JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 59, No. 7, Nov. 2024, pp. 3500–3530

doi:10.1017/S0022109023000984

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

Don H. Kim
Division of Monetary Affairs, Board of Governors of the Federal Reserve System
don.h.kim@frb.gov

Marcel A. Priebsch
Division of Monetary Affairs, Board of Governors of the Federal Reserve System
marcel.a.priebsch@frb.gov (corresponding author)

Abstract

We examine the structural stability of Gaussian shadow rate term structure models in a sample of Treasury yields that includes the "effective lower bound" (ELB) period from 2008 to 2015. After highlighting the challenges of testing for structural breaks in a latent-factor model, we proceed to document various pieces of empirical evidence for a structural break. As one of several practical implications, the expected policy rate paths during ELB years are notably shallower in our model that accommodates a structural break compared with a model that imposes structurally stability.

I. Introduction

From 2008 to 2015, the U.S. policy rate was set at its effective lower bound (ELB) for the first time in recent history.[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), Christensen and Rudebusch (2014), Bauer and Rudebusch (2016), and 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 is conceptual: It has a parallel with macroeconomic models incorporating an ELB in which the policy rate

This article has evolved from our earlier preliminary working paper entitled "Estimation of Multi-Factor Shadow Rate Term Structure Models." The analysis and conclusions set forth in this article 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 an anonymous referee, Hendrik Bessembinder (the editor), Michiel De Pooter, Marco Giacoletti (a referee), Edith Liu, Andrew Meldrum, Hiroatsu Tanaka, Jonathan Wright, and seminar participants at the Federal Reserve Board for helpful comments.

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

is analogously described in a censored form as $\max\{s_t, \underline{r}\}$, 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. Indeed, as further discussed later in the article, principal component analysis (PCA) of yields has substantially different results for the ELB period versus 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 a 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.

However, there are several reasons to believe that even shadow rate models might not be structurally stable: First, economic intuition suggests that the dynamics of key variables such as inflation and output gap may change once the economy enters the ELB regime, as conventional monetary policy can no longer provide policy accommodation. In turn, 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. Second, 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. Third, 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 risk that described the pre-ELB years. Lastly, it is possible that the financial crisis that 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 article, 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), Ahn, Dittmar, and Gallant (2002), and Duffee (2002)), raises new challenges that have not been encountered in the existing literature on structural break tests. In latent-factor models, the state variables do not have unique, well-defined meaning; 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.[3] Therefore, investigating the change in a specific parameter of the model may not shed much

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

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

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: 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 (such as a reordering of factors).

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 complementary 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 estimation based on the post-break subsample alone, which is relatively short, and entails greater estimation uncertainties. With these estimation difficulties as caveat, we do also perform an estimation with the post-break subsample, to further characterize the properties of a potential structural break. In this context, a helpful feature of the shadow rate models in this article that facilitates the examination of a structural break is that, while the model-implied observed yields follow nonlinear processes, the so-called "shadow yields" follow structurally stable Gaussian processes under the null hypothesis of no structural break.

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 independently and identically distributed (IID) under the null hypothesis of structural stability, display not only contemporaneous correlations but also serial correlations. Likelihood-score-based tests 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 those implied by a model that allows for a structural break.

This article complements several recent and contemporaneous studies that have analyzed the behavior of U.S. interest rates over samples including the ELB period, and found evidence suggestive of structural change. Andreasen and Meldrum (2019) note that, unlike an affine-Gaussian model estimated on pre-ELB data, a shadow rate model estimated on a sample of yields that includes the ELB period no longer matches standard empirical patterns of excess bond return predictability regression coefficients, even when yields are away from the lower bound. They attribute this finding to a change in the time-series dynamics of the pricing factors after 2008. Similarly, Andreasen, Jørgensen, and Meldrum (2019)

document changing patterns of excess bond return predictability regressions and long-horizon short rate expectations in the ELB period. They find that a model featuring both regime switching in the $\mathbb{P}$ parameters and a permanent structural break in the level of the pricing factors is able to replicate the documented empirical patterns, whereas a single-regime shadow rate model is not. While this finding strongly points to the presence of structural change, Andreasen et al. (2019) do not conduct an explicit structural break test. Of note, because their excess bond return predictability regressions are based on actual yields (as opposed to shadow yields), the regression coefficients would be expected to change near the ELB even within a single-regime shadow rate model. In our framework of analyzing shadow yields, we are able to quantitatively distinguish the changes in excess return predictability patterns due to ELB effects from those due to structural change, and to formally test the null hypothesis of structural stability. Hördahl and Tristani (2019) analyze U.S. term structure dynamics using a regime-switching model, motivated by their observation that the speed of policy rate normalization after liftoff in 2015 was much slower than in prior tightening episodes. However, in light of the flexibility of latent-factor shadow rate term structure models, it is not a priori obvious that the behavior of the yield curve in the post-ELB period, including the slower pace of policy rate normalization, cannot be captured by *some* set of latent factors in a structurally stable shadow rate model. Our work, in which the structurally stable shadow rate model is a special case of the structurally broken model, allows us to formally gauge the improvement in the statistical goodness of fit when the structural stability restriction is relaxed. Indeed, a key feature of our structurally broken model is that it produces less steep paths of the expected short rate than the nested structurally stable model, which lends support to Hördahl and Tristani's (2019) motivation. There are also a number of empirical studies outside the no-arbitrage framework suggesting significant changes in the dynamics of the yield curve and its relation to macro variables around the ELB period (e.g., Swanson and Williams (2014), Liu, Theodoridis, Mumtaz, and Zanetti (2019)). Such changes might be consistent with either a structural break in model parameters or a single-regime shadow rate model in which relationships change near the ELB. Our results below provide statistical evidence of the former.

II. Model and Data

A. Shadow Rate Term Structure Model and Implementation

We begin with a discussion of the shadow rate term structure model used in this article. The observed short rate is specified as

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

where $\underline{r}$ is the ELB, and the shadow rate s_t is an affine function of the N-dimensional vector of latent variables x_t:

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

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 $\mathbb{Q}$. The $\mathbb{P}$-measure and $\mathbb{Q}$-measure dynamics of the state vector x_t are specified as stationary multivariate Ornstein–Uhlenbeck processes:

$$dx_t = (k_0^{\mu} + K_1^{\mu} x_t)dt + \Sigma dW_t^{\mu}, \tag{3}$$

where $\mu \in \{\mathbb{P}, \mathbb{Q}\}$. Equation (3) implies that the market price of risk vector λ_t takes the affine form

$$\lambda_t = \lambda_0 + \Lambda_1 x_t. \tag{4}$$

The $\mathbb{P}$-measure parameters $k_0^{\mathbb{P}}$, $K_1^{\mathbb{P}}$ and the $\mathbb{Q}$-measure parameters $k_0^{\mathbb{Q}}$, $K_1^{\mathbb{Q}}$ are linked via market price of risk parameters λ_0, Λ_1 as

$$k_0^{\mathbb{Q}} = k_0^{\mathbb{P}} - \Sigma\lambda_0, \tag{5}$$

$$K_1^{\mathbb{Q}} = K_1^{\mathbb{P}} - \Sigma\Lambda_1. \tag{6}$$

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

$$P_{t,\tau} = E_t^{\mathbb{Q}}\left[\exp\left(-\int_t^{t+\tau} r_s ds\right)\right], \tag{7}$$

with associated zero-coupon bond yield

$$y_{t,\tau} = -\frac{1}{\tau}\log P_{t,\tau}. \tag{8}$$

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 (2023).

