Loss Functions for Forecasting Treasury Yields

Hitesh Doshi   Kris Jacobs   Rui Liu

University of Houston

October 20, 2015

Abstract

Many recent advances in the term structure literature have focused on model specification and estimation. Forecasting the yield curve is critically important, but it has thus far not been explicitly taken into account at the estimation stage. We propose to estimate term structure models by aligning the loss functions for in-sample estimation and out-of-sample forecast evaluation. We document the resulting differences in forecasting performance using three-factor affine term structure models with and without stochastic volatility. We confirm that aligning loss functions provides substantial improvements in out-of-sample forecasting performance, especially for long forecast horizons. We document the trade-off between in-sample and out-of-sample fit. The resulting parameter estimates imply factors that differ from the traditional term structure factors, especially in the case of the third (curvature) factor. This suggests that the improvement in out-of-sample fit results from identification of the third factor, which captures information otherwise hidden to conventional in-sample loss functions.

JEL Classification: G12, E43

Keywords: term structure; forecasting; loss function; state variables; identification; hidden factor.

1   Introduction

Modeling and predicting government bond yields is a topic of great practical importance for both investors and monetary policy makers. It is therefore not surprising that the literature on forecasting Treasury yields is very extensive, but existing studies focus almost exclusively on comparisons of the forecasting performance for alternative specifications of the term structure model itself.[1] The forecasting exercise is not explicitly taken into account at the estimation stage. We take a different perspective and analyze how the choice of loss function affects a given model's out-of-sample forecasting performance. We investigate if it is possible to improve out-of-sample forecasting performance by aligning the loss function at the estimation stage with the out-of-sample evaluation measure. We analyze this question using the class of Affine Term Structure Models (ATSMs). These models are popular tools for term structure modeling because they deliver essentially closed-form expressions for bond prices and yields.[2]

It is well known in the statistics literature that the specification of the loss function is critical for model estimation and evaluation. Indeed, the specification of a loss function implicitly amounts to the specification of a statistical model, because the loss function determines how different forecast errors are valued (see Engle, 1993; Granger, 1993; Weiss, 1996; Elliott and Timmermann, 2008). The loss function is an important element in the process of delivering a forecast, and is therefore an integral part of model specification. Estimating a model under one loss function and evaluating it under another amounts to changing the model specification without allowing the parameter estimates to adjust. If a particular criterion is used to evaluate forecasts, it should also be used at the estimation stage.[3]

Motivated by these insights, we align the loss functions for in-sample estimation and out-of-sample evaluation of ATSMs. We propose to estimate the model by minimizing the squared forecasting errors for a given forecast horizon, and we refer to these estimates as based on the forecasting loss function. We compare the out-of-sample performance of these estimates with the performance of estimates obtained by minimizing the mean-squared error loss function based on

---

[1]See Duffee (2002), Ang and Piazzesi (2003), Diebold and Li (2006), Bowsher and Meeks (2008), and Christensen, Diebold, and Rudebusch (2011) for examples of studies that focus on point forecasts. See Hong, Li, and Zhao (2004), Egorov, Hong, and Li (2006), and Shin and Zhong (2013) for studies that focus on density forecasts.

[2]The empirical literature on ATSMs is very extensive. See Vasicek (1977), Cox, Ingersoll, and Ross (1985), Chen and Scott (1992), Longstaff and Schwartz (1992), Duffie and Kan (1996), and Dai and Singleton (2000) for important contributions.

[3]An extensive literature studies the theoretical properties of optimal forecasts under asymmetric loss functions and documents that forecast errors have different properties under different loss functions. See for example Patton and Timmermann (2007a, 2007b), Elliott, Komunjer, and Timmermann (2005, 2008), and Christoffersen and Diebold (1996, 1997). Christoffersen and Jacobs (2004) highlight the importance of aligning the loss function for the purpose of option valuation, using the Dumas, Fleming, and Whaley (1998) implied volatility model.

current yields, which we refer to as the standard loss function.

We focus on three-factor ATSMs because of their importance in the existing literature and their tractability. Despite the popularity of this class of models, it is well-known that the presence of latent state variables gives rise to identification problems that may complicate comparisons of out-of-sample performance. We therefore provide an additional analysis of the Gaussian three-factor model. Identification in Gaussian ATSMs is facilitated by the new canonical form proposed by Joslin, Singleton, and Zhu (2011, henceforth referred to as JSZ), in which the state variables are restricted to be the first three principal components. The JSZ normalization is also particularly well suited for out-of-sample model evaluation with recursive estimation, because it provides substantial computational advantages.

We compare the out-of-sample forecasting performance using the forecasting loss function with the performance using the standard loss function. We first compare the performance using Gaussian and stochastic volatility models with three latent factors, which we implement using the Kalman filter. We then repeat the exercise for the Gaussian model using the JSZ canonical form. JSZ restrict the state variables to be the first three principal components, because for in-sample estimation the weights corresponding to the principal components provide the best possible fit. We confirm this result, but we also find that for out-of-sample forecasting, these weights are not optimal. We therefore provide an alternative implementation of the JSZ canonical form in which we allow the portfolio weights to be free parameters. We specify the state variables as weighted averages of the yields, but rather let the data determine the best possible weights from a forecasting perspective. This approach is motivated by the literature on predicting bond returns. Cochrane and Piazzesi (2005) and Duffee (2011a), among others, argue that a hidden factor not captured by the traditional level, slope, and curvature factors helps in predicting excess bond returns.

We find substantial improvements in the out-of sample forecasting performance of all three-factor models we studied when using the forecasting loss function in estimation, especially for longer forecast horizons and shorter maturities. For example, using the JSZ canonical specification for the Gaussian model, the improvement in the root mean square error (RMSE) for short maturity yields is about 11% on average across different forecast horizons, which corresponds to an out-of-sample R-square of 23%. For the six-month forecast horizon, the improvement is about 7% on average across maturities, which corresponds to an out-of-sample R-square of 15%. The improvements obtained using the Gaussian latent factor model are similar in magnitude. We also find substantial improvements in the out-of sample forecasting performance of the stochastic volatility models with three latent factors, especially for longer forecast horizons. For example, in the $A_1(3)$ model, for the six-month forecast horizon, the improvement in the forecasting RM-

SEs is approximately 15% on average across maturities, which corresponds to an out-of-sample R-square of 28%.

These results confirm the insights of Granger (1993) and Engle (1993) that aligning the estimation loss function with the loss function used for out-of-sample model evaluation improves out-of-sample forecasting performance. Based on these insights, we also expect the parameters estimated using the forecasting loss function not to improve on the in-sample fit based on the parameters obtained using the standard loss function. We confirm that this is the case using the estimates for the JSZ canonical specification. The differences in in-sample fit are relatively small but show up at longer maturities.

We compare the state variables implied by the forecasting loss function with the state variables based on a standard loss function in the JSZ canonical form. The forecasting loss function implies a different linear combination of yields compared to the traditional level, slope, and curvature factors, especially for the curvature (third) factor. The changes in the portfolio weights capture the information hidden from the term structure, which is uncovered in the forecasting exercise.

Our paper contributes to the literature on the estimation of ATSMs. Much of the recent literature on these models focuses on innovative estimation approaches to address the well-known identification problems inherent in the estimation of ATSMs.[4] We do not focus on new estimation techniques, and we do not directly focus on identification problems. Our contribution is therefore complementary to most of the recent literature on ATSMs, because the insight that estimation using the forecasting loss function will lead to better out-of-sample performance is valid regardless of the estimation method. The closest related work is by Adrian, Crump, and Moench (2013) and Sarno, Schneider, and Wagner (2014), who estimate model parameters in ATSMs using an objective function that takes into account excess returns for different horizons. This approach is similar to ours in the sense that the implied loss function is different from the standard loss function based on yields. However, their implied loss function is different from ours, and therefore not necessarily optimal from a forecasting perspective.

The paper proceeds as follows. Section 2 compares the forecasting loss function with the standard loss function based on yields. Section 3 presents the data. Section 4 compares the forecasting performance of different loss functions based on the estimation of Gaussian and stochastic volatility models with latent factors. Section 5 repeats this exercise for the Gaussian model using the JSZ canonical specification. Section 6 documents the trade-off between in-sample and out-of-sample fit, and discusses the differences in implied state variables and parameter

---

[4]On identification problems in these models, see for example Duffee (2011b), Duffee and Stanton (2012), and Hamilton and Wu (2012). For examples of methods that help address these identification problems, see JSZ (2011), Hamilton and Wu (2012), Adrian, Crump, and Moench (2013), Diez de los Rios (2014), Bauer, Rudebusch, and Wu (2012), and Creal and Wu (2015).

estimates. Section 7 concludes.

