Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

Are Shadow Rate Models of the Treasury Yield Curve Structurally Stable?

Don H. Kim and Marcel A. Priebsch

2020-061

Please cite this paper as:
Kim, Don H., and Marcel A. Priebsch (2020). "Are Shadow Rate Models of the Treasury Yield Curve Structurally Stable?," Finance and Economics Discussion Series 2020-061. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2020.061.

Are Shadow Rate Models of the Treasury Yield Curve Structurally Stable?

Don H. Kim and Marcel A. Priebsch*

June 15, 2020

Abstract

We examine the structural stability of Gaussian shadow rate term structure models of Treasury yields over a sample that includes the period from 2008 onwards during which the U.S. policy rate was at its effective lower bound. After a conceptual discussion of several potential sources of a structural break in the context of the shadow rate model, we document various pieces of evidence for structural instability based on predictive tests and Lagrange multiplier tests, as well as with separate estimations of the pre-ELB and post-ELB subsamples. In order to overcome the difficulties associated with the latent-factor nature of the model in testing for a structural break, we focus on objects that can be given intuitive interpretation, such as principal components, or that are constructed to be invariant to factor rotations.

Keywords: shadow rate term structure models, Treasury yields, ELB, structural break, structural instability, factor rotations, principal components

JEL classification: E43, E44, G12

*Division of Monetary Affairs, Board of Governors of the Federal Reserve System; don.h.kim@frb.gov; marcel.a.priebsch@frb.gov. This paper has evolved from our earlier preliminary paper entitled "Estimation of Multi-Factor Shadow Rate Term Structure Models." The analysis and conclusions set forth in this talk are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors of the Federal Reserve System. We thank Michiel De Pooter, Hiroatsu Tanaka, Jonathan Wright, and seminar participants at the Federal Reserve Board for helpful comments.

1 Introduction

From 2008 to 2015, the U.S. policy rate was set at its effective lower bound (ELB).[1] This experience has provided an impetus for studying yield dynamics using term structure models that respect the ELB constraint. Much of that effort has been through the use of shadow rate term structure models, which have risen in popularity in recent years, e.g., Kim and Singleton (2012), Krippner (2013), Bauer and Rudebusch (2016), Christensen and Rudebusch (2014), Wu and Xia (2016). These models can capture, in a natural way, some of the key qualitative features of the ELB yield dynamics, such as the shape of the yield curve near the ELB and the compression of yield volatilities for short maturities. Another part of the attraction of these models has been conceptual: It has a parallel with macroeconomic models incorporating an ELB in which the policy rate is analogously described in a "censored form" $\max\{s_t, \underline{r}\}$, where $\underline{r}$ is the ELB and s_t follows some variant of a Taylor rule.[2]

Intuitively, many properties of yields can be expected to change once the short rate is at or near the ELB. For example, as further discussed later in the paper, principal component analysis (PCA) of yields leads to substantially different results for the ELB period vs. the pre-ELB period. This is obviously problematic for standard affine-Gaussian models, in particular those that use observed principal components as state variables (e.g., Joslin, Priebsch, and Singleton (2014)). However, one might still hope that the nonlinearity of the ELB regime that gives rise to structural break patterns in affine-Gaussian models and PCA can be captured with a *structurally stable* shadow rate term structure model. In this case, the same set of shadow rate model parameters describes the yield dynamics in the non-ELB and ELB periods, and therefore the parameters of the shadow rate model that have been estimated with a pre-ELB sample can be used for analyzing the ELB sample. This is particularly convenient, as the shadow rate model can be well approximated by a (more tractable) affine model in the pre-ELB regime. In fact, to our knowledge, *all* existing studies with shadow rate term structure models, including those mentioned above, have assumed structural stability across non-ELB and ELB periods.[3]

However, there are several reasons to be concerned about potential structural instability of shadow rate term structure models:

1. State variables in term structure models, even in latent-factor models where they are not explicitly equated to macro variables, are thought to have macroeconomic underpinnings. Economic intuition suggests that the dynamics of key variables such as inflation and output gap may change once the economy enters the ELB regime.[4]

[1] The Federal Open Market Committee (FOMC) specified a target range for the federal funds rate of 0 to 0.25 percent. Our sample ends before the most recent ELB episode in response to the COVID-19 outbreak.

[2] See, for example, Reifschneider and Williams (2000), Eggertsson and Woodford (2003), Nakata and Tanaka (2016), and Johannsen and Mertens (2016).

[3] In concurrent independent work, Andreasen, Jørgensen, and Meldrum (2019) document patterns in bond excess return regression that are suggestive of structural instability of shadow rate models, and propose a regime-switching shadow rate model.

[4] In Section 2, we provide a more detailed discussion referencing several macro models incorporating an ELB.

2. Even if the data-generating ($\mathbb{P}$-measure) state vector dynamics were stable, their market price of risk may not necessarily remain stable across the non-ELB/ELB regimes. Indeed, as we shall discuss in detail below, certain mechanisms may imply a structural change in the dynamics of the market price of risk.

3. While the Federal Reserve provided accommodation during the ELB years in the form of unconventional monetary policy (specifically, forward guidance and asset purchases), it is not clear that the effects of these unconventional tools are well captured by the same dynamics of the shadow rate and market price of risks that described the pre-ELB years.

4. Lastly, it is possible that the financial crisis which led to the ELB regime was so severe that the structure of the economy has changed in a substantial way since the pre-ELB years, with consequential implications for yield curve dynamics.

In this paper, we empirically examine the potential presence of a structural break in shadow rate models of the U.S. Treasury yield curve. An investigation of structural breaks in shadow rate models, and more generally latent-factor term structure models—including Dai and Singleton (2000), Duffee (2002), and Ahn, Dittmar, and Gallant (2002)—, raises new challenges that have not been encountered in the existing literature on structural break tests. In these latent-factor models, the state variables do not have well-defined meaning, because a given set of factors can be "rotated" to another set of factors, with corresponding changes in the model parameters, in a way that keeps the model's empirical content the same.[5] Therefore, investigating the change in a specific parameter of the model may not shed much helpful light on the structural stability question. Furthermore, Wald-type tests (e.g., Andrews and Fair (1988)), in which the parameter vector θ_1 estimated from one subsample is compared to the parameter vector θ_2 estimated from another subsample (i.e., a test of the hypothesis $\theta_1 = \theta_2$) run into further difficulties, because, even with normalization restrictions that are put in place to guarantee econometric identification, there are multiple vectors in the parameter space (with the same empirical contents) linked by discrete transformations.

In testing for a structural break in shadow rate term structural models, we address the issues arising from the latent nature of the factors with several approaches. One is to construct rotation-invariant test statistics, i.e., test statistics that are unaffected by invariant transformations. Another is to rotate the factors so that they can be given empirical meaning, specifically, we can transform the factors such that the transformed factors have interpretation as principal components (level, slope, and curvature). We also extensively utilize GMM techniques with likelihood scores based on the estimation with either the pre-break subsample ("predictive" tests) or the full sample (Lagrange multiplier tests). These tests do not run into the problem of multiple equivalent parameter vectors, as they involve only local changes in the parameter space surrounding an optimum. They also have the attraction of avoiding the estimation of the post-break subsample, which is relatively short, and entails greater estimation uncertainties. With these estimation difficulties as caveat, we do also perform

[5] Dai and Singleton (2000) contains an extensive discussion of such invariant transformations.

an estimation with the post-break subsample, and further examine the consequences of a potential structural break.

While the techniques outlined above have applicability to the study of structural breaks in broader latent-factor term structure models, we also utilize a particular feature of the shadow rate model we are analyzing (an affine-Gaussian shadow rate model), namely that, although the true yields implied by the model follow nonlinear processes, the factors themselves follow Gaussian processes, and the so-called "shadow yields" therefore follow Gaussian processes as well. Therefore, under the assumption of no structural break, the factors and shadow yields follow structurally stable VAR(1) processes, a hypothesis that we examine in multiple ways in the paper.

To preview our results, we find extensive evidence pointing to structural instability, and this instability does not appear to be confined to any readily identifiable subset of parameters. For example, the innovation vectors for the ELB period implied by pre-ELB subsample parameter estimates, which should be approximately i.i.d. under the null hypothesis of structural stability, display not only contemporaneous correlations but also serial correlations. Likelihood score–based tests also soundly reject the hypothesis of structural stability. Our empirical findings indicate a change not only in the $\mathbb{P}$-dynamics of the model, but also in the $\mathbb{Q}$-dynamics and in risk pricing. This has important practical implications; for example, ignoring the structural change (i.e., using a structurally stable shadow rate model) leads to implied policy rate paths during the ELB years that are notably steeper than implied by the model that allows for a structural break.

The remainder of this paper is organized as follows. Section 2 spells out the shadow rate term structure model we are analyzing, and briefly describes the data and estimation method. Section 3 provides a conceptual discussion of various potential channels of instability in the shadow rate model. Section 4 discusses the techniques we use to address the problems arising from the latent-factor nature of the model in testing for a structural break. Section 5 describes the evidence for structural instability based on predictive and LM approaches. Section 6 provides further evidence from the comparison of pre-ELB subsample estimation and post-break subsample estimation, and Section 7 concludes.

2 Model and Data

2.1 Shadow Rate Term Structure Model

We begin with the discussion of the shadow rate term structure model used in this paper. Let $W_t^{\mathbb{P}}$ be N-dimensional standard Brownian motion under the real-world probability measure $\mathbb{P}$. Assume there is a pricing measure $\mathbb{Q}$, equivalent to $\mathbb{P}$, and denote by $W_t^{\mathbb{Q}}$ Brownian motion under $\mathbb{Q}$. Suppose N latent factors (state varibles) describing the term structure dynamics follow the stationary multivariate Ornstein-Uhlenbeck process

$$dx_t = (k_0^{\mu} + K_1^{\mu} x_t)dt + \Sigma dW_t^{\mu} \tag{1}$$

where $\mu \in \{\mathbb{P}, \mathbb{Q}\}$. Let the shadow short rate be

$$s_t = \rho_0 + \rho_1' x_t, \tag{2}$$

and the observed short rate the shadow rate censored at the ELB $\underline{r}$:

$$r_t = \max\{s_t, \underline{r}\}. \tag{3}$$

The arbitrage-free time t price of a zero-coupon bond with time to maturity τ is then given by $\mathbb{Q}$-measure expectation

$$P_{t,\tau} = E_t^{\mathbb{Q}}\left[\exp\left(-\int_t^{t+\tau} r_s\, ds\right)\right], \tag{4}$$

with associated zero-coupon bond yield

$$y_{t,\tau} = -\frac{1}{\tau}\log P_{t,\tau}. \tag{5}$$

The bond yields $y_{t,\tau}$ in the model will, in general, be nonlinear functions of x_t. We approximate this function using the second-order method proposed in Priebsch (2013).

Shadow bond prices and yields are defined analogously, with the shadow short rate s_t in place of the observed short rate r_t in (4). The shadow yields $y^s_{t,\tau}$ correspond to the arbitrage-free yields in the underlying Gaussian model not constrained by the ELB, and take the affine form

$$y^s_{t,\tau} = a(\tau) + b(\tau)' x_t, \tag{6}$$

where a and b and given by the usual recursive formulas (Duffie and Kan (1996)).[6] As the lower bound becomes less binding, yields in the shadow rate model will converge to their Gaussian counterpart.[7] In this sense, away from the lower bound, the shadow rate model is approximated by the underlying Gaussian model.

On the other hand, as the lower bound becomes more binding, yields and $\mathbb{P}$-expectations of average future short rates will both converge to $\underline{r}$: If the short rate is currently constrained by the lower bound and is expected with high probability to remain at the lower bound for an extended period, there is little uncertainty about the path of the short rate going forward, and therefore forward rates and $\mathbb{P}$-expected future short rates will trivially be close to $\underline{r}$. By implication, term premiums—the difference between observed yields and $\mathbb{P}$-expected average future short rates—will be close to zero. In this way, the shadow rate model is able to capture periods of forward guidance during which policymakers commit to keeping rates at the ELB for an extended period. Conversely, in the Gaussian model, the term structure of uncertainty about the future short rate is time-invariant (under both $\mathbb{P}$ and $\mathbb{Q}$).

[6] See Bauer and Rudebusch (2016) for an extensive analysis based on shadow yields.

[7] Formally, this follows from an application of the Monotone Convergence Theorem to (4), taking $\underline{r} \downarrow -\infty$.

2.2 Sample Period and Break Point

We use data from the beginning of 1990 to the middle of 2019. Where applicable, we assume that the FOMC's announcement on December 16, 2008, that it would establish a target range for the federal funds rate of zero to a quarter percent represents an unanticipated structural break point. We will somewhat loosely label the first period (observations from 1990 to mid-December 2008) as "pre-ELB" and the second period (observations from mid-December 2008 to June 2019) as "post-ELB." Below, we also discuss the possibility of other break points (for example, around "liftoff" of the federal funds rate from the ELB in December 2015). Our empirical designs in this paper focus on structural change with a single break point; the case of multiple breaks would be even more challenging empirically, in part because of the limited amount of post-liftoff data.

2.3 Treasury Yield Data

We use continuously compounded zero-coupon Treasury yield data for maturities of 3 months, 6 months, 1, 2, 4, 7, and 10 years, sampled weekly on Wednesdays (or the prior trading day if Wednesday is a holiday), from January 3, 1990, to June 26, 2019. For the 3- and 6-month maturities, we use (appropriately transformed) secondary market Treasury bill rates from the Federal Reserve's H.15 release. For maturities of 1 year and longer, we use the updated zero coupon yields based on Gürkaynak, Sack, and Wright (2007).

2.4 Survey Data

To ameliorate small-sample problems associated with persistent series such as yields, we also include forecasts for the 3-month Treasury bill rate from Blue Chip Financial Forecasts (see Kim and Orphanides (2012)).[8] The consensus forecasts are linearly interpolated to constant horizons of 6 months, 12 months, and 5-to-10 years. Short-range forecasts are available monthly and long-range forecasts are available semiannually. We line up survey data with yield data using the strategy described in Appendix B of Kim and Orphanides (2012). For weeks with no matched survey data, we treat surveys as missing observations.

2.5 Implementation of Shadow Rate Models

For estimation purposes, we express the model as a standard discrete-time state space model with non-linear observation equation. In particular, assume that both observed yields ($\tilde{y}_t$) and survey forecasts ($\tilde{z}_t$) are measured with error, giving rise to the observation equation

$$\begin{pmatrix} \tilde{y}_t \\ \tilde{z}_t \end{pmatrix} = h(x_t) + \begin{pmatrix} e_{y,t} \\ e_{z,t} \end{pmatrix}. \tag{7}$$

We set the number of factors (the dimension of x_t) to $N = 3$. Measurement errors are assumed to be mutually independent and independent across time. For yields, we assume

[8] Wolters Kluwer Legal and Regulatory Solutions U.S., Blue Chip Financial Forecasts.

that the measurement error variance δ_y^2 is the same for all maturities. For surveys, we allow different measurement error variances $\delta_{z,6m}^2$, $\delta_{z,12m}^2$, and $\delta_{z,5-10y}^2$ for the different forecast horizons. The measurement error variances are treated as estimated parameters. For the 5-to-10-year survey horizon, we impose a lower bound of 50 bps on $\delta_{z,5-10y}$ which is binding in all our estimated models.[9] The $\mathbb{P}$-measure version of (1), sampled discretely at weekly intervals ($\Delta t = 7/365.25$), is the transition equation

$$x_t = m_0 + M_1 x_{t-1} + \varepsilon_t, \tag{8}$$

where $\varepsilon_t \sim$ i.i.d. $N(0, \Omega)$ and Ω, m_0, M_1 are known functions of the underlying (continuous-time) parameters.[10]

When $h(\cdot)$ is linear, the standard linear Kalman filter can be used to filter state variables and compute a likelihood value, for given parameters. Parameters can then be estimated by maximum likelihood. In the shadow rate model, $h(\cdot)$ is non-linear, and we analogously use the unscented Kalman filter and quasi–maximum likelihood for estimation.