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 equation (7). The shadow yields $y_{t,\tau}^s$ correspond to the arbitrage-free yields in the underlying Gaussian model not constrained by the ELB, and take the affine form

$$y_{t,\tau}^s = a(\tau) + b(\tau)' x_t, \tag{9}$$

where a and b are given by the usual recursive formulas (Duffie and Kan (1996)).[4] As the lower bound becomes less binding, yields in the shadow rate model will approach their Gaussian counterpart.[5] In this sense, away from the lower bound, the shadow rate model is approximated by the underlying Gaussian model.

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

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

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 0. 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 $\mathbb{Q}$).

To estimate the model, we proceed analogously to Kalman filter-based maximum-likelihood estimation in studies such as Kim and Wright (2005), Christensen, Diebold, and Rudebusch (2011), and Duffee (2011), except that our observation equation is nonlinear in the state vector; therefore, we use the unscented Kalman filter (see, e.g., Wan and van der Merwe (2001)). While the model can in principle be estimated with yield data alone, the well-known small sample problem associated with persistent time series such as bond yields presents significant challenges. To ameliorate this problem, we follow the approach of Kim and Orphanides (2012) and augment our estimation sample with survey forecasts of the 3-month Treasury bill rate.[6] Therefore, our observation equation takes the form

$$\begin{pmatrix} \tilde{y}_t \\ \tilde{z}_t \end{pmatrix} = h(x_t) + \begin{pmatrix} e_{y,t} \\ e_{z,t} \end{pmatrix}, \tag{10}$$

where $\tilde{y}_t$ and $\tilde{z}_t$ are vectors of observed Treasury yields and survey forecasts, respectively, $h(x_t)$ is a vector of model-implied counterparts of $[\tilde{y}_t, \tilde{z}_t]'$ (generally nonlinear functions of x_t), and $e_{y,t}$ and $e_{z,t}$ are measurement errors for Treasury yields and survey forecasts, respectively. For the survey forecasts, we use a couple of near-term horizons (6 months and 12 months) as well as a longer horizon (5–10 years).

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 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 basis points on $\delta_{z,5-10y}$ which is binding in all our estimated models.[7]

[6] See also Li, Meldrum, and Rodriguez (2017) for the use of survey forecast data for addressing the small-sample problem. Other approaches include adjusting the estimation procedure (Bauer, Rudebusch, and Wu (2012)), and imposing parameter restrictions (Christensen et al. (2011), Joslin et al. (2014), and Bauer (2018)); see also Bauer and Rudebusch (2020) for a related discussion.

[7] 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.

B. Data, Sample Period, and Break Point

As sample period, we take the beginning of 1990 to the middle of 2019. Where applicable, we assume that the FOMC's announcement on Dec. 16, 2008, that it would establish a target range for the federal funds rate of 0 to a quarter percent represents an unanticipated structural break point. We will somewhat loosely label the first period (observations from 1990 to mid-Dec. 2008) as "pre-ELB" and the second period (observations from mid-Dec. 2008 to June 2019) as "(post-)ELB." Below, we discuss the sensitivity of our findings to this choice of break date. Our empirical designs in this article focus on structural change with a single break point at a known time; the cases of a break at an unknown time or multiple breaks would be even more challenging empirically, in part because of the limited amount of data.

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

For the survey forecast data, we use the 3-month Treasury bill rate from Blue Chip Financial Forecasts, linearly interpolated to constant horizons of 6 months, 12 months, and 5–10 years.[8] Short-range forecasts are available monthly and long-range forecasts are available semiannually. Each survey is lined up with the sample date closest to the actual date on which the survey was likely conducted. For weeks with no matched survey data, we treat surveys as missing observations.[9]

III. Empirical Strategies

A. Challenges in 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 equations (1)–(3), note that we can obtain an observationally equivalent model by transforming the N-dimensional state vector x_t to

$$x_t^{\dagger} = l + L x_t \tag{11}$$

[8] 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. We treat the survey consensus as a noisy observation of the model-implied *modal* expectation.

[9] The addition of the survey data occurs at the estimation stage and leaves the theoretical shadow rate model setup in Section II.A unaffected.

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

$$K_1^{\mu\dagger} = LK_1^\mu L^{-1}, \quad k_0^{\mu\dagger} = Lk_0^\mu - LK_1^\mu L^{-1} l, \tag{12}$$
$$\Sigma^\dagger = (L\Sigma\Sigma' L')^{1/2}, \quad \rho_0^\dagger = \rho_0 - \rho_1' L^{-1} l, \quad \rho_1^\dagger = (L^{-1})' \rho_1.$$

Because the most general 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 is imposed to identify the parameters and thus ensure that the Hessian matrix $\partial^2 \log \mathcal{L} / \partial\theta\partial\theta'$ is not singular. In the present article, the normalization restrictions we impose for estimation purposes are

$$k_0^{\mathbb{P}} = 0, \quad K_1^{\mathbb{P}} = \text{lower triangular matrix}, \quad \Sigma = cI, \tag{13}$$

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 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[10]

$$\lambda_T = T(\hat{\theta}_1 - \hat{\theta}_2)' (\pi^{-1} \hat{V}_1 + (1-\pi)^{-1} \hat{V}_2)^{-1} (\hat{\theta}_1 - \hat{\theta}_2) \tag{14}$$

provides a valid statistic for the test of structural stability. However, even after imposing normalization restrictions that guarantee a nonsingular Hessian of the likelihood function around an estimate (such as equation (13)), discrete invariant transformations (such as permutation and reflection) still remain, which creates multiple equivalent parameter vectors.

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 equation (14) 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 such as 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 small-sample estimates when the true parameters are close. One could also entertain the idea of

$$\lambda_T = \min_i T(\hat{\theta}_1 - \hat{\theta}_2^{(i)})' (\pi^{-1} \hat{V}_1 + (1-\pi)^{-1} \hat{V}_2^{(i)})^{-1} (\hat{\theta}_1 - \hat{\theta}_2^{(i)}), \tag{15}$$

where the $\hat{\theta}_2^{(i)}$s denote 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

[10] See, e.g., Andrews and Fair (1988).

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.[11]

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

B. Tests for Structural Change

Aside from the issue of latent factors, 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 potentially multiple breaks at unknown times. 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.
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.

In addition to the identification issue discussed previously, Wald-type tests can be expected to have low power in our application, as they involve a separate estimation of the post-ELB subsample. A sample of about 10 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 estimation. 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, the asymmetric distribution of the short rate process for horizons beyond the predicted liftoff date (see footnote 8) means that the measurement of survey forecasts would likely be less reliable than usual. Lastly, we do not wish to rely on survey data to such a degree as to confound the questions of whether the structural break occurred in yields or in survey forecasts.[12]

For evidence regarding structural instability, we therefore focus on predictive tests and LM tests. Predictive tests are particularly attractive in our context, as we

[11] 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.

[12] We examine the robustness of our findings to the omission of surveys below.

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 V we also estimate a model based on the post-ELB subsample, as this estimation may cast helpful light on *how* the model structure might have changed.