2   Loss Functions for Term Structure Estimation

Given term structure data for months $t = 1, ..., T$ on maturities $n = 1, ..., N$, the parameters Θ of a term structure model are typically estimated using a loss function that minimizes a well-defined distance between the observed yields y_t^n and the model yield, which we denote here by $\widehat{y}_{t|t}^n(\Theta)$ to emphasize that the model yield is computed using the state variables at time t. In general, the notation $\widehat{y}_{t+k|t}^n$ indicates a model-implied yield at time $t + k$ computed using information up to time t. We use this type of loss function as a benchmark. Several such loss functions can in principle be used, but we limit ourselves to loss functions that are based on the difference between observed and model yields.[5] We estimate the term structure parameters Θ by minimizing the root-mean-squared-error based on observed and model yields:[6]

$$RMSE(\Theta) = \sqrt{\frac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} (\widehat{y}_{t|t}^n(\Theta) - y_t^n)^2}. \tag{2.1}$$

Estimating the model parameters by optimizing the log likelihood or the root-mean-squared-error provides the best possible in-sample fit. Our focus is not on in-sample fit but rather on forecasting. To improve forecasting performance, we deviate from the benchmark implementation by aligning the loss functions for in-sample and out-of-sample evaluation, as suggested by Granger (1993) and Weiss (1996). The choice of loss function at the estimation stage should therefore reflect that out-of-sample forecasting is the objective of the empirical exercise. The out-of-sample forecasting performance for the n-maturity yield with forecast horizon k is evaluated using

$$RMSE\_OS_{n,k} = \sqrt{\frac{1}{T-k} \sum_{t=1}^{T-k} (\widehat{y}_{t+k|t}^n(\Theta) - y_{t+k}^n)^2}, \tag{2.2}$$

where y_{t+k}^n is the observed n-maturity yield at time $t + k$ and $\widehat{y}_{t+k|t}^n(\Theta)$ is the model-predicted k-period ahead n-maturity yield based on the parameter set Θ, which is estimated at time t.

To align the loss function at the estimation stage with the out-of-sample loss function, we

---

[5]Alternatively, loss functions based on relative errors or other transformations of yields can be studied, but in the term structure literature this is less critical than for other applications, such as derivative securities.

[6]In-sample estimation of term structure models usually maximizes the log likelihood. We use the root-mean-squared error instead to facilitate the comparison with the forecasting loss function. If the measurement errors are normally distributed and constant across maturities, the likelihood simply scales the mean-squared error. For other cases, optimizing the likelihood and the mean squared error gives very similar results.

therefore estimate the models for a given forecast horizon k by minimizing the following loss function:

$$OS\_RMSE_k(\Theta) = \sqrt{\frac{1}{N(T-k)}\sum_{n=1}^{N}\sum_{t=k+1}^{T}(\widehat{y}_{t|t-k}^n(\Theta) - y_t^n)^2}. \qquad (2.3)$$

3   Data

We use monthly data on continuously compounded zero-coupon bond yields with maturities of three and six months, and one, two, three, four, five, ten and twenty years, for the period April 1953 to December 2012. The three- and six-months yields are obtained from the Fama CRSP Treasury Bill files, and the one- to five-year bond yields are obtained from the Fama CRSP zero coupon files. The ten- and twenty-year maturity zero-coupon yields are obtained from the H.15 data release of the Federal Reserve Board of Governors.[7]

Table 1 shows that, on average, the yield curve is upward sloping, and the volatility of yields is relatively lower for longer maturities. The yields for all maturities are highly persistent, with slightly higher autocorrelation for long-term yields than for short-term yields. Yields exhibit mild excess kurtosis and positive skewness for all maturities.

4   Results for Models with Latent Factors

We compare the forecasting performance of estimation based on the benchmark loss function equation (2.1) and the forecasting loss function in equation (2.3). Our argument about the choice of loss function applies in principle to all term structure models, but we limit ourselves to a comparison based on three-factor affine term structure models with and without stochastic volatility. This choice is mainly motivated on the one hand by the popularity of affine term structure models, as well as by their tractability.

It is always important to be mindful of identification problems, but it is especially critical for our analysis, because these problems can easily affect the comparison of the loss functions. Recently, important advances have been made in the estimation of the Gaussian three-factor model $A_0(3)$ that facilitate a meaningful comparison of loss functions for this choice of model (JSZ, 2011). For the $A_0(3)$ model, we can therefore investigate the implications of the loss

---

[7]The Federal Reserve database provides constant maturity treasury (CMT) rates for different maturities. The ten- and twenty-year CMT rates are converted into zero-coupon yields using the piecewise cubic polynomial. Data on 20-year yields are not available from January 1987 through September 1993. We fill this gap by computing the 20-year CMT forward yield using 10-year and 30-year CMT yields.

function using a traditional implementation of this model with latent factors, but also using the canonical specification proposed by JSZ (2011).

In this section we report on the loss function comparison based on the Gaussian and the stochastic volatility models with latent factors. In the next section we investigate the robustness of our findings using the canonical specification by JSZ (2011) for the Gaussian model, which addresses the identification problems by mapping these latent variables into observables.

4.1 Three-factor Affine Models

In the term structure literature, affine term structure models (ATSMs) have received significant attention because of their rich structure and tractability. The existing literature has concluded that at least three factors are needed to explain term structure dynamics (see for example Litterman and Scheinkman, 1991; Knez, Litterman, and Scheinkman, 1994). Accordingly, we use an ATSM with three state variables.

Using the classification of Dai and Singleton (2000), we focus on $A_j(3)$ models with $j = 0, 1, 2$ or 3 factors driving the conditional variance of the state variables, which are given by

$$dX_t = (K_{0\Delta}^P + K_{1\Delta}^P X_t)dt + \Sigma\sqrt{S_t}dW_{t+1}^P, \tag{4.1}$$

$$dX_t = (K_{0\Delta}^Q + K_{1\Delta}^Q X_t)dt + \Sigma\sqrt{S_t}dW_{t+1}^Q, \tag{4.2}$$

$$r_t = \rho_0 + \rho_1 X_t, \tag{4.3}$$

where W_{t+1}^P and W_{t+1}^Q are three-dimensional independent standard Brownian motions under physical measure P and risk-neutral measure Q respectively, r_t is the instantaneous spot interest rate, and $\Sigma S_t \Sigma'$ is the conditional covariance matrix of X_t. S_t is a 3×3 diagonal matrix with the ith diagonal element given by

$$[S_t]_{ii} = \alpha_i + \beta_i' X_t, \tag{4.4}$$

where α_i is a scalar, and β_i is a 3×1 vector. $\alpha = [\alpha_1, \alpha_2, \alpha_3]'$ is a 3×1 vector. $\beta = [\beta_1, \beta_2, \beta_3]$ is a 3×3 matrix. We follow the Dai and Singleton identification scheme to ensure the $[S_t]_{ii}$ are strictly positive for all i. Under this identification scheme, Σ is an identity matrix.[8] In the $A_0(3)$ model, α is a vector of ones and β_i is a vector of zeros for all i. In the $A_1(3)$ model, β_i is a vector of zeros for $i = 2$ and $i = 3$, and in the $A_2(3)$ model, β_i is a vector of zeros for $i = 3$.

The model-implied continuously compounded yields $\widehat{y}_t$ are given by (see Duffie and Kan,

---

[8]The identification constraints can be applied either on P- or Q- parameters, see Dai and Singleton (2000) and Singleton (2006).

1996)

$$\widehat{y}_t = A(\Theta^Q) + B(\Theta^Q)X_t, \tag{4.5}$$

where the $N \times 1$ vector $A(\Theta^Q)$, and the $N \times 3$ matrix $B(\Theta^Q)$ are functions of the parameters under the Q-dynamics, $\Theta^Q = \{K_{0\Delta}^Q, K_{1\Delta}^Q, \rho_0, \rho_1, \Sigma, \alpha, \beta\}$, through a set of Ricatti ordinary differential equations. Recall that N denotes the number of available yields in the term structure. We adopt the essentially affine specification for the price of risk, as in Duffee (2002).