When using survey expectations, we face the question what the appropriate model counterpart is. In a Gaussian model, it is justifiable to match the model-implied $\mathbb{P}$-expected future rate with the survey consensus forecast.[11] At or near the ELB, the distribution of the future short rate is arguably skewed (in the shadow rate model, it is simply a censored Gaussian distribution), so that its mean, median, and mode will not necessarily coincide. Unless asked explicitly, a forecaster stating their point forecast may have any one (or none) of those objects in mind, and different forecasters may have in mind different concepts of central tendency. Matching the model-implied mean-expectation with the conventional survey consensus forecast is thus less compelling. At the same time, in a new regime with limited historical data and with the short end of the term structure constrained in its information content, survey data are potentially particularly valuable to pin down the data-generating dynamics of state variables. Mindful of this tradeoff, we proceed by treating the survey consensus as a noisy observation of the model-implied *modal* expectation in the post-ELB sample.

3 Potential Sources of Structural Instability

In this section, we provide a conceptual discussion of potential sources of structural instability of the shadow rate term structure model.

[9] See Kim and Orphanides (2012) for a discussion of the rationale for imposing a conservative lower bound on the error variance for long-term surveys.

[10] To first order, $m_0 \approx k_0 \Delta t$, $M_1 \approx K_1 \Delta t + I_N$, and $\Omega \approx \Sigma\Sigma' \Delta t$. We use the exact relationships; see, for example, Priebsch (2013) Appendix A.

[11] The "consensus" is defined as the arithmetic mean of point forecast across survey respondents. If the future short rate is normally distributed, and survey respondents receive i.i.d. noisy signals of its unknown mean μ_r, the average of respondents' signals is a sufficient statistic for μ_r.

3.1 State Vector Dynamics

One potential source of instability in ELB term structure models is that the $\mathbb{P}$-measure dynamics of state variables could change once the economy enters the ELB regime. The intuition is that conventional monetary policy, implemented by adjusting the target federal funds rate, can no longer respond to macroeconomic developments and provide monetary stimulus once the economy enters the ELB regime. In turn, this would affect the evolution and interplay of macro variables. This idea is reflected in the stylized model of Reifschneider and Williams (2000), in which the output gap g_t, inflation π_t and nominal short rate i_t are given by

$$g_t = \delta g_{t-1} + \theta(i_{t-1} - \pi_{t-1} - r^*) + \epsilon_t, \tag{9}$$

$$\pi_t = \pi_{t-1} + \phi g_{t-1} + \nu_t, \tag{10}$$

$$i_t = \max[0, r^* + \pi_t + \alpha(\pi_t - \pi^*) + \beta g_t]. \tag{11}$$

It can be seen that the dynamics of g_t and π_t are different between ELB and non-ELB periods, since the $\theta(\cdot)$ term in eq. (9) is $-\theta(\pi_{t-1}+r^*)$ in the former and $\theta(\alpha(\pi_{t-1}-\pi^*)+\beta g_{t-1})$ in the latter. Later works have often modeled ELB dynamics within a new-Keynesian framework. For example, from Eggertsson and Woodford's (2003) study of monetary policy at the ELB under commitment, it can be seen that the dynamics of inflation and output gap are described by a VAR with *time-dependent* parameters, even when those variables are described by VAR with *constant* parameters in the normal regime. Similarly, the new-Keynesian model of Nakata and Tanaka (2016), in which variables such as inflation and output gap are driven by a preference shock and a productivity shock, has the feature that those macro variables behave differently in ELB vs. non-ELB periods.[12] In addition, a DSGE model study of Gust, Herbst, López-Salido, and Smith (2017) concludes that the effective lower bound had a significant effect on macro variables.

These theoretical considerations raise the question whether it is reasonable to assume stable state vector dynamics in shadow rate term structure models. One potentially alleviating consideration is that in the ELB regime there can be more monetary accommodation than is implied by eq. (9), as central banks could employ unconventional tools to provide monetary accommodation. That said, as discussed more below, it is not clear whether that would be sufficient to ensure structural stability across the ELB boundary.

In light of the above discussion, the structural stability issue could be especially concerning for those shadow rate term structure models that have macro variables—such as inflation and GDP gap—as part of the state vector.[13] However, even latent-factor models, in which all of the state variables are latent variables, may not be free from structural instability concerns, as latent variables are typically thought to embody macroeconomic risks.

[12] See, for example, Nakata and Tanaka's (2016) Figure 1.

[13] Ang and Piazzesi (2003) and Joslin et al. (2014) are examples of term structure models with macro factors. ELB term structure models with macro factors include Bauer and Rudebusch (2016) and Wu and Xia (2016).

3.2 Market Price of Risk

A feature of the affine-Gaussian shadow rate models analyzed here is that λ_t, the vector of market price of risk of x_t, is affine in x_t:

$$\lambda_t = \lambda_0 + \Lambda_1 x_t. \tag{12}$$

The constant vector λ_0 and the constant matrix Λ_1 are related to the parameters describing the $\mathbb{P}$-measure and $\mathbb{Q}$-measure dynamics as[14]

$$\begin{aligned} k_0^{\mathbb{Q}} &= k_0^{\mathbb{P}} - \Sigma\lambda_0, & (13)\\ K_1^{\mathbb{Q}} &= K_1^{\mathbb{P}} - \Sigma\Lambda_1. & (14) \end{aligned}$$

An implication of the structurally stable shadow rate model is that eq. (12) is also structurally stable, and the relations (13) and (14) hold with the same set of $k_0^{\mathbb{P}}, K_1^{\mathbb{P}}, k_0^{\mathbb{Q}}, K_1^{\mathbb{Q}}, \lambda_0, \Lambda_1, \Sigma$ parameters in the pre-ELB and post-ELB periods.

For the sake of argument, one could still think of latent-factor ELB term structure models as "statistical models" and posit that there exists a set of latent state variables whose ($\mathbb{P}$-measure) dynamics are stable across the ELB boundary. But is it reasonable to assume that the structure of market price of risk would also remain stable across the ELB boundary? Could it be that the nature of interest rate risk compensation changes once the economy enters the ELB regime? Unfortunately, the micro foundations for modeling the market price of risk are not sufficiently well developed to provide a lot of theoretical guidance in the normal regime, let alone the ELB regime. Hördahl, Tristani, and Vestin (2006) try to get around this problem in their new-Keynesian term structure model by specifying the market price of risk as a flexible affine function of the variables that enter the structural VAR for new-Keynesian dynamics. Recently, Sakurai (2016) estimated a new-Keynesian ELB term structure model with essentially the same specification of market price of risk as in Hördahl et al. (2006), thereby assuming that the structure of the market price of risk is stable across the ELB boundary.

There is at least one known mechanism for the variation in the market price of risk that implies a *change* in the dynamics of market price of risk at the ELB (from the perspective of Gaussian shadow rate models): In King's (2019) model of supply effects in an ELB regime, the market price of risk vector is an affine function of Gaussian state variables in the normal regime, but its dynamics become non-Gaussian in the ELB regime. More specifically, it can be shown that in King's model, the market price of risk of the state vector $x_t = [s_t, \beta_t]'$, with shadow rate s_t and supply factor β_t following structurally stable Gaussian processes under the $\mathbb{P}$-measure ($dx = (k + Kx)dt + \Sigma dW^{\mathbb{P}}$), is given by

$$\lambda_t = a \int_0^T du\, \zeta_t^{(u)} \Sigma' \frac{1}{P^{(u)}} \frac{\partial P^{(u)}}{\partial x_t} \tag{15}$$

where a is the risk aversion of arbitrageurs and $\zeta_t^{(u)}$, the arbitrageurs' position in bonds with maturity u, is an affine function of the supply factor β_t. Because the log bond price $\log P$

[14] This follows from, for example, eq. (6) in Kim and Singleton (2012).

is approximately an affine function of the state vector x away from the ELB (the normal regime), the market price of risk simplifies to Greenwood and Vayanos's (2014) affine form (their equation (11)) in the normal regime. But in the ELB regime, $\log P$ is a non-affine function of x, leading to a non-affine market price of risk and non-Gaussian $\mathbb{Q}$-measure dynamics of x.[15]

3.3 Unconventional Monetary Policy

Another complication is the change in monetary policy tools: As conventional monetary policy is constrained at the ELB, monetary policymakers in the U.S. and abroad have employed *unconventional* monetary policy tools, namely large scale asset purchases (LSAPs) and forward guidance.[16] These unconventional tools may have helped to soften the blow of the ELB constraint to the real economy, but to what extent they have done so is still an actively debated topic.[17] Furthermore, the way they work could be different from conventional policy. In particular, LSAPs are believed to have a more direct effect on longer-term yields (relative to conventional monetary policy) through the suppression of term premiums. While forward guidance may be the more conventional tool of the two—and could be viewed as having a precedent before the ELB period (in particular in 2003)—, the extent of the guidance during the recent U.S. ELB period was unprecedented. It is thus not clear whether the effect of these unconventional policy tools can be captured by some regions of the state space within a structurally stable model, or whether it requires different dynamics (a different model structure).

On a related note, in Sakurai's (2016) new-Keynesian shadow rate term structure model, the output gap dynamics depend on the shadow rate rather than the nominal short rate that is standard in the literature (as in eq. (9)). Besides the tractability issue, the economic justification given by Sakurai is that at the ELB, monetary policy can still play a role by affecting *expectations* about the policy rate path, and the level of the shadow rate (below the ELB) can capture this effect. However, such a specification appears to be an extension of conventional monetary policy to the ELB regime, as opposed to being a model of unconventional policy.[18]

[15] Because of the non-affine nature of the market price of risk in the ELB regime, King's model cannot be exactly matched by affine-Gaussian shadow rate models even with a structural break. But allowing for a structural break in the affine-Gaussian shadow rate models may help *approximate* such nonlinear behavior of market price of risk.

[16] For recent reviews, see Bernanke (2020) and Kuttner (2018).

[17] Chung, Laforte, Reifschneider, and Williams (2012), for example, find that the Federal Reserve's asset purchases, while materially improving macroeconomic conditions, did not prevent the ELB constraint from having first-order adverse effects on real activity and inflation.

[18] LSAPs have greater effect on longer-term yields, which are less closely tied to the level of the shadow rate than short-term yields. Furthermore, a low level of the shadow rate (below the ELB), as in Sakurai's (2016) model, by itself may not necessarily imply a substantially long period of policy rates at the ELB (as under forward guidance), if the state variable responsible for the low level of the shadow rate is fast mean-reverting.

3.4 Structural Change at the End of the ELB Period?

Some of the potential mechanisms discussed above would also imply a potential structural change at the end of the ELB period, i.e., when the Federal Reserve began to raise the target federal funds rate in December 2015. However, the prospect for substantial structural change around that time is likely weaker than around the *beginning* of the ELB period, for the following reasons:

1. It is possible that the Financial Crisis had significant impact on the structure of the economy, altering interest rate dynamics well beyond the ELB regime. For example, mortgage lending conditions for households have been persistently tighter relative to those prevailing before the Financial Crisis, and this and other "headwinds" have often been mentioned as factors behind the relatively slow pace of federal funds rate normalization in the tightening cycle that began in December 2015. In addition, many macro models, including Holston, Laubach, and Williams (2017), indicate that the neutral rate of real interest, often called "r-star," declined noticeably around the time of Financial Crisis, and has remained relatively low.

2. Furthermore, the effects of LSAPs—an unconventional policy first implemented around the time of the onset of the ELB—may persist beyond the ELB period: Normalization of the Fed's balance sheet was not announced until September 2017 (well past liftoff in December 2015), and the subsequent shrinking of the balance sheet proceeded at a gradual pace. As of 2019, the balance sheet remains notably larger than before the Financial Crisis, and is expected to grow again, with the policy implementation in the "ample reserves" framework.

3. Market participants generally had not been pricing in ELB risk until close to the event in 2008,[19] but the Financial Crisis and the extended ELB period that followed may have sensitized market participants to the ELB risk, and they may now be pricing in significant ELB risk even in the "new normal" regime, analogous to downside jump risk getting priced into equity options after the "Black Friday" stock market crash in 1987.[20] Data from the New York Fed's surveys of primary dealers and market participants provide some support for this. Since liftoff in December 2015, respondents have consistently judged the probability of moving to the ELB at some point within the next 2 to 3 years as substantial (with typical median responses in the 20 to 30 percent range).[21]

[19] Analogously, Chung et al. (2012) note that prior to the Financial Crisis, most economists probably did not view the ELB as a major problem for central banks, and that previous research tended to understate the ELB risk.

[20] See, for example, Bates (2000).

[21] The surveys are available at `https://www.newyorkfed.org/markets/primarydealer_survey_questions` and `https://www.newyorkfed.org/markets/survey_market_participants`.

4 Empirical Strategies

4.1 Challenges of Dealing with Latent Factor Models

We start this section with a discussion of the challenges posed by the latent nature of the factors in the study of structural stability. Let us consider the possibility that a latent-factor term structure model is described collectively by parameter vector θ_1 in period 1 ($t = 1, \ldots, T_1$), and by another parameter vector θ_2 in period 2 ($t = T_1+1, \ldots, T_1+T_2(\equiv T)$); for later use, define $\pi = T_1/T$. The null hypothesis of structural stability is $\theta_1 = \theta_2$.

With the model given in eq. (1,2,3), note that the model retains exactly the same content if the N-dimensional state vector x_t is transformed to

$$x_t^\dagger = l + Lx_t \tag{16}$$

where l is a constant N-vector and L is an invertible $N \times N$ constant matrix, and the model parameters $\theta = (K_1^{\mathbb{P}}, K_1^{\mathbb{Q}}, k_0^{\mathbb{P}}, k_0^{\mathbb{Q}}, \Sigma, \rho_0, \rho_1)$ are transformed to $\theta^\dagger = (K_1^{\mathbb{P}\dagger}, K_1^{\mathbb{Q}\dagger}, k_0^{\mathbb{P}\dagger}, k_0^{\mathbb{Q}\dagger}, \Sigma^\dagger, \rho_0^\dagger, \rho_1^\dagger)$, where

$$\begin{aligned} K_1^{\mu\dagger} &= LK_1^{\mu}L^{-1}, \quad k_0^{\mu\dagger} = Lk_0^{\mu} - LK_1^{\mu}L^{-1}l, \\ \Sigma^\dagger &= (L\Sigma\Sigma'L')^{1/2}, \quad \rho_0^\dagger = \rho_0 - \rho_1' L^{-1} l, \quad \rho_1^\dagger = L^{-1'}\rho_1. \end{aligned} \tag{17}$$

Because the unrestricted model has an infinite number of equivalent parameters linked by such invariant transformations, to estimate these models by maximizing a log likelihood function (or some analogous GMM criterion) $\log \mathcal{L}$, a set of normalization restrictions are imposed to ensure that the Hessian matrix $\partial^2 \log \mathcal{L}/\partial\theta\partial\theta'$ is not singular. In the present paper, the normalization restrictions we impose are

$$k_0^{\mathbb{P}} = 0, \quad K_1^{\mathbb{P}} = \text{lower triangular matrix}, \quad \Sigma = cI \tag{18}$$

where I is an identity matrix, and $c = 0.01$.

In a model with typical normalization (such as the one just mentioned), factors do not have a clear empirical meaning; therefore, examining the change in a specific parameter, e.g., $[K_1^{\mathbb{P}}]_{11}$, across subsamples may not be very meaningful. Nonetheless, one may still hope that the Wald test[22]

$$\lambda_T = T(\hat{\theta}_1 - \hat{\theta}_2)'(\pi^{-1}\hat{V}_1 + (1-\pi)^{-1}\hat{V}_2)^{-1}(\hat{\theta}_1 - \hat{\theta}_2) \tag{19}$$

(based on entire parameter vectors θ_1 and θ_2) provides a valid statistic for the test of structural stability. However, even after normalization restrictions that guarantee a nonsingular Hessian of the likelihood function around an estimate (such as eq. (18)), discrete invariant transformations still remain, which creates multiple equivalent parameter vectors. One such transformation is permutation, an example being the transformation $x_t = [x_{1t}, x_{2t}, x_{3t}]'$ $\to x_t^\dagger = [x_{2t}, x_{1t}, x_{3t}]'$, in other words,

$$l = 0, L = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{20}$$

[22] See, for example, Andrews and Fair (1988).