C. Diagnostics 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 (10)) 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. 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 the discretized transition equation under $\mathbb{P}$ derived from equation (3) implied by the θ_{pre} estimate, which would be a closer yet still-imperfect analog of Chow's regression residuals. A normalized innovation vector $\eta_t \equiv \Omega^{-1/2}\varepsilon_t$ (where Ω denotes the covariance matrix of ε_t) has the theoretical property:

$$\eta_t \sim N(0_N, I_{N\times N}). \tag{16}$$

Furthermore, these innovation vectors are serially uncorrelated, that is,

$$E(\eta_t \eta_{t+j}') = 0_{N\times N}, \quad j \neq 0. \tag{17}$$

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.

Statistics can be constructed that test the conditions in equations (16) and (17). A complication in our setting, discussed previously in Section III.A, is that state variables and their innovation vectors in a generically normalized model (such as the normalization conditions in equation (13)) 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., $\text{corr}(\eta_{1t}, \eta_{2t}) \neq 0$) 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 Section IA.I of the Supplementary Material, 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:[13]

$$T\frac{1}{2}\left\|T^{-1}\sum_t \eta_t \eta'_t - I\right\|_F^2 \sim \chi^2_{(N^2+N)/2}, \tag{18}$$

$$T\left\|T^{-1}\sum_t \eta_t \eta'_{t-u}\right\|_F^2 \sim \chi^2_{N^2}, \quad u \neq 0. \tag{19}$$

Moreover, since the innovation vectors in different subsamples are independent, the *ratio* of χ^2 statistics for different periods follows an F-distribution under the null hypothesis and may give an indication of the *relative* degree to which implied innovation vectors deviate from their theoretical distribution.

While these statistics provide useful summary diagnostics regarding structural stability, in the event they point to a structural change in the post-ELB period, more granular statistics that shed light on the nature of structural change would be useful. Therefore, we also examine more disaggregated 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 to instantaneous change in the level, slope, and curvature of the shadow yield curve.[14] 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 1 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 in equations (16) and (17).

D. Tests Based on Likelihood Scores

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

$$E(\partial \log \mathcal{L}/\partial\theta) = 0 \tag{20}$$

[13]The notation $\|A\|_F$ denotes the Frobenius norm of the matrix A. 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 $\frac{1}{2}(T^{-1}\sum_t \eta_{1t}^2 - 1)^2 = \frac{1}{2}(T^{-1}\sum_t \eta_{1t}^{\dagger 2} - 1)^2$.

[14]This procedure is described in greater detail in Sections V.A and V.B.

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 in equation (20) will hold in each subsample.

The predictive approach (testing the condition in equation (20) 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,

$$\frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T_1+T_2}\frac{\partial \ell_t(\hat{\theta}_1)}{\partial \theta'}\hat{V}_2^{-1}\frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T_1+T_2}\frac{\partial \ell_t(\hat{\theta}_1)}{\partial \theta} \sim \chi^2_{\dim(\theta)}, \tag{21}$$

where $\hat{V}_2$ is a consistent estimator of $\text{plim}_{T\to\infty}\text{Var}(T_2^{-1/2}\sum_{t=T_1+1}^{T_1+T_2}\partial \ell_t(\hat{\theta}_1)/\partial\theta)$.[15]

The LM approach (testing the condition in equation (20) with parameter estimates $\hat{\theta}$ based on the full sample) can be viewed as an analog of likelihood score-based LM tests of Nyblom (1989) and Hansen (1990), (1992). The relevant testing statistic is given by[16]

$$\text{LM} = \frac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1}\frac{\partial \ell_t(\hat{\theta})}{\partial \theta'}\hat{V}^{-1}\frac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1}\frac{\partial \ell_t(\hat{\theta})}{\partial \theta}, \tag{22}$$

where $\hat{V}$ is a consistent estimate of $\text{plim}_{T\to\infty}\text{Var}(T_1^{-1/2}\sum_{t=1}^{T_1}\partial \ell_t(\hat{\theta})/\partial\theta)$.

These likelihood-score-based approaches do not face the problems discussed in connection with a Wald test of the hypothesis $\theta_1 = \theta_2$ using the statistic in equation (14), as the score-based tests only look at 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 derivative-based asymptotic standard errors of a parameter estimate.

IV. Empirical Evidence Regarding Structural Instability

A. Yield Fitting Errors and Innovation Vectors

To build intuition and get a preliminary sense of the existence and possible nature of a regime change, we first evaluate how well a model with pre-ELB parameters is able to fit the cross-sectional and time-series properties of yields in the post-ELB period once the lower bound constraint is imposed. For this purpose, we analyze yield fitting errors and implied innovation vectors from 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 (i.e., with state variables filtered from post-ELB data using pre-ELB parameters).[17]

[15] Ghysels and Hall (1990) discuss several easily computable candidates for the variance estimator.

[16] See Hansen (1990) and Andrews (1993).

[17] 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 0. 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

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.[18] However, in the later years of the post-ELB period, beginning in early 2014, there is a more persistent 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.

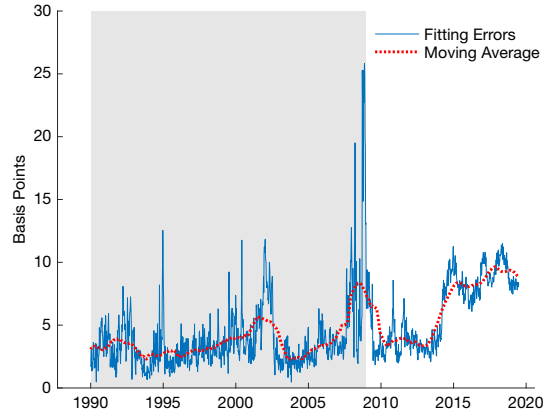
2. Innovation Vectors

Using the same model as in Section IV.A.1, we can compute empirical implied innovation vectors $\hat{\eta}_t$ as defined in Section III.C. Recall from the discussion there that these vectors theoretically have a contemporaneous covariance matrix equal to the identity matrix, and no autocorrelation. Since the empirical innovation covariance matrices depend on the chosen model rotation, we rotate the model as proposed by Duffee (2011) and discussed in detail in Section V.B, such that the factors can be interpreted as level, slope, and curvature of shadow yields. Table 1 reports the contemporaneous empirical innovation covariance matrices after this rotation, separately for the pre-ELB and post-ELB subsample periods. While the covariance

FIGURE 1

Yield Fitting Errors

Figure 1 displays the 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) and the 1-year moving average. The pre-ELB period is shaded.



when estimating a single-regime shadow rate model on the full sample. We find that other sensible choices of r lead to broadly similar results.

[18]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.

TABLE 1

Innovation Vector Covariances

Table 1 displays the empirical covariance matrix of implied innovation vectors (model rotated such that state variables can be interpreted as principal components).

Panel A. Pre-ELB		
0.96	0.00	0.00
0.00	0.93	−0.08
0.00	−0.08	0.68
Panel B. Post-ELB		
0.69	0.39	0.22
0.39	0.66	0.23
0.22	0.23	0.64

matrix for the pre-ELB period shows some deviation from the identity matrix (most notably in the third diagonal entry, that is, low variance for the implied innovations in curvature), the covariance matrix for the post-ELB period displays a number of more notable departures: All three elements of the innovation vector have contemporaneous variances well below 1, and the off-diagonal elements of the covariance matrix are non-negligible, with particularly pronounced positive association between level and slope innovations (whose contemporaneous correlation evaluates to about 0.6).[19]

Figure 2 shows autocorrelograms of the rotated innovation vectors up to lags of 52 weeks. In the pre-ELB period, autocorrelations are generally insignificant or at most marginally significant, with no clear pattern across lags; the most notable excursions are visible in curvature innovations. Conversely, in the post-ELB subsample, the implied innovations to the slope factor (and to a lesser extent the curvature factor) display significant and prolonged positive temporal dependency. In other words, the model persistently mispredicts the slope and curvature factors.

