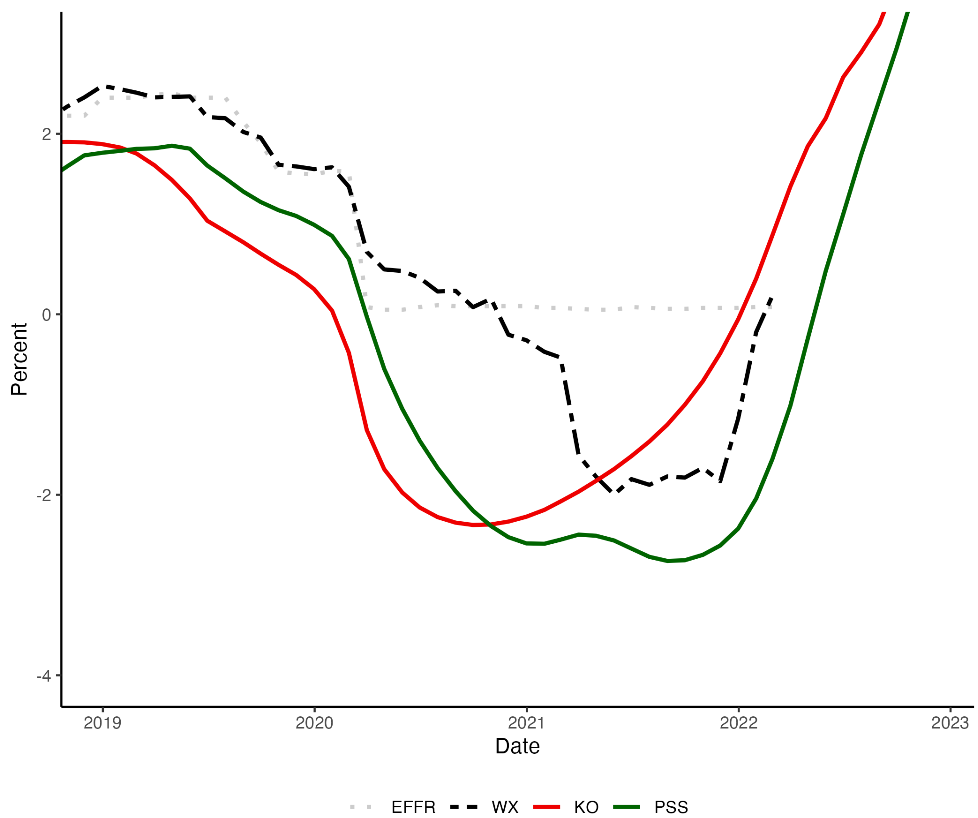


Figure 10: Smoothed estimates of effective federal funds rate and shadow rate estimates during and after the COVID-19 recession. YO3 is not reported since it is never negative.

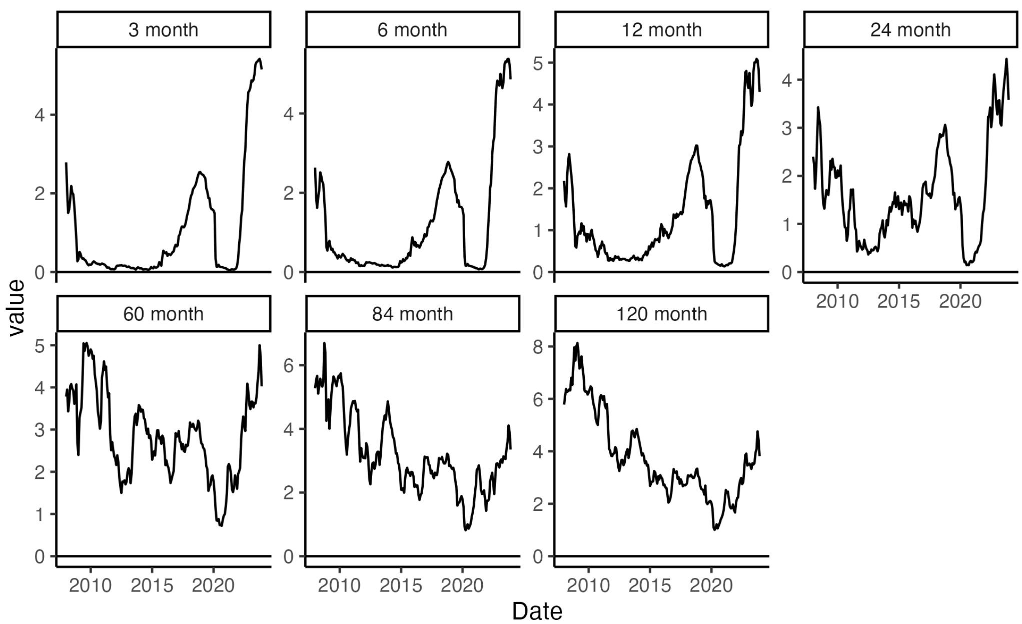


Figure 11: Liu and Wu (2021) zero coupon yield curve estimates, 2006-2023.

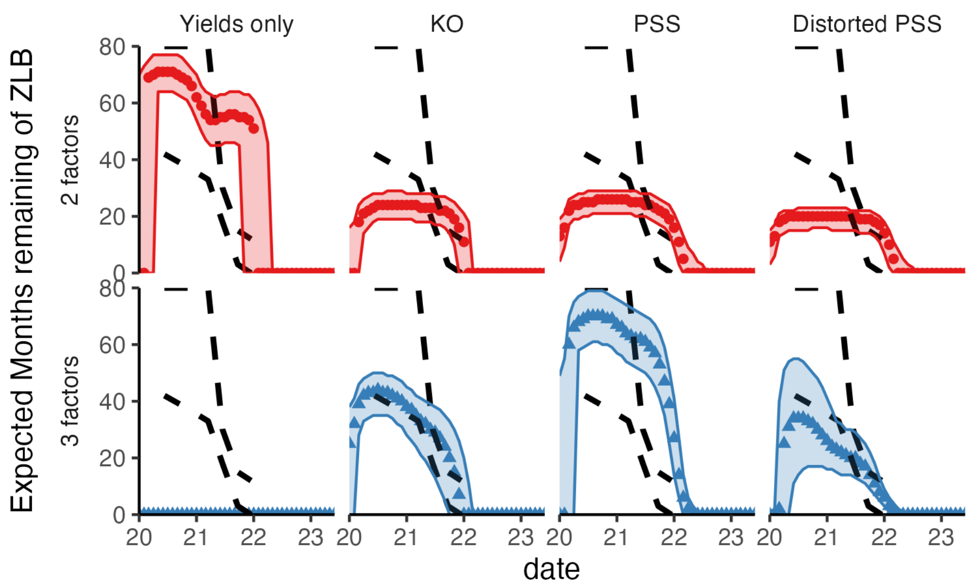


Figure 12: Real-time implied mean duration of ZLB period during and following the COVID-19 pandemic. Bands indicate 99th percentile of liftoff dates. Dashed lines indicate forward guidance implied by median respondent to FOMC Statement of Economic Projections.

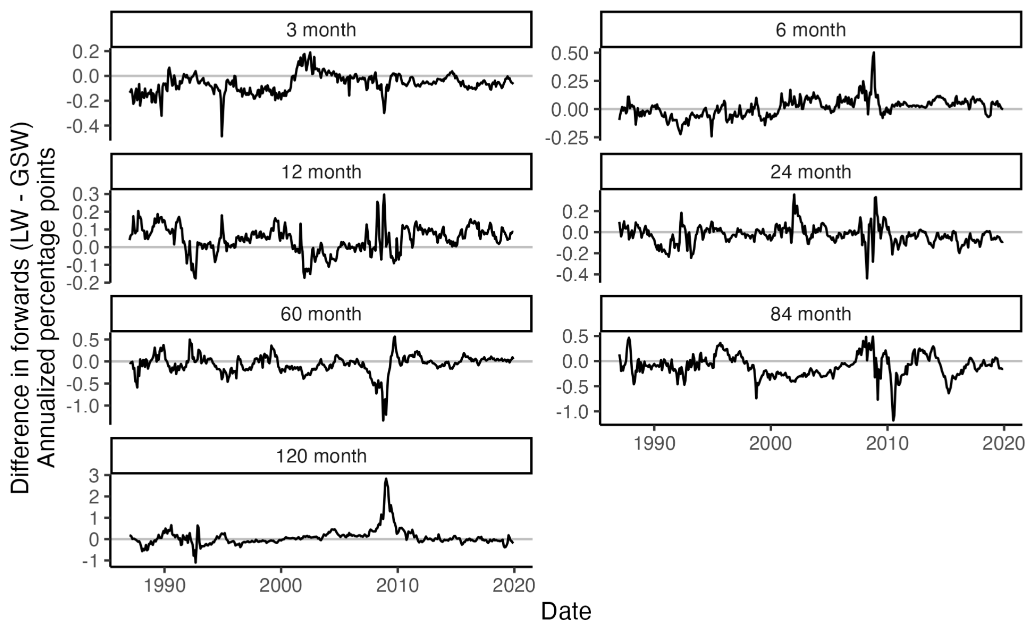


Figure 13: Difference in forward rates (annualized percentage points) for Liu and Wu (2021) zero coupon yield curve estimates relative to Gürkaynak et al. (2007) estimates.

References

Adrian, Tobias, Richard K Crump, and Emanuel Moench (2013) "Pricing the term structure with linear regressions," *Journal of Financial Economics*, Vol. 110, pp. 110–138.

Barakchian, S. Mahdi and Christopher Crowe (2013) "Monetary policy matters: Evidence from new shocks data," *Journal of Monetary Economics*, Vol. 60, pp. 950–966.

Barillas, Francisco and Kristoffer Nimark (2018) "Speculation and the bond market: An empirical no-arbitrage framework," *Management Science*, DOI: `http://dx.doi.org/10.1287/mnsc.2018.3027`.

Bauer, Michael D. and Glenn D. Rudebusch (2014) "The Signaling Channel for Federal Reserve Bond Purchases," *International Journal of Central Banking*, Vol. 10, pp. 233–290.

——— (2016) "Monetary Policy Expectations at the Zero Lower Bound," *Journal of Money, Credit and Banking*, Vol. 48, pp. 1439–1465, URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/jmcb.12338`, DOI: `http://dx.doi.org/10.1111/jmcb.12338`.

Bernanke, Ben S., Jean Boivin, and Piotr Eliasz (2005) "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive

(FAVAR) Approach," *The Quarterly Journal of Economics*, Vol. 120, pp. 387–422.

Black, Fischer (1995) "Interest Rates as Options," *The Journal of Finance*, Vol. 50, pp. 1371–1376.

Boivin, Jean, Michael T. Kiley, and Frederic S. Mishkin (2010) "How How Has the Monetary Transmission Mechanism Evolved Over Time?," *Handbook of Macroeconomics*, Vol. 3, pp. 369–422.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer (2018) "Overreaction in Macroeconomic Expectations," *Working Paper*.

Buraschi, Andrea, Ilaria Piatti, and Paul Whelan (2018) "Rationality and Subjective Bond Risk Premia," *Said Business School Research Papers*, Vol. 36.