For an N-factor model, there are $N!$ permutations. Another discrete transformation is reflection, for example, $x_t = [x_{1t}, x_{2t}, x_{3t}]' \to x_t^\dagger = [x_{1t}, -x_{2t}, x_{3t}]'$, in other words,

$$l = 0, L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{21}$$

Discrete transformations can make it difficult to determine how close a parameter estimate θ_1 for one subsample is from a parameter estimate θ_2 for the other subsample. For example, even when the Euclidean distance $\theta_2 - \theta_1$ in eq. (19) is large, it could be that θ_1 and θ_2 have quite similar empirical contents, which would be the case if θ_2 is close to one of the permuted versions of θ_1. One might consider eliminating discrete transformation degrees of freedom like permutation by imposing a specific ordering of certain parameters, e.g., $[K_1^{\mathbb{P}}]_{11} > [K_1^{\mathbb{P}}]_{22} > [K_1^{\mathbb{P}}]_{33}$. However, there is no guarantee that such ordering is preserved in a case of structural break. One could also entertain the idea of

$$\lambda_T = \min_i T(\hat{\theta}_1 - \hat{\theta}_2^{(i)})'(\pi^{-1}\hat{V}_1 + (1-\pi)^{-1}\hat{V}_2^{(i)})^{-1}(\hat{\theta}_1 - \hat{\theta}_2^{(i)}), \tag{22}$$

where $\hat{\theta}_2^{(i)}$ denotes all possible discrete invariant transformations of $\hat{\theta}_2$, and $\hat{V}_2^{(i)}$ denotes the corresponding covariance matrix, i.e., pick the $\hat{\theta}_2^{(i)}$ that produces the smallest test statistic. However, besides being cumbersome both in implementation and statistical evaluation, such a procedure can be unsatisfactory unless one of the $\theta_2^{(i)}$'s stands out sufficiently low in terms of the test statistic value.[23]

Below we consider various test statistics that overcome the problems associated with the latent nature of the factors in the model.

4.2 Tests for Structural Change

Our structural break test problem is classical in the sense that we consider a single potential break at a known time (around the onset of the ELB), rather than searching for break(s) at unknown time(s). In this setting, the available tests for the kinds of nonlinear estimation problems such as ours can be summarized as follows:

1. **Predictive tests:** These types of tests are based on the *estimation of the pre-ELB subsample*, and examine the properties of objects in the post-ELB subsample implied by the pre-ELB parameter estimates.

2. **Lagrange multiplier (LM) tests**: These types of tests are based on the estimation of the *full sample*, and examine moment restrictions across the two subsamples.

[23] As an illustration, consider a 2-factor model whose $\mathbb{Q}$-measure dynamics are given by $r = \rho_0 + \rho_1'[x_{1t}, x_{2t}]'$, $dx_{it} = (k_{0i}^{\mathbb{Q}} + k_i^{\mathbb{Q}} x_{it})dt + dW_{it}^{\mathbb{Q}}$ $(i = 1, 2)$. Suppose that $(k_{11}^{\mathbb{Q}}, k_{12}^{\mathbb{Q}})$ in the first subsample is estimated to be (0.1, 0.5), and $(k_{11}^{\mathbb{Q}}, k_{12}^{\mathbb{Q}})$ in the second subsample is estimated to be (0.45, 0.15). Then it would be reasonable to view that the appropriate version of the second subsample estimate to compare is (0.15, 0.45), i.e., the permuted version $(x_{1t}, x_{2t}) \to (x_{2t}, x_{1t})$. However, if we obtained (0.1, 0.5) for the first subsample and (0.3, 0.32) for the second subsample, it would be less clear. Note that we have not spelled out the estimates of other parameters, which can lessen or increase the ambiguity.

3. **Wald tests:** These types of tests involve *separate estimations with pre-ELB and post-ELB subsamples*, and analyze potential differences between the two estimates.

Of these, Wald-type tests can be expected to have low power, as they involve an estimation of the post-ELB subsample. A sample of about ten years (late 2008 to mid-2019) may not be long enough to precisely estimate the parameters of the model, in view of the well-known small-sample problems with term structure model estimations. Moreover, this particular sample period contains only one tightening cycle (near the end of the period) and limited variability in short rate movements. While Kim and Orphanides (2012) found that the use of survey forecasts for Treasury bill yields helps ameliorate some of these concerns, the near-horizon survey forecast data may be less informative in this subsample, as a substantial portion of this period had fairly flat forecasts due to the FOMC's forward guidance; furthermore, as discussed above, the asymmetric distribution of the short rate process for horizons beyond the predicted liftoff date means that the measurement of survey forecasts would arguably be less reliable than usual.

For evidence regarding structural instability, we therefore focus on predictive tests and LM tests. Predictive tests are particularly attractive in our context, as we only need to estimate a model based on pre-ELB data, where the shadow rate model is adequately approximated by the affine-Gaussian model. Under the null hypothesis that there is no structural break, the parameters thus estimated should produce a shadow rate model fitting the post-ELB period adequately. That said, in Section 6 we also estimate a model based on the post-ELB subsample, as the estimation may throw helpful light on how the model structure might have changed.

4.3 Tests Based on Fitting Errors or Innovation Vectors

A well-known example of predictive tests for structural breaks in a classical regression is the Chow (1960) test that examines the residuals from the second subsample computed with regression coefficients from the first subsample. Though an exact analog of regression residuals does not exist in our a latent-factor term structure model setup, we can consider the following:

Fitting errors. While it may be tempting to regard yield fitting error ($e_{y,t}$ in equation (7)) as an analog, yield fitting errors are generally not a sufficient diagnostic, especially in the case of flexibly specified latent-factor models such as ours. Indeed, a small overall fitting error may not necessarily indicate a well-specified model, as it could be a consequence of the fact that a model with N latent factors can fit N yields exactly and that yield curves tend to be smooth. Nonetheless, a meaningful *change* in the pattern of fitting errors could be indicating a structural change in $\mathbb{Q}$ dynamics; we shall therefore examine fitting errors as part of our structural break diagnostics.

Innovation vectors. As can be seen in eq. (8), if the shadow rate model is structurally stable, the state variables will follow a standard VAR(1) process. One way to examine this is to look at the innovation vector ε_t based on equation (8) implied by the θ_{pre} estimate, which would be a closer yet still-imperfect analog of Chow's regression residuals. A normalized innovation vector $\eta_t \equiv \Omega^{-1/2}\varepsilon_t$ has the theoretical property

$$\eta_t \sim N(0_N, I_{N\times N}). \tag{23}$$

Furthermore, these innovation vectors are serially uncorrelated, i.e.,

$$E(\eta_t \eta'_{t+j}) = 0_{N\times N}, \qquad j \neq 0. \tag{24}$$

The innovation vectors implied by the pre-ELB parameter estimate should satisfy these conditions well in the post-ELB sample (provided, of course, that the model is reasonably well specified). But if there is a structural change, they may not satisfy these conditions in the post-ELB sample; therefore, they could provide another useful diagnostic check for structural change.

Test statistics can therefore be constructed that examine these conditions. A complication in our setting, discussed above in Section 4.1, is that state variables and their innovation vectors in a generically normalized model (such as the normalization conditions in eq. (18)) do not have specific economic meaning, as they can be invariantly transformed so that the model's content remains the same. Therefore, departures from these conditions for individual elements of the innovation vector, e.g., $corr(\eta_{1t}, \eta_{2t}) \neq 0$, are difficult to interpret. We address this problem in two ways.

First, we construct aggregate test statistics whose values are unchanged if the model is re-written with different (transformed) state vectors. In particular, in Appendix A, we show that the objects $\| T^{-1}\sum_t \eta_t \eta'_{t-u} \|_F$ and $\| T^{-1}\sum_t \eta_t \eta'_t - I \|_F$ remain unchanged under invariant transformations and have the following asymptotic distributions:[24,25]

$$T\frac{1}{2} \parallel T^{-1}\sum_t \eta_t \eta'_t - I \parallel_F^2 \ \sim \ \chi^2_{(N^2+N)/2}, \tag{25}$$

$$T \parallel T^{-1}\sum_t \eta_t \eta'_{t-u} \parallel_F^2 \ \sim \ \chi^2_{N^2}, \quad u \neq 0. \tag{26}$$

While these statistics provide useful summary diagnostics regarding structural stability, in the event they point to a structural change in the post-ELB period, additional statistics that shed light on the nature of structural change would be useful. Therefore, we also examine more detailed test statistics based on state vectors that have an intuitive interpretation. To this end, we transform the state vectors in our original normalization to a new set of state variables which can be viewed as level, slope, and curvature factors. More precisely, we rotate the model to create the state vector $x_t^\dagger$, with the property that the instantaneous changes $dx_t^\dagger$ are mutually independent (i.e., $dx_t^\dagger dx_t^{\dagger\prime}$ is a diagonal matrix), and correspond

[24]The notation $\| A \|_F$ denotes the Frobenius norm of the matrix A.

[25]An example of a test statistic that is *not* unchanged under invariant transformations is $\frac{1}{2}(T^{-1}\sum_t \eta_{1t}^2 - 1)^2$ (which is made up of only the first element of the η_t vector). It can be shown that if $\eta_t^\dagger$ is the innovation vector corresponding to an invariant transformation, in general we do not have $\frac{1}{2}(T^{-1}\sum_t \eta_{1t}^2 - 1)^2 = \frac{1}{2}(T^{-1}\sum_t \eta_{1t}^{\dagger 2} - 1)^2$.

to instantaneous change in the level, slope, and curvature of the shadow yield curve.[26] In discrete time, the one-period innovation $\varepsilon_t^\dagger$ has variance-covariance matrix $\Omega^\dagger$ which is not exactly diagonal but almost diagonal if a single period is sufficiently short, as in our case (one period being one week). Therefore, the innovation vector $\eta_t^\dagger(\equiv \Omega^{\dagger-1/2}\varepsilon_t^\dagger)$ can still be well-interpreted as changes in level, slope, and curvature factors. We examine whether the individual elements of $\eta_t^\dagger$ have the contemporaneous correlation and serial correlation properties discussed in eqs. (23) and (24).

4.4 Tests Based on Likelihood Scores

We can also explore the parameter stability of the model directly by testing the moment condition

$$E(\partial \log \mathcal{L}/\partial\theta) = 0 \tag{27}$$

within a GMM framework, either with a predictive approach or with the LM approach. The idea is that, if there is no change in parameters ($\theta_{pre} = \theta_{post}$), the first-order condition eq. (27) will hold in each subsample.

The predictive approach—testing condition (27) in the post-ELB sample using the parameter estimates from the pre-ELB sample—corresponds to the technique of Ghysels and Hall (1990), who showed that, asymptotically,

$$\frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T_1+T_2}\frac{\partial \ell_t(\hat{\theta}_1)}{\partial\theta'}\hat{V}_2^{-1}\frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T_1+T_2}\frac{\partial \ell_t(\hat{\theta}_1)}{\partial\theta} \sim \chi^2_{dim(\theta)} \tag{28}$$

where $\hat{V}_2$ is a consistent estimate of $\lim_{T\to\infty} Var(\sum_{t=T_1+1}^{T_1+T_2}\partial\ell_t(\theta^*)/\partial\theta)$.[27]

The LM approach—testing condition (27) with parameter estimates $\hat{\theta}$ based on the full sample—can be viewed as an analogue of likelihood score–based LM tests of Nyblom (1989), Hansen (1990), and Hansen (1992). The relevant testing statistic is given by[28]

$$LM = \frac{T}{\pi(1-\pi)}\frac{1}{T_1}\sum_{t=1}^{T_1}\frac{\partial \ell_t(\hat{\theta})}{\partial\theta'}\hat{V}^{-1}\frac{1}{T_1}\sum_{t=1}^{T_1}\frac{\partial \ell_t(\hat{\theta})}{\partial\theta} \tag{29}$$

where, similarly to the predictive test, $\hat{V}$ is a consistent estimate of

$$\lim_{T\to\infty} Var(\sum_{t=1}^{T_1}\partial\ell_t(\theta^*)/\partial\theta).$$

These likelihood score–based approaches do not face the problems discussed in in connection with a Wald test of the hypothesis $\theta_1 = \theta_2$ (eq. (19)), as the score-based tests look at only local changes around an optimum in the parameter space. This is analogous to the presence of discrete invariant transformations not causing any difficulties in computing standard errors of a parameter estimate.

[26] This procedure is described in greater detail in Sections 6.1 and 6.2.
[27] Ghysels and Hall (1990) discuss several easily computable candidates for the variance estimator.
[28] See Hansen (1990) and Andrews (1993).

5 Empirical Evidence Regarding Structural Instability

5.1 Yield Fitting Errors and Innovation Vectors

Our first battery of tests uses a model whose parameters are those of an affine-Gaussian model estimated on pre-ELB data, extended to the post-ELB period as a shadow rate model.[29] Because $\underline{r}$ is poorly identified by pre-ELB data, unless otherwise noted, we set $\underline{r}$ to approximately 7.3 basis points in these predictive exercises. This is the estimate we obtain when estimating a single-regime shadow rate model on the full sample. We find that other sensible choices of $\underline{r}$ lead to broadly similar results.

5.1.1 Yield Fitting Errors

Figure 1 plots a time series of root-mean-squared yield fitting errors (the difference between observed and model-implied yields, averaged across maturities on a given date). Fitting errors vary somewhat over time and tend to spike during periods of financial stress, most notably the Financial Crisis.[30] However, in the later years of the ELB period, beginning in early 2014, there is a level shift in fitting errors—the model appears to have more difficulty fitting the yield curve based on pre-ELB risk-neutral parameters (which determine the model-implied cross-sectional relationships between yields), even with state variables filtered through the shadow rate model structure. This could be indicative of a structural change in the $\mathbb{Q}$ parameters of the model.

5.1.2 Innovation Vectors

Using the same model as in the previous subsection, we can compute empirical implied innovation vectors $\hat{\eta}_t$ as defined in Section 4.3. Recall from the discussion there that these vectors theoretically have a contemporaneous covariance matrix equal to the identity matrix, and no autocorrelation. Table 1 shows the empirical covariance matrices, separately for the pre-ELB and post-ELB subsample periods. While the covariance matrix for the pre-ELB period shows moderate deviations from the identity matrix—most notably, diagonal entries that are somewhat below 1 and some off-diagonal entries that modestly deviate from zero—, the covariance matrix for the post-ELB period displays a number of more notable departures: The diagonal entries for the second and third state variables are only about 0.5, and there is substantial negative correlation between the first and third state variables.

The overall magnitude of the deviation from the identity matrix and the statistical significance of any such deviation are hard to eyeball. Recall, moreover, that the innovation covariance matrices depend on the chosen model rotation. The rotation-invariant χ^2 statistics introduced in Section 4.3 and more formally justified in Appendix A allow us to quantify

[29] We have confirmed that it makes little difference to the estimated parameters whether we estimate an affine-Gaussian model or shadow rate model for the pre-ELB subsample, as long as the lower bound $\underline{r}$ is set to a plausible value near zero.

[30] Hu, Pan, and Wang (2013) find that errors in fitting flexible functional forms to Treasury yields carry meaningful information about liquidity conditions in the market.