The rotation-invariant χ^2 statistics introduced in Section III.C and more rigorously justified in Section IA.I of the Supplementary Material allow us to quantify the overall magnitude and significance of the deviations of the implied innovation vectors from their theoretical properties more formally. As shown in Table 2, these statistics indicate significant misspecification in both periods at the lags considered. The overall *ratio* of post-ELB and pre-ELB statistics is significantly larger than 1 at the 1% level (suggesting a greater relative degree of misspecification in the post-ELB period), but the ratios computed separately for lags 0, 1, and 2–12 are not individually significant. In particular, despite the patterns shown in Table 1 and the noticeably larger χ^2 statistic at lag 0 in the post-ELB period, the evidence based on

[19]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. More generally, in the affine-Gaussian model, the short rate is an affine function of the factors ($r_t = \rho_0 + \rho_1' x_t$); therefore, the short rate being stuck at the ELB would imply that a linear combination of innovation vectors has to sum up to 0, i.e., the implied innovation vectors have to be contemporaneously correlated during the ELB period. Our results here suggest that simply imposing a shadow rate model structure does not fix this manifestation of model misspecification.

FIGURE 2

Innovation Autocorrelations

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

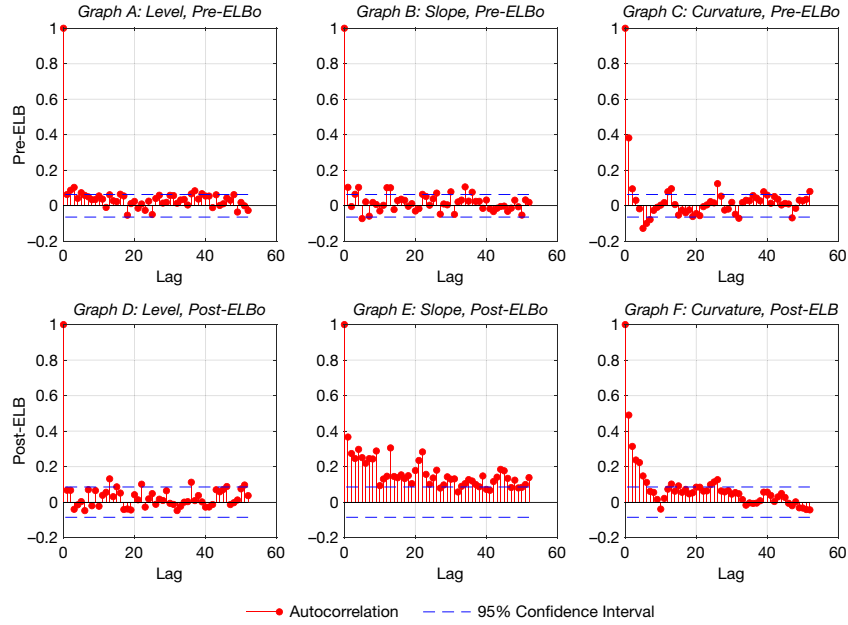


TABLE 2

Rotation-Invariant Statistics

Table 2 presents rotation-invariant χ^2 statistics of implied vectors. Lag 0 corresponds to equation (18), and lags ≥ 1 correspond to equation (19).

Lag	Pre-ELB	Post-ELB	d.f.	1% Cutoff	Post/Pre	1% Cutoff
0	55.98	206.73	6	16.81	3.69	8.47
1	108.19	115.59	9	21.67	1.07	5.35
2–12	281.02	380.10	99	134.64	1.35	1.60
Sum	445.19	702.42	114	152.04	1.58	1.55

the empirical instantaneous covariance matrix of implied innovation vectors alone is not significantly indicative of misspecification.

3. Innovation Vectors in Post-ELB Subsamples

In light of the somewhat mixed evidence emerging from the post-ELB implied innovation vectors, and the pattern of yield fitting errors in Figure 1 that is not uniform over our post-ELB sample, it might be natural to ask how sensitive our results are to our assumption about the *timing* of the candidate structural break in late 2008. Indeed, regarding the U.S. experience, some have suggested that the first calendar-based forward guidance announcement by the FOMC in Aug. 2011 may

TABLE 3

Innovation Vector Covariances by Subperiod

Table 3 shows the empirical covariances matrix of implied innovation vectors (model rotated such that state variables can be interpreted as principal components), 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.

Panel A. Period I		
1.03	0.70	0.32
0.70	0.85	0.41
0.32	0.41	0.79
Panel B. Period II		
0.84	0.40	0.34
0.40	0.67	0.54
0.34	0.54	0.74
Panel C. Period III		
0.65	0.14	0.26
0.14	0.55	0.04
0.26	0.04	0.71
Panel D. Period IV		
0.35	0.24	0.13
0.24	0.33	0.19
0.13	0.19	0.34

TABLE 4

Rotation-Invariant Statistics by Subperiod

Table 4 provides the 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.

Lag	I	II	III	IV	d.f.	1% Cutoff
0	96.85	52.74	23.80	125.50	6	16.81
1	41.07	21.22	93.25	10.64	9	21.67
2–12	192.58	53.78	536.24	54.04	99	134.64
Sum	330.50	127.74	653.29	190.18	114	152.04

have had a more significant impact on yield curve dynamics than the arrival at the ELB in late 2008.[20] Our own Figure 1 shows a deterioration in yield curve fit from early 2014 onward. Therefore, in Tables 3 and 4, we also examine the empirical innovation vector covariances as well as the statistics in equations (18) and (19) based on segments of the post-ELB sample, in particular Dec. 2008 to mid-Aug. 2011 (before date-based forward guidance, segment I), mid-Aug. 2011 to May 2013 (before the "taper tantrum," segment II), June 2013 to Dec. 2015 (before liftoff, segment III), and Jan. 2016 to June 2019 (segment IV).

This more granular analysis indicates that the evidence for a structural break, at least according to this metric, varies in strength over the different subperiods and reveals different facets of the misspecification of the structurally stable model. For example, the Dec. 2008 to Aug. 2011 and Jan. 2016 to June 2019 subperiods show the most notable signs of misspecification in the contemporaneous (lag 0)

[20]Swanson and Williams (2014), e.g., note that 1-year and 1-year Treasury yields appeared to be "unconstrained" from 2008 to 2010, and only became more constrained from late 2011 onward.

properties of innovation vectors. We can glean from Table 3 that shocks in period I (the immediate aftermath of the financial crisis) were highly positively correlated, while in period IV, after liftoff in 2015, shocks had unusually small variances as the Fed raised rates at a much more measured pace than in previous tightening cycles. Meanwhile, in the June 2013 to Dec. 2015 period III, the serial correlation of innovation vectors shows prominent signs of misspecification. During this period, the Fed tapered and eventually ended its purchases of Treasury and mortgage-backed securities (thus reducing policy accommodation at the long end of the yield curve) while keeping the federal funds rate at the ELB. The structurally stable model appears to have difficulty capturing this process, and indeed the autocorrelations in the post-ELB period shown in Figure 2 are particularly prominent for the shocks to the slope factor.[21]

B. Likelihood-Score-Based Tests

While our findings in Section IV.A point to some signs of greater model misspecification in the post-ELB period than the pre-ELB period, the results are arguably not clear-cut and uniform across the considered metrics. Therefore, in this section, we test for a structural break explicitly, using variants of the moment restriction in equation (20) based on both the pre-ELB parameter estimate (the predictive framework) and the full-sample, single-regime estimate (the LM framework). For this purpose, in addition to the model introduced in Section IV.A, 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 0, as a first-order optimality condition of estimation. Intuitively, if there is no structural break, the

TABLE 5

Ghysels–Hall Statistics

Table 5 displays the Ghysels–Hall statistics for null hypothesis of structural stability. In Graphs B and C, r is set to 7.3 basis points. 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.