We use monthly data on continuously compounded zero-coupon bond yields with nine different maturities for the period April 1953 to December 2012 to estimate the models. The affine dynamic for X_t in equation (4.1) implies that the one-period ahead conditional expectation of X_t under the P measure, $\widehat{X}_{t+\Delta|t} = \text{constant} + \Delta e^{K_{1\Delta}^P} X_t$, where $\Delta = 1/12$. Thus X_t follows a first order VAR when sampled monthly. Similarly, the affine dynamic in equation (4.2) under the Q measure implies a first order VAR for X_t sampled at the monthly frequency. For estimation based on the forecasting loss function in equation (2.3), we need the model's prediction of the k-period ahead n-maturity yield, based on parameter estimates at time t. This is given by

$$\begin{aligned} \widehat{y}_{t+k|t}^n(\Theta) &= A_n(\Theta^Q) + B_n(\Theta^Q)\widehat{X}_{t+k|t} \\ &= A_n(\Theta^Q) + B_n(\Theta^Q)f(X_t, k; K_0^P, K_1^P), \end{aligned} \tag{4.6}$$

where $A_n(\Theta^Q)$ is the nth element of $A(\Theta^Q)$, $B_n(\Theta^Q)$ is the nth row of $B(\Theta^Q)$, and f is given by

$$f(X_t, k; K_0^P, K_1^P) = K_0^P(I_3 + K_1^P + ... + (K_1^P)^{k-1}) + (K_1^P)^k X_t.$$

where K_0^P and K_1^P are the parameters for the VAR(1) process of X_t under the P measure, which can be mapped to $K_{0\Delta}^P$ and $K_{1\Delta}^P$ respectively in equation (4.1) through the nonlinear relations $K_1^P = e^{\Delta K_{1\Delta}^P}$ and $K_0^P = K_{0\Delta}^P \int_0^\Delta e^{sK_{1\Delta}^P} ds$. In particular, for small Δ, $K_0^P \approx \Delta K_{0\Delta}^P$ and $K_1^P \approx I_3 + \Delta K_{1\Delta}^P$. We can view K_0^P and $K_{0\Delta}^P$, and K_1^P and $K_{1\Delta}^P$ interchangeably. Similarly, K_0^Q and $K_{0\Delta}^Q$, and K_1^Q and $K_{1\Delta}^Q$ are interchangeable.[9]

A three factor latent model can be expressed using a state-space representation. Using equation (4.1) and an Euler discretization, the state equation can be written as $X_{t+1} = K_0^P + K_1^P X_t + \varepsilon_{t+1}^P$, where $\varepsilon_{t+1|t}^P$ is assumed to be distributed $N(0, \Sigma S_t \Sigma')$. The observed yield curve $y_t = \widehat{y}_t + e_t$ is the measurement equation, where $\widehat{y}_t$ is the model-implied yield as specified in equation (4.5), and e_t is a vector of measurement errors that is assumed to be $i.i.d.$ normal with diagonal covariance matrix R. The estimates of the P-parameters, $\Theta^P = \{K_0^P, K_1^P\}$ are related to the Q-parameters,

---

[9]Since our data frequency is monthly, it is more convenient to focus on K_0^P, K_1^P, K_0^Q and K_1^Q in the empirical analysis.

since the pricing model is required to filter the latent factors X_t. We therefore need to estimate the P- and Q-parameters simultaneously. We do this by applying the Kalman filter to the state-space representation.[10] We estimate the parameters $\Theta = \{K_0^P, K_1^P, K_0^Q, K_1^Q, \rho_0, \rho_1, \Sigma, \alpha, \beta\}$ and filter the state variables X_t by minimizing the forecasting loss function, equation (2.3). We compare the results obtained from the forecasting loss function with the estimation of the fully latent models based on the standard loss function equation (2.1).

When estimating these models with latent factors, the numerical implementation is important because of the existence of identification problems. We discuss our implementation in Appendix A.

4.2 The Forecasting Performance of the Latent Gaussian Model

We compare the out-of-sample forecasting performance of the latent $A_0(3)$ model with forecasting loss function equation (2.3) relative to the latent $A_0(3)$ model with standard loss function equation (2.1) by computing the out-of-sample forecast RMSEs for the one-month to six-month forecast horizons, for all nine maturities used in estimation.

Our procedure for examining the out-of-sample forecasts of the model with forecasting loss function is as follows. We proceed recursively with estimation and forecasting, each time adding one month to the estimation sample. At each time t and for each forecast horizon k, we estimate the specification using data up to and including t. Our first estimation uses the first half of the data, up to December 1982. The estimation is based on the forecasting loss function as expressed in equation (2.3). We estimate the parameters $\Theta_t^k = \{K_0^P, K_1^P, K_0^Q, K_1^Q, \rho_0, \rho_1, \Sigma\}$ by minimizing the k-period ahead squared forecasting errors, applying the Kalman filter to the state-space representation of the $A_0(3)$ model with latent factors, and filtering the state variables X_t. Subsequently, we forecast the k-period ahead yields $\widehat{y}_{t+k|t}^n(\Theta_t^k)$, $n = 1, ..., N$.

The recursion then proceeds: we add one month of data, re-estimate the parameters and re-filter the latent factors using information up to and including time $t + 1$, and forecast the k-period ahead yields $\widehat{y}_{t+1+k|t+1}^n(\Theta_{t+1}^k)$. We continue to update the sample in this way until time $T - k$, where T is the end of the sample, December 2012. Note that the estimation based on the forecasting loss function is forecast-horizon specific. At each time t, we have a different parameter set Θ_t^k for each k.

The procedure for the latent model with the standard loss function equation (2.1) follows the same recursion, but this procedure is by definition not horizon-specific, instead, one set of

---

[10]See Duffee and Stanton (2012) and Christoffersen, Dorion, Jacobs and Karoui (2014) for estimation using Kalman filter.

parameters is estimated that is used to generate forecasts for different horizons.

Panels A and B of Table 2 present the RMSEs for the forecasting loss function equation (2.3) and the standard loss function equation (2.1). Panel C presents the RMSE ratios. The RMSE ratios are defined as the ratio of the RMSE obtained using the forecasting loss function and the RMSE obtained using the standard loss function. An RMSE ratio less than one indicates that the forecasting loss function provides improvements in forecasting relative to the benchmark standard loss function.

The improvements in forecast performance are greatest for longer forecast horizons and shorter maturities. For the six-month forecast horizon, the improvement in the forecasting RMSEs from using the forecasting loss function equation (2.3) is on average across maturities approximately 12%. In the forecasting literature, the out-of-sample R-square is often considered, which is defined as $1 - (MSE^{FL}/MSE^{SL})$, where SL refers to the benchmark model with standard loss function and FL to the alternative model with forecasting loss function. For the six-month forecasting horizon in Table 2, this gives $1 - (1 - 0.12)^2 = 0.22$. The improvement in forecasting RMSE therefore corresponds to an out-of-sample R-square of 22%. For the three-month yield, the improvement in RMSE is approximately 10% on average across forecast horizons, which corresponds to an out-of-sample R-square of 19%.

4.3 The Forecasting Performance of the Latent Stochastic Volatility Models

The procedure for examining the out-of-sample forecasts of the models with stochastic volatility is the same as that of the $A_0(3)$ model. At each time t and for each forecast horizon k, we estimate the parameters $\Theta_t^k = \{K_0^P, K_1^P, K_0^Q, K_1^Q, \rho_0, \rho_1, \Sigma, \alpha, \beta\}$ by minimizing the forecasting loss function as expressed in equation (2.3), applying the Kalman filter to the state-space representation of the $A_j(3)$ models with latent factors, and filtering the state variables X_t. Subsequently, we forecast the k-period ahead yields $\widehat{y}_{t+k|t}^n(\Theta_t^k)$, $n = 1, ..., N$. We report the out-of-sample forecast RMSEs of the $A_1(3)$ model in Table 3, the $A_2(3)$ model in Table 4, and the $A_3(3)$ model in Table 5. In each table, Panels A and B present the RMSEs for the forecasting loss function equation (2.3) and the standard loss function equation (2.1). Panels C present the RMSE ratios.

The results based on the stochastic volatility models are consistent with the results from the Gaussian model. Aligning loss functions for in-sample estimation and out-of-sample forecast evaluation provides improvements in out-of-sample forecasting performance. The improvements are more pronounced for long forecast horizons in the stochastic volatility models. In the $A_1(3)$ model, for the six-month forecast horizon, the improvement in the forecasting RMSEs from using

the forecasting loss function equation (2.3) is on average across maturities approximately 15%, which corresponds to an out-of-sample R-square of 28%. The improvements of the $A_2(3)$ model and the $A_3(3)$ model are very similar to that of the $A_1(3)$ model. The out-of-sample R-square is on average across maturities approximatly 28% for both the $A_2(3)$ model and the $A_3(3)$ model at six-month forecast horizon.