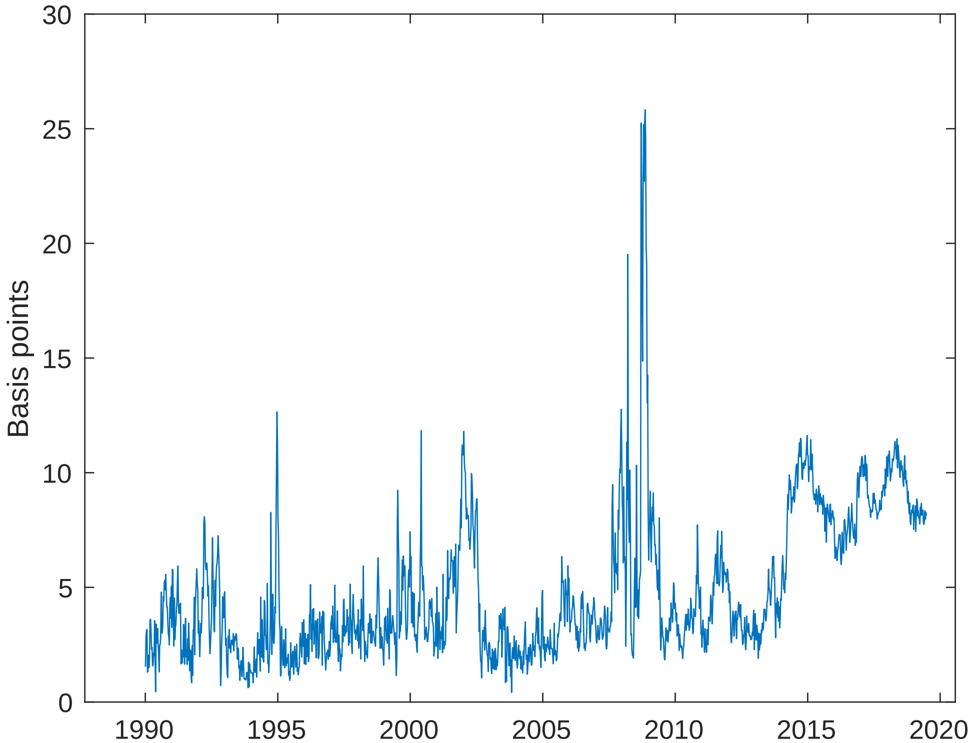

Figure 1: Time series of mean-root-squared yield fitting errors in the single-regime model estimated on pre-ELB data (extended to the post-ELB subsample as a shadow rate model).

Pre-ELB				Post-ELB		
0.86	0.14	0.03		1.01	0.01	−0.36
0.14	0.79	−0.05		0.01	0.47	−0.04
0.03	−0.05	0.91		−0.36	−0.04	0.56

Table 1: Sample covariance matrix of implied innovation vectors $\hat{\eta}_t$.

Lag	Pre-ELB	Post	d.f.	5% cuttoff	1% cutoff
0	56.66	164.80	6	12.59	16.81
1	110.62	136.50	9	16.92	21.67
2–12	294.60	594.59	99	123.23	134.64
13–52	572.97	1,352.01	360	405.24	425.35

Table 2: Rotation-invariant χ^2 statistics of implied innovation vectors. Lag 0 corresponds to eq. (25), and lags ≥ 1 correspond to eq. (26).

any misspecification more objectively. As shown in Table 2, these statistics indicate a degree of misspecification in both periods, but much more dramatically in the post-ELB subsample. In particular, over the post-ELB period, the χ^2 statistics indicate a much more pronounced degree of misspecification across the time-series dimension (i.e., autocorrelation of the innovation vectors).

To gain a better intuitive understanding of the source of the misspecification in the post-ELB period, we rotate the model as proposed by Duffee (2011) and discussed in detail Section 6.2, such that the factors can be interpreted as level, slope, and curvature of shadow yields. Table 3 displays the contemporaneous empirical innovation covariance matrices after this rotation. Figure 2 shows autocorrelograms up to lags of 52 weeks. In the pre-ELB period, the contemporaneous misspecification appears to primarily manifest in low variance for the implied innovations in curvature. In addition, the off-diagonal elements of the contemporaneous covariance matrix are small, and autocorrelations are generally insignificant or at most marginally significant, with no clear pattern across lags. Conversely, in the post-ELB subsample, all three elements of the innovation vector have contemporaneous variances notably below 1. Moreover, off-diagonal elements of the covariance matrix are non-negligible, with particularly pronounced positive association between level and slope (whose contemporaneous correlation evaluates to about 0.5). This would not be surprising if we tried to fit an *affine-Gaussian* model to the post-ELB subsample: With the short end of the empirical yield curve constrained at the ELB, any shocks to longer-term yields could only be captured by that model with offsetting movements in level and curvature (contrary to their model-implied co-movement), so as to keep the model-implied short rate unchanged.[31] Our results

[31] More generally, in the affine-Gaussian model, the short rate is an affine function of the factors ($r_t =$

Pre-ELB				Post-ELB		
0.97	0.01	−0.01		0.77	0.35	0.22
0.01	0.90	−0.07		0.35	0.64	0.20
−0.01	−0.07	0.68		0.22	0.20	0.62

Table 3: Empirical covariance matrix of implied innovation vectors (model rotated such that state variables can be interpreted as principal components).

here suggest that simply imposing a shadow rate model structure does not fix this manifestation of model misspecification. Note furthermore from Figure 2 that the implied innovations to the slope factor display significant and prolonged positive temporal dependency. In other words, the model persistently mispredicts the slope factor.

5.1.3 Timing of Structural Break

It is worth noting that large values of these test statistics based on the entire post-ELB subsample indicate a likely structural break *somewhere* in that period, but the break point might not necessarily be located at its beginning. Indeed, regarding the U.S. experience, some have suggested that the first calendar-based forward guidance announcement by the FOMC in August 2011 may have had a more significant impact on yield curve dynamics than the arrival at the ELB in late 2008.[32] Therefore, in Table 4, we also examine the statistics (25) and (26) based on segments of the post-ELB subsample, in particular December 2008 to mid-August 2011 (before date-based forward guidance), mid-August 2011 to May 2013 (before the "taper tantrum"), June 2013 to December 2015 (before liftoff), and January 2016 to June 2019.

This more granular analysis indicates that the evidence for a structural break, at least according to this metric, is not limited to just one of the four subperiods. That said, different subperiods reveal different facets of the misspecification of the structurally stable model. In particular, in the December 2008 to August 2011 and January 2016 to June 2019 subperiods, there are notable signs of misspecification in the contemporaneous covariance of innovation vectors, while in the June 2013 to December 2015 period, the serial correlation of innovation vectors shows prominent signs of misspecification.

5.2 Likelihood Score–Based Tests

In this section, we test variants of the moment restriction (27) based on both the pre-ELB parameter estimate (the predictive framework) and the full-sample, single-regime estimate

$\rho_0 + \rho_1' x_t$); therefore, the short rate being stuck at ELB would imply that a linear combination of innovation vectors have to sum up to zero, i.e., the implied innovation vectors have to be contemporaneously correlated during the ELB period.

[32] Swanson and Williams (2014), for example, note that one-year and two-year Treasury yields appeared to be "unconstrained" from 2008 to 2010, and only became more constrained from late 2011 onward.

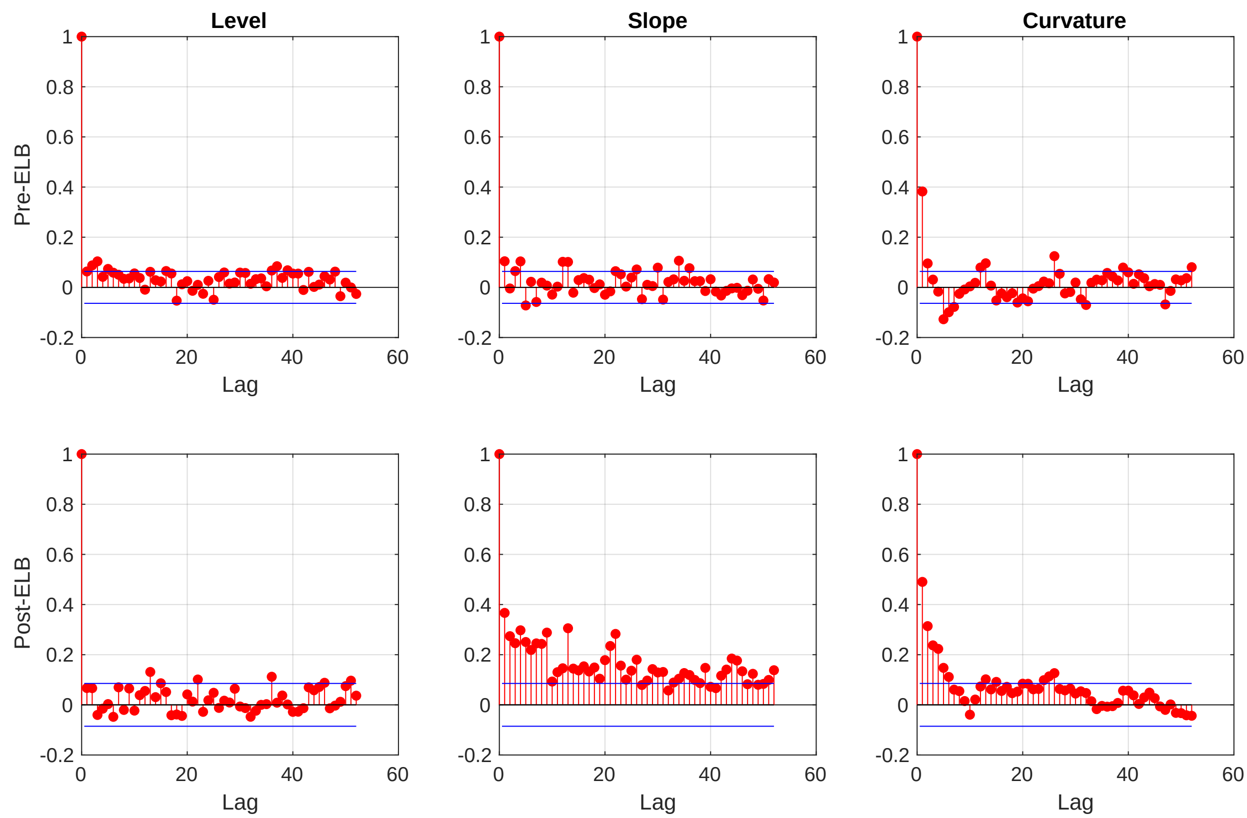

Figure 2: Autocorrelations of level, slope, curvature innovations (lags are in weeks). The blue lines represent approximate 95 percent confidence bounds under the hypothesis of no autocorrelation.

Lag	I	II	III	IV	d.f.	5% cuttoff	1% cutoff
0	107.38	46.71	23.76	117.43	6	12.59	16.81
1	49.03	23.14	155.87	14.02	9	16.92	21.67
2–12	184.02	49.21	1,159.70	82.10	99	123.23	134.64
13–52	526.08	176.00	3,051.97	205.96	360	405.24	425.35

Table 4: Rotation-invariant χ^2 statistics of innovation vectors by post-ELB subperiod. Periods I, II, III, and IV denote 12/2008–8/2011, 8/2011–5/2013, 6/2013–12/2015, and 1/2016–6/2019, respectively.

$\underline{r}$	χ^2
12.5	299.36
7.3	326.12
5.5	325.20

Subsample	χ^2
I	117.18
II	86.42
III	120.53
IV	179.58

Surveys	χ^2
Included	326.12
Excluded	331.28

Table 5: Ghysell-Hall statistics for null hypothesis of structural stability. In the middle table ("Subsample") and the rightmost table ("Surveys"), $\underline{r}$ is set to 7.3 basis points.

(the LM framework). For this purpose, in addition to the model introduced in Section 5.1, we estimate a single-regime, full-sample shadow rate model.

Consider first the predictive framework. For a model estimated on pre-ELB data, the pre-ELB-subsample average of empirical scores will be zero, as a first-order optimality condition of estimation. Intuitively, if there is no structural break, the post-ELB-subsample average of empirical scores based on pre-ELB parameters should be *close to* zero, with known asymptotic distribution (Ghysels and Hall (1990)). As discussed in Section 5.1, this exercise requires an assumption on the level of the lower bound $\underline{r}$, since this is not identified from pre-ELB data. The left panel of Table 5 displays χ^2 statistics for a joint test of all moment restrictions, for different values of the lower bound $\underline{r}$.[33] The statistics have 25 degrees of freedom—corresponding to the number of estimated model parameters—, with a 5 percent cutoff of 37.65 and a 1 percent cutoff of 44.31. The null hypothesis of structural stability is thus firmly rejected, with a p-value of virtually zero, and with only modest sensitivity to the exact chosen value of $\underline{r}$. For the remaining tests, we use $\underline{r} = 7.3$. The middle panel of the table shows statistics for the various segments of the post-ELB subsample separately. Using the same four subdivisions as in Section 5.1.3, the null hypothesis that a model based on pre-ELB parameters captures the behavior of post-ELB yields is solidly rejected for each of the segments separately. Lastly, when computing the post-ELB scores, we can include both yield and survey data, or yield data alone. Indeed, one natural question is whether poor forecasting performance by the survey respondents drives the rejection of the null hypothesis. This does not appear to be the case, as the rightmost panel of the table shows. When we compute post-ELB scores using only yield data, the statistic changes little. We thus conclude that the statistical behavior of post-ELB *yield* data is sufficient to reject the null hypothesis of structural stability.

Next, we perform LM tests based on the full-sample, single-regime shadow rate model parameter estimate. For such a model, the full-sample average of empirical scores will be identically equal to zero as a first-order condition. Without a structural break, the pre- and post-ELB-subsample averages of empirical scores based on full-sample parameters should

[33] A lower bound of 12.5 basis points corresponds to the mid-point of the federal funds target range during the ELB years. Lower bounds of approximately 7.3 and 5.5 basis points are suggested by our estimations below that include post-ELB data.

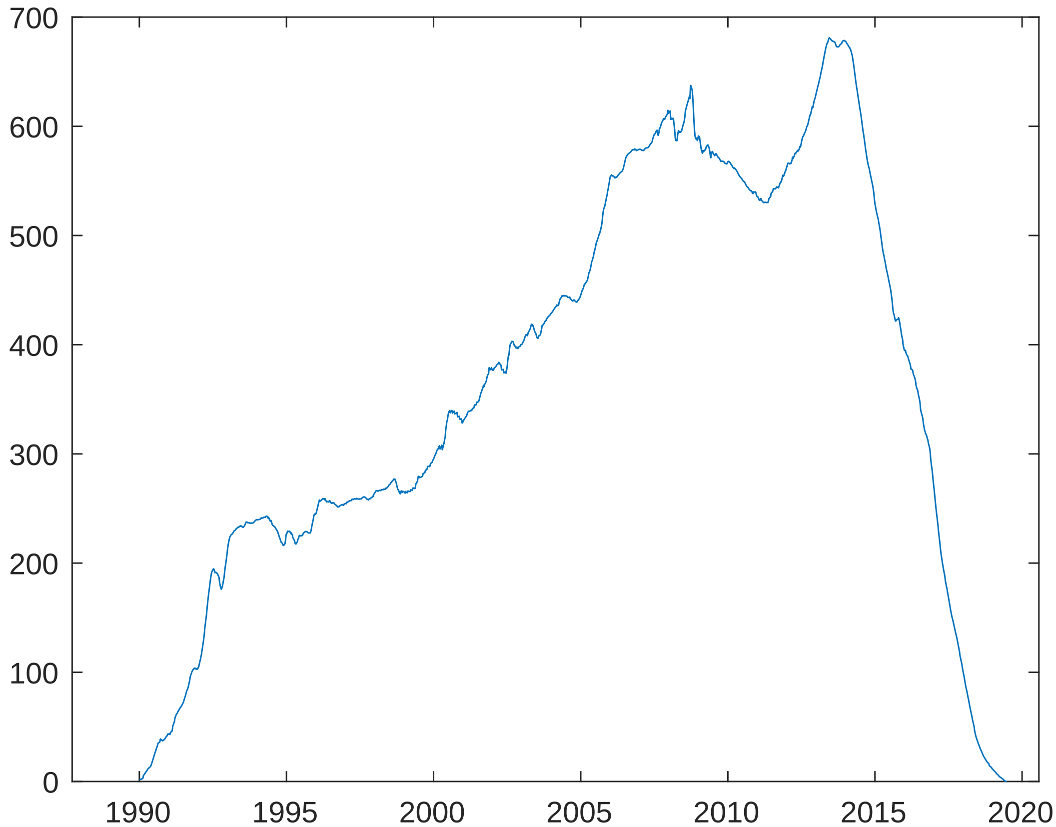

Figure 3: Time series of Lagrange multiplier statistics. A large statistic rejects the hypothesis that the full-sample GMM conditions are satisfied separately in the subsample before the given date and the subsample after the given date.

individually be close to zero, again with known asymptotic distribution.[34] Since this statistic is based on a single-regime model estimated on the entire sample, it is computationally cheap to evaluate the statistic for different structural break times. As shown in Figure 3, there is strong evidence of a structural break for a wide range of break point times, suggesting that our finding is not sensitive to the exact choice of break point date.