Level of Lower Bound	
Panel A. Level of Lower Bound (r)	χ^2
12.5	299.36
7.3	326.12
5.5	325.20
Panel B. Subsample Period	χ^2
I	117.18
II	86.42
III	120.53
IV	179.58
Panel C. Survey Inclusion	χ^2
Included	326.12
Excluded	331.28

[21] We are grateful to an anonymous referee for suggesting possible linkages between the misspecification we document and concurrent monetary policy decisions.

post-ELB sample average of empirical scores (based on pre-ELB parameters) should be *close to* 0, with known asymptotic distribution (Ghysels and Hall (1990)). This exercise requires an assumption on the level of the lower bound $\underline{r}$, since this is not identified from pre-ELB data (see footnote 17). Panel A of Table 5 displays χ^2 statistics for a joint test of all moment restrictions, for different values of the lower bound $\underline{r}$.[22] The statistics have 25 degrees of freedom (corresponding to the number of estimated model parameters), with a 5% cutoff of 37.65 and a 1% cutoff of 44.31. The null hypothesis of structural stability is thus firmly rejected, with a p-value of virtually 0, and with only modest sensitivity to the exact chosen value of $\underline{r}$. Therefore, for the remaining tests, we use $\underline{r} = 7.3$ basis points. Panel B shows statistics for the various segments of the post-ELB subsample separately. Using the same four subdivisions as in Section IV.A.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 in equation (10), 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 Panel C 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 0 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 *individually* be close to 0, again with known asymptotic distribution.[23] Since this statistic is based on a single-regime model estimated on the entire sample, it is computationally cheap to evaluate the statistic for a range of candidate structural break times around our assumed break point. As shown in Figure 3, there is strong evidence of a structural break at our assumed break time, and the statistic is relatively flat in the years before and after, suggesting that our finding is not locally sensitive to the exact choice of break point date.[24] We show in Section IA.III of the Supplementary Material that this finding is again robust to the omission of survey data.

[22] 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.

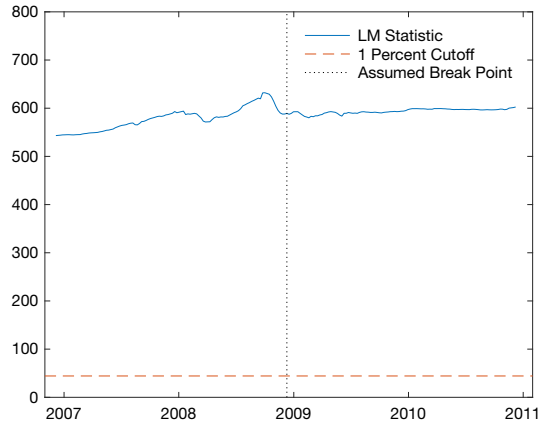
[23] The pre- and post-ELB subsample averages are linked, as $\sum_{t=1}^{T_{pre}} \partial \ell_t(\hat{\theta})/\partial\theta = -\sum_{t=T_{pre}+1}^{T} \partial \ell_t(\hat{\theta})/\partial\theta$.

[24] In this article, we do not attempt to estimate the time of structural break empirically. One straightforward such estimator is the date on which the LM test statistic is largest (the "sup LM" statistic described by Andrews (1993)). Based on the window considered in Figure 3, this estimator would place the sample break a few months before our assumed break in Dec. 2008. However, this estimator depends on the chosen window, with reliance on asymptotic properties becoming increasingly dubious as we expand the range of candidate break points toward the beginning and end of the sample.

FIGURE 3

Lagrange Multiplier Statistics

Figure 3 charts the time series of Lagrange Multiplier Statistics from 2 years before to 2 years after our assumed break point. 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.



V. 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 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.

Our strategy allows us to analyze the characteristics of the structural break in the greatest possible generality within the shadow rate model framework. However, this flexibility comes at the cost of doubling the number of free parameters, and therefore, we do not intend to propose the structurally broken model in this section as a general-purpose term structure model. In Section IA.IV of the Supplementary Material, we consider a shadow rate model with a structural break restricted to a subset of parameters, and we are also cognizant of the possibility that a different class of nonlinear models (e.g., regime-switching models) might ultimately turn out to be more promising as a description of U.S. yield dynamics in the post-ELB era.

A. Comparison of PCs

We begin with a discussion of PCA decomposition. To motivate this analysis, Panels A and B of Table 6 show the PCA loadings for the first three principal components of weekly yield changes, for the pre-ELB sample (Panels A and C) and the post-ELB sample (Panels B and D). The pre-ELB sample loadings show typical behaviors, consistent with the frequent "level," "slope," and "curvature"

TABLE 6

Principal Component Loadings

Panels A and B of Table 6 show the PC1, PC2, PC3 loadings for the PC decomposition of weekly changes in yields in the pre-ELB period (Panels A and C) and in the post-ELB period (Panels B and D). Panels C and D show the loadings for the PC decomposition of instantaneous changes in shadow yields implied by the pre-ELB subsample estimate $\hat{\theta}_{pre}$ (Panel C) and by the post-ELB subsample estimate $\hat{\theta}_{post}$ (Panel D). These PCs are based on maturities (τ) of 0.25, 0.5, 1, 2, 4, 7, 10 years. Standard errors are shown in parentheses.

	Panel A. Observed Yields, Pre-ELB			*Panel B. Observed Yields, 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
	Panel C: Model-Implied Shadow Yields, Pre-ELB			*Panel D. Model-Implied Shadow Yields, 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)

designations of PC1, PC2, and PC3: For example, the PC1 loadings are roughly flat (at least in the 1-year to 10-year range), while PC2 loadings monotonically increase with maturity. In the case of the post-ELB sample, 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, that is, the PCA decomposition of the changes in *shadow yields* (which are unaffected by the ELB) should be the same in the pre-ELB and post-ELB periods.