Caldara, Dario, Etienne Gagnon, Enrique Martinez-Garcia, and Christopher J. Neely (2020) "Monetary Policy and Economic Performance since the Financial Crisis," *Finance and Economics Discussion Series*.

Carriero, Andrea, Todd E. Clark, Massimiliano Marcellino, and Elmar Mertens (2021) "Forecasting with Shadow-Rate VARs," Working Papers 21-09, Federal Reserve Bank of Cleveland.

.

Christensen, Jens H. E. and Glenn D. Rudebusch (2014) "Estimating Shadow-Rate Term Structure Models with Near-Zero Yields," *Journal of Financial Econometrics*, Vol. 13, pp. 226–259.

Cieslak, Anna (2018) "Short-Rate Expectations and Unexpected Returns in Treasury Bonds," *The Review of Financial Studies*, Vol. 31 (9), pp. 3265–3306.

Clarida, Richard H. (2019) "Models, Markets, and Monetary Policy," May, URL: `https://www.federalreserve.gov/newsevents/speech/clarida20190503a.htm`.

Cochrane, John H. (2018) *The Radical Implications of Stable Quiet Inflation at the Zero Bound*: Hoover Institution, Stanford University.

Cochrane, John H. and Monika Piazzesi (2005) "Bond Risk Premia," *American Economic Review*, Vol. 95, pp. 138–160.

Coibion, Olivier and Yuriy Gorodnichenko (2015) "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review*, Vol. 105, pp. 2644–78.

Colacito, Riccardo, Eric Ghysels, Jinghan Meng, and Wasin Siwasarit (2016) "Skewness in Expected Macro Fundamentals and the Predictability of Equity Returns: Evidence and Theory," *Review of Financial Studies*.

Corrado, Luisa, Stefano Grassi, and Enrico Minnella (2021) "The Transmission Mechanism of Quantitative Easing: A Markov-Switching FAVAR Approach," *CEIS Working Paper No. 520*.

Crump, Richard, Stefano Eusepi, and Emanuel Moench (2018) "The Term Structure of Expectations and Bond Yields," *Federal Reserve Bank of New York Staff Reports No. 775*.

D'Amico, Stefania, William English, David Lopez-Salido, and Edward Nelson (2012) "The Federal Reserve's Large-Scale Asset Purchase Programmes: Rationale and Effects," *The Economic Journal*, Vol. 122, pp. F415–F446.

D'Amico, Stefania and Thomas B. King (2012) "Flow and stock effects of large-scale asset purchases: evidence on the importance of local supply," Finance and Economics Discussion Series 2012-44, Board of Governors of the Federal Reserve System (U.S.).

De La O, Ricardo and Sean Myers (2021) "Subjective Cash Flow and Discount Rate Expectations," *The Journal of Finance*, Vol. n/a, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.13016, DOI: http://dx.doi.org/https://doi.org/10.1111/jofi.13016.

De Rezende, Rafael B. and Annukka Ristiniemi (2023) "A shadow rate without a lower bound constraint," *Journal of Banking & Finance*, Vol. 146, p. 106686, URL: https://www.sciencedirect.com/science/article/pii/S0378426622002667, DOI: http://dx.doi.org/https://doi.org/10.1016/j.jbankfin.2022.106686.

Dick, Christian, Maik Schmeling, and Andreas Schrimpf (2013) "Macro-expectations, aggregate uncertainty, and expected term premia," *European Economic Review*, Vol. 58, pp. 58–80.

Douc, Randal, Eric Moulines, Jimmy Olsson, and Ramon van Handel (2011) "Consistency of the maximum likelihood estimator for general hidden Markov models," *Annals of Statistics*, Vol. 39, pp. 474–513.

Douc, Randal, Éric Moulines, and Tobias Rydén (2004) "Asymptotic Properties of the Maximum Likelihood Estimator in Autoregressive Models with Markov Regime," *The Annals of Statistics*, Vol. 32, pp. 2254–2304.

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

Duffee, Gregory R. (2018) "Expected Inflation and Other Determinants of Treasury Yields," *The Journal of Finance*, Vol. 73 (5), pp. 2139–2180.

Durbin, James and Siem Jan Koopman (2012) *Time Series Analysis by State Space Methods* in , Oxford Statistical Science, No. 9780199641178: Oxford University Press.

D'Amico, Stefania, James Egelhof, Steve Friedman, Mike Kiley, Don Kim, Marcel Priebsch, Matt Raskin, Francisco Vazquez-Grande, and Min Wei (2013) "Assessing the Risk of a Substantial Increase in Long-term Interest Rates," memo, Board of Governors of the Federal Reserve System (U.S.).

Eichenbaum, Martin S (2018) *Comments on The Zero Lower Bound*, United States: Hoover Institution.

Farmer, Leland E. (2017) "Discrete Methods for the Estimation of Nonlinear Economic Models," Doctoral Dissertation, University of California, San Diego.

——— (2021) "The discretization filter: A simple way to estimate nonlinear state space models," *Quantitative Economics*, Vol. 12, pp. 41–76.

Farmer, Leland E. and Alexis Akira Toda (2017) "Discretizing nonlinear, non-Gaussian Markov processes with exact conditional moments," *Quantitative Economics*, Vol. 8, pp. 651–683, URL: `https://onlinelibrary.wiley.com/doi/abs/10.3982/QE737`, DOI: `http://dx.doi.org/https://doi.org/10.3982/QE737`.

Gagnon, Joseph, Matthew Raskin, Julie Remache, and Brian Sack (2011) "The Financial Market Effects of the Federal Reserve's Large-Scale Asset Purchases," *International Journal of Central Banking*, Vol. 7, pp. 3–44.

Gospodinov, Nikolay and Damba Lkhagvasuren (2014) "A Moment-Matching Method For Approximating Vector Autoregressive Processes By Finite-State Markov Chains," *Journal of Applied Econometrics*, Vol. 29, pp. 843–859.

Greenwood, Robin and Andrei Shleifer (2014) "Expectations of Returns and

Expected Returns," *The Review of Financial Studies*, Vol. 27 (3), pp. 714–746.

Gust, Christopher, Edward Herbst, David López-Salido, and Matthew E. Smith (2017) "The Empirical Implications of the Interest-Rate Lower Bound," *American Economic Review*, Vol. 107, pp. 1971–2006, URL: http://www.aeaweb.org/articles?id=10.1257/aer.20121437, DOI: http://dx.doi.org/10.1257/aer.20121437.

Gürkaynak, Refet, Brian Sack, and Jonathan Wright (2007) "The U.S. Treasury yield curve: 1961 to the present," *Journal of Monetary Economics*, Vol. 54, pp. 2291–2304.

Hamilton, James D (1989) "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, Vol. 57, pp. 357–384.

Hamilton, James D. (2018) "The Efficacy of Large-Scale Asset Purchases When the Short-term Interest Rate is at its Effective Lower Bound," *Brookings Papers on Economic Activity*.

——— .

Hanson, Samuel G. and Jeremy Stein (2015) "Monetary Policy and Long-Term Real Rates," *Journal of Financial Economics*, Vol. 115, pp. 429–448.

Huber, Florian and Manfred Fischer (2018) "A Markov Switching Factor-

Augmented VAR Model for Analyzing US Business Cycles and Monetary Policy," *Oxford Bulletin of Economics and Statistics*, Vol. 80, pp. 575–604.

Huther, Jeffrey, Jane Ihrig, and Elizabeth Klee (2017) "The Federal Reserve's Portfolio and its Effect on Interest Rates," *Finance and Economics Discussion Series*.

Johannsen, Benjamin K. and Elmar Mertens (2021) "A Time-Series Model of Interest Rates with the Effective Lower Bound," *Journal of Money, Credit and Banking*, Vol. 53, pp. 1005–1046.

Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu (2011) "A New Perspective on Gaussian Dynamic Term Structure Models," *Review of Financial Studies*, Vol. 24, pp. 926–970.

Kiley, Michael and John Roberts (2017) "Monetary Policy in a Low Interest Rate World," *Brookings Papers on Economic Activity*, Vol. 48, pp. 317–396.

Kim, Chang-Jin (1994) "Dynamic linear models with Markov-switching," *Journal of Econometrics*, Vol. 60, pp. 1–22.

Kim, Don H. and Athanasios Orphanides (2012) "Term Structure Estimation with Survey Data on Interest Rate Forecasts," *Journal of Financial and Quantitative Analysis*, Vol. 47, pp. 241–272.

Kim, Don H. and Marcel Priebsch (2020) "Are Shadow Rate Models of the

Treasury Yield Curve Structurally Stable?," Finance and Economics Discussion Series 61, Board of Governors of the Federal Reserve System (U.S.).

Kim, Don H. and Jonathan H. Wright (2005) "An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates," *Federal Reserve Board Finance and Economics Discussion Series*, Vol. 33.

King, Thomas B. (2019) "Expectation and duration at the effective lower bound," *Journal of Financial Economics*, Vol. 134, pp. 736–760.

Koijen, Ralph S.J., Maik Schmeling, and Evert B. Vrugt (2015) "Survey expectations of returns and asset pricing puzzles," *Working Paper*.

Krippner, Leo (2015a) *Zero Lower Bound Term Structure Modeling: A Practitioner's Guide*, Applied Quantitative Finance: Palgrave Macmillian.

——— (2015b) "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.

——— (2020) "A Note of Caution on Shadow Rate Estimates," *Journal of Money, Credit and Banking*, Vol. 52, pp. 951–962, URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/jmcb.12613`, DOI: `http://dx.doi.org/https://doi.org/10.1111/jmcb.12613`.

Krishnamurthy, Arvind and Annette Vissing-Jorgensen (2011) "The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy," *Brookings Papers on Economic Activity*, Vol. 42, pp. 215–287.

Kuttner, Kenneth N (2018) "Outside the box: Unconventional monetary policy in the great recession and beyond," *Journal of Economic Perspectives*, Vol. 32, pp. 121–46.

Lenza, Michele and Giorgio E. Primiceri (2022) "How to estimate a vector autoregression after March 2020," *Journal of Applied Econometrics*, Vol. 37, pp. 688–699, DOI: `http://dx.doi.org/10.1002/jae.2895`.

Li, Canlin, Andrew C. Meldrum, and Marius del Giudice Rodriguez (2017) "Robustness of Long-Maturity Term Premium Estimates," FEDS Notes 2017-04-03, Board of Governors of the Federal Reserve System (U.S.).

Li, Canlin and Min Wei (2013) "Term Structure Modeling with Supply Factors and the Federal Reserve's Large-Scale Asset Purchase Progarms," *International Journal of Central Banking*, Vol. 9, pp. 3–39.

Litterman, Scheinkman Josè, Robert B (1991) "Common Factors Affecting Bond Returns," Vol. 1.

Liu, Yan and Jing Cynthia Wu (2021) "Reconstructing the yield curve," *Journal of Financial Economics*, Vol. 142, pp. 1395–1425, DOI: `http://dx.doi.org/10.1016/j.jfineco.2021.05`.

Martin, Christopher and Costas Milas (2012) "Quantitative easing: a sceptical survey," *Oxford Review of Economic Policy*, Vol. 28, pp. 750–764.

McCracken, Michael W. and Serena Ng (2015) "Fred-MD: A Monthly Database for Macroeconomic Research," *FRB St. Louis Working Paper No. 2015-12*.