6 Shadow Rate Model with a Structural Break

The evidence for structural instability discussed in the previous section naturally raises the question *how* the structure of the model has changed in the post-ELB period. While some of

[34]The pre- and post-ELB subsample averages are linked, as $\sum_{t=1}^{T_{pre}} \partial \ell_t(\hat{\theta})/\partial\theta = -\sum_{t=T_{pre}+1}^{T} \partial \ell_t(\hat{\theta})/\partial\theta$.

the diagnostics already shed light on this issue, we further investigate by estimating a model in which the yield dynamics in the pre-ELB and post-ELB periods are both described by the shadow rate model, but with different parameters, θ_{pre} and θ_{post}. Practically, this amounts to separate estimations of the model with pre-ELB and post-ELB subsamples.

6.1 Comparison of PCs

We begin with a discussion of PCA decomposition. To motivate the discussion, the upper panel of Table 6 shows the PCA loadings for the first three principal components of weekly yield changes, for pre-ELB sample (left side) and post-ELB sample (right side). The pre-ELB sample loadings show typical behaviors, consistent with the frequent "level," "slope," and "curvature" designation of PC1, PC2, and PC3: For example, the PC1 loadings are roughly similar (at least in the 1-year to 10-year range), while PC2 loadings monotonically increase with maturity. In the post-ELB sample case, however, the loadings do not display these typical patterns: the PC1 loadings increase with maturity for short and intermediate maturities, while the PC2 loadings are no longer monotonically increasing in maturity. The presence of the ELB likely explains much of this atypical behavior. In particular, the ELB compresses the volatility of short-maturity yields, and the usual notions of "level" and "slope" get intermingled, as any rise in the level of the yield curve during the ELB would likely also be associated with an increase in the slope of the yield curve.

But is the ELB the whole story behind the qualitative difference between the pre-ELB and post-ELB PCs? *Structurally stable* shadow rate models, including those that have been estimated in the literature, would imply so, i.e., the PCA decomposition of the changes in *shadow yields* (which are unaffected by the ELB) should be the same between the pre-ELB and post-ELB periods.

The lower panel of Table 6 shows the loadings for the first three components of changes in *shadow yields*, implied by the first subsample estimate $\hat{\theta}_{pre}$ and by the second subsample estimate $\hat{\theta}_{post}$. Although shadow yields are unobserved, the PC loadings based on instantaneous changes in shadow yields implied by the model parameters can be tractably evaluated. Recall from eq. (6) that shadow yields are affine in the state vector. Denoting the vector of shadow yields at time t for a given set of m maturities $(\tau_1, \tau_2, ..., \tau_m)$ as y_t^s, we have, schematically,

$$y_t^s = a + Bx_t, \tag{30}$$

where a is an m-dimensional vector, and B is an $m \times N$ matrix. This, together with eq. (1), gives

$$(dy_t^s dy_t^{s'})/dt = B\Sigma\Sigma' B'. \tag{31}$$

Therefore, the singular value decomposition of this matrix

$$B\Sigma\Sigma' B' = P\Psi P', \tag{32}$$

where Ψ is a diagonal matrix, gives the PC1, PC2, PC3 loadings implied by the estimated model parameters (the first three columns of the P matrix). The lower panel of Table 6 also

		pre-ELB			post-ELB	
τ	PC1	PC2	PC3	PC1	PC2	PC3
0.25	0.2578	-0.7150	0.5128	0.0224	0.1032	0.6425
0.5	0.2829	-0.4273	-0.1066	0.0538	0.2031	0.5739
1	0.3601	-0.1749	-0.4435	0.1318	0.4302	0.2891
2	0.4338	0.0093	-0.4428	0.2900	0.5748	-0.1426
4	0.4599	0.1930	-0.1212	0.4819	0.3552	-0.3192
7	0.4244	0.3167	0.2841	0.5756	-0.1872	-0.0318
10	0.3795	0.3715	0.4873	0.5759	-0.5210	0.2261
		pre-ELB			post-ELB	
τ	PC1	PC2	PC3	PC1	PC2	PC3
0.25	0.22(0.03)	-0.62(0.02)	0.49(0.03)	0.11(0.04)	-0.58(0.02)	0.35(0.04)
0.5	0.28(0.02)	-0.48(0.01)	0.10(0.03)	0.14(0.04)	-0.53(0.01)	0.19(0.04)
1	0.37(0.01)	-0.28(0.03)	-0.34(0.02)	0.18(0.04)	-0.44(0.02)	-0.08(0.05)
2	0.44(0.01)	-0.03(0.04)	-0.52(0.01)	0.27(0.04)	-0.27(0.06)	-0.43(0.04)
4	0.46(0.01)	0.21(0.03)	-0.19(0.01)	0.42(0.02)	-0.01(0.08)	-0.57(0.02)
7	0.43(0.01)	0.34(0.02)	0.27(0.02)	0.56(0.02)	0.21(0.05)	-0.08(0.06)
10	0.39(0.02)	0.38(0.02)	0.50(0.02)	0.61(0.04)	0.29(0.03)	0.57(0.02)

Table 6: The upper panel shows the PC1, PC2, PC3 loadings for the PC decomposition of weekly changes in yields in the pre-ELB period (left hand side) and in the post-ELB period (right hand side). The lower panel shows the loadings for the PC decomposition of instantaneous changes in shadow yields implied by the pre-ELB subsample estimate $\hat{\theta}_{pre}$ (left hand side) and by the post-ELB subsample estimate $\hat{\theta}_{post}$ (right hand side). These PCs are based on maturities (τ) of 0.25, 0.5, 1, 2, 4, 7, 10 years. Standard errors are shown in parentheses.

provides standard errors based on the delta method formula. As Duffee (2011) notes, this decomposition depends only on the parameters determining the risk-neutral dynamics of the model $(K_1^{\mathbb{Q}}, k_0^{\mathbb{Q}}, \Sigma, \rho_0, \rho_1)$, which are generally more precisely estimated than the $\mathbb{P}$-measure parameters. Therefore, even in the second subsample where the estimates are more uncertain than the first subsample, the standard errors associated with these loadings are of modest sizes.

Not surprisingly, the pre-ELB period PC decompositions based on $\hat{\theta}_{pre}$, shown on the left side of the lower panel of Table 6, are quite similar to those on the left side of the upper panel, as $\hat{\theta}_{pre}$ fits the pre-ELB period data reasonably well. On the other hand, the post-ELB period shadow yield PCs based on $\hat{\theta}_{post}$ do show differences compared to the the loadings on the right hand side of the upper panel; notably, the PC2 loadings are now monotonic in maturities. However, the post-ELB period shadow yield PCs still do not match the pre-ELB period shadow yield PCs well. In particular, the post-ELB period shadow yield PC1 loadings have notable slope, i.e., increase with maturities. In other words, PC1 in the post-ELB periods does not look like a standard "level shock." Qualitative difference between shadow yield PCs in the pre-ELB and post-ELB periods, like this one, adds to our evidence that yield curve dynamics have changed materially.

6.2 PC-Rotated Parameters

As discussed in Section 4.1, the examination of how much $\hat{\theta}_2$ differs from $\hat{\theta}_1$ is complicated by the fact that there are multiple images of $\hat{\theta}_2$ with identical empirical content (discrete invariant transformations). Rotating the state variables based on the PCA decomposition of shadow yield changes discussed above provides a natural means to surmount this problem, as shocks to the transformed the state vector $x_t^\dagger$ now have a specific meaning as "level," "slope," and "curvature" shocks, thus allowing for an "apples-to-apples" comparison.

Specifically, we perform the transformation in eq. (16), with L given by[35]

$$L = P'B, \tag{33}$$

with the P and B matrices in eq. (32). The transformed parameters ($\theta^\dagger$ in eq. (17)) from the first subsample estimation and from the second subsample estimation (i.e., $\hat{\theta}^\dagger_{pre}$ and $\hat{\theta}^\dagger_{post}$) can then be compared on an equal footing. One caveat is that the PCs are defined only up to their sign. For example, if the $x^\dagger_{2t}$ is a slope factor, $-x^\dagger_{2t}$ is also a slope factor. To eliminate this ambiguity, we define the factors such that the "level factor" loadings are generally positive (as opposed to generally negative), the "slope factor" loadings increase with maturity (as opposed to decreasing with maturity), and the "curvature factor" loadings are a U-shaped function of maturity (instead of an inverted U-shape). Implicitly, this imposes an ordering based on principal components, and assumes that the ordering is preserved in a structural change, e.g., the level factor in the first subsample does not become the slope factor in the second subsample. We argue this is a weaker assumption than imposing a more artificial ordering (based on parameters that are less intuitive).

Table 7 shows the PCA-rotated parameters, $\hat{\theta}^\dagger_{pre}$ for the pre-ELB sample and $\hat{\theta}^\dagger_{post}$ for post-ELB sample, which were originally estimated based on a generic normalization (eq. (18)); the standard errors are computed with the delta method.[36] Note that the table shows more parameters than are necessary to estimate the model, as some of these are linked to each other; for example, knowing two of $K_1^{\mathbb{P}}$, $K_1^{\mathbb{Q}}$, and Λ_1 determines the third; recall eq. (14).

It can be seen that pre-ELB period parameters are often more precisely estimated than the post-ELB period parameters, i.e., $\theta^\dagger_{pre}$ tends to have smaller standard errors than $\theta^\dagger_{post}$. In addition, the $\mathbb{Q}$-measure parameters tend to be estimated with smaller standard errors than the $\mathbb{P}$-measure counterparts, reflecting the fact that $\mathbb{Q}$-measure parameters are determined in large part from cross-sectional information. Indeed, the standard errors for most elements of the $K_1^{\mathbb{Q}\dagger}$ matrices for both $\hat{\theta}_{pre}$ and $\hat{\theta}_{post}$ are small enough to indicate statistically significant changes in $\mathbb{Q}$-measure dynamics. This adds to the indication from the fitting errors (Section 5.1.1) that the $\mathbb{Q}$-measure dynamics have changed.

[35] Here, we follow Duffee (2011), who considered the PCs of instantaneous changes in yields in affine-Gaussian models.

[36] The original parameter estimates are given in Appendix B.

	$\hat{\theta}^{\dagger}_{pre}$	$\hat{\theta}^{\dagger}_{post}$		$\hat{\theta}^{\dagger}_{pre}$	$\hat{\theta}^{\dagger}_{post}$
$[K_1^{\mathbb{P}\dagger}]_{11}$	-0.3245(0.0992)	-0.1039(0.1176)	$\rho_0^{\dagger}$	0.0450(0.0023)	0.0374(0.0097)
$[K_1^{\mathbb{P}\dagger}]_{21}$	-0.5771(0.1076)	-0.3488(0.0839)	$[\rho_1^{\dagger}]_1$	0.1372(0.0415)	0.0892(0.0426)
$[K_1^{\mathbb{P}\dagger}]_{31}$	0.0703(0.0327)	0.0317(0.0623)	$[\rho_1^{\dagger}]_2$	-0.7881(0.0552)	-0.6272(0.0420)
$[K_1^{\mathbb{P}\dagger}]_{12}$	0.1105(0.1725)	0.5613(0.3275)	$[\rho_1^{\dagger}]_3$	1.0756(0.0469)	0.5382(0.0320)
$[K_1^{\mathbb{P}\dagger}]_{22}$	-1.2233(0.3176)	-0.8334(0.3421)	$[\Sigma^{\dagger}]_{11}$	0.0209(0.0007)	0.0370(0.0119)
$[K_1^{\mathbb{P}\dagger}]_{32}$	0.0094(0.0266)	0.0372(0.0844)	$[\Sigma^{\dagger}]_{22}$	0.0086(0.0007)	0.0060(0.0007)
$[K_1^{\mathbb{P}\dagger}]_{13}$	2.5916(0.8509)	0.9523(1.0423)	$[\Sigma^{\dagger}]_{33}$	0.0032(0.0002)	0.0030(0.0004)
$[K_1^{\mathbb{P}\dagger}]_{23}$	5.6326(0.4849)	3.5576(0.6313)	$[\lambda_0^{\dagger}]_1$	-0.2102(0.1844)	-0.3395(0.4170)
$[K_1^{\mathbb{P}\dagger}]_{33}$	-0.7590(0.2928)	-0.4871(0.3646)	$[\lambda_0^{\dagger}]_2$	0.4708(0.3557)	1.2801(1.2667)
$[K_1^{\mathbb{Q}\dagger}]_{11}$	0.5469(0.0523)	0.2749(0.0122)	$[\lambda_0^{\dagger}]_3$	-0.9817(0.1722)	0.8064(1.3404)
$[K_1^{\mathbb{Q}\dagger}]_{21}$	-0.7085(0.1127)	-0.2958(0.0515)	$[\Lambda_1^{\dagger}]_{11}$	-41.7789(6.8946)	-10.2476(4.3675)
$[K_1^{\mathbb{Q}\dagger}]_{31}$	0.1410(0.0555)	-0.0508(0.0174)	$[\Lambda_1^{\dagger}]_{21}$	15.2818(8.5456)	-8.9001(9.0878)
$[K_1^{\mathbb{Q}\dagger}]_{12}$	1.4295(0.2204)	0.5185(0.1330)	$[\Lambda_1^{\dagger}]_{31}$	-22.0198(11.4565)	27.3390(17.6313)
$[K_1^{\mathbb{Q}\dagger}]_{22}$	-1.4008(0.2523)	-0.6274(0.1834)	$[\Lambda_1^{\dagger}]_{12}$	-63.2374(15.2839)	1.1569(10.2985)
$[K_1^{\mathbb{Q}\dagger}]_{32}$	0.2006(0.0207)	-0.0539(0.0345)	$[\Lambda_1^{\dagger}]_{22}$	20.6275(15.0579)	-34.6007(27.7273)
$[K_1^{\mathbb{Q}\dagger}]_{13}$	-2.9793(0.2268)	-0.6228(0.2122)	$[\Lambda_1^{\dagger}]_{32}$	-59.5334(8.9792)	30.2072(24.7968)
$[K_1^{\mathbb{Q}\dagger}]_{23}$	4.7374(0.1870)	2.2668(0.0781)	$[\Lambda_1^{\dagger}]_{13}$	267.0819(36.9095)	42.6123(30.4754)
$[K_1^{\mathbb{Q}\dagger}]_{33}$	-1.3247(0.2714)	-0.1417(0.1303)	$[\Lambda_1^{\dagger}]_{23}$	104.0556(59.7746)	216.8494(86.7756)
$[k_0^{\mathbb{Q}\dagger}]_1$	0.0044(0.0038)	0.0126(0.0146)	$[\Lambda_1^{\dagger}]_{33}$	176.1143(63.0105)	-114.5137(85.4108)
$[k_0^{\mathbb{Q}\dagger}]_2$	-0.0041(0.0031)	-0.0076(0.0078)			
$[k_0^{\mathbb{Q}\dagger}]_3$	0.0032(0.0004)	-0.0024(0.0040)			

Table 7: PC-rotation ($\theta^{\dagger}_{pre}$, $\theta^{\dagger}_{post}$) of the estimated parameters for the pre-ELB subsample and post-ELB subsample. Standard errors are shown in parentheses. The parameter vector $k_0^{\mathbb{P}}$ is zero by normalization.