Panels C and D of Table 6 show 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 equation (9) 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 $(\tau_1, \tau_2, \ldots, \tau_m)$ as y_t^s, we have, schematically,

$$y_t^s = a + Bx_t, \tag{23}$$

where a is an m-dimensional vector, and B is an $m \times N$ matrix. This, together with equation (3), gives

$$(dy_t^s dy_t^{s'})/dt = B\Sigma\Sigma' B'. \tag{24}$$

Therefore, the singular value decomposition of this matrix

$$B\Sigma\Sigma' B' = P\Psi P', \tag{25}$$

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).[25] Panels C and D of Table 6 also 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 in Panel C of Table 6, are quite similar to those in Panel A, 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 loadings in Panel B; 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 maturity). In other words, PC1 in the post-ELB periods does not look like a standard "level shock." The qualitative difference between shadow yield PC loadings in the pre-ELB and post-ELB periods documented here adds to our evidence that yield curve dynamics have changed materially. Specifically, this evidence points to a change in the risk-neutral ($\mathbb{Q}$) parameters, because, as noted by Duffee (2011), this construction of PCs involves only risk-neutral parameters.

B. PC-Rotated Parameters

As discussed in Section III.A, 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 previously provides a natural means to surmount this problem, as shocks to the transformed 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 equation (11), with L given by[26]

$$L = P' B, \tag{26}$$

[25]Note, therefore, that Graphs A and B of Table 6 shows model-implied *population* loadings (which allows us to compute standard errors), whereas the top panel shows *sample* loadings derived from observed yields.

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

TABLE 7

Parameter Estimates (PC Rotation)

Table 7 displays the 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 0 by normalization.

	$\widehat{\theta}^{\dagger}_{pre}$	$\widehat{\theta}^{\dagger}_{post}$		$\widehat{\theta}^{\dagger}_{pre}$	$\widehat{\theta}^{\dagger}_{post}$
$[K_1^{\mathbb{P}\dagger}]_{11}$	−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)			

with the P and B matrices in equation (25). The transformed parameters ($\theta^{\dagger}$ in equation (12)) 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 $x^{\dagger}_{2t}$ 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 given by equation (13); the standard errors are computed with the delta method.[27] Note that the table shows more parameters than are necessary to estimate the model, as some of these are linked to each other (e.g., knowing two of $K_1^{\mathbb{P}}$, $K_1^{\mathbb{Q}}$, and Λ_1 determines the third (recall equation (6))).

It can be seen that pre-ELB period parameters are often more precisely estimated than the post-ELB period parameters; in other words, $\theta^{\dagger}_{pre}$ 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 IV.A.1) and PC loadings (Section V.A) that the $\mathbb{Q}$-measure dynamics have changed.

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 the natural rate "r-star" has declined since financial crisis, albeit with large estimation uncertainty.

C. EH Regressions for Shadow Yields

Additional insights into how the model has changed can be gleaned from model-implied expectations hypothesis (EH) regression coefficients.[28]

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

$$y^s_{t+1,n-1} - y^s_{t,n} = \alpha + \beta \frac{1}{n-1}(y^s_{t,n} - y^s_{t,1}) + e_{t+1,n}, \tag{27}$$

and the Fama (1984) regression:

$$y^s_{t+n,1} - y^s_{t,1} = \alpha + \beta(f^s_{t,n} - y^s_{t,1}) + e_{t+n,n}, \tag{28}$$

[27] The original parameter estimates are given in Section IA.II of the Supplementary Material.

[28] 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).

TABLE 8

Expectations Hypothesis Regression Coefficients

Table 8 gives the model-implied expectations hypothesis regression coefficients. Panel A shows the Campbell–Shiller regression coefficients implied by the pre-ELB and post-ELB subsample parameter estimates, $\hat{\theta}_{pre}$ and $\hat{\theta}_{post}$. Panel B shows the implied Fama regression coefficients. Standard errors are shown in parentheses.

Panel A. 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)

Panel B. 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)

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., $f^s_{t,n} = \log(P^s_{t,n}/P^s_{t,n+1})$).[29] 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 βs from the pre-ELB period should equal the βs from the post-ELB period in these regressions.

Table 8 shows model-implied population EH regression coefficients along with their standard errors (calculated with the delta method) for both the pre- and post-ELB parameters.[30] 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

[29]It can be shown that for $n = 2$, the Campbell–Shiller and Fama regressions contain the same information. However, for $n > 2$, the two regressions probe somewhat different aspects of the departures from the expectations hypothesis.

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

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_n^{CS} and 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 instead of shadow yields.

Somewhat unsurprisingly, 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 Section V.B that the parameters that describe the market price of risk $(\lambda_0^\dagger, \Lambda_1^\dagger)$ have changed meaningfully between the two periods.

D. Expectations and Term Premiums

Here, we compare some of the predictive outputs from our various models to illustrate their 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. Still, as a central ingredient in the model, the estimates of shadow rates are worth taking a look at. 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 a structural break is generally less negative than

FIGURE 4

Shadow Rate Estimates

Figure 4 charts the time series of Estimated Shadow Rates 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.

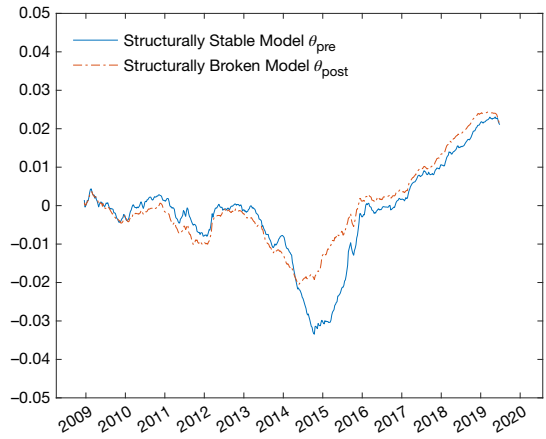
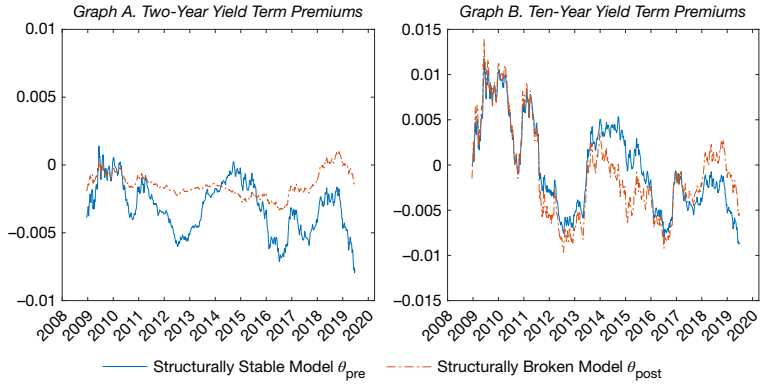


FIGURE 5

Term Premium Estimates

Figure 5 charts the 2-year (Graph A) and 10-year (Graph B) yield Term Premiums implied by our different models, shown for the post-ELB subsample.



the shadow rate assuming structural stability. In particular, the structurally stable model produces a notably more negative shadow rate during the 2014–2016 period.

Figure 5 shows the time series of 2-year and 10-year yield term premiums implied by both models. An especially notable feature in Graph A is that the 2-year term premium implied by the structurally broken model is generally closer to 0 than that implied by the structurally stable model; in other words, the structurally broken model is closer to the expectations hypothesis, consistent with the finding in Section V.C 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 $\hat{\theta}_{pre}$. For most of the post-ELB period, the 2-year term premium implied by the structurally stable model is substantially more negative, implying that the near-term expected path of the short rate implied by the structurally stable model is *steeper* than the structurally broken model counterpart.[31] On the other hand, as shown in Figure 5, Graph B, 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.

Due to space constraints, additional results comparing the outputs from the structurally stable and broken models are provided in Section IA.V of the Supplementary Material.