Nakamura, Emi and Jón Steinsson (2018a) "High-frequency identification of monetary non-neutrality: the information effect," *The Quarterly Journal of Economics*, Vol. 133, pp. 1283–1330.

Nakamura, Emi and Jón Steinsson (2018b) "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect*," *The Quarterly Journal of Economics*, Vol. 133, pp. 1283–1330.

Nyholm, Ken (2021) *A Practitioner's Guide to Discrete-Time Yield Curve Modelling: With Empirical Illustrations and MATLAB Examples*, Elements in Quantitative Finance: Cambridge University Press, DOI: `http://dx.doi.org/10.1017/9781108975537`.

Olson, Peter and David Wessel (2017) "We asked Fed watchers to rate the Fed's communications—here's what we found," November, URL: `https://www.brookings.edu/research/we-asked-fed-watchers-to-rate-the-feds-communications/`.

Piazzesi, Monika, Juliana Salomao, and Martin Schneider (2015) "Trend and cycle in bond premia,"Technical report.

Priebsch, Marcel (2013) “Computing Arbitrage-Free Yields in Multi-Factor Gaussian Shadow-Rate Term Structure Models,” Finance and Economics Discussion Series 63, Board of Governors of the Federal Reserve System (U.S.).

——— (2017) “”A Shadow Rate Model of Intermediate-Term Policy Rate Expectations,”,” feds notes, Board of Governors of the Federal Reserve System (U.S.).

Priebsch, Marcel A. (2017) “A Shadow Rate Model of Intermediate-Term Policy Rate Expectations,” DOI: http://dx.doi.org/10.17016/2380-7172.2056.

Ramey, V.A. (2016) “Chapter 2 - Macroeconomic Shocks and Their Propagation,” Vol. 2 of Handbook of Macroeconomics: Elsevier, pp. 71–162, URL: https://www.sciencedirect.com/science/article/pii/S1574004816000045, DOI: http://dx.doi.org/https://doi.org/10.1016/bs.hesmac.2016.03.003.

Ramey, Valerie A. (2016) “Macroeconomic Shocks and Their Propagation,” *Handbook of Macroeconomics*, Vol. 2, pp. 71–162.

Rubin, Donald B. (1976) “Inference and Missing Data,” *Biometrika*, Vol. 63, pp. 581–592.

Schmeling, Maik and Andreas Schrimpf (2011) “Expected inflation, expected stock returns, and money illusion: What can we learn from survey expectations?" *European Economic Review*, Vol. 55, pp. 702–719.

Sims, Christopher A. (1980) "Macroeconomics and Reality," *Econometrica*, Vol. 48, pp. 1–48.

Speekenbrink, Maarten and Ingmar Visser (2021) "Ignorable and non-ignorable missing data in hidden Markov models," *arXiv preprint arXiv:2109.02770*.

Stavrakeva, Vania and Jenny Tang (2018) "Survey-based Exchange Rate Decomposition: New Methodology and New Facts," *Working Paper*.

Struby, Ethan (2018) "Macroeconomic Disagreement in Treasury Yields," working paper, Carleton College Departmetn of Economics.

Swanson, Eric T. (2018) "Measuring the Effects of Federal Reserve Forward Guidance and Asset Purchases on Financial Markets," *NBER Working Paper*.

Swanson, Eric T. and John C. Williams (2014) "Measuring the Effect of the Zero Lower Bound on Medium- and Longer-Term Interest Rates," *American Economic Review*, Vol. 104, pp. 3154–3185.

Wang, Chen (2019) "Under-and Over-Reaction in Yield Curve Expectations."

Wright, Jonathan H. (2011) "Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset," *American Economic Review*, Vol. 101, pp. 1514–1534.

——— (2012) "What does Monetary Policy do to Long-term Interest Rates at the Zero Lower Bound?" *Economic Journal*, Vol. 122, pp. 447–466.

Wu, Jing Cynthia and Fan Dora Xia (2016) "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound," *Journal of Money, Credit and Banking*, Vol. 48, pp. 253–291.

Xia, Dora and Jing Cynthia Wu (2018) "The negative interest rate policy and the yield curve," BIS Working Papers 703, Bank for International Settlements.

Xu, Zhengyang (2019) "Expectation Formation in the Treasury Bond Market," *Working Paper*.

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.

$$\bar{a_n} = \delta_0 + \delta_1 \sum_{k=0}^{n-1} (\rho^{\mathbb{Q}})^k \mu^{\mathbb{Q}} \tag{20}$$

$$a_n = \bar{a_n} - \frac{1}{2} \sum_{j=0}^{n-1} \delta_1 \left[(\rho^{\mathbb{Q}})^j \Sigma \Sigma' ((\rho^{\mathbb{Q}})')^j \right] \delta_1' \tag{21}$$

$$b_n = \delta_1 (\rho^{\mathbb{Q}})^n \tag{22}$$

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.[19] The shift in the long end of the yield curve conflates this news about short term interest rates and economic conditions which affect risk premia.

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

[19]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 greater support to mortgage lending and housing markets, the Committee decided today to increase the size of the Federal Reserve's balance sheet further by purchasing up to an additional $750 billion of agency mortgage-backed securities, bringing its total purchases of these securities to up to $1.25 trillion this year, and to increase its purchases of agency debt this year by up to $100 billion to a total of up to $200 billion. Moreover, to help improve conditions in private credit markets, the Committee decided to purchase up to $300 billion of longer-term Treasury securities over the next six months."

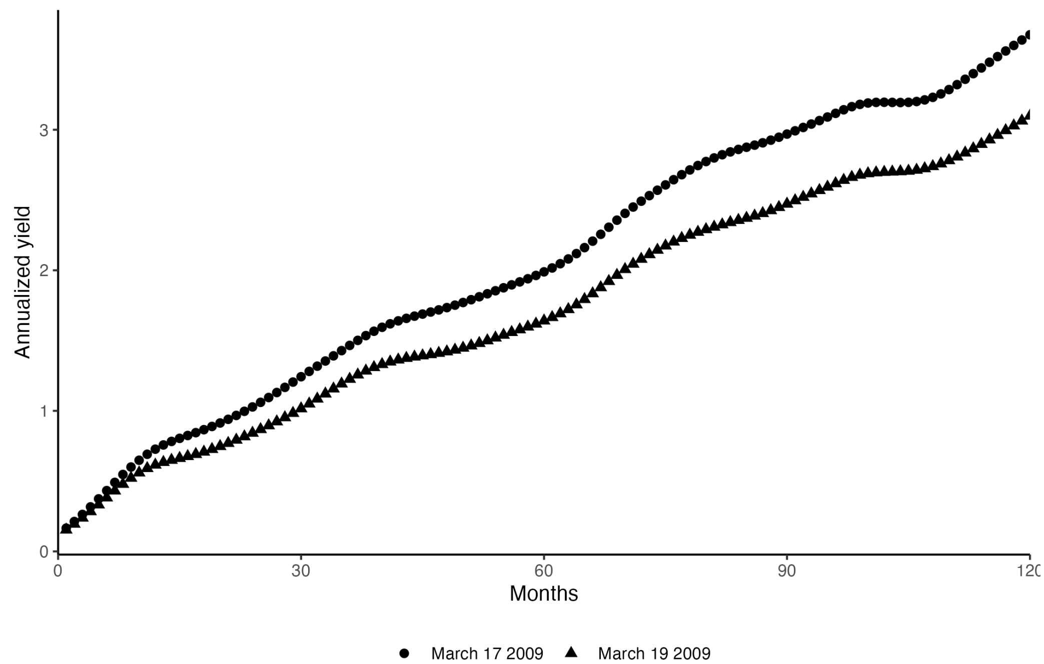


Figure B.1: US Treasury yield curve on March 17 (dashed) and March 19 (circles), 2009. Data from Liu and Wu (2021).

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.

Third, and most importantly, our approach in this paper is neither to

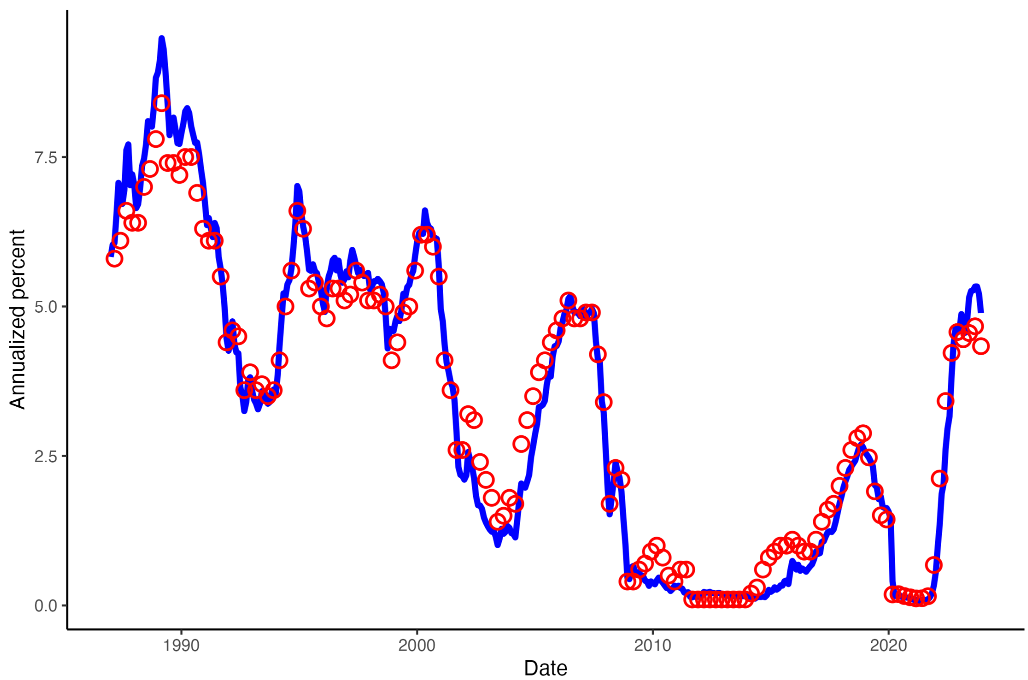


Figure B.2: Yields on 12 month Treasuries (solid line) and average expected short rates of 12 month Treasuries from the Blue Chip Financial Survey (circles). 12 month Treasury yields from Liu and Wu (2021).

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 $\mathrm{E}_t[\bar{r}_{t+n-2,t+n}]$ to be the consensus forecast made in month t of the average level of short rates between months $t+n-2$ and $t+n$.[20] Because Additionally, call $FE_t(\bar{r}_{t+n}) = \bar{r}_{t+n} - \mathrm{E}_t[\bar{r}_{t+n-2,t+n}]$ the forecast *error* from month t to month $t+n$ and $FR_t(\bar{r}_{t+n}) = \mathrm{E}_t[\bar{r}_{t+n-2,t+n}] - \mathrm{E}_{t-1}[\bar{r}_{t+n-2,t+n}]$ the forecast revision between months $t-1$ and t.

We regress forecast errors across horizons n on forecast revisions:

$$FE_t(\bar{r}_{t+n}) = \alpha(n) + \beta(n) FR_t(\bar{r}_{t+n}) + \epsilon_{t+n} \tag{23}$$

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 *underestimate* 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

[20]Further details of the construction of Blue Chip forecasts are provided in appendix C.