The $\mathbb{P}$-measure parameter estimates $K_1^{\mathbb{P}\dagger}$ also indicate substantial change between the pre-ELB and post-ELB periods. The standard errors are quite sizable, especially for the post-ELB period estimate, but some of the elements of $K_1^{\mathbb{P}\dagger}$ still show statistically significant change. Similarly, the matrix $\Lambda_1^{\dagger}$ shows substantial changes, with some of the elements even flipping signs (although the standard errors here are also fairly large). This suggests significant changes in the structure of market price of risk, in addition to the changes in $\mathbb{P}$ and $\mathbb{Q}$ dynamics. Below, we examine how these changes in market price of risk translate to different implications for expectations hypothesis regressions.

Finally, note that the estimate of the long-run mean of the shadow rate ($\rho_0^{\dagger}$) is somewhat lower in $\theta^{\dagger}_{post}$ than in $\theta^{\dagger}_{pre}$, adding further credence to the view that r-star has declined since Financial Crisis (recall the discussion in Section 3.4), albeit with large estimation uncertainty.

6.3 EH Regressions for Shadow Yields

Additional insights on how the model has changed can be gleaned from model-implied expectations hypothesis (EH) regression coefficients.[37]

Campbell-Shiller			
	n	$\beta(\hat{\theta}_{pre})$	$\beta(\hat{\theta}_{post})$
	2	-0.2181 (0.1386)	1.0367 (0.6128)
	3	-0.2527 (0.1322)	0.9087 (0.6903)
	6	-0.3469 (0.1150)	0.5429 (0.9432)
	12	-0.4933 (0.0903)	-0.0865 (1.2645)
	24	-0.6543 (0.0755)	-0.9392 (1.0551)
	36	-0.7203 (0.0918)	-1.4068 (0.4604)
	48	-0.7527 (0.1198)	-1.6552 (0.6097)
	60	-0.7777 (0.1518)	-1.7798 (1.2152)
	84	-0.8393 (0.2200)	-1.8352 (2.2844)
	120	-0.9813 (0.3225)	-1.6951 (3.5157)
Fama			
	n	$\beta(\hat{\theta}_{pre})$	$\beta(\hat{\theta}_{post})$
	6	0.3173 (0.0668)	0.6428 (0.5004)
	12	0.2582 (0.0662)	0.3518 (0.5240)
	18	0.2269 (0.0676)	0.1815 (0.4152)
	24	0.2120 (0.0699)	0.0820 (0.2855)
	30	0.2051 (0.0722)	0.0226 (0.1761)
	36	0.2019 (0.0744)	-0.0141 (0.0983)

Table 8: Model-implied expectations hypothesis regression coefficients. The upper panel shows the Campbell-Shiller regression coefficients implied by the pre-ELB and post-ELB subsample parameter estimates, $\hat{\theta}_{pre}$ and $\hat{\theta}_{post}$. The lower panel shows the implied Fama regression coefficients. Standard errors are shown in parentheses.

We consider two such regressions, namely the Campbell and Shiller (1991) regression

$$y^s_{t+1,n-1} - y^s_{t,n} = \alpha + \beta \frac{1}{n-1}(y^s_{t,n} - y^s_{t,1}) + e_{t+1,n}, \tag{34}$$

and the Fama (1984) regression,

$$y^s_{t+n,1} - y^s_{t,1} = \alpha + \beta(f^s_{t,n} - y^s_{t,1}) + e_{t+n,n}, \tag{35}$$

[37]Studies like Backus, Foresi, Mozumdar, and Wu (2001) and Dai and Singleton (2002) have examined term structure model–implied beta coefficients in EH regressions as a part of model evaluation, while our interest with these model-implied coefficients is in characterizing the change between the two sample periods (pre-ELB and post-ELB).

where $y^s_{t,n}$ is the n-period shadow yield at time t, and $f^s_{t,n}$ is n-period ahead shadow forward rate at time t, i.e., $f^s_{t,n} = \log(P^s_{t,n}/P^s_{t,n+1})$. We are defining these regressions in terms of shadow yields/forward rates, rather than true yields and forward rates, to facilitate the investigation of potential differences between the pre-ELB and post-ELB periods: Because shadow yields/forward rates are unaffected by the ELB if the model is structurally stable, the β's from the pre-ELB period should equal the β's from the post-ELB period in these regressions.

It can be shown that for $n = 2$, the two regressions contain the same information, as the two regression coefficients are related by the formula[38]

$$\beta^{CS} = 2\beta^F - 1. \tag{36}$$

However, for $n > 2$, the two regressions probe somewhat different aspects of the departures from the expectations hypothesis. The Campbell-Shiller regression is closely linked to the excess bond return predictability regression

$$er^s_{t+1,n} = \alpha + \beta^{er}(y^s_{t,n} - y^s_{t,1}) + e_{t+1,n} \tag{37}$$

where the shadow bond excess return $er^s_{t+1,n}$ is given by $er^s_{t+1,n} = \log(P^s_{t+1,n-1}/P^s_{t,n}) - y^s_{t,1}$. It is straightforward to show that

$$\beta^{CS} = 1 - \beta^{er}. \tag{38}$$

The Fama regression probes the departure from the forward rate expectations hypothesis, hence the forward term premium $f_{t,n} - E_t[y_{t+n,1}]$. Note that in both Campbell-Shiller regressions and Fama regressions, the expectations hypothesis holding means that $\beta = 1$, while in the excess return predictability regressions, the expectations hypothesis corresponds to $\beta^{er} = 0$.

Table 8 shows model-implied EH regression coefficients along with their standard errors (calculated with the delta method) for both the pre-ELB sample and post-ELB sample.[39] The existing literature on expectations hypothesis tests has found that the Campbell-Shiller regression coefficient is often negative, while the Fama regression coefficient is typically less than 1 but larger than 0. The implied regression coefficients for the pre-ELB sample are consistent with these patterns. Meanwhile, the implied coefficients in the post-ELB period display interesting qualitative differences from the pre-ELB sample. In particular, both with β^{CS}_n and with β^F_n, for low n, the coefficients are close to 1; in other words, shadow yield/forward rate dynamics are close to the EH for short maturities in the post-ELB sample. But as n gets larger, the departures from EH based on the post-ELB sample estimation get even more pronounced than those based on the pre-ELB sample estimation, with β^{CS}_n and

[38] Backus et al. (2001) have shown that what they call "forward rate regression" (their eq. (4)) has identical content to the the Campbell-Shiller regression (eq. (34) here) for $n = 2$.

[39] The model-implied regression coefficients are straightforward to compute. For example, in the case of the Campbell-Shiller coefficient, it is given by $\beta^{CS}_n = (n-1)\,Cov(y^s_{t+1,n-1} - y^s_{t,n}, y^s_{t,n} - y^s_{t,1})/\,Var(y^s_{t,n} - y^s_{t,1}) = (n-1)\frac{(b'_{n-1}e^{(1/12)K^{\mathbb{P}}_1} - b_n)V_x(b_n - b_1)}{(b'_n - b'_1)V_x(b_n - b_1)}$ for n measured in months, where V_x is the unconditional variance-covariance matrix of the state variables (implied by the model parameters).

β_n^F for the post-ELB sample taking lower values than those for the pre-ELB sample. The Campbell-Shiller beta for larger n's (such as $n = 120$ months) taking a more negative value in the post-ELB period than in the pre-ELB period is reminiscent of Andreasen et al. (2019), who obtain similar results with actual yields[40]

Somewhat not surprisingly, with limited amount of post-ELB period data, the standard errors for that period are fairly large. While these results therefore might not constitute a sufficient body of evidence by themselves, they are suggestive of a change in the pricing of interest rate risk, and further add to the finding in the previous section (Section 6.2) that the parameters that describe the market price of risk ($\Lambda^\dagger$ parameters) have changed meaningfully between the two periods.

6.4 Expectations and Term Premiums

Here we compare some of the predictive outputs from our various models to illustrate the differences concretely in terms of quantities of economic interest.

We start with implied time series of the shadow rate s_t. The shadow rate is not a quantity of economic interest per se. It is an unobserved variable, and its value does not have a simple relationship with quantities of economic interest such as *actual* bond yields, interest rate expectations, and term premiums. Moreover, its value can depend materially on the model specification: For example, Kim and Singleton (2012) find that in the case of Japanese data, the quadratic-Gaussian specification of the shadow rate leads to a different behavior of the shadow rate than the affine-Gaussian specification of the shadow rate. Meanwhile, studies including Krippner (2013) and Wu and Xia (2016) with U.S. data indicate that the shadow rate estimate varies significantly with the number of factors in the model. Nonetheless, as a central ingredient in the model, closer examination of the shadow rate might shed at least some light on the structural change in the model.

Figure 4 shows the estimates of the shadow rate over the post-ELB period based on the pre-ELB sample parameter estimate (structural stability assumption) as well as post-ELB sample parameter estimate (structural break assumption). Interestingly, the shadow rate assuming structural break is generally less negative than the shadow rate assuming structural stability. However, a specific ordering of shadow rates on a given date does not imply the same ordering for expected *future* shadow rates. Figure 4 illustrates this, with model-implied expected shadow rate paths (the dashed lines) as of December 17, 2008 (the first date in the post-ELB subsample), March 11, 2015 (the sample date on which the difference between the two shadow rate estimates is largest), and June 26, 2019 (the end of the sample). The first and last dates illustrate that a similar shadow rate estimate can imply substantially different expected paths; at the end of the sample, these paths even point in different directions. The paths for March 11, 2015, show that, even though the two models initially imply dramatically different shadow rates, the expected path from the model assuming structural stability catches up quickly with the expected path based on the

[40] They consider the excess return regression (eq. (37) with actual yields instead of shadow yields). Recall this is related to the Campbell-Shiller regressions we examine by eq. (38).

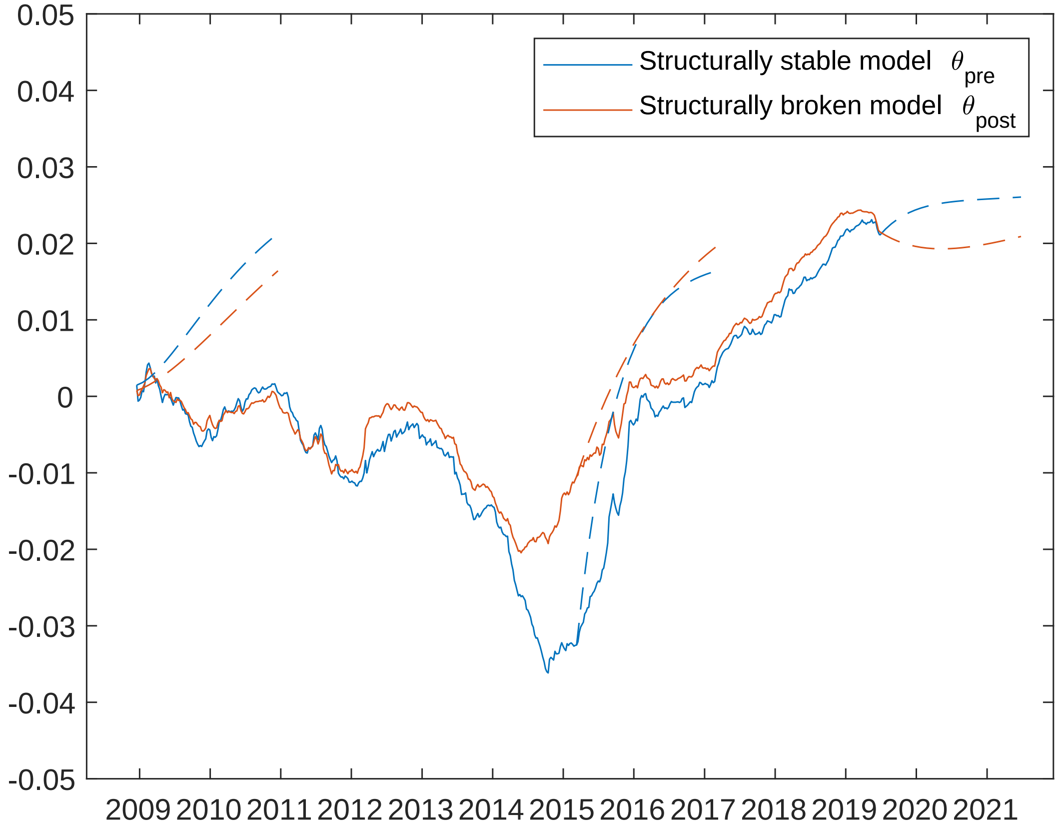

Figure 4: Time series of estimated shadow rates (solid lines) from the structurally stable model based on pre-ELB parameter estimates (and subsequently extended to the post-ELB period as a shadow rate model), and the structurally broken shadow rate model. Dashed lines show model-implied expected shadow rate paths out to two-year horizons for the following three dates: 12/17/2008, 03/11/2015, and 06/26/2019.

structural break assumption; the two paths are virtually on top of each other for a period, only for the path from the structurally stable model to flatten out *faster* than the path from the structurally broken model.

Turning to the expected path of the short rate ($E_t^{\mathbb{P}}(r_{t+\tau})$), we find that the model with structural stability assumption generally (although not universally) implies a steeper path of the short rate during the post-ELB period, i.e., faster normalization of the policy rate (compared to the model with structural break). Figure 5 illustrates this for a couple of dates (after the introduction of date-based forward guidance in August 2011, and after liftoff in December 2015). This feature seems to be related to our earlier finding in Section 5.1.2 that the model estimated with pre-ELB sample data produces slope innovation vectors that have substantial serial correlation in the post-ELB sample. Another notable feature in Figure 5 is that the Treasury forward rate curves are generally lower than the model-implied expected paths of the short rate, implying that term premiums are negative. Indeed, as shown in the top panel of Figure 6, the time series of 2-year yield term premiums implied by both models are generally negative (the chart also shows the term premium implied by the structurally stable, full-sample shadow rate model). For most of the post-ELB period, the 2-year term premium implied by the model with structure break is less negative than the models with structural stability assumption, consistent with the latter model generating steeper paths of the expected short rate, as discussed above. The fact that the model with structural break generates near-term term premiums that are closer to 0 in the post-ELB period also seems to be consistent with the finding in Section 6.3 that the implied Fama regression coefficients based on $\hat{\theta}_{post}$ are closer to 1 (as would be implied by the EH) for short horizons, relative to $\hat{\theta}_{pre}$. On the other hand, as shown in the bottom panel of Figure 6, the 10-year term premiums based on the models with and without structural break are qualitatively more similar to each other. The 10-year term premium based on the model with structural break is somewhat lower than that from the structurally stable model in the 2011–2016 period and somewhat higher in the period since 2018.

To better understand the different orderings of 2- and 10-year term premiums implied by the different models, we compare the model-implied longer-horizon short rate expectations in Figure 7. The figure shows times series of 5-to-10-year-ahead expected average short rates implied by models with and without structural break. The 5-to-10-year Blue Chip survey expectations are also shown. The model-implied expectations show a fair amount of variation over time, and roughly track the contour of survey expectations. However, the longer-horizon expectations based on the models do display some interesting differences. While both the structurally stable full-sample model and the model with structural break use the surveys as estimation inputs, the model with structural break tracks them much more closely, suggesting that structurally stable models have greater difficulty reconciling short rate expectations over the entire sample. From mid-2011 to 2016, the structurally stable model based on pre-ELB data implies substantially lower long-run expectations than the model with structural break and surveys. This more than offsets the the steeper expected path for near-term horizons relative to the model with structural break, leading to the somewhat less negative 10-year term premium for the structurally stable model in this period, as we saw in Figure 6. From

2018 onward, the expectations implied by the model with structural break are notably below those implied by the other two models.