5   Results Based on the JSZ Canonical Specification

The estimation of ATSMs is challenging due to the high level of nonlinearity in the parameters (Duffee, 2011b; Duffee and Stanton, 2012). Dai and Singleton (2000) argue that not all parameters are well identified, and that rotation and normalization restrictions need to be imposed. Even with the Dai-Singleton normalization, it is possible to end up within a parameter space that is locally unidentified. See for instance the discussions in Hamilton and Wu (2012), Collin-Dufresne, Goldstein, and Jones (2008) and Aït-Sahalia and Kimmel (2010).

This implies that we need to be careful about the interpretation of the results in Section 4. Most critically, if the estimation using the standard loss function equation (2.1) does not lead to the global optimum, we may overestimate the advantages provided by the forecasting loss function equation (2.3). The opposite is of course also possible.

In recent work, JSZ (2011) developed a canonical representation that allows for stable and tractable estimation of the $A_0(3)$ model and addresses these identification problems. In this section we repeat the analysis using their representation of the model. We first provide the main aspects of the $A_0(3)$ canonical representation in JSZ. Subsequently, we present the empirical results.

5.1   The JSZ Canonical Form

We now provide the main aspects of the $A_0(3)$ canonical representation in JSZ. For further details, we refer to Appendix B and JSZ (2011). The state variables under the JSZ normalization are the perfectly priced portfolios of yields, $PO_t = Wy_t$. W denotes the portfolio weights, a $3 \times N$ matrix. PO_t is governed by the same dynamics as the latent state variable X_t, as specified in equations (4.1)-(4.3).[11] The model-implied continuously compounded yields $\widehat{y}_t$ are given by

$$\widehat{y}_t = A(\Theta^Q) + B(\Theta^Q)PO_t. \tag{5.1}$$

---

[11]Note that the $A_0(3)$ canonical representation in JSZ (2011) is presented in discrete time. In our setup, the continuous-time affine dynamics in equations (4.1)-(4.2) imply a first order VAR for PO_t at the monthly frequency. The parameters for the VAR(1) process of PO_t can be mapped to the continuous-time parameters.

JSZ show that $A(\Theta^Q)$ and $B(\Theta^Q)$ are ultimately functions of $\Theta^Q = \{r_\infty^Q, \lambda^Q, \Sigma\}$, where r_∞^Q is a scalar related to the long-run mean of the short rate under risk neutral measure and λ^Q, a 3×1 vector, represents the ordered eigenvalues of K_1^Q. Appendix B provides further details about this transformation.

Note that the state variables under the JSZ normalization are observable, and thus the parameters governing the P-dynamics $\Theta^P = \{K_0^P, K_1^P\}$ can be estimated separately from the parameters governing the Q-dynamics. JSZ demonstrate that the ordinary least squares (OLS) estimates of K_0^P and K_1^P from the observed factors PO_t nearly recover the maximum likelihood (ML) estimates of K_0^P and K_1^P from the P- and Q-dynamics jointly, to the extent that $W y_t \approx W \widehat{y}_t$. As noted by JSZ, the best approximation is obtained by choosing W_0 such that $W_0 y_t = PC_t$, the first three principal components of the observed term structure of yields.[12]

The JSZ normalization results in substantial computational advantages, which arise because of the smaller number of Q-parameters to be estimated through maximum likelihood. For a three-factor model, there are in total $1 + 3 + 6 + N = 10 + N$ parameters to be estimated (1 for r_∞^Q, 3 for λ^Q, 6 for Σ, and N for the variance-covariance matrix of the measurement errors).

The model-predicted k-period ahead n-maturity yield given the estimated parameter set Θ at time t can be defined as follows

$$
\begin{aligned}
\widehat{y}_{t+k|t}^n(\Theta) &= A_n(\Theta^Q) + B_n(\Theta^Q)\widehat{PC}_{t+k|t} \qquad\qquad (5.2)\\
&= A_n(\Theta^Q) + B_n(\Theta^Q)f(PC_t, k; K_0^P, K_1^P),
\end{aligned}
$$

where $A_n(\Theta^Q)$ is the nth element of $A(\Theta^Q)$, $B_n(\Theta^Q)$ is the nth row of $B(\Theta^Q)$, and f is given by

$$
f(PC_t, k; K_0^P, K_1^P) = K_0^P(I_3 + K_1^P + ... + (K_1^P)^{k-1}) + (K_1^P)^k PC_t.
$$

When implementing the JSZ canonical form using the forecasting loss function, we estimate the parameters $\Theta = \{\Theta^P, \Theta^Q\}$ by minimizing the forecasting loss function, equation (2.3). The P-parameters determine the properties of the state variables, which are important for forecasting yields, as seen in equation (5.2). In contrast, these parameters do not play a role in the standard loss function equation (2.1) under the JSZ normalization.[13] This is a critical difference between the loss functions. The forecasting loss function takes into account the properties of the state

---

[12]Strictly speaking, the OLS estimates are exactly the ML estimates only if one assumes that the yields are measured without errors. Empirically, JSZ show that the use of the principal components ensures that the OLS estimates and ML estimates are nearly identical.

[13]Note that JSZ do not minimize the mean-squared error but instead use maximum likelihood. However, the same argument applies: these parameters play no role in the standard likelihood function under the JSZ normalization.

variables. When using the forecasting loss function, we therefore cannot determine K_0^P and K_1^P from the OLS estimates, because the forecasting loss function depends on all parameters simultaneously.

5.2   The Role of the Loss Function with Fixed Portfolio Weights

We now provide an empirical comparison of the forecasting performance of the forecasting loss function equation (2.3) and the standard loss function equation (2.1). Both loss functions are based on the JSZ canonical form of the $A_0(3)$ specification with observed factors. As mentioned above, the JSZ canonical form provides important computational advantages, because it allows the estimation to be performed directly on the principal components of the observed yields, which in turn allows factorization of the likelihood and isolates the subset of parameters governing the Q-dynamics. This canonical form therefore dramatically reduces the difficulties that typically arise in the search for the global optimum. Note that in the JSZ canonical form, the portfolio weights W in $PO_t = Wy_t$ are given by W_0 such that $W_0 y_t = PC_t$.

With fixed weights W_0, it is straightforward to use the method proposed by JSZ to estimate the parameters under both the standard loss function equation and the forecasting loss function. For the standard loss function, we perform a recursive estimation that uses all yields. In the case of the forecasting loss function, for each month t and each forecast horizon k, we estimate the JSZ using data up to and including t. By minimizing the k-period ahead squared forecasting errors, we get the estimated parameter sets Θ^P and Θ^Q for forecast horizon k, and we forecast the k-period ahead yields based on equation (5.2). Table 6 reports the RMSEs in Panels A and B and the RMSE ratios in Panel C.

Note that a comparison of Panel B of Table 6 with Panel B of Table 2 indicates that the JSZ canonical form provides important computational advantages. The RMSE for the JSZ specification in Table 6 is smaller than the RMSE for the latent Gaussian model in Table 2 for almost all maturities and forecast horizons. This confirms that the findings in JSZ (2011) also hold in an out-of-sample setting.

Panel C of Table 6 indicates that the improvement in out-of-sample forecasting performance when using the forecasting loss function is smaller than in the case of the latent model in Table 2. For example, for the six-month forecast horizon, the improvement in the RMSEs is approximately 3% on average across different maturities. This corresponds to an out-of-sample R-square of 5%.

One possible interpretation of these results is that the findings in Table 2, obtained in a model with latent factors, are due to identification problems. Once we adopt the more robust JSZ canonical form, the advantages from aligning loss functions seem to be much more modest.

However, the exercise in Table 6 imposes a very important restriction. We use fixed portfolio weights W_0, which means that we are restricted to using the first three principal components as the state variables at each recursion. We now investigate the importance of this restriction.

5.3 The Role of the Loss Function with Variable Portfolio Weights

The forecasting loss function does not help much in improving the forecasting performance of the JSZ normalization with fixed portfolio weights W_0, as documented in Section 5.2. However, this implementation implicitly assumes that the state variables are equal to the first three principal components at each recursion. JSZ show that this restriction does not affect the results of in-sample estimation much.[14] However, from a forecasting perspective, imposing these restrictions may mean that the parameters governing the dynamics of the state variables, K_0^P and K_1^P, do not have a strong incentive to move away from the OLS estimates, even though the OLS estimates may not be optimal in terms of the out-of-sample forecasting performance.