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.

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 in 1997).[21] 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

[21]The Blue Chip also publishes long-horizon forecasts semi-annually, which we do not utilize due to the sparse time series.

a quarter already contain past realizations of yields. To address this issue, we adjust forecasts for prior yields within a given quarter.[22] 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 $\mathrm{E}_t[\bar{r}_{t+1,t+2}] = (3 \times \mathrm{E}_t[\bar{r}_{t,t+2}] - \bar{r}_t)/2$. 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: $\mathrm{E}_t[\bar{r}_{t+1}] = 3 \times \mathrm{E}_t[\bar{r}_{t-1,t+1}] - \bar{r}_t - \bar{r}_{t-1}$. In both cases, the average of the first three weeks of the month is assumed to be approximately equals to the monthly average.

[22] This procedure is identical to Xu (2019), except that we use a slightly different forecast horizon convention.

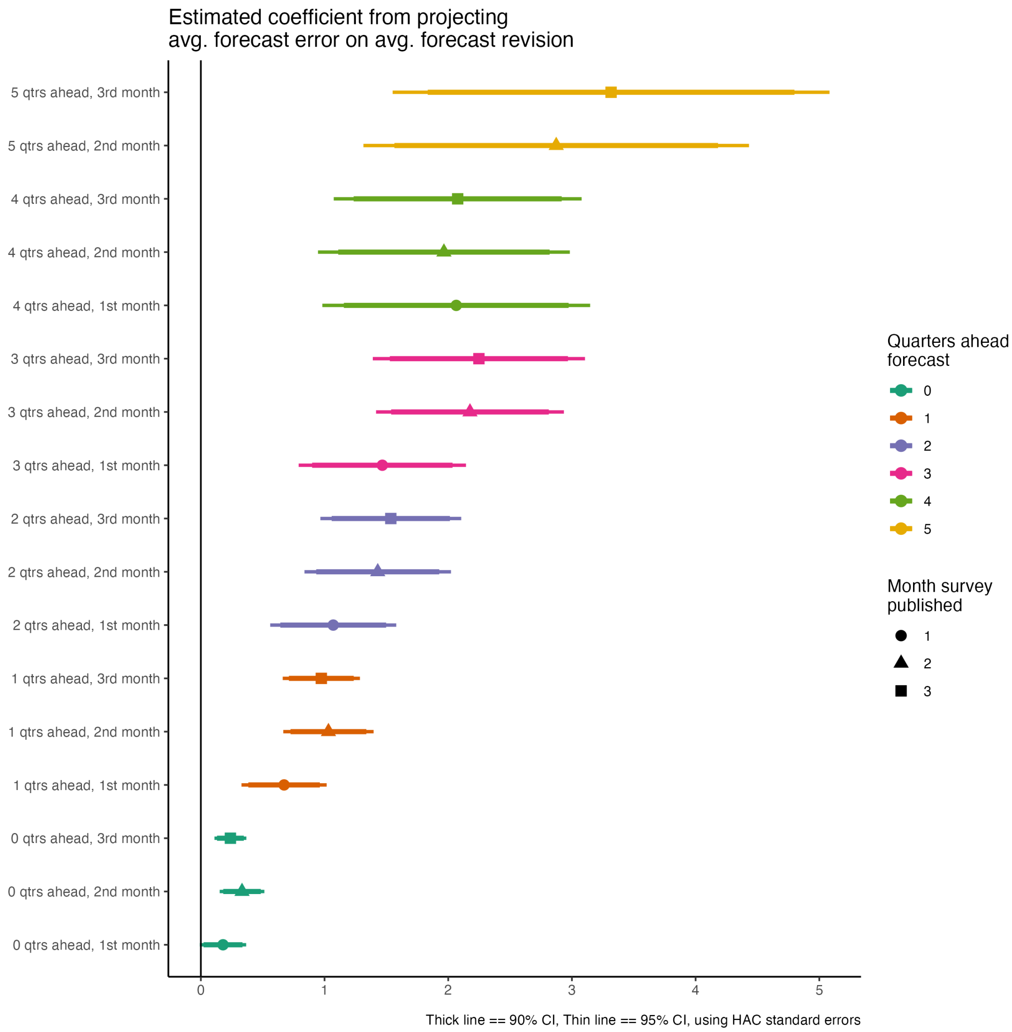


Figure B.3: Each line represents confidence intervals for coefficient estimates of forecast error on forecast revision as in equation (23), by forecast horizon and month within the quarter.

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-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.

1200μ	-0.3432	-0.2369	0.0310
	(0.0231)	(3.6536)	(0.3069)
ρ	0.9638	-0.0069	0.5228
	(0.1366)	(1.3324)	(18.1090)
	-0.0225	0.9385	1.0152
	(0.4016)	(1.6757)	(48.6500)
	0.0031	0.0028	0.8708
	(0.0422)	(0.1105)	(2.3400)
$\text{diag}(\rho^{\mathbb{Q}})$	0.9978	0.9526	0.9586
	(0.0282)	(0.5029)	(0.6506)
1200Σ	0.2938		
	(5.2643)		
	-0.1789	0.2109	
	(4.1672)	(2.5707)	
	-0.0082	0.0032	0.0329
	(0.1417)	(0.2507)	(0.4953)
$1200\ \underline{\text{r}}$	0.2500		
	(1.8187)		
$1200\ \delta_0$	13.5658		
	(37.3359)		
1200 (yield meas. err)	0.6667		
	(0.7202)		
Log Likelihood		17647.7522	

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

1200μ	-0.3043	0.0002
	(0.0000)	(0.0000)
ρ	0.9693	0.0797
	(0.0000)	(0.0000)
	-0.0171	0.8880
	(0.0000)	(0.0000)
diag($\rho^{\mathbb{Q}}$)	0.9931	0.9793
	(0.0000)	(0.0000)
1200Σ	0.8677	0.0000
	(0.0000)	(0.0000)
	-0.6246	0.5324
	(0.0000)	(0.0000)
1200 $\underline{\mathrm{r}}$	0.2111	
	(0.0000)	
1200 δ_0	10.2188	
	(0.0000)	
1200 (yield meas. err)	0.5423	
	(0.0000)	
Log Likelihood		17842.4535

Table D.2: Estimated parameters for 2 factor model without forecasts (YO model). QMLE standard errors in parentheses

1200μ	-0.0578	-0.0487	0.0253
	(0.5868)	(0.9084)	(0.0339)
ρ	0.9888	0.0183	-0.1387
	(0.0885)	(0.1027)	(2.5589)
	-0.0029	0.9793	-0.2899
	(0.0642)	(0.4759)	(3.6838)
	0.0019	0.0117	0.8643
	(0.0059)	(0.0761)	(0.8439)
$\text{diag}(\rho^{\mathbb{Q}})$	0.9956	0.9615	0.8591
	(0.0045)	(0.0949)	(0.4511)
1200Σ	0.3823		
	(0.6024)		
	-0.3718	0.3891	
	(0.1596)	(0.9604)	
	-0.0263	-0.0017	0.0125
	(0.1055)	(0.0124)	(0.2534)
$1200\ \underline{\text{r}}$	0.1669		
	(2.0782)		
$1200\ \delta_0$	10.0723		
	(5.0516)		
1200 (yield meas. err)	1.4327		
	(1.7711)		
1200 (fcst meas. err)	1.3394		
	(1.4785)		
Log Likelihood		29007.0952	

Table D.3: Estimated parameters for 3 factor model including forecasts (KO model). QMLE standard errors in parentheses

1200μ	-0.2730	0.0818
	(0.1037)	(0.7580)
ρ	0.9537	0.0145
	(0.0183)	(0.0485)
	0.0225	0.9179
	(0.1165)	(0.1584)
$\mathrm{diag}(\rho^{\mathbb{Q}})$	0.9942	0.9758
	(0.0029)	(0.0088)
1200Σ	0.5960	
	(0.4049)	
	-0.4729	0.4948
	(0.3337)	(0.1461)
1200 $\underline{\mathrm{r}}$	0.1224	
	(0.0504)	
1200 δ_0	10.1321	
	(0.1900)	
1200 (yield meas. err)	0.9494	
	(1.8260)	
1200 (fcst meas. err)	0.6446	
	(1.1019)	
Log Likelihood		31343.3806

Table D.4: Estimated parameters for 2 factor model including forecasts (KO model). QMLE standard errors in parentheses

1200μ	-0.2289	-0.0220	0.0594
	(0.1187)	(0.1608)	(0.0075)
ρ	0.9838	0.0074	0.3458
	(0.0000)	(0.0026)	(0.5633)
	0.0001	0.9114	0.0521
	(0.0008)	(0.0256)	(0.2183)
	0.0018	0.0056	0.8655
	(0.0064)	(0.0129)	(0.0017)
$\text{diag}(\rho^{\mathbb{Q}})$	0.9958	0.9228	0.8537
	(0.0027)	(0.0625)	(0.2167)
1200Σ	0.5002		
	(0.7863)		
	-0.3072	0.3696	
	(0.0261)	(0.0725)	
	-0.1043	-0.0557	0.0974
	(0.4335)	(0.0399)	(0.2302)
1200 $\underline{\text{r}}$	0.1741		
	(0.1027)		
1200 δ_0	10.0292		
	(1.5995)		
k	18.7778	14.8334	-90.4622
	(0.0018)	(0.0016)	(0.0002)
	64.2452	-67.9476	94.2916
	(0.0012)	(0.0001)	(0.0006)
	4.4544	60.4822	94.2916
	(0.0007)	(0.0003)	(0.0006)
1200 (yield meas. err)	1.0907		
	(1.3479)		
1200 (fcast meas. err)	0.8522		
	(0.9588)		
Log Likelihood		30560.2868	

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

eig($\rho - \Sigma k$)	0.9930
	0.9353
	0.8382

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

1200μ	-0.0319	-0.1410
	(0.0151)	(0.0110)
ρ	0.9571	0.0505
	(0.0220)	(0.0553)
	0.0282	0.9144
	(0.0496)	(0.0877)
diag($\rho^{\mathbb{Q}}$)	0.9961	0.9660
	(0.0052)	(0.0053)
1200Σ	0.5174	
	(0.4285)	
	-0.4412	0.5980
	(0.4593)	(0.0352)
1200 $\underline{r}$	0.2051	
	(0.0532)	
1200 δ_0	10.1156	
	(3.1573)	
k	-0.4713	-27.7791
	(0.6283)	(8.1201)
	19.9101	47.5442
	(27.3278)	(84.2557)
1200 (yield meas. err)	0.9272	
	(1.3466)	
1200 (fcast meas. err)	0.8635	
	(1.0583)	
Log Likelihood		30931.7551