7 Conclusion

In this paper, we examined the structural stability of shadow rate term structure models as applied to U.S. Treasury yield data. We found various pieces of evidence pointing to structural instability, including diagnostics based on innovation vectors and likelihood score–based tests. We further elaborated on the changes in yield dynamics with estimation of a shadow rate model in which the pre-ELB and post-ELB periods are described by different sets of parameters. Overall, the results presented in this paper point to extensive changes in the shadow yield dynamics since around the Financial Crisis. The structural change does not appear to be confined to a single aspect of the model (say, only the $\mathbb{P}$-dynamics), but instead spans many facets of the model, including the principal components (of shadow yields), $\mathbb{P}$-dynamics, $\mathbb{Q}$-dynamics, and risk pricing. The results also indicate that ignoring structural change can lead to notable differences in quantities of practical interest, e.g., ignoring structural change implies generally steeper expected paths of the short rate and generally more negative near-term term premiums in the post-ELB period.

In sum, we find that the yield dynamics in the post-ELB regime, at least in the case of U.S. data, is not as simple as suggested by structurally stable shadow rate models, and that more material changes in the dynamics of Treasury yields took place after the federal funds rate hit the ELB in 2008. Our paper has examined this issue from the perspective of a shadow rate model with a structural break, but the empirical results herein may be also be consistent with broader misspecification of shadow rate models. It could be that a fundamentally different model would yield a more satisfying description and richer insights into the dynamics of yields near the ELB.

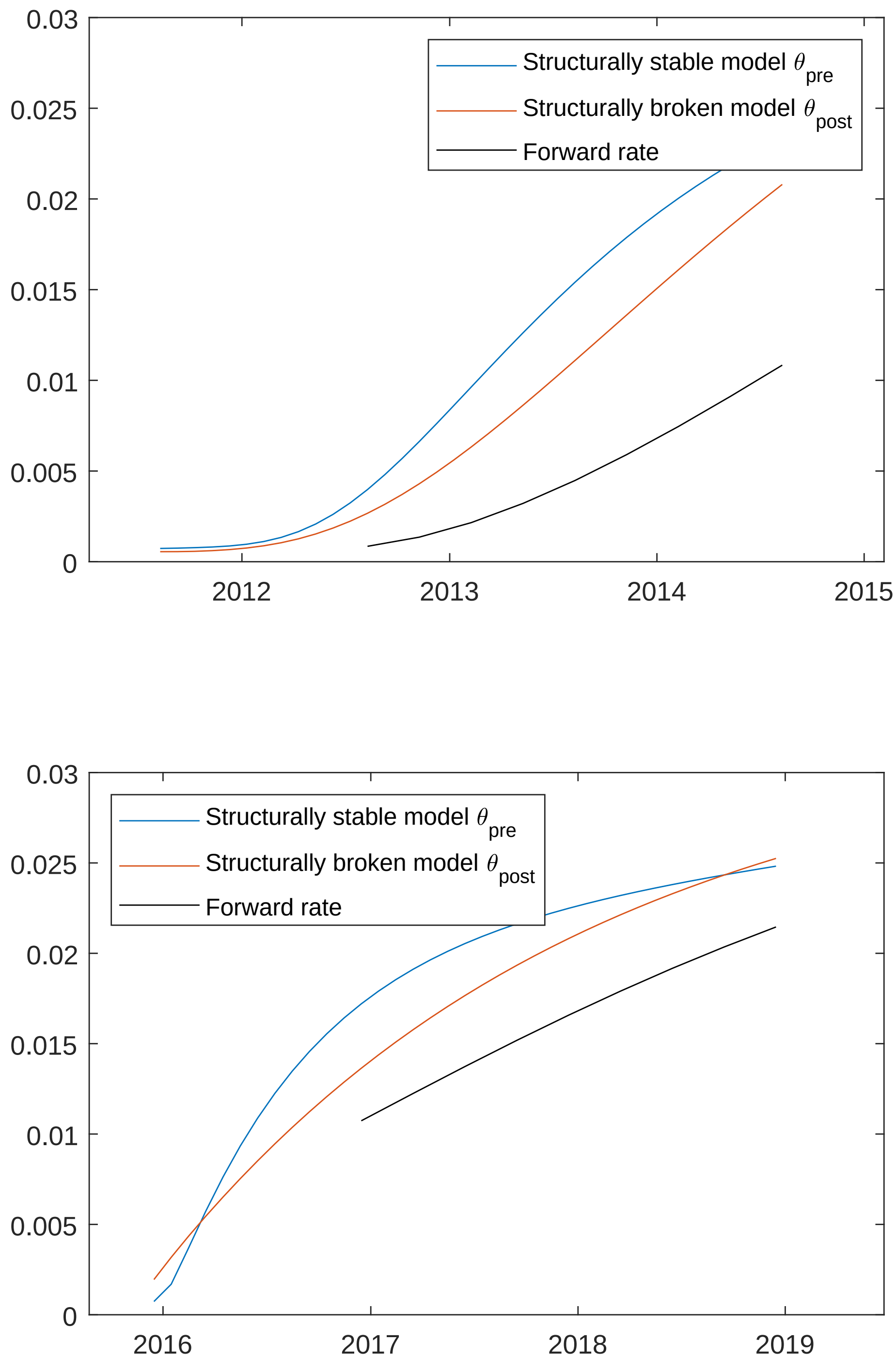

Figure 5: Model-implied expected short rate paths after introduction of date-based forward guidance in August 2011 (top panel) and after liftoff in December 2015 (bottom panel). The figures also plot the instantaneous Treasury forward rates on those dates (based on the Svensson curve as in Gürkaynak et al. (2007)), starting at the one-year horizon.

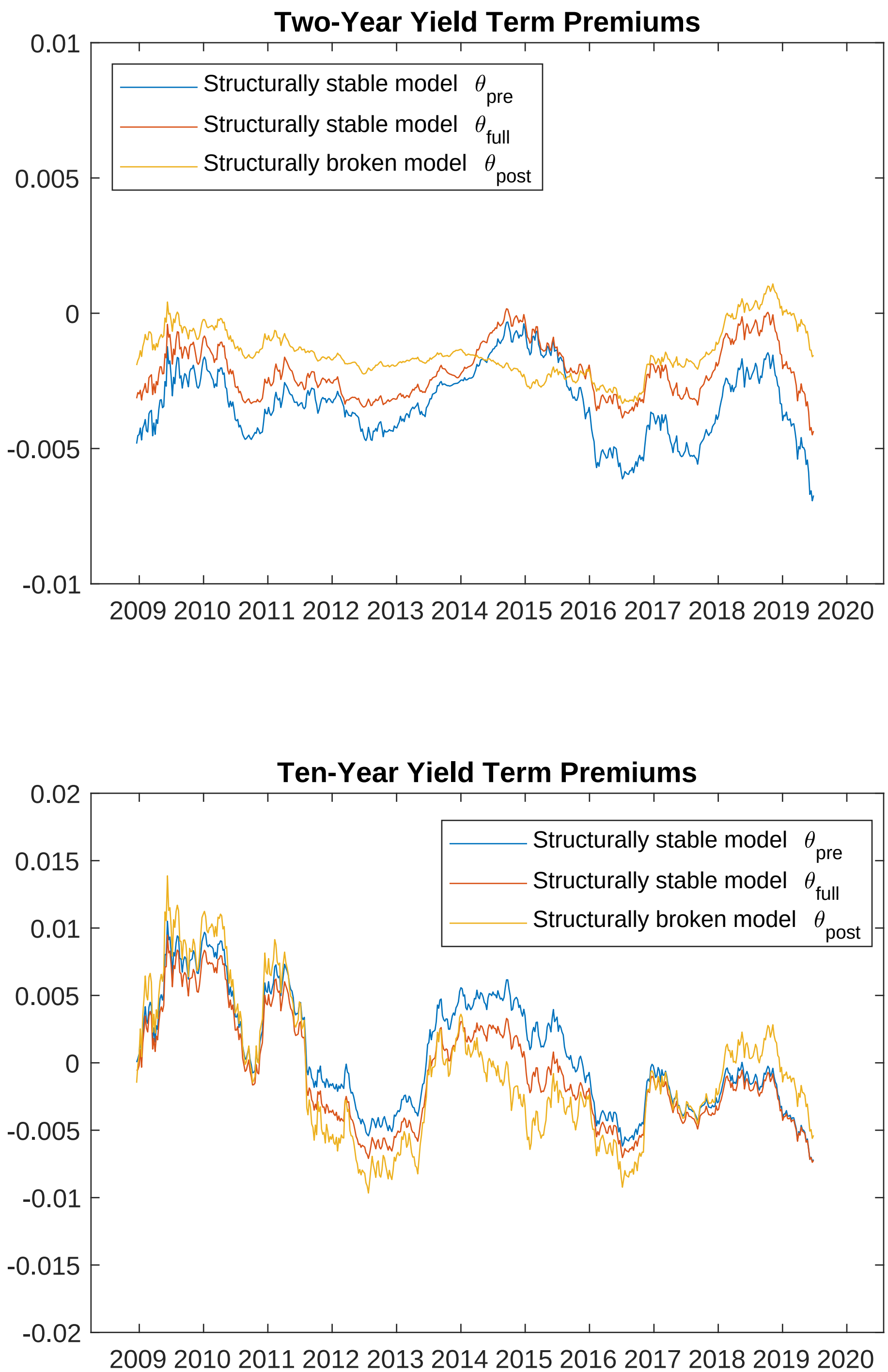

Figure 6: Two- (top panel) and ten-year (bottom panel) yield term premiums implied by our different models, shown for the post-ELB subsample.

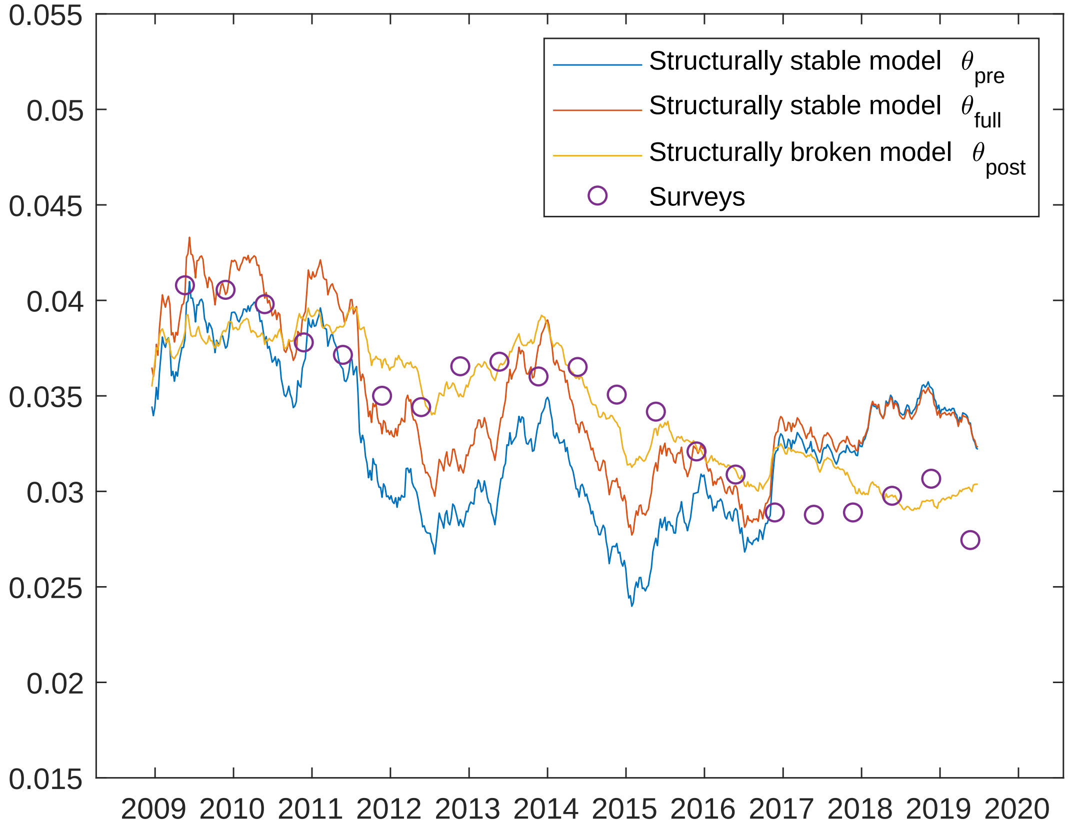

Figure 7: Five-to-ten-year expected average short rates as implied by our models during the post-ELB subsample, as well as Blue Chip survey expectations (Wolters Kluwer Legal and Regulatory Solutions U.S., Blue Chip Financial Forecasts).

A Innovation Statistics

This appendix proves invariance to model rotation and asymptotic distribution of the innovation-based statistics discussed in Sections 4.3 and 5.1.2.

Consider a general first-order Gaussian VAR. That is, suppose the N-dimensional state vector x_t follows the transition equation $x_t = m_0 + M_1 x_{t-1} + \varepsilon_t$, where the ε_t are i.i.d. $N(0, \Omega)$, with $\Omega > 0$. Choose any matrix decomposition such that $SS' = \Omega$ (for example, the Cholesky decomposition) and let $\eta_t = S^{-1}\varepsilon_t = S^{-1}(x_t - m_0 - M_1 x_{t-1})$. Then by construction, the η_t are i.i.d. $N(0, I)$. Next, take an arbitrary invariant transformation, that is, an N-vector l and invertible $N \times N$-matrix L, and let $\tilde{x}_t = l + Lx_t$. Then $\tilde{x}_t = \tilde{m}_0 + \tilde{M}_1 \tilde{x}_{t-1} + \tilde{\varepsilon}_t$, where the $\tilde{\varepsilon}_t$ are i.i.d. $N(0, \tilde{\Omega})$, with $\tilde{m}_0 = (I - LM_1L^{-1})l + Lm_0$, $\tilde{M}_1 = LM_1L^{-1}$, and $\tilde{\Omega} = L\Omega L'$. If we choose any $\tilde{S}$ such that $\tilde{S}\tilde{S}' = \tilde{\Omega}$ and let $\tilde{\eta}_t = \tilde{S}^{-1}\tilde{\varepsilon}_t = \tilde{S}^{-1}(\tilde{x}_t - \tilde{m}_0 - \tilde{M}_1\tilde{x}_{t-1})$, then, denoting the Frobenius matrix norm by $||\cdot||_F$, for $u = 1, 2, \ldots$,

$$\begin{aligned}
||T^{-1}{\textstyle\sum_t}\tilde{\eta}_t\tilde{\eta}'_{t-u}||_F^2 &= ||T^{-1}{\textstyle\sum_t}\tilde{S}^{-1}\tilde{\varepsilon}_t\tilde{\varepsilon}'_{t-u}\tilde{S}^{-1\prime}||_F^2 \\
&= T^{-2}\mathrm{tr}(\tilde{S}^{-1}{\textstyle\sum_t}\tilde{\varepsilon}_{t-u}\tilde{\varepsilon}'_t\tilde{S}^{-\prime}\tilde{S}^{-1}{\textstyle\sum_t}\tilde{\varepsilon}_t\tilde{\varepsilon}'_{t-u}\tilde{S}^{-1\prime}) \\
&= T^{-2}\mathrm{tr}(\tilde{S}^{-1}L{\textstyle\sum_t}\varepsilon_{t-u}\varepsilon'_tL'(\tilde{S}\tilde{S}')^{-1}L{\textstyle\sum_t}\varepsilon_t\varepsilon'_{t-u}L'(\tilde{S}^{-1\prime}) \\
&= T^{-2}\mathrm{tr}(\Omega^{-1}{\textstyle\sum_t}\varepsilon_{t-u}\varepsilon'_t\Omega^{-1}{\textstyle\sum_t}\varepsilon_t\varepsilon'_{t-u}) \\
&= T^{-2}\mathrm{tr}(S^{-1}{\textstyle\sum_t}\varepsilon_{t-u}\varepsilon'_tS^{-1\prime}S^{-1}{\textstyle\sum_t}\varepsilon_t\varepsilon'_{t-u}S^{-1\prime}) \\
&= ||T^{-1}{\textstyle\sum_t}\eta_t\eta'_{t-u}||_F^2.
\end{aligned}$$

The statistic $||T^{-1}\sum_t \eta_t\eta'_{t-u}||_F^2$ is thus invariant both to the original choice of S (say, whether the Cholesky decomposition or a symmetric matrix square root is used), and to observationally equivalent model rotations. By an analogous argument, the same is true for $||T^{-1}\sum_t \eta_t\eta'_t - I||_F^2$.