This insight is motivated by the literature on forecasting bond returns. Cochrane and Piazzesi (2005) suggest that the fourth principal component of the yield curve explains a large portion of bond return predictability. Moreover, the literature on the predictability of bond excess return shows that other variables, such as forward rates (Cochrane and Piazzesi (2005)), macroeconomic variables (Ludvigson and Ng (2009), Cooper and Priestley (2009), Cieslak and Povala, (2015), Joslin, Priebsch, and Singleton (2014)), and a hidden factor (Duffee (2011a)) also help predict bond excess returns. By allowing the weights to be free parameters, the estimation based on the forecasting loss function has more flexibility to search for the best possible state variables for the purpose of forecasting. This parameterization thus provides more flexibility to the forecasting loss function to determine the state variables that are best suited for out-of-sample forecasting.

The resulting econometric problem is somewhat more complex, and it is worth outlining it in more detail. First, consider the model-predicted k-period ahead n-maturity yield given parameter estimates Θ at time t, which can be written as follows

$$
\begin{aligned}
\widehat{y}_{t+k|t}^n(\Theta) &= A_n(\Theta^Q) + B_n(\Theta^Q)\widehat{PO}_{t+k|t} \\
&= A_n(\Theta^Q) + B_n(\Theta^Q)f(y_t, k; K_0^P, K_1^P, W),
\end{aligned}
\tag{5.3}
$$

---

[14]We confirm this by performing a full sample one-time estimation of the JSZ with standard loss function and variable weights. The portfolio weights W converge to W_0. The first three principal components provide the best in-sample fit.

where $A_n(\Theta^Q)$ is the nth element of $A(\Theta^Q)$, $B_n(\Theta^Q)$ is the nth row of $B(\Theta^Q)$, and f is given by

$$f(y_t, k; K_0^P, K_1^P, W) = K_0^P(I_3 + K_1^P + ... + (K_1^P)^{k-1}) + (K_1^P)^k W y_t.$$

We estimate the JSZ representation with variable portfolio weights for each forecast horizon k by minimizing the forecasting loss function, equation (2.3), with respect to $\Theta = \{\Theta^P, \Theta^Q, W\}$. By varying W, we construct the state variables as linear combinations of the observed term structure of yields, but they are not restricted to be the first three principal components of the observed yields.

We implement this estimation using a two-step procedure, taking full advantage of the estimation method proposed by JSZ, which typically converges in a few seconds. We start our estimation based on the forecasting loss function in equation (2.3) by using the converged JSZ estimates from the standard loss function in equation (2.1) as initial values. Given these initial Θ^P and Θ^Q, the estimation is performed using the following steps.

1. For a given Θ^P and Θ^Q, we search for the best possible weights W among the linear combinations of yields that provide the lowest squared forecasting error in equation (2.3).

2. Once we obtain a W in step 1, we fix it and solve for the parameter set Θ^P and Θ^Q by minimizing the squared forecasting error.

3. Once we obtain the converged Θ^P and Θ^Q from the previous step, we go back to the first step, and the optimization goes back and forth between the two steps until it converges.

Table 7 provides the empirical results. Panel A of Table 7 provides the RMSEs resulting from the JSZ canonical specification with forecasting loss function equation (2.3) and variable portfolio weights. Panel B presents RMSEs from the JSZ empirical implementation with fixed portfolio weights and the standard loss function equation (2.1). Panel B of Table 7 is therefore identical to Panel B of Table 6.

One might argue that the benchmark specification should also allow the portfolio weights to be free parameters. However, we know from JSZ that this is irrelevant under the standard loss function, since W_0 gives the optimal results for in-sample fit.[15] This suggests that allowing the portfolio weights to be free parameters under the standard loss function yields the same parameter estimates as the JSZ model with fixed weights, and therefore also the same out-of-sample performance. We verified that this is indeed the case.

---

[15]Hamilton and Wu (2014) also find that the first three pricncipal components lead to a better fit than any other linear combination of yields.

Panel C of Table 7 presents the ratio of the out-of-sample RMSEs. The improvements in forecasting RMSE are substantial for three-month to six-month forecast horizons. The improvement in the RMSEs is about 7% on average across different maturities for the six-month forecast horizon. This corresponds to an out-of-sample R-square of 15%. For short maturity yields (3-month, 6-month and 1-year yields), the forecasting loss function outperforms the standard loss function at all forecast horizons. The improvement in the RMSEs is about 11% on average, which corresponds to an out-of-sample R-square of 23% .

These results are different from the results in Table 6, which are based on fixed portfolio weights. This suggests that when using the JSZ canonical form, the time-series properties of the state variables are critically important to achieve better out-of-sample forecasting performance, which can be achieved using the forecasting loss function. It is imperative to free the portfolio weights to give the forecasting loss function more power to search for the best possible state variables for the purpose of out-of-sample forecasting. This contrasts with in-sample estimation, where fixing the portfolio weights is optimal, as demonstrated by JSZ.

Most importantly, we conclude that the results in Table 7 confirm the results from Table 2, obtained using the latent three-factor $A_0(3)$ model. Aligning the loss functions for in-sample estimation and out-of-sample evaluation allows us to determine the best possible state variables and model parameters for the purpose of out-of-sample forecasting.

6   In-Sample and Out-of-Sample Fit

We find that out-of-sample forecasting can be substantially improved by aligning the loss functions for in-sample and out-of-sample evaluation, as suggested by Granger (1993) and Weiss (1996). Presumably this finding results from differences in parameter estimates and implied state variables. In this section we document and discuss these differences. It is also to be expected that the parameter estimates based on the forecasting loss function give rise to an in-sample fit that is worse than that for the standard loss-function, because the latter loss function selects the parameters to provide the best possible in-sample fit. We document this trade-off between in-sample and out-of-sample fit. In this section, we illustrate these issues using the estimates for the JSZ canonical specification, because these estimates are arguably more reliable than the estimates obtained using the model with latent factors.[16]

---

[16]We find similar results for the $A_1(3)$, $A_2(3)$ and $A_3(3)$ stochastic volatility models with latent factors. Because of space constraints, we report these results in Table A1 and Figures A1-A3 in the Appendix.

6.1 In-Sample Fit

Table 8 reports the in-sample RMSEs for the JSZ model with forecasting loss function and variable weights, the JSZ model with forecasting loss function and fixed weights, and the JSZ model with standard loss function.[17] To be consistent with the out-of-sample experiment, we recursively estimate these specifications each month using data up to and including time t and compute the model error at time t. We compute the in-sample RMSE from the resulting time series. Recall that the resulting estimates for the two specifications with forecasting loss function are forecast-horizon specific. For these models, we therefore report RMSEs for each forecast horizon.

The results in Panels A and C of Table 8 indicate a clear trade-off between in-sample and out-of-sample fit. While the JSZ model with forecasting loss function and variable weights (in Panel A) provides a better in-sample fit than the model with standard loss function for short maturities, it provides a higher RMSE for medium and long maturities (in Panel C). Overall, the RMSEs in Panel C are on average smaller than those in Panel A. This result is of course not surprising, since the parameters for the JSZ model with forecasting loss function and variable weights are chosen to optimally fit yields k periods ahead. These results therefore simply reflect a trade-off between in-sample and out-of-sample fitting. Interestingly, the in-sample fit in Panel A is rather similar for different forecast horizons.

The in-sample RMSE for the JSZ model with forecasting loss function and fixed weights in Panel B of Table 8 is similar to that of the model with standard loss function in Panel C. The fit in Panel B is also similar across forecast horizons.

These findings are consistent with the out-of-sample results in Table 6. Both in-and out-of-sample, the JSZ model with forecasting loss function and fixed weights performs similarly to the model with standard loss function. When using variable portfolio weights and the forecasting loss function however, results strongly differ both in-and out-of-sample. Presumably these differences are due to differences in estimated parameters and implied state variables. We now investigate these differences in more detail.

6.2 Loss Functions and State Variables

We examine the time-series properties of the state variables for the models with standard and forecasting loss functions. Figure 1 is based on the JSZ with standard loss function. Panel A shows the time series of the first three principal components PC, level, slope and curvature.

---

[17]The JSZ model with standard loss function can also be implemented with fixed and variable weights. As mentioned before, the results are nearly identical, and we therefore only report results for fixed weights.