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

eig($\rho - \Sigma k$)	0.9699
	0.8678

Table D.8: Subjective physical dynamics, 2 factor PSS model

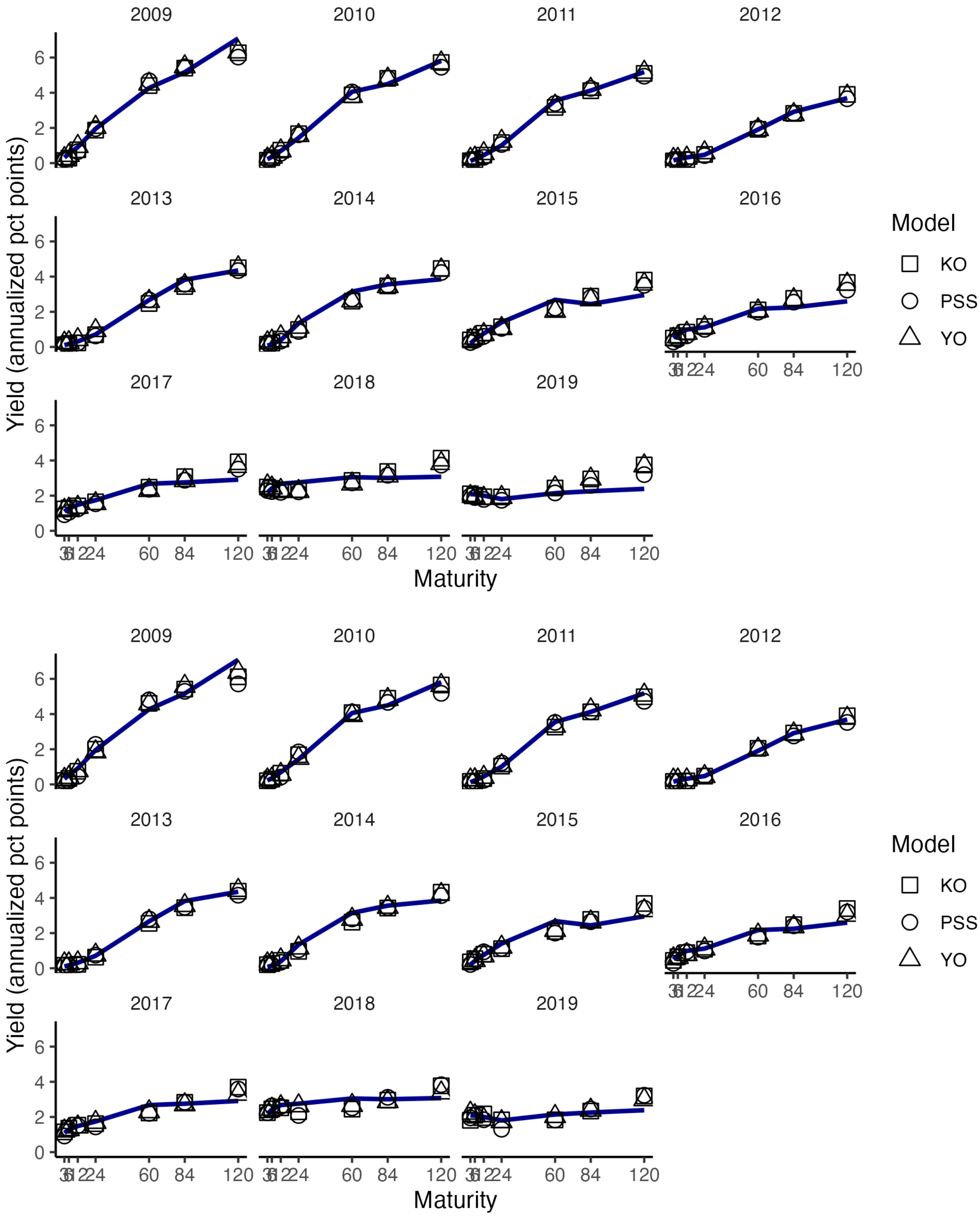


Figure D.1: Average fit of 2-figure (top) and 3-figure (bottom) model across years, using smoothed state estimates. Line indicates average yield curve in the indicated year.

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.

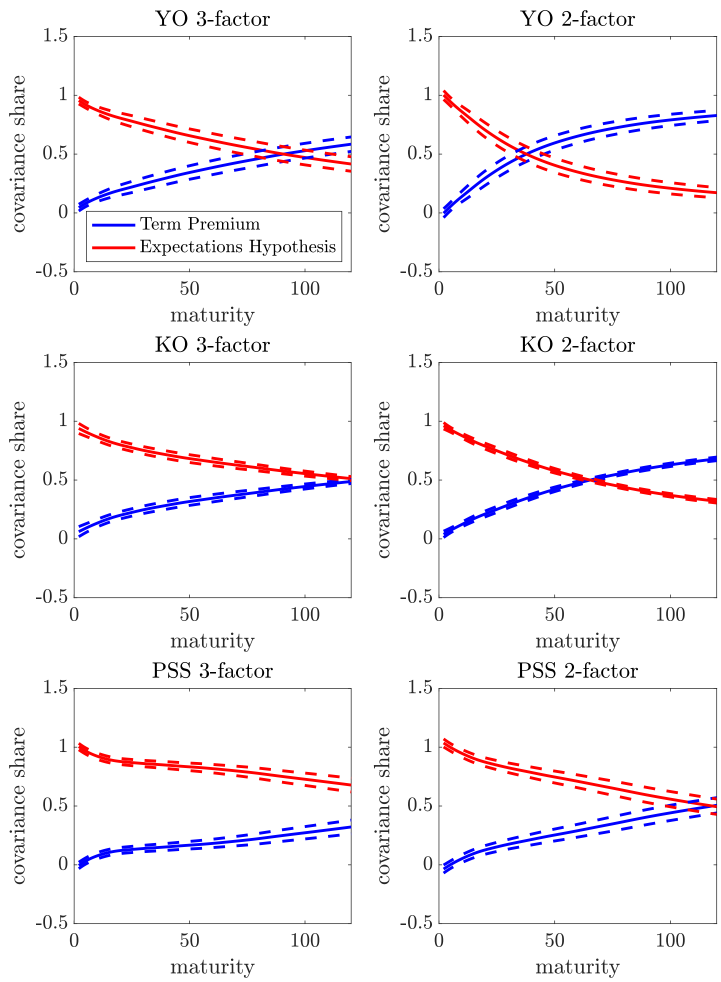


Figure E.1: Decomposition of unconditional variance of the level of yields into Term Premium and Expectations Hypothesis components for yields only, KO, and PSS models for the full sample (1987-2018). Newey-West 99% confidence bands are shown for point estimates. Maturity is reported in months.

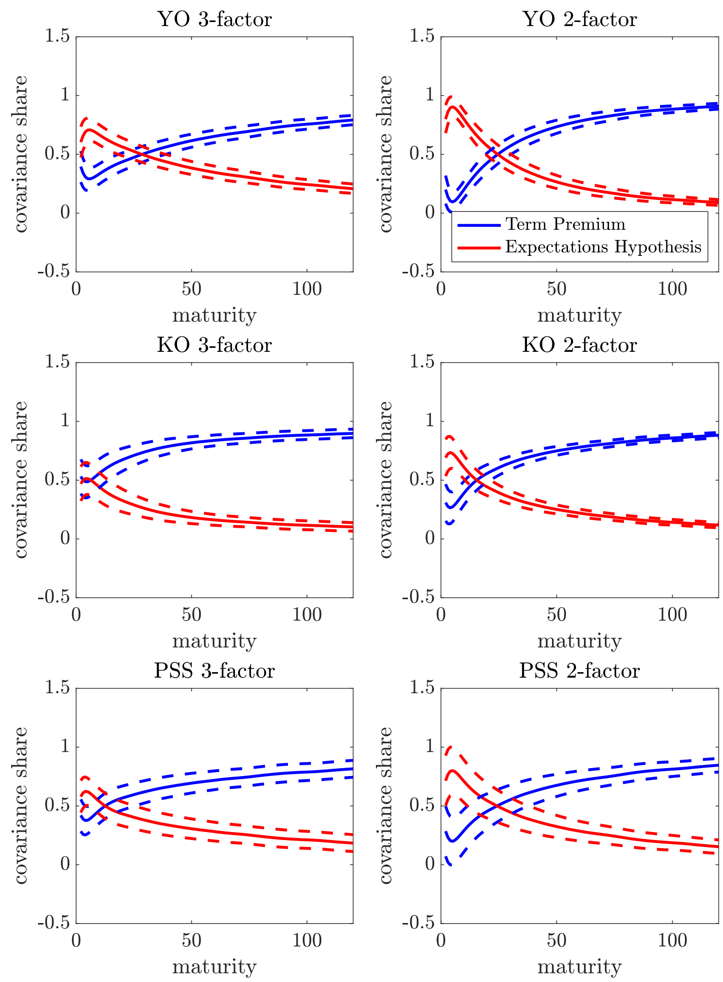


Figure E.2: Decomposition of unconditional variance of the change in yields into Risk Premium and Expectations Hypothesis components for yields only, KO, and PSS models for the ZLB sample (1987-2018). Newey-West 99% confidence bands are shown for point estimates. Maturity is reported in months.

F Two and three factor results

In this appendix, we report analogous results for both two- and three-factor models for policy applications.

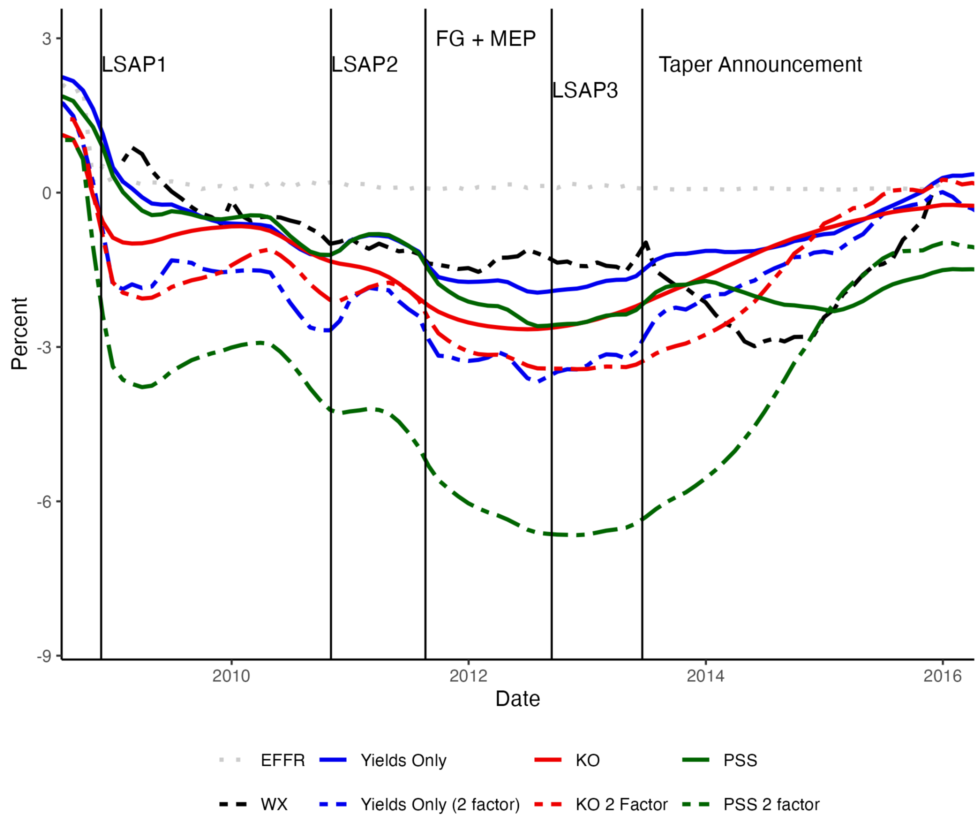


Figure F.1: Smoothed estimates of shadow rates for both two- and three-factor models during/post Great Recession, with event dates (three rounds of Large Scale Asset Purchases (LSAPs), the introduction of calendar-based forward guidance and the Maturity Extension Program (FG+MEP), Taper Announcement). FG and MEP were introduced in August and September 2011, respectively, but are shown in August 2011.