Moreover, since the elements of η_t are independent both contemporaneously and across time, $E(\eta_{t,i}\eta_{t-u,j})$ is equal to 1 if $u = 0$ and $i = j$, and is equal to 0 otherwise. Similarly, $Cov(\eta_{t,i}\eta_{t-u,j}, \eta_{t,p}\eta_{t-v,q})$ is 2 if $u = v = 0$ and $i = j = p = q$; is equal to 1 if $u = v = 0$ and either $i = p \neq j = q$ or $i = q \neq j = p$, or if $u = v > 0$ and either $i = p, j = q$ or $i = q, j = p$; and, is equal to 0 otherwise. Thus, by the Law of Large Numbers for random vectors,

$$\sqrt{T}\begin{pmatrix} \mathrm{vech}\left[(T^{-1}{\textstyle\sum_t}\eta_t\eta'_t - I) \otimes (I + J)^{\circ -1/2}\right] \\ \mathrm{vec}\left(T^{-1}{\textstyle\sum_t}\eta_t\eta'_{t-1}\right) \\ \vdots \\ \mathrm{vec}\left(T^{-1}{\textstyle\sum_t}\eta_t\eta'_{t-u}\right) \end{pmatrix} \xrightarrow{d} N(0, I)$$

as $T \to \infty$. Note $\eta_t\eta'_{t-u}$ is symmetric only when $u = 0$, which is why we use half-vectorization in that case; furthermore, the diagonal elements in that matrix have variance 2, which is accounted for by the Hadamard operation $\otimes(I + J)^{\circ -1/2}$ which divides the diagonal elements by $\sqrt{2}$. Using this convergence result, the Continuous Mapping Theorem implies that

$$\begin{aligned}
\tfrac{1}{2}T||T^{-1}{\textstyle\sum_t}\eta_t\eta'_t - I||_F^2 &\xrightarrow{d} \chi^2_{(N^2+N)/2} \\
T||T^{-1}{\textstyle\sum_t}\eta_t\eta'_{t-u}||_F^2 &\xrightarrow{d} \chi^2_{N^2} \qquad u = 1, 2, \ldots
\end{aligned}$$

as $T \to \infty$. Furthermore, as the individual statistics are independent under their asymptotic distribution, they can be added to obtain a joint statistic with asymptotic χ^2 distribution and degrees of freedom equal to the sum of the individual degrees of freedom.

B Parameter Estimates

This appendix provides the parameter estimates, in original normalization (eq. (18)), based on the pre-ELB sample ($\hat{\theta}_{pre}$), post-ELB sample ($\hat{\theta}_{post}$), and full sample ($\hat{\theta}_{full}$). The PC-rotated parameter estimates $\hat{\theta}^{\dagger}_{pre}$ and $\hat{\theta}^{\dagger}_{post}$ in Section 6.2 are invariant transformations of $\hat{\theta}_{pre}$ and post-ELB sample $\hat{\theta}_{post}$, respectively.

Recall that, to achieve econometric identification, we impose the normalization restrictions $k_0^{\mathbb{P}} = 0_{3\times 1}$, $\Sigma = 0.01I$, $[K_1^{\mathbb{P}}]_{i,j(i<j)} = 0$.

A few properties of our estimates are of note: In $\hat{\theta}_{pre}$, the estimate of $[K_1^{\mathbb{P}}]_{33}$ does not have a standard error because the estimate is at the boundary ($[K_1^{\mathbb{P}}]_{33} = [K_1^{\mathbb{P}}]_{22}$); that is, the matrix has a repeated eigenvalue. In $\hat{\theta}_{post}$, the estimate of δ_y does not have a standard error as it is at the minimum bound (4 bps) that we imposed. If this parameter is left unconstrained in the post-ELB sample estimation, the estimated $\mathbb{Q}$ parameters display strong signs of overfitting.

In all three models, we impose a lower bound of 50 bps on the measurement error estimate for the 5-to-10-year survey expectation.

	$\hat{\theta}_{pre}$	$\hat{\theta}_{post}$	$\hat{\theta}_{full}$
ρ_0	0.0450 (0.0023)	0.0374 (0.0097)	0.0465 (0.0028)
$[\rho_1]_1$	0.4051 (0.1051)	0.2541 (0.0824)	0.3512 (0.0315)
$[\rho_1]_2$	-0.1366 (0.0803)	0.1596 (0.0600)	0.0449 (0.0485)
$[\rho_1]_3$	0.6915 (0.0711)	0.4295 (0.0615)	0.7472 (0.0533)
$[K_1^{\mathbb{P}}]_{11}$	-0.0637 (0.0192)	-0.1109 (0.4311)	-0.0917 (0.0167)
$[K_1^{\mathbb{P}}]_{21}$	0.0584 (0.1125)	0.1390 (0.3667)	0.1235 (0.0603)
$[K_1^{\mathbb{P}}]_{31}$	-0.0744 (0.2951)	-0.2366 (0.7887)	-0.0151 (0.0582)
$[K_1^{\mathbb{P}}]_{22}$	-1.1215 (0.0538)	-0.6351 (0.0956)	-0.8391 (0.2096)
$[K_1^{\mathbb{P}}]_{32}$	2.5296 (0.3192)	2.8466 (0.9364)	1.8229 (0.2069)
$[K_1^{\mathbb{P}}]_{33}$	-1.1215	-0.6784 (0.0576)	-0.9855 (0.0375)
$[k_0^{\mathbb{Q}}]_1$	0.0073 (0.0015)	-0.0069 (0.0098)	0.0049 (0.0002)
$[k_0^{\mathbb{Q}}]_2$	-0.0068 (0.0016)	0.0063 (0.0086)	-0.0022 (0.0014)
$[k_0^{\mathbb{Q}}]_3$	0.0048 (0.0038)	0.0124 (0.0127)	0.0045 (0.0012)
$[K_1^{\mathbb{Q}}]_{11}$	0.1194 (0.1783)	-0.3428 (0.1242)	0.0138 (0.0066)
$[K_1^{\mathbb{Q}}]_{21}$	0.1212 (0.1129)	0.3423 (0.1345)	0.1212 (0.0089)
$[K_1^{\mathbb{Q}}]_{31}$	0.3998 (0.3284)	0.0872 (0.4703)	0.1682 (0.0074)
$[K_1^{\mathbb{Q}}]_{12}$	1.4357 (0.2536)	-0.5689 (0.3800)	0.6421 (0.0099)
$[K_1^{\mathbb{Q}}]_{22}$	-0.9906 (0.1593)	0.3044 (0.2306)	-0.4941 (0.0166)
$[K_1^{\mathbb{Q}}]_{32}$	2.5270 (0.3326)	2.1732 (0.6383)	1.7591 (0.0097)
$[K_1^{\mathbb{Q}}]_{13}$	-0.6656 (0.0794)	0.0465 (0.0371)	-0.2775 (0.0072)
$[K_1^{\mathbb{Q}}]_{23}$	0.1049 (0.0692)	-0.0081 (0.0272)	-0.0464 (0.0081)
$[K_1^{\mathbb{Q}}]_{33}$	-1.3073 (0.1178)	-0.4557 (0.0813)	-0.8964 (0.0097)
δ_y	0.0006 (0.0000)	0.0004	0.0006 (0.0000)
$\delta_{z,6m}$	0.0017 (0.0001)	0.0011 (0.0001)	0.0018 (0.0001)
$\delta_{z,12m}$	0.0023 (0.0001)	0.0017 (0.0001)	0.0025 (0.0007)
$\delta_{z,5-10y}$	0.0050	0.0050	0.0050
$\underline{r}$		0.0006 (0.0000)	0.0007 (0.0000)

Table 9: Parameter estimates for the three models used in the paper.

References

Ahn, Dong-Hyun, Robert F. Dittmar, and A. Ronald Gallant (2002) "Quadratic Term Structure Models: Theory and Evidence," *The Review of Financial Studies*, 15 (1), 243–288, 10.1093/rfs/15.1.243.

Andreasen, Martin M., Kasper Jørgensen, and Andrew Meldrum (2019) "Bond Risk Premiums at the Zero Lower Bound," Working Paper 2019-040, Board of Governors of the Federal Reserve System, Washington, 10.17016/FEDS.2019.040.

Andrews, Donald W. K. (1993) "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61 (4), 821–856.

Andrews, Donald W. K. and Ray C. Fair (1988) "Inference in Nonlinear Econometric Models with Structural Change," *The Review of Economic Studies*, 55 (4), 615–639.

Ang, Andrew and Monika Piazzesi (2003) "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables," *Journal of Monetary Economics*, 50 (4), 745–787, 10.1016/S0304-3932(03)00032-1.

Backus, David, Silverio Foresi, Abon Mozumdar, and Liuren Wu (2001) "Predictable changes in yields and forward rates," *Journal of Financial Economics*, 59 (3), 281–311.

Bates, David S. (2000) "Post-'87 crash fears in the S&P 500 futures option market," *Journal of Econometrics*, 94 (1), 181–238, 10.1016/S0304-4076(99)00021-4.

Bauer, Michael D. and Glenn D. Rudebusch (2016) "Monetary Policy Expectations at the Zero Lower Bound," *Journal of Money, Credit and Banking*, 48 (7), 1439–1465, 10.1111/jmcb.12338.

Bernanke, Ben S. (2020) "The New Tools of Monetary Policy," *American Economic Review*, 110 (4), 943–83, 10.1257/aer.110.4.943.

Campbell, John Y. and Robert J. Shiller (1991) "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *The Review of Economic Studies*, 58 (3), 495–514.

Chow, Gregory C. (1960) "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28 (3), 591–605.

Christensen, Jens H. E. and Glenn D. Rudebusch (2014) "Estimating Shadow-Rate Term Structure Models with Near-Zero Yields," *Journal of Financial Econometrics*, 13 (2), 226–259, 10.1093/jjfinec/nbu010.

Chung, Hess, Jean-Philippe Laforte, David Reifschneider, and John C. Williams (2012) "Have We Underestimated the Likelihood and Severity of Zero Lower Bound Events?" *Journal of Money, Credit and Banking*, 44 (s1), 47–82, 10.1111/j.1538-4616.2011.00478.x.

Dai, Qiang and Kenneth J. Singleton (2000) "Specification Analysis of Affine Term Structure Models," *The Journal of Finance*, 55 (5), 1943–1978, 10.1111/0022-1082.00278.

——— (2002) "Expectation puzzles, time-varying risk premia, and affine models of the term structure," *Journal of Financial Economics*, 63 (3), 415–441, 10.1016/S0304-405X(02)00067-3.

Duffee, Gregory R. (2002) "Term Premia and Interest Rate Forecasts in Affine Models," *The Journal of Finance*, 57 (1), 405–443, 10.1111/1540-6261.00426.

——— (2011) "Information in (and not in) the Term Structure," *The Review of Financial Studies*, 24 (9), 2895–2934, 10.1093/rfs/hhr033.

Duffie, Darrell and Rui Kan (1996) "A yield-factor model of interest rates," *Mathematical Finance*, 6 (4), 379–406, 10.1111/j.1467-9965.1996.tb00123.x.

Eggertsson, Gauti B. and Michael Woodford (2003) "The Zero Bound on Interest Rates and Optimal Monetary Policy," *Brookings Papers on Economic Activity*, 2003 (1), 139–211.

Fama, Eugene F. (1984) "The information in the term structure," *Journal of Financial Economics*, 13 (4), 509–528, 10.1016/0304-405X(84)90013-8.

Ghysels, Eric and Alastair Hall (1990) "A Test for Structural Stability of Euler Conditions Parameters Estimated Via the Generalized Method of Moments Estimator," *International Economic Review*, 31 (2), 355–364.

Greenwood, Robin and Dimitri Vayanos (2014) "Bond Supply and Excess Bond Returns," *The Review of Financial Studies*, 27 (3), 663–713, 10.1093/rfs/hht133.

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright (2007) "The U.S. Treasury yield curve: 1961 to the present," *Journal of Monetary Economics*, 54 (8), 2291–2304, 10.1016/j.jmoneco.2007.06.029.

Gust, Christopher, Edward Herbst, David López-Salido, and Matthew E. Smith (2017) "The Empirical Implications of the Interest-Rate Lower Bound," *American Economic Review*, 107 (7), 1971–2006, 10.1257/aer.20121437.

Hansen, Bruce E. (1990) "Lagrange Multiplier Tests for Parameter Instability in Non-Linear Models," Working Paper October 1990, University of Rochester.

——— (1992) "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business & Economic Statistics*, 10 (3), 321–335, 10.1080/07350015.1992.10509908.

Holston, Kathryn, Thomas Laubach, and John C. Williams (2017) "Measuring the natural rate of interest: International trends and determinants," *Journal of International Economics*, 108, S59–S75, 10.1016/j.jinteco.2017.01.004.

Hördahl, Peter, Oreste Tristani, and David Vestin (2006) "A joint econometric model of macroeconomic and term-structure dynamics," *Journal of Econometrics*, 131 (1), 405–444, 10.1016/j.jeconom.2005.01.012.

Hu, Grace Xing, Jun Pan, and Jiang Wang (2013) "Noise as Information for Illiquidity," *The Journal of Finance*, 68 (6), 2341–2382, 10.1111/jofi.12083.

Johannsen, Benjamin K. and Elmar Mertens (2016) "A Time Series Model of Interest Rates With the Effective Lower Bound," Working Paper 2016-033, Board of Governors of the Federal Reserve System, Washington, 10.17016/FEDS.2016.033.

Joslin, Scott, Marcel Priebsch, and Kenneth J. Singleton (2014) "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," *The Journal of Finance*, 69 (3), 1197–1233, 10.1111/jofi.12131.

Kim, Don H. and Athanasios Orphanides (2012) "Term Structure Estimation with Survey Data on Interest Rate Forecasts," *Journal of Financial and Quantitative Analysis*, 47 (1), 241–272, 10.1017/S0022109011000627.

Kim, Don H. and Kenneth J. Singleton (2012) "Term structure models and the zero bound: An empirical investigation of Japanese yields," *Journal of Econometrics*, 170 (1), 32–49, 10.1016/j.jeconom.2011.12.005.

King, Thomas B. (2019) "Expectation and duration at the effective lower bound," *Journal of Financial Economics*, 134 (3), 736–760, 10.1016/j.jfineco.2019.05.009.

Krippner, Leo (2013) "Measuring the stance of monetary policy in zero lower bound environments," *Economics Letters*, 118 (1), 135–138, 10.1016/j.econlet.2012.10.011.

Kuttner, Kenneth N. (2018) "Outside the Box: Unconventional Monetary Policy in the Great Recession and Beyond," *Journal of Economic Perspectives*, 32 (4), 121–46, 10.1257/jep.32.4.121.

Nakata, Taisuke and Hiroatsu Tanaka (2016) "Equilibrium Yield Curves and the Interest Rate Lower Bound," Working Paper 2016-085, Board of Governors of the Federal Reserve System, Washington, 10.17016/FEDS.2016.085.

Nyblom, Jukka (1989) "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84 (405), 223–230.

Priebsch, Marcel A. (2013) "Computing Arbitrage-Free Yields in Multi-Factor Gaussian Shadow-Rate Term Structure Models," Working Paper 2013-063, Board of Governors of the Federal Reserve System, Washington.

Reifschneider, David and John C. Williams (2000) "Three Lessons for Monetary Policy in a Low-Inflation Era," *Journal of Money, Credit and Banking*, 32 (4), 936–966.

Sakurai, Yuji (2016) "How Does the Bond Market Perceive Macroeconomic Risks under Zero Lower Bound?" Doctoral Dissertation, University of California, Los Angeles.

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*, 104 (10), 3154–85, 10.1257/aer.104.10.3154.

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*, 48 (2-3), 253–291, 10.1111/jmcb.12300.