Panel B presents the factor loadings $B(\Theta^Q)$ on the yield curve. Panel C shows the portfolio weights W_0 that ensure $W_0 y_t = PC_t$. For the JSZ with standard loss function, we obtain the customary level, slope and curvature factors.

Figures 2-4 are based on the JSZ with forecasting loss function and variable weights. To emphasize the differences resulting from the use of different loss functions, we present the resulting differences between the state variables, factor loadings, and portfolio weights, rather than the levels. Because the estimation is forecast-horizon specific, each figure has six panels, one for each forecast horizon k.

Figure 2 shows the differences in the time series of the state variables, $Wy_t - PC_t$, where W is estimated using the forecasting loss function. Note that the magnitude of the third factor is on average smaller than that of the curvature factor in the JSZ with standard loss function, regardless of the forecast horizon. The magnitudes of the first two factors on average are larger than the level and slope factors in the JSZ with standard loss function, especially for longer forecast horizons.

Figure 3 plots the differences between the estimated factor loadings $B(\Theta^Q)$ from the JSZ with forecasting loss function and variable weights and the loadings from JSZ with standard loss function. For the first factor, the loadings are exactly the same for all forecast horizon estimations. For the second factor, the estimated factor loadings are very similar, except for long maturity yields for longer forecast horizons. The most pronounced differences are observed for the third factor. For all forecast horizons, the estimated loadings for the JSZ with forecasting loss function and variable weights are smaller than those for the JSZ with standard loss function for intermediate maturity yields, but larger for short- and long-maturity yields.

Figure 4 shows the differences in portfolio weights, $W - W_0$. The differences between the weights are similar across forecast horizons. The JSZ with forecasting loss function and variable weights implies a different linear combination of yields, and the resulting time series of the state variables differs from the traditional level, slope and curvature factors. Differences are especially pronounced for the third factor. We find that the third factor in the JSZ with forecasting loss function and variable weights is correlated with the fourth principal component of the yield curve. This result is in line with Cochrane and Piazzesi (2005), who find that the fourth principal component explains a large part of the bond return predictability, even though it explains only a small part of in-sample variability. The third factor in the JSZ with forecasting loss function and variable weights captures information that is hidden from the current yield curve, and this results in gains in out-of-sample forecasting performance.

6.3 Loss Functions and Parameter Estimates

We now compare the parameter estimates from the JSZ model with forecasting loss function and variable weights with those from the JSZ model with standard loss function. Table 9 presents the estimates of the parameters governing the state variables under the P- and Q-measures (K_0^P, K_1^P, K_0^Q, K_1^Q) for both specifications. Panel A of Table 9 reports the estimates for the JSZ model with forecasting loss function and variable portfolio weights, which are different for each forecast horizon k. In the JSZ model with standard loss function, K_0^P and K_1^P are the OLS estimates, as shown in Panel B of Table 9.

The most interesting observations are related to the dynamic properties of the model. Regardless of the model and the forecast horizon, under both measures the first factor is the most persistent and the third factor is the least persistent. To assess the persistence properties of the model, we need to inspect the eigenvalues rather than the diagonal elements of K_1. The eigenvalues are generally higher under the Q-measure than under the P-measure, in both Panels A and B. However, in Panel B the dominant eigenvalue under the Q-measure is equal to one, whereas under the P-measure it is slightly smaller than one. In Panel A it is slightly smaller than one under both the P- and Q-measures.

Another difference between Panels A and B is the $(1,3)$ entry of the feedback matrix, which governs how the third factor this period forecasts the first factor next period. The relative impact of the third factor on the first factor is higher in the model with forecasting loss function. A similar finding obtains for the $(2,3)$ entry of the feedback matrix.[18] These results are consistent with the results in Figure 2: the third factor behaves differently under the two loss functions.

Panel A of Table 10 reports the same parameters for the JSZ model with forecasting loss function and fixed weights. Panel B again reports the estimates from the JSZ with standard loss function. The differences between Panels A and B are much smaller than in Table 9, but once again the largest eigenvalue under the Q-measure in Panel B is one, in contrast with the estimate in Panel A.

We conclude that the analysis of the state variables and the parameter estimates confirms that the improvement in forecasting performance is driven by both the variable weights and the use of the forecasting loss function. The differences between Panels A and B are much more significant in Table 9, because the use of variable weights allows the forecasting loss function to play a more important role. The most important observation in Tables 9 and 10 is that the dominant eigenvalue under the Q-measure differs in a qualitative sense between Panels A and B.

---

[18]Joslin and Le (2013) discuss estimation of the feedback matrix in ATSMs with stochastic volatility. They show that the implicit restriction on the relation betwee K_1^P and K_1^Q causes the estimates of K_1^P to differ from the OLS estimates.

A plausible explanation for this finding is that when computing yields, the dominant eigenvalue can be larger than one under Q if it is not too high. For in-sample fitting, it may be worthwhile for the model to have a dominant eigenvalue of one to fit yields better. When the yield forecast is explicitly considered in the loss function, the loss function constrains the dominant eigenvalue to be smaller than one.

7   Conclusion

We propose estimating term structure parameters by aligning the loss functions for in-sample estimation and out-of-sample evaluation, instead of the traditional optimization of the likelihood criterion or the mean-squared error based on yields. We compare the resulting forecasting performance using three-factor affine term structure models. Aligning the loss functions provides substantial improvements in out-of-sample forecasting performance, especially for long forecast horizons. We document a trade-off between in-sample and out-of-sample fit.

Our approach amounts to letting the data determine the state variables that are best suited for out-of-sample forecasting. The resulting parameter estimates based on the JSZ canonical form imply factors that differ from the traditional level, slope, and curvature factors, especially for curvature. This suggests that the improvement in out-of-sample fit results from identification of the third factor, which captures information hidden from the current term structure of yields.

Our results may be extended in several ways. Most importantly, the question arises if our results generalize to other affine and non-affine term structure models. It may be challenging to address this issue because of the presence of identification problems. Using currently available estimation techniques, addressing these identification problems is harder than in the case of the affine model. Development of improved estimation methods for these models is therefore critically important.

Appendix A: Estimation of Models with Latent Factors: Numerical Implementation

We follow the implementation of Hamilton and Wu (2012). We extract the first three principal components from the observed term structure of yields, and normalize each principal component to have zero mean and unit variance. We estimate the dynamics of the normalized first three principal components through OLS, and use the OLS estimates of K_0^P and K_1^P as initial values.

We obtain the initial values for ρ_0 and ρ_1 by regressing one-month yields on the normalized principal components. To get the initial value for Q parameters, we regress the observed term structure of yields on the normalized principal components to get estimated loadings $\widehat{A}$ and $\widehat{B}$. The Q parameters enter the loadings $A(\Theta^Q)$ and $B(\Theta^Q)$ through recursive equation. Subsequently we obtain an initial guess for the Q parameters by minimizing the distances from $A(\Theta^Q)$ and $B(\Theta^Q)$ to the estimated loadings $\widehat{A}$ and $\widehat{B}$.

With this set of initial values, we find Θ by optimizing the standard log likelihood function using the fminsearch algorithm in MATLAB. We compute the 99% confidence interval, $[\Theta_-, \Theta^-]$, for the converged values of Θ. Then we generate another 100 different sets of Θ from the uniform distributions $U[\Theta_-, \Theta^-]$. We rank these different sets of Θ by the implied likelihood, and use the top 10 ranked sets of Θ as initial values for another round of numerical search. We choose the converged sets of Θ based on the likelihood, and form the new range of the parameter set using the chosen sets of Θ. We continue generating different sets of initial values until they converge to very similar values.

Appendix B: The JSZ Canonical Form

Given the dynamics in equations (4.1)-(4.3), the model-implied continuously compounded yields $\widehat{y}_t$ are given by

$$\widehat{y}_t = A(\Theta^Q) + B(\Theta^Q)PO_t.$$

Now consider M linear combinations of N yields, $PO_t = Wy_t$, that are priced without error. We focus on a simple case where the eigenvalues of K_1^Q are real, distinct, and nonzero. This follows Joslin, Singleton and Zhu (2011), who demonstrate the result for all cases including zero, repeated and complex eigenvalues.