[31]Because the sum of term premium and short rate expectation equals forward rate, and because fitting errors for forward rates and yields are relatively small in latent-factor models, more negative term premiums translate to higher (i.e., more positive) short rate expectations. Interestingly, the more negative level of shadow rate does not necessarily translate to a lower level of the near-term expected short rate path. For example, during 2014–2016, the shadow rate implied by the structurally stable model was lower than that of the structurally broken model, but (as can be seen in Figure 5) the 2-year term premium implied by the structurally stable model was less negative than that of the structurally broken model. The reason is that the short rate expectations depend not only on the current (spot) shadow rate but also on other factors that affect expectations.

VI. Economic Interpretations of Structural Break Evidence

The results presented in Sections IV and V indicate that the evidence for a structural break is *broad* and covers multiple facets of yield curve dynamics, including the $\mathbb{P}$-measure dynamics of factors, risk premiums, and the $\mathbb{Q}$-measure dynamics. We now discuss these aspects in turn.

1. *Changes in $\mathbb{P}$-measure dynamics*: The comparison of PC-rotated parameters in Table 7 indicates significant changes in the $\mathbb{P}$-measure dynamics of the state vector. In addition, the innovation vector statistics presented in Section IV.A.2 also point to a change in $\mathbb{P}$-measure dynamics. This change in $\mathbb{P}$-measure dynamics is consistent with the economic intuition that the evolution and interplay of key macro variables will change once the economy arrives at the ELB, since conventional monetary policy can no longer respond to macroeconomic developments and provide monetary stimulus in the ELB regime.[32] This structural instability 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.[33] However, even pure latent-factor models, in which all of the state variables are latent, may not be free from these concerns, as latent variables are typically thought to embody macroeconomic risks. One potentially alleviating consideration is that unconventional monetary policy 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. 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.

   It is possible also that the financial crisis and the Great Recession resulted in significant structural changes in the economy, which would be (partly) reflected through changes in $\mathbb{P}$-measure parameters. Persistent "headwinds" have been often mentioned in policy discussions in the post-ELB period.

2. *Changes in risk premiums*: Table 7 also indicates a significant change in risk premium behavior in the ELB regime; the market price of risk parameters (λ_0, Λ_1) in the pre- and post-ELB subsamples differ notably, some even have different signs. Results in Section V, including the EH regression coefficients and term premiums implied by pre- and post-ELB subsample estimates, further add to the evidence of a change in the behavior of risk premiums.

   In terms of economic forces behind the change, the state of being in the ELB regime could be consequential enough to change risk premium behaviors

[32] This idea is reflected in simple stylized models like Reifschneider and Williams (2000). Studies of ELB dynamics within a new-Keynesian framework, including Eggertsson and Woodford (2003), Nakata and Tanaka (2016), and Gust, Herbst, López-Salido, and Smith (2017), also indicate the macro variables behave differently in ELB versus non-ELB periods.

[33] 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).

(e.g., by affecting economic agents' risk attitudes and hedging behaviors). But the onset of ELB in the U.S. also roughly coincided with the start of LSAPs (an unconventional policy tool which likely contributed substantially to the change in the behavior of risk premiums) and the transition to an "ample reserves" regime. A stylized model of King (2019) (a simple shadow rate model with supply factor) indeed implies that the risk premium behavior changes as the economy goes into the ELB regime.[34] More broadly, we can expect LSAPs to significantly alter the behavior of risk premiums in bond markets, as the asset purchases are also believed to directly affect longer-term maturities of the yield curve through the suppression of term premiums (as opposed to conventional policy, which affects longer-term maturities through its effect on expected short-term rates).[35] In addition, forward guidance (the other unconventional policy tool that the Fed has employed over much of the ELB period) may also have affected risk premium behavior especially at near- and intermediate-term horizons. Indeed, as Graph A of Figure 5 shows, short-dated term premiums implied by the structurally broken model are, on average, smaller in magnitude and less volatile in the ELB period.

3. *Changes in $\mathbb{Q}$-measure dynamics*: The estimated parameters in Table 7 also indicate a change in $\mathbb{Q}$-measure dynamics. This is consistent with the patterns in fitting errors we saw in Section IV.A (the pre-ELB parameters imply notably larger yield fitting errors in the post-ELB period, especially after 2014). The change in $\mathbb{Q}$-measure parameters is also manifested in the changes in shadow yield PC patterns, as we saw in Section V.A. Recall from Section II that $\mathbb{P}$-measure parameters $k_0^{\mathbb{P}}$ and $K_1^{\mathbb{P}}$ and $\mathbb{Q}$-measure parameters $k_0^{\mathbb{Q}}$ and $K_1^{\mathbb{Q}}$ are linked by market price of risk parameters (see equation (6)). As we have noted previously, empirical evidence and economic intuition support changes in both the $\mathbb{P}$-measure dynamics of state variables and the behavior of risk premiums. Therefore, the $\mathbb{Q}$-measure parameters are naturally expected to change as well, unless the changes in $\mathbb{P}$-measure parameter $K_1^{\mathbb{P}}$ and in market price of risk Λ_1 exactly offset each other. There appears to be no obvious economic argument for such cancelation.[36]

VII. Conclusion

In this article, 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 yield fitting errors, as well as likelihood score-based tests. To

[34]In King's model, 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.

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

[36]Some related studies, including Andreasen et al. (2019) and Giacoletti, Laursen, and Singleton (2021), have assumed that the $\mathbb{Q}$-measure dynamics remain structurally stable even as the $\mathbb{P}$-measure parameters are subject to a structural break. The increased tractability of such a model may justify the assumption of stable $\mathbb{Q}$ parameters in some applications even as our statistical evidence indicates a structural break in them. We discuss this point further in Section IA.IV of the Supplementary Material.

characterize the changes in yield dynamics as generally as possible, we estimate 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 article 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 $\mathbb{P}$-dynamics, $\mathbb{Q}$-dynamics (and hence the principal components of shadow yields), 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 change in yield dynamics in the post-ELB period is not as simple as suggested by structurally stable shadow rate models, at least in the case of U.S. data, and that more material changes in the dynamics of Treasury yields took place after the federal funds rate hit the ELB in 2008. Our article has examined this issue from the perspective of a shadow rate model with a structural break, but the empirical results herein may 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.

Supplementary Material

To view supplementary material for this article, please visit http://doi.org/10.1017/S0022109023000984.

References

Ahn, D.-H.; R. F. Dittmar; and A. R. Gallant. "Quadratic Term Structure Models: Theory and Evidence." *Review of Financial Studies*, 15 (2002), 243–288.

Andreasen, M. M.; K. Jørgensen; and A. Meldrum. *Bond Risk Premiums at the Zero Lower Bound*. Washington, DC: Board of Governors of the Federal Reserve System (2019).

Andreasen, M. M., and A. Meldrum. "A Shadow Rate or a Quadratic Policy Rule? The Best Way to Enforce the Zero Lower Bound in the United States." *Journal of Financial and Quantitative Analysis*, 54 (2019), 2261–2292.

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

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

Ang, A., and M. Piazzesi. "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables." *Journal of Monetary Economics*, 50 (2003), 745–787.

Backus, D.; S. Foresi; A. Mozumdar; and L. Wu. "Predictable Changes in Yields and Forward Rates." *Journal of Financial Economics*, 59 (2001), 281–311.