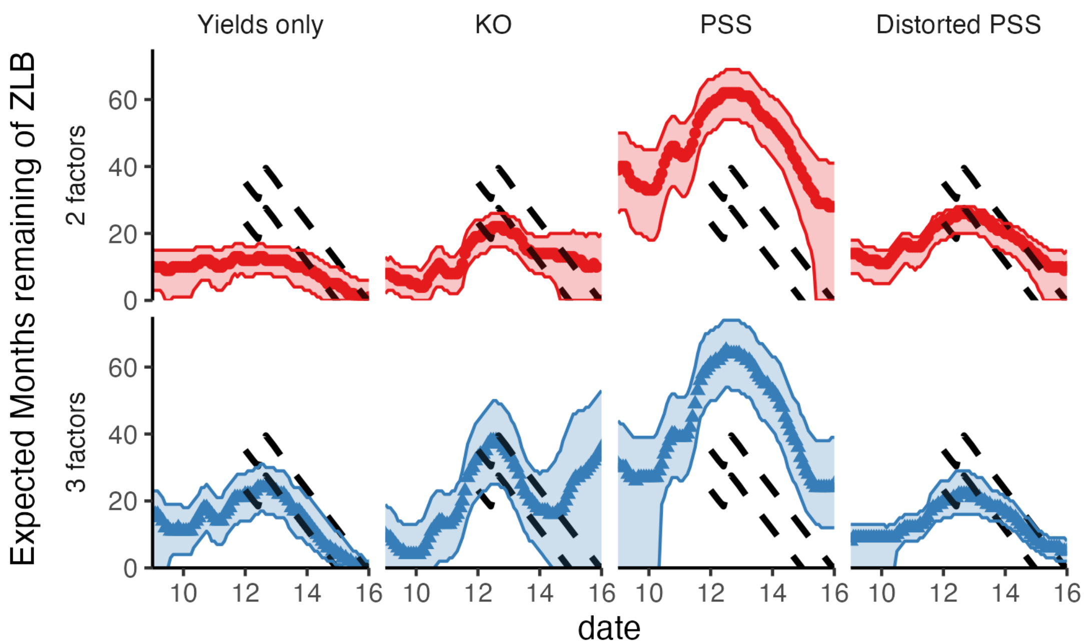


Figure F.2: Real-time implied mean duration of ZLB period. Bands indicate 99th percentile of liftoff dates.

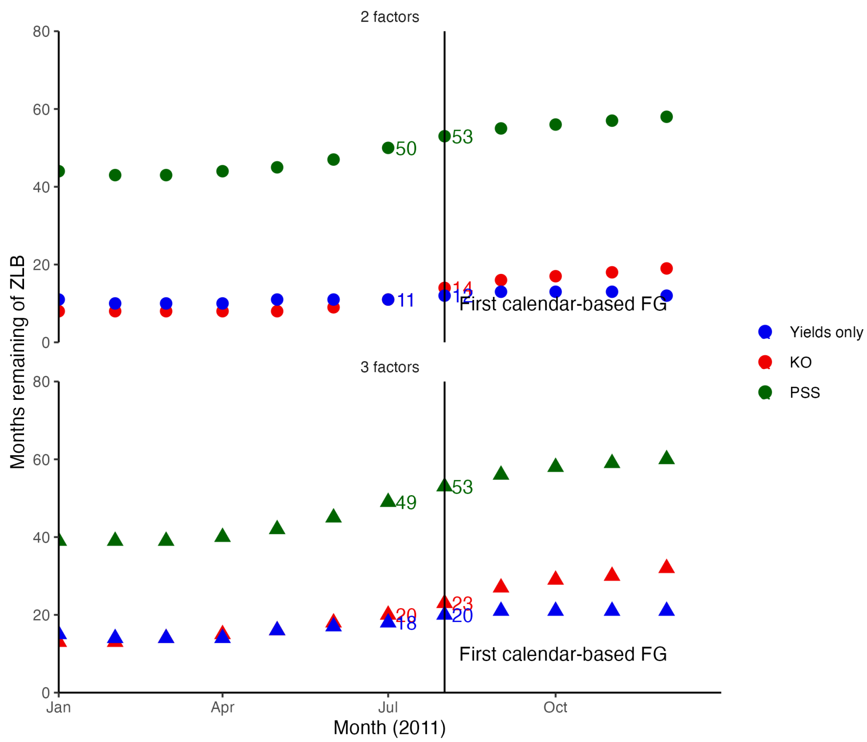


Figure F.3: Real-time implied mean duration of 2011 using the same estimates as those in Figure 3.

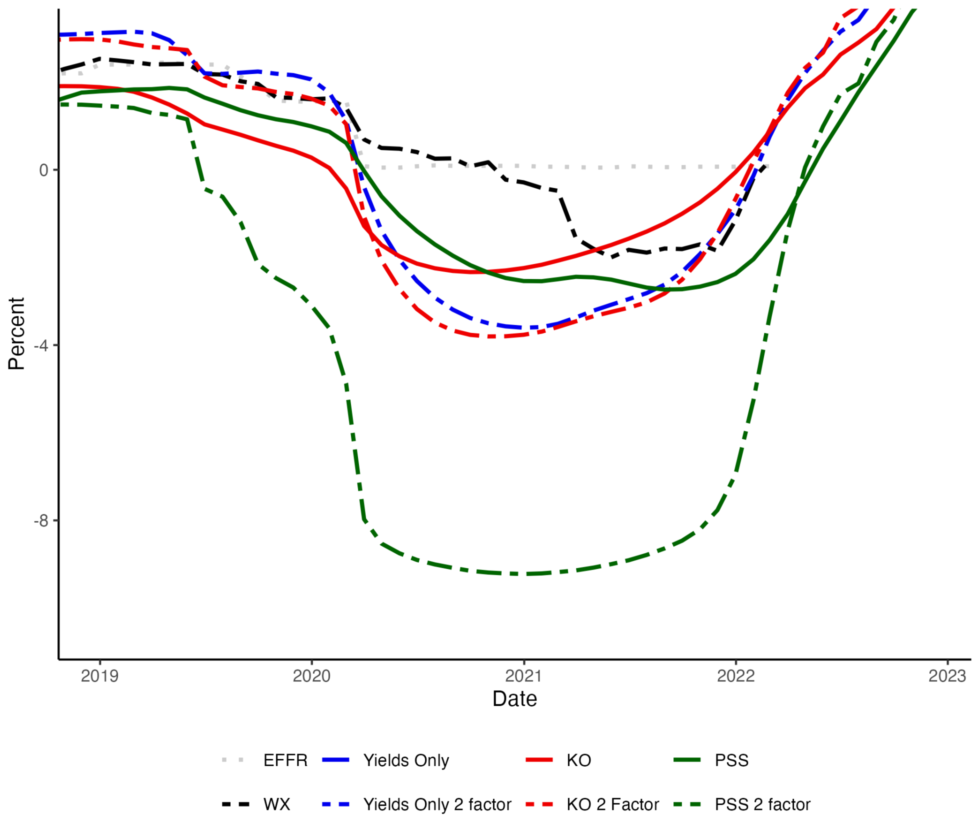


Figure F.4: Smoothed estimates of effective federal funds rate and shadow rate estimates during and after the COVID-19 recession.

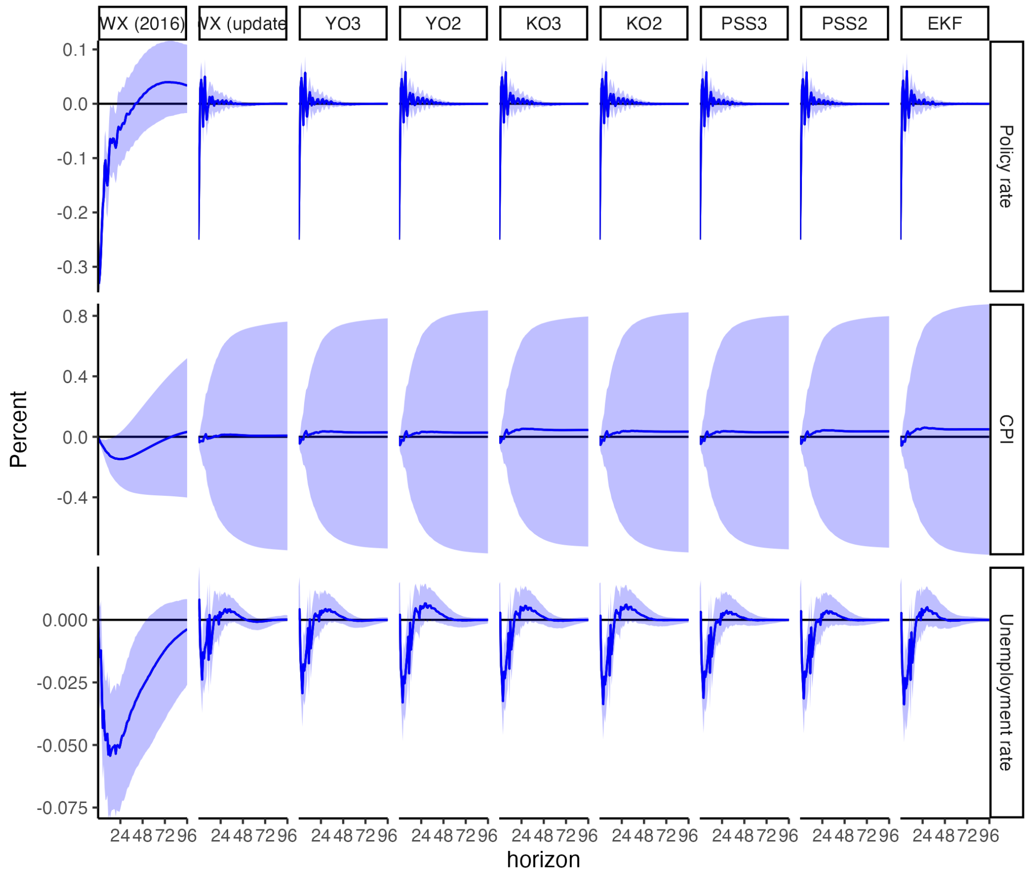


Figure F.5: Estimated impulse response of unemployment to a 25 basis point decrease in the policy rate (Federal funds rate/shadow rate). Solid lines indicate median response and bands indicate 90% confidence intervals. The difference across columns is the underlying estimate of the shadow rate. "WX (2016)" indicates the original impulse response reported in Wu and Xia (2016); "WX (updated)" uses the most recent vintage of data as of December 2024. "EKF" is the YO3 factor model estimated using Li and Wei (2013) yields and the extended Kalman filter.

G EKF 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 $\underline{r}$. We then restrict $\underline{r} = 0.25$ 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

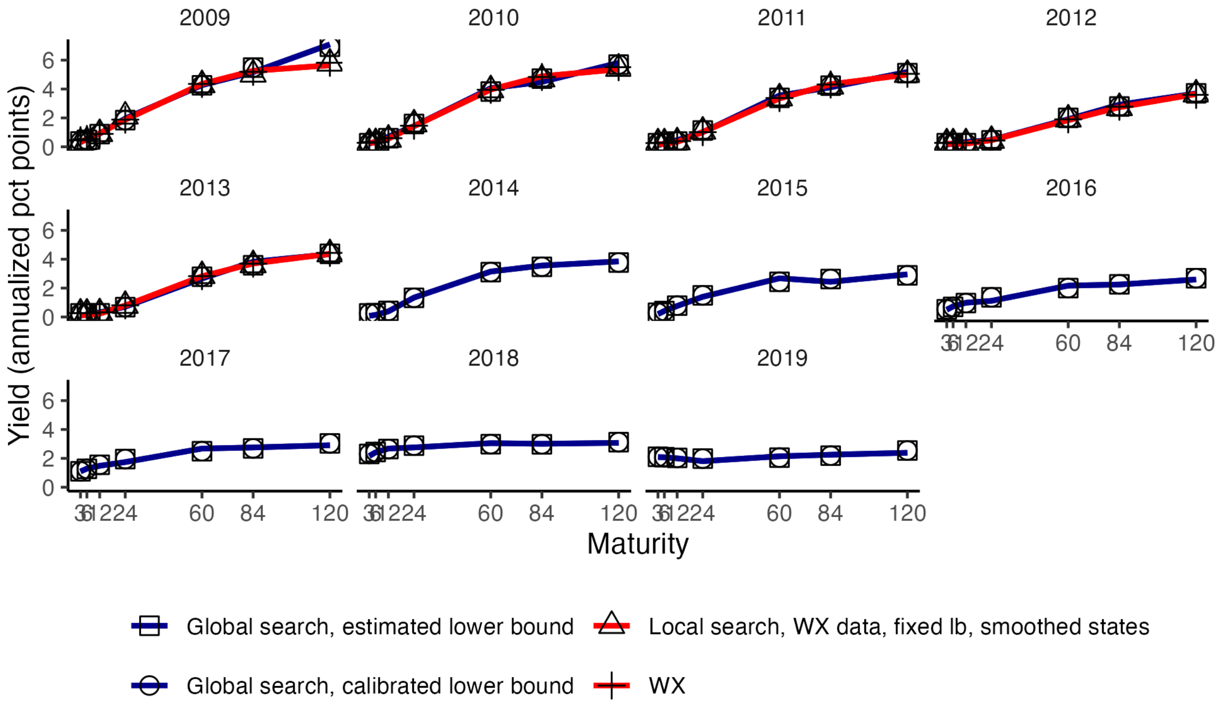


Figure G.1: Predicted and actual forward curves for 3 factor YO model estimated with extended Kalman filter. Blue line: Gürkaynak et al. (2007) forward rate curves. Square: estimated with global search and without fixing $\underline{r}$. Triangle: Global search with $\underline{r} = 0.25$. Circle: local search with $\underline{r} = 0.25$. Plus sign: Results from Wu and Xia (2016).

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

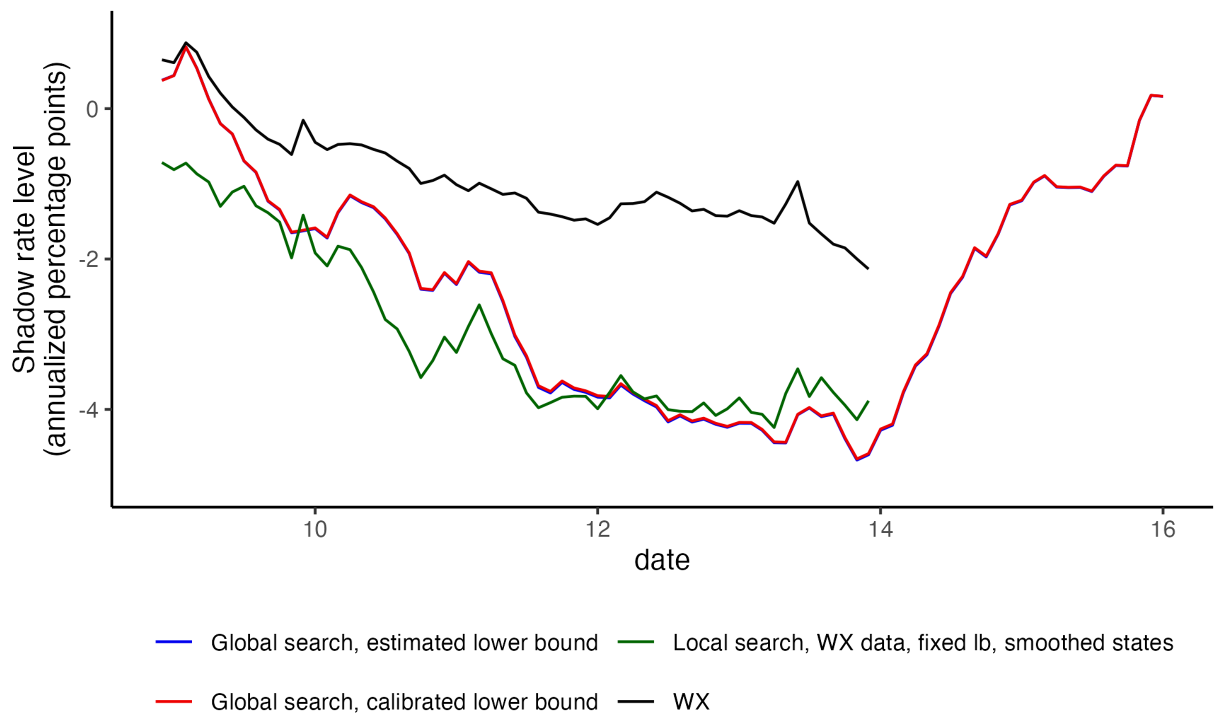


Figure G.2: Estimated shadow rates from 3 factor YO model estimated with extended Kalman filter. Blue: estimated with global search and without fixing $\underline{r}$. Red: Global search with $\underline{r} = 0.25$. Green: local search with $\underline{r} = 0.25$. Black: Results from Wu and Xia (2016).

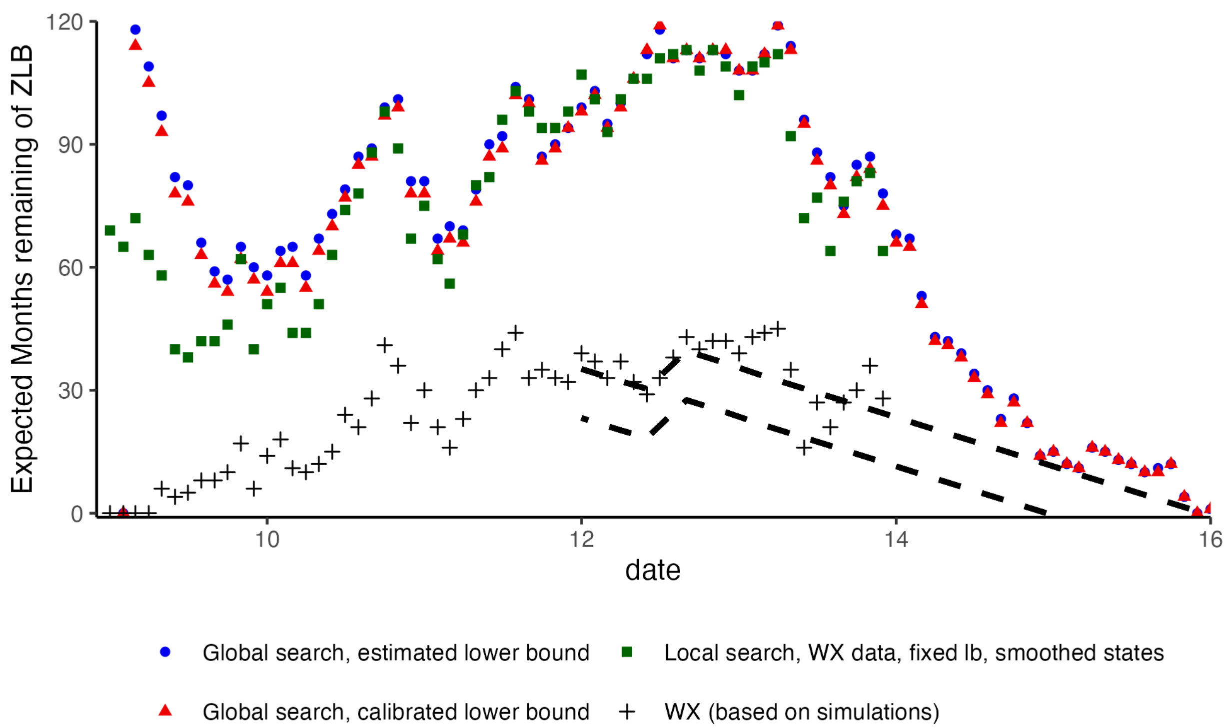


Figure G.3: Implied duration of ZLB from 3 factor YO model estimated with extended Kalman filter. Blue circle: estimated with global search and without fixing $\underline{r}$. Red triangle: Global search with $\underline{r} = 0.25$. Green square: local search with $\underline{r} = 0.25$. Black plus sign: 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.

.

results with EKF results estimated using global methods and a fixed lower bound at $\underline{r} = .25$. 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 Smaller, with the difference being largest at the long end of the yield curve (22 bp). On the other hand, we achieve reasonably similar average fits despite the grid approximation method of the discretization filter, particularly at the short end. . 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 outperforms the YO3 model on average, although the gains are marginal at the short end. During the COVID period, the YO three-factor model does better for short horizons and similarly at long horizons. The performance of the models with forecasts at the short end is superior for both the sub-periods of the forecasting exercise.

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 (396 months). Panel B contains estimates for the out-of-sample fit, which estimates the model each December from 2007-2023, and calculates forecasts for 1- to 12-months ahead. Panels C and D report subcomponents of the out-of-sample forecasts splitting the sample before and after the COVID-19 pandemic. MAE and RMSE are reported across all horizons (10 sets of forecasts at 12 horizons each).

Statistic	MAE	RMSE
Panel A: In-Sample Fit (N=396)		
3mo	0.12	0.14
6mo	0.06	0.09
12mo	0.11	0.15
24mo	0.12	0.16
60mo	0.15	0.19
84mo	0.15	0.21
120mo	0.10	0.14
Panel B: Out-of-Sample Fit: 1-12 month-ahead forecasts (N=180)		
3mo	0.31	0.60
6mo	0.33	0.62
12mo	0.37	0.60
24mo	0.42	0.56
60mo	0.49	0.68
84mo	0.52	0.71
120mo	0.51	0.68
Panel B: Out-of-Sample Fit, 2007-2019, 1-12 month-ahead forecasts (N=132)		
3mo	0.20	0.31
6mo	0.21	0.33
12mo	0.25	0.34
24mo	0.34	0.45
60mo	0.49	0.67
84mo	0.55	0.76
120mo	0.53	0.70
Panel B: Out-of-Sample Fit, 2020-2023,-1-12 month-ahead forecasts (N=48)		
3mo	0.31	0.60
6mo	0.33	0.62
12mo	0.37	0.60
24mo	0.42	0.56
60mo	0.49	0.68
84mo	0.52	0.71
120mo	0.51	0.68

H Additional figures from COVID-19 period

This appendix contains parameter estimates and measures of model fit for each model when we extend the estimation period to the end of 2023.

1200μ	-0.0597	-0.2151	0.2046
	(3.2668)	(0.0061)	(6.8541)
ρ	0.9132	0.0127	-0.0267
	(0.4008)	(0.0594)	(0.3640)
	-0.0862	0.9088	-0.0476
	(1.0136)	(1.0024)	(0.1787)
	0.5249	0.4330	0.9143
	(6.7627)	(6.1827)	(0.1319)
$\text{diag}(\rho^{\mathbb{Q}})$	0.8985	0.9243	0.9428
	(0.6319)	(0.0466)	(0.0324)
1200Σ	0.4249		
	(2.9425)		
	-0.1358	0.6236	
	(1.7546)	(0.9734)	
	-0.4900	0.4665	0.6459
	(0.5240)	(0.7956)	(0.9900)
1200 $\underline{\text{r}}$	-0.2403		
	(3.8958)		
1200 δ_0	13.5130		
	(12.8653)		
1200 (yield meas. err)	2.8574		
	(6.5310)		
Log Likelihood		14689.9476	

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

1200μ	-0.1974	-0.2251
	(0.0000)	(0.0000)
ρ	0.9448	0.0329
	(0.0000)	(0.0000)
	0.0548	0.8500
	(0.0000)	(0.0000)
$\text{diag}(\rho^{\mathbb{Q}})$	0.9980	0.9531
	(0.0000)	(0.0000)
1200Σ	0.5523	0.0000
	(0.0000)	(0.0000)
	-0.3358	0.5556
	(0.0000)	(0.0000)
$1200\ \underline{\text{r}}$	0.2314	
	(0.0000)	
$1200\ \delta_0$	11.0194	
	(0.0000)	
1200 (yield meas. err)	0.5890	
	(0.0000)	
Log Likelihood		19488.0762

Table H.2: Estimated parameters for 2 factor model without forecasts (YO model). QMLE standard errors in parentheses

1200μ	-0.0582	-0.0479	0.0488
	(0.2308)	(0.1663)	(0.0014)
ρ	0.9893	0.0162	-0.0127
	(0.0146)	(0.0150)	(0.0414)
	-0.0024	0.9796	-0.2913
	(0.0179)	(0.1264)	(0.9352)
	0.0075	0.0140	0.8695
	(0.0134)	(0.0395)	(0.0494)
$\text{diag}(\rho^{\mathbb{Q}})$	0.9960	0.9616	0.8586
	(0.0058)	(0.0383)	(0.0876)
1200Σ	0.3824		
	(0.0452)		
	-0.3702	0.3727	
	(0.1154)	(0.1200)	
	-0.0238	-0.0027	0.0100
	(0.0832)	(0.0163)	(0.0350)
$1200\ \underline{\mathrm{r}}$	0.2847		
	(0.0749)		
$1200\ \delta_0$	10.0468		
	(11.5554)		
1200 (yield meas. err)	1.3493		
	(1.7248)		
1200 (fcst meas. err)	1.2400		
	(1.4425)		
Log Likelihood		32923.7398	

Table H.3: Estimated parameters for 3 factor model including forecasts (KO model). QMLE standard errors in parentheses

1200μ	-0.2730	0.0818
	(0.6159)	(0.9105)
ρ	0.9537	0.0145
	(0.1030)	(0.1802)
	0.0225	0.9179
	(0.1536)	(0.6665)
diag$(\rho^{\mathbb{Q}})$	0.9942	0.9758
	(0.0033)	(0.0572)
1200Σ	0.5960	
	(0.1927)	
	-0.4729	0.4948
	(0.1978)	(0.4130)
1200 $\underline{\text{r}}$	0.1224	
	(0.5777)	
1200 δ_0	10.1321	
	(0.9655)	
1200 (yield meas. err)	0.9494	
	(1.0457)	
1200 (fcst meas. err)	0.6446	
	(0.8210)	
Log Likelihood		35157.4959

Table H.4: Estimated parameters for 2 factor model including forecasts (KO model). QMLE standard errors in parentheses

1200μ	-0.0351	-0.1054	-0.0943
	(0.2585)	(0.6623)	(0.2894)
ρ	0.9157	0.0380	-0.0075
	(0.4349)	(0.1263)	(0.0257)
	0.2138	0.8799	-0.0303
	(0.8587)	(0.1907)	(0.0736)
	0.1754	-0.0355	0.8502
	(0.0936)	(0.1847)	(0.1174)
$\text{diag}(\rho^{\mathbb{Q}})$	0.9999	0.9613	0.9468
	(0.0102)	(0.0044)	(0.0157)
1200Σ	0.2429		
	(1.2436)		
	-0.4031	0.5716	
	(2.6091)	(0.0659)	
	0.0274	0.1088	0.2250
	(0.2231)	(0.2740)	(0.3756)
1200 $\underline{\text{r}}$	0.4961		
	(0.4035)		
1200 δ_0	10.0084		
	(0.4093)		
k	-69.1248	-64.9436	-53.6921
	(80.0687)	(78.2228)	(63.1861)
	6.2826	21.9106	-13.4181
	(14.4430)	(37.5347)	(72.4396)
	-58.8692	-45.0318	-13.4181
	(11.0421)	(10.1943)	(72.4396)
1200 (yield meas. err)	2.6259		
	(4.8342)		
1200 (fcast meas. err)	0.7964		
	(1.1710)		
Log Likelihood		31592.0330	

Table H.5: Estimated parameters for 3 factor model with distorted forecaster dynamics (PSS model). QMLE standard errors in parentheses

$\text{eig}(\rho - \Sigma k)$	0.7544
	0.9845
	0.8936

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

1200μ	-0.5000	0.0584
	(0.2669)	(0.0576)
ρ	0.9707	-0.0149
	(0.0101)	(0.0343)
	0.0016	0.9303
	(0.0114)	(0.0674)
$\text{diag}(\rho^{\mathbb{Q}})$	0.9947	0.9685
	(0.0009)	(0.0357)
1200Σ	0.6616	
	(0.0379)	
	0.0226	0.9699
	(0.0950)	(0.0755)
$1200\ \underline{\text{r}}$	0.5000	
	(0.5455)	
$1200\ \delta_0$	16.6512	
	(3.9729)	
k	-31.5626	-40.9760
	(28.3081)	(30.9323)
	37.6006	-13.8504
	(30.9087)	(31.0041)
1200 (yield meas. err)	1.1259	
	(1.6982)	
1200 (fcast meas. err)	0.9111	
	(1.2025)	
Log Likelihood		33854.8993

Table H.7: Estimated parameters for 2 factor model with distorted forecaster dynamics (PSS model). QMLE standard errors in parentheses

$\text{eig}(\rho - \Sigma k)$	0.9827
	0.9477

Table H.8: Subjective physical dynamics, 2 factor PSS model

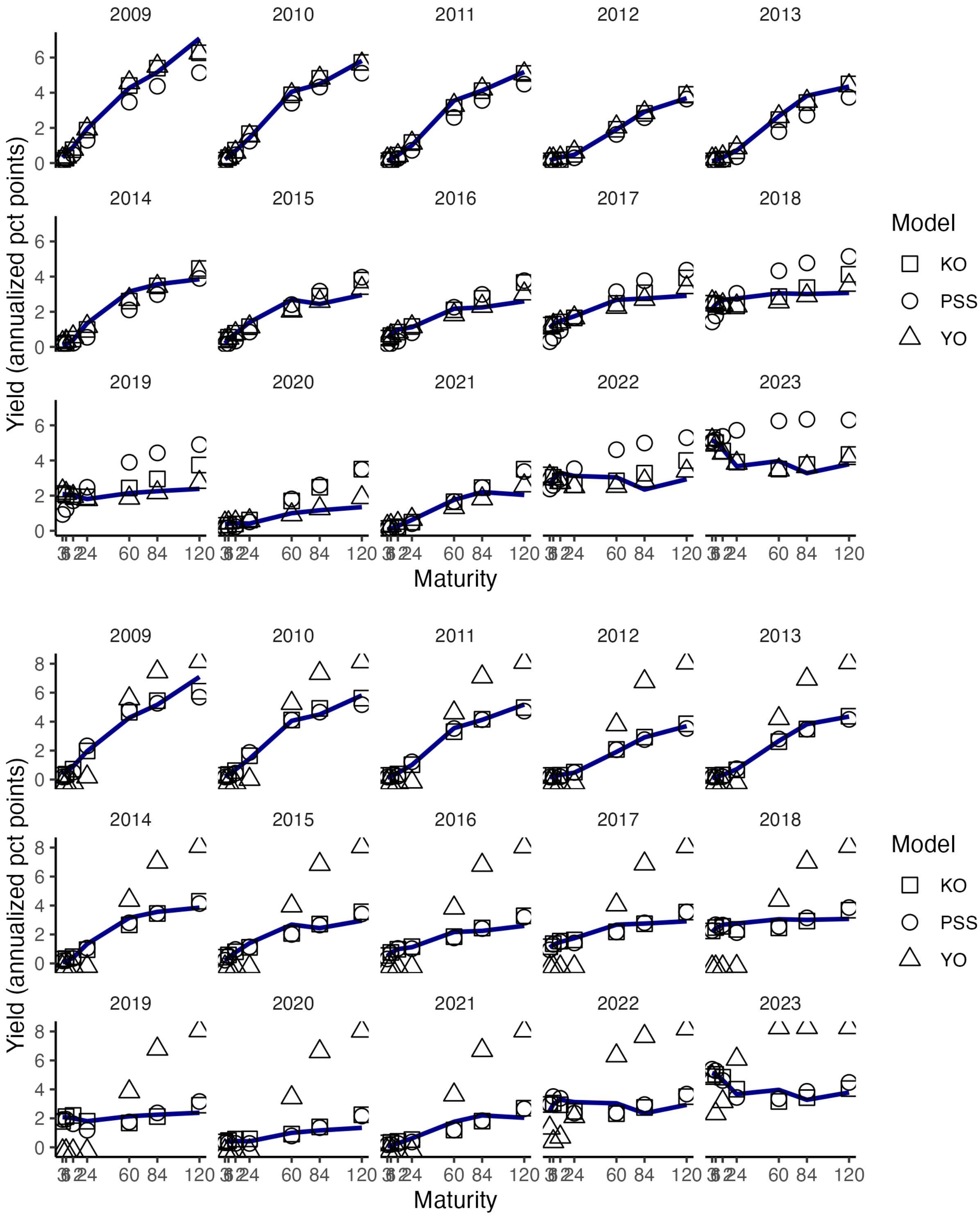


Figure H.1: Average fit of 2-figure (top) and 3-figure (bottom) model across years for extended sample, using smoothed state estimates. Line indicates average yield curve in the indicated year.