There exists a matrix C such that $K_1^Q = Cdiag(\lambda^Q)C^{-1} + I_M$. Define $D = Cdiag(\rho_1)C^{-1}$, $D^{-1} = Cdiag(\rho_1)^{-1}C^{-1}$ and

$$
\begin{aligned}
L_t &= D(PO_t + (K_1^Q - I_M)^{-1}K_0^Q), \\
&\Rightarrow PO_t = D^{-1}L_t - (K_1^Q - I_M)^{-1}K_0^Q.
\end{aligned}
$$

Then we have the dynamic of L_t under the Q-measure

$$
\begin{aligned}
\Delta L_{t+1} &= D\Delta PO_{t+1} \\
&= D[K_0^Q + (K_1^Q - I_M)(D^{-1}L_t - (K_1^Q - I_M)^{-1}K_0^Q) + \Sigma\varepsilon_{t+1}^P] \\
&= diag(\lambda^Q)L_t + D\Sigma\varepsilon_{t+1}^P,
\end{aligned}
\tag{B.1}
$$

and the dynamic of L_t under the P-measure

$$
\begin{aligned}
\Delta L_{t+1} &= D\Delta PO_{t+1} \\
&= D[K_0^P + (K_1^P - I_M)(D^{-1}L_t - (K_1^Q - I_M)^{-1}K_0^Q) + \Sigma\varepsilon_{t+1}^P] \\
&= DK_0^P + D(K_1^P - I_M)D^{-1}L_t - D(K_1^P - I_M)(K_1^Q - I_M)^{-1}K_0^Q + D\Sigma\varepsilon_{t+1}^P,
\end{aligned}
\tag{B.2}
$$

The dynamic of the short rate is

$$
\begin{aligned}
r_t &= \rho_0 + \rho_1 PO_t \\
&= \rho_0 + \rho_1(D^{-1}L_t - (K_1^Q - I_M)^{-1}K_0^Q) \\
&= \rho_0 - \rho_1(K_1^Q - I_M)^{-1}K_0^Q + \rho_1 D^{-1}L_t \\
&= r_\infty^Q + \tau L_t,
\end{aligned}
\tag{B.3}
$$

where $r_\infty^Q = \rho_0 - \rho_1(K_1^Q - I_M)^{-1}K_0^Q$, and τ is a row of M ones. Given the dynamics in equations

(B.1)-(B.3), the model-implied continuously compounded yields $\widehat{y}_t$ is given by

$$\widehat{y}_t = A(\Theta_L^Q) + B(\Theta_L^Q)L_t,$$

where $\Theta_L^Q = \{r_\infty^Q, \lambda^Q, \Sigma\}$. The M linear combinations of N yields are perfectly priced and can be written as

$$
\begin{aligned}
PO_t &= Wy_t \\
&= W(A(\Theta_L^Q) + B(\Theta_L^Q)L_t)
\end{aligned}
$$

If the model is non-redundant, $WB(\Theta_L^Q)$ is invertible, and we have

$$L_t = (WB(\Theta_L^Q))^{-1}PO_t - (WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q).$$

Then we can rewrite the dynamic of PO_t under the Q-measure as follows

$$
\begin{aligned}
\Delta PO_{t+1} &= WB(\Theta_L^Q)\Delta L_{t+1} \\
&= WB(\Theta_L^Q)(diag(\lambda^Q)L_t + D\Sigma\varepsilon_{t+1}^P) \\
&= WB(\Theta_L^Q)\{diag(\lambda^Q)[(WB(\Theta_L^Q))^{-1}PO_t - (WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q)] + D\Sigma\varepsilon_{t+1}^P\}.
\end{aligned}
\tag{B.4}
$$

Comparing the coefficients in equations (B.4) and (4.2), we have

$$
\begin{aligned}
K_1^Q &= WB(\Theta_L^Q)diag(\lambda^Q)(WB(\Theta_L^Q))^{-1} + I_M, \\
K_0^Q &= -WB(\Theta_L^Q)diag(\lambda^Q)(WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q).
\end{aligned}
\tag{B.5}
$$

We can also rewrite the dynamic of the short rate as follows

$$
\begin{aligned}
r_t &= r_\infty^Q + \tau L_t \\
&= r_\infty^Q + \tau[(WB(\Theta_L^Q))^{-1}PO_t - (WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q)] \\
&= r_\infty^Q - \tau(WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q) + \tau(WB(\Theta_L^Q))^{-1}PO_t.
\end{aligned}
\tag{B.6}
$$

Comparing the coefficients in equations (B.6) and (4.3), we have

$$
\begin{aligned}
\rho_0 &= r_\infty^Q - \tau(WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q), \\
\rho_1 &= \tau(WB(\Theta_L^Q))^{-1}.
\end{aligned}
\tag{B.7}
$$

References

[1] Adrian, T., Crump, R. K., and Moench, E. (2013). Pricing the term structure with linear regressions. Journal of Financial Economics, 110(1), 110-138.

[2] Ang, A., and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics, 50(4), 745-787.

[3] Aït-Sahalia, Y., and Kimmel, R. L. (2010). Estimating affine multifactor term structure models using closed-form likelihood expansions. Journal of Financial Economics, 98(1), 113-144.

[4] Bauer, M. D., Rudebusch, G. D., and Wu, J. C. (2012). Correcting estimation bias in dynamic term structure models. Journal of Business and Economic Statistics, 30(3), 454-467.

[5] Bowsher, C. G., and Meeks, R. (2008). The dynamics of economic functions: modeling and forecasting the yield curve. Journal of the American Statistical Association, 103(484), 1419-1437.

[6] Chen, R. R., and Scott, L. (1992). Pricing interest rate options in a two-factor Cox-Ingersoll-Ross model of the term structure. Review of Financial Studies, 5(4), 613-636.

[7] Christensen, J. H., Diebold, F. X., and Rudebusch, G. D. (2011). The affine arbitrage-free class of Nelson–Siegel term structure models. Journal of Econometrics, 164(1), 4-20.

[8] Christoffersen, P. F., and Diebold, F. X. (1996). Further results on forecasting and model selection under asymmetric loss. Journal of Applied Econometrics, 11(5), 561-571.

[9] Christoffersen, P. F., and Diebold, F. X. (1997). Optimal prediction under asymmetric loss. Econometric Theory, 13(6), 808-817.

[10] Christoffersen, P., Dorion, C., Jacobs, K., and Karoui, L. (2014). Nonlinear Kalman filtering in affine term structure models. Management Science, 60(9), 2248-2268.

[11] Christoffersen, P., and Jacobs, K. (2004). The importance of the loss function in option valuation. Journal of Financial Economics, 72(2), 291-318.

[12] Cieslak, A., and Povala, P. (2015). Expected returns in Treasury bonds. Review of Financial Studies.

[13] Cochrane, J. H., and Piazzesi, M. (2005). Bond risk premia. American Economic Review, 95(1), 138-160.

[14] Cooper, I., and Priestley, R. (2009). Time-varying risk premiums and the output gap. Review of Financial Studies, 22(7), 2801-2833.

[15] Collin-Dufresne, P., Goldstein, R. S., and Jones, C. S. (2008). Identification of maximal affine term structure models. The Journal of Finance, 63(2), 743-795.

[16] Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. Econometrica, 53(2), 385-408.

[17] Creal, D. D., and Wu, J. C. (2015). Estimation of affine term structure models with spanned or unspanned stochastic volatility. Journal of Econometrics, 185(1), 60-81.

[18] Dai, Q., and Singleton, K. J. (2000). Specification analysis of affine term structure models. The Journal of Finance, 55(5), 1943-1978.

[19] Diez de Los Rios, A. (2014). A new linear estimator for Gaussian dynamic term structure models. Journal of Business and Economic Statistics, 33(2), 282-295.

[20] Diebold, F. X., and Li, C. (2006). Forecasting the term structure of government bond yields. Journal of Econometrics, 130(2), 337-364.

[21] Diebold, F. X., and Mariano, R. (1995). Comparing predictive accuracy. Journal of Business and Economic Statistics 13(3), 253-63.

[22] Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. The Journal of Finance, 57(1), 405-443.

[23] Duffee, G. R. (2011a). Information in (and not in) the term structure. Review of Financial Studies, 24(9), 2895-2934.

[24] Duffee, G. R. (2011b). Forecasting with the term structure: The role of no-arbitrage restrictions. Working Paper. Johns Hopkins University.

[25] Duffee, G. R., and Stanton, R. H. (2012). Estimation of dynamic term structure models. The Quarterly Journal of Finance, 2(2).

[26] Duffie, D., and Kan, R. (1996). A yield-factor model of interest rates. Mathematical Finance, 6(4), 379-406.

[27] Dumas, B., Fleming, J., and Whaley, R. E. (1998). Implied volatility functions: Empirical tests. The Journal of Finance, 53(6), 2059-2106.

[28] Egorov, A. V., Hong, Y., and Li, H. (2006). Validating forecasts of the joint probability density of bond yields: Can affine models beat random walk? Journal of Econometrics, 135(1), 255-284.

[29] Elliott, G., Komunjer, I., and Timmermann, A. (2005). Estimation and testing of forecast rationality under flexible loss. The Review of Economic Studies, 72(4), 1107-1125.

[30] Elliott, G., Komunjer, I., and Timmermann, A. (2008). Biases in macroeconomic forecasts: irrationality or asymmetric loss? Journal of the European Economic Association, 6(1), 122-157.

[31] Elliott, G., and Timmermann, A. (2008). Economic forecasting. Journal of Economic Literature, 46(1), 3-56.

[32] Engle, R. (1993). A comment on Hendry and Clements on the limitations of comparing mean square forecast errors. Journal of Forecasting 12(8), 642-644.

[33] Granger, C. W. (1993). On the limitations of comparing mean square forecast errors: Comment. Journal of Forecasting, 12(8), 651-652.

[34] Gourieroux, C., Monfort, A., and Polimenis, V. (2006). Affine models for credit risk analysis. Journal of Financial Econometrics, 4(3), 494-530.

[35] Hamilton, J. D., and Wu, J. C. (2012). Identification and estimation of Gaussian affine term structure models. Journal of Econometrics, 168(2), 315-331.

[36] Hamilton, J. D., and Wu, J. C. (2014). Testable implications of affine term structure models. Journal of Econometrics, 178, 231-242.

[37] Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. International Journal of Forecasting, 13(2), 281-291.

[38] Hong, Y., Li, H., and Zhao, F. (2004). Out-of-sample performance of discrete-time spot interest rate models. Journal of Business & Economic Statistics, 22(4), 457-473.

[39] Joslin, S., and Le, A. (2013). Interest rate volatility and no-arbitrage affine term structure models. Working Paper. University of Southern California.

[40] Joslin, S., Priebsch, M., and Singleton, K. J. (2014). Risk premiums in dynamic term structure models with unspanned macro risks. The Journal of Finance, 69(3), 1197-1233.

[41] Joslin, S., Singleton, K. J., and Zhu, H. (2011). A new perspective on Gaussian dynamic term structure models. Review of Financial Studies, 24(3), 926-970.

[42] Knez, P. J., Litterman, R., and Scheinkman, J. (1994). Explorations into factors explaining money market returns. The Journal of Finance, 49(5), 1861-1882.

[43] Le, A., Singleton, K. J., and Dai, Q. (2010). Discrete-time affine term structure models with generalized market prices of risk. Review of Financial Studies, 23(5), 2184-2227.

[44] Litterman, R. B., and Scheinkman, J. (1991). Common factors affecting bond returns. The Journal of Fixed Income, 1(1), 54-61.

[45] Longstaff, F. A., and Schwartz, E. S. (1992). Interest rate volatility and the term structure: A two-factor general equilibrium model. The Journal of Finance, 47(4), 1259-1282.

[46] Ludvigson, S. C., and Ng, S. (2009). Macro factors in bond risk premia. Review of Financial Studies, 22(12), 5027-5067.

[47] Patton, A. J., and Timmermann, A. (2007a). Properties of optimal forecasts under asymmetric loss and nonlinearity. Journal of Econometrics, 140(2), 884-918.

[48] Patton, A. J., and Timmermann, A. (2007b). Testing forecast optimality under unknown loss. Journal of the American Statistical Association, 102(480), 1172-1184.

[49] Sarno, L., Schneider, P., and Wagner, C. (2014). The economic value of predicting bond risk premia: Can anything beat the expectations hypothesis? Working Paper. University of Lugano.

[50] Shin, M., and Zhong, M. (2013). Does realized volatility help bond yield density prediction? Working Paper. University of Pennsylvania.

[51] Singleton, Kenneth J. (2006). Empirical dynamic asset pricing. Princeton University Press.

[52] Vasicek, O. (1977). An equilibrium characterization of the term structure. Journal of Financial Economics, 5(2), 177-188.

[53] Weiss, A. A. (1996). Estimating time series models using the relevant cost function. Journal of Applied Econometrics, 11(5), 539-560.

Figure 1: State Variables for the JSZ Canonical Form with Standard Loss Function.
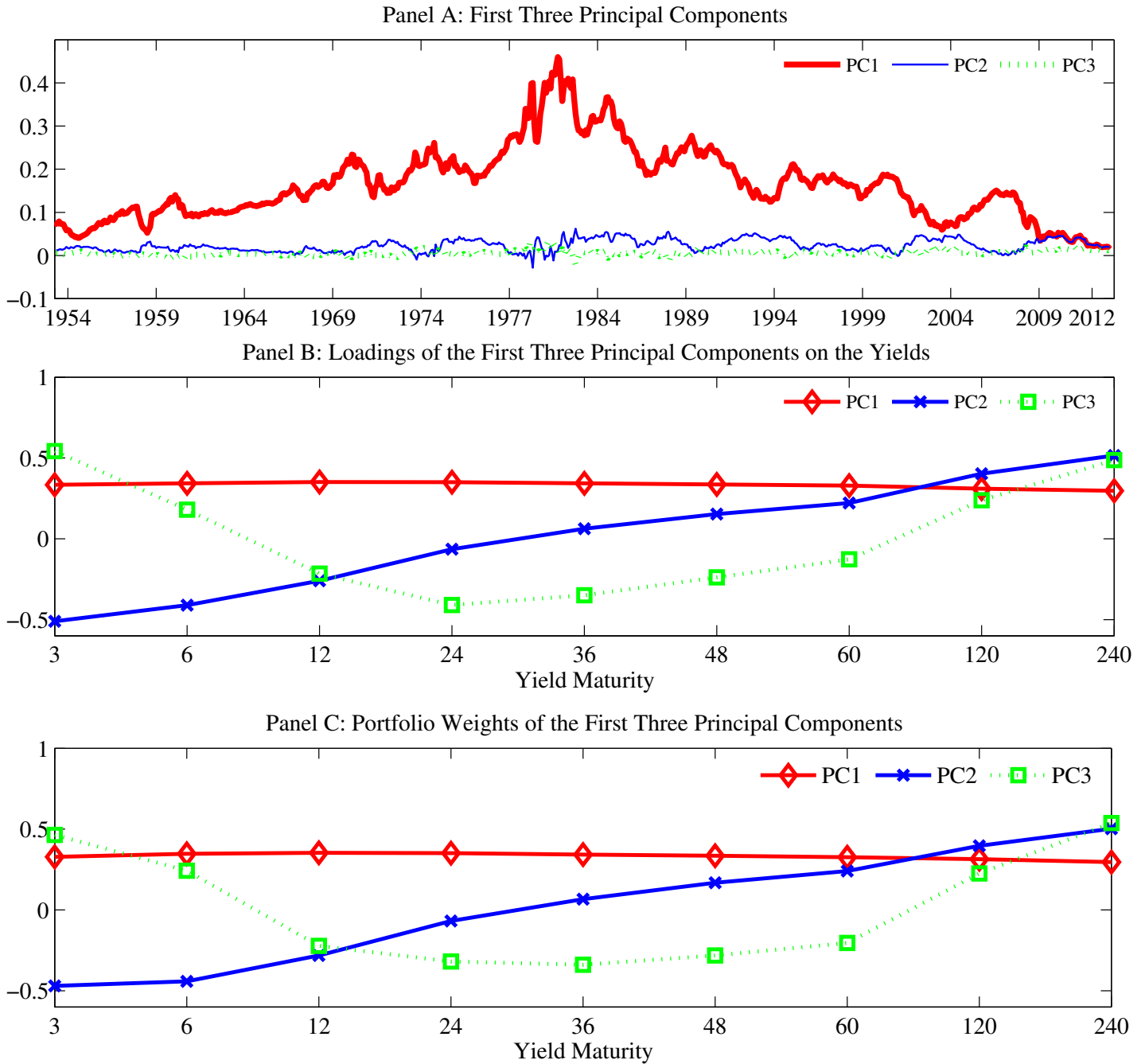
Notes to Figure: This figure is based on the in-sample estimation of the JSZ canonical form with standard loss function. Panel A shows the time series of the first three principal components PC. Panel B shows the factor loadings $B(\Theta^Q)$ on the yield curve. Panel C shows the portfolio weights W_0.

Figure 2: Differences Between