Bauer, M. D. "Restrictions on Risk Prices in Dynamic Term Structure Models." *Journal of Business & Economic Statistics*, 36 (2018), 196–211.

Bauer, M. D., and G. D. Rudebusch. "Monetary Policy Expectations at the Zero Lower Bound." *Journal of Money, Credit and Banking*, 48 (2016), 1439–1465.

Bauer, M. D., and G. D. Rudebusch. "Interest Rates Under Falling Stars." *American Economic Review*, 110 (2020), 1316–1354.

Bauer, M. D.; G. D. Rudebusch; and J. C. Wu. "Correcting Estimation Bias in Dynamic Term Structure Models." *Journal of Business & Economic Statistics*, 30 (2012), 454–467.

Bernanke, B. S. "The New Tools of Monetary Policy." *American Economic Review*, 110 (2020), 943–983.
Campbell, J. Y., and R. J. Shiller. "Yield Spreads and Interest Rate Movements: A Bird's Eye View." *Review of Economic Studies*, 58 (1991), 495–514.
Chow, G. C. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." *Econometrica*, 28 (1960), 591–605.
Christensen, J. H.; F. X. Diebold; and G. D. Rudebusch. "The Affine Arbitrage-Free Class of Nelson–Siegel Term Structure Models." *Journal of Econometrics*, 164 (2011), 4–20.
Christensen, J. H. E., and G. D. Rudebusch. "Estimating Shadow-Rate Term Structure Models with Near-Zero Yields." *Journal of Financial Econometrics*, 13 (2014), 226–259.
Chung, H.; J.-P. Laforte; D. Reifschneider; and J. C. Williams. "Have We Underestimated the Likelihood and Severity of Zero Lower Bound Events?" *Journal of Money, Credit and Banking*, 44 (2012), 47–82.
Dai, Q., and K. J. Singleton. "Specification Analysis of Affine Term Structure Models." *Journal of Finance*, 55 (2000), 1943–1978.
Dai, Q., and K. J. Singleton. "Expectation Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure." *Journal of Financial Economics*, 63 (2002), 415–441.
Duffee, G. R. "Term Premia and Interest Rate Forecasts in Affine Models." *Journal of Finance*, 57 (2002), 405–443.
Duffee, G. R. "Information in (and not in) the Term Structure." *Review of Financial Studies*, 24 (2011), 2895–2934.
Duffie, D., and R. Kan. "A Yield-Factor Model of Interest Rates." *Mathematical Finance*, 6 (1996), 379–406.
Eggertsson, G. B., and M. Woodford. "The Zero Bound on Interest Rates and Optimal Monetary Policy." *Brookings Papers on Economic Activity*, 2003 (2003), 139–211.
Fama, E. F. "The Information in the Term Structure." *Journal of Financial Economics*, 13 (1984), 509–528.
Ghysels, E., and A. Hall. "A Test for Structural Stability of Euler Conditions Parameters Estimated Via the Generalized Method of Moments Estimator." *International Economic Review*, 31 (1990), 355–364.
Giacoletti, M.; K. T. Laursen; and K. J. Singleton. "Learning from Disagreement in the U.S. Treasury Bond Market." *Journal of Finance*, 76 (2021), 395–441.
Gürkaynak, R. S.; B. Sack; and J. H. Wright. "The U.S. Treasury Yield Curve: 1961 to the Present." *Journal of Monetary Economics*, 54 (2007), 2291–2304.
Gust, C.; E. Herbst; D. López-Salido; and M. E. Smith. "The Empirical Implications of the Interest-Rate Lower Bound." *American Economic Review*, 107 (2017), 1971–2006.
Hansen, B. E. "Lagrange Multiplier Tests for Parameter Instability in Non-Linear Models." Working Paper, University of Rochester (1990).
Hansen, B. E. "Tests for Parameter Instability in Regressions with I(1) Processes." *Journal of Business & Economic Statistics*, 10 (1992), 321–335.
Hördahl, P., and O. Tristani. "Modelling Yields at the Lower Bound Through Regime Shifts." Working Paper, European Central Bank, Frankfurt (2019).
Hu, G. X.; J. Pan; and J. Wang. "Noise as Information for Illiquidity." *Journal of Finance*, 68 (2013), 2341–2382.
Johannsen, B. K., and E. Mertens. "A Time Series Model of Interest Rates with the Effective Lower Bound." Working Paper, Board of Governors of the Federal Reserve System, Washington (2016).
Joslin, S.; M. Priebsch; and K. J. Singleton. "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks." *Journal of Finance*, 69 (2014), 1197–1233.
Kim, D. H., and A. Orphanides. "Term Structure Estimation with Survey Data on Interest Rate Forecasts." *Journal of Financial and Quantitative Analysis*, 47 (2012), 241–272.
Kim, D. H., and K. J. Singleton. "Term Structure Models and the Zero Bound: An Empirical Investigation of Japanese Yields." *Journal of Econometrics*, 170 (2012), 32–49.
Kim, D. H., and J. H. Wright. "An Arbitrage-Free Three-Factor Term Structure Model and the Recent Behavior of Long-Term Yields and Distant-Horizon Forward Rates." Working Paper, Board of Governors of the Federal Reserve System, Washington (2005).
King, T. B. "Expectation and Duration at the Effective Lower Bound." *Journal of Financial Economics*, 134 (2019), 736–760.
Krippner, L. "Measuring the Stance of Monetary Policy in Zero Lower Bound Environments." *Economics Letters*, 118 (2013), 135–138.
Kuttner, K. N. "Outside the Box: Unconventional Monetary Policy in the Great Recession and Beyond." *Journal of Economic Perspectives*, 32 (2018), 121–146.

Li, C.; A. Meldrum; and M. Rodriguez. "New Financial Market Measures of the Neutral Real Rate and Inflation Expectations." Working Paper, Board of Governors of the Federal Reserve System, Washington (2017).
Liu, P.; K. Theodoridis; H. Mumtaz; and F. Zanetti. "Changing Macroeconomic Dynamics at the Zero Lower Bound." *Journal of Business & Economic Statistics*, 37 (2019), 391–404.
Nakata, T., and H. Tanaka. "Equilibrium Yield Curves and the Interest Rate Lower Bound." Working Paper, Board of Governors of the Federal Reserve System, Washington (2016).
Nyblom, J. "Testing for the Constancy of Parameters over Time." *Journal of the American Statistical Association*, 84 (1989), 223–230.
Priebsch, M. A. "Computing Arbitrage-Free Yields in Multi-Factor Gaussian Shadow-Rate Term Structure Models." *Quarterly Journal of Finance*, 13 (2023), 2350013.
Reifschneider, D., and J. C. Williams. "Three Lessons for Monetary Policy in a Low-Inflation Era." *Journal of Money, Credit and Banking*, 32 (2000), 936–966.
Swanson, E. T., and J. C. Williams. "Measuring the Effect of the Zero Lower Bound on Medium- and Longer-Term Interest Rates." *American Economic Review*, 104 (2014), 3154–3185.
Wan, E., and R. van der Merwe. "The Unscented Kalman Filter." In *Kalman Filtering and Neural Networks*, S. Haykin, ed. New York, NY: Wiley (2001), 221–280.
Wu, J. C., and F. D. Xia. "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound." *Journal of Money, Credit and Banking*, 48 (2016), 253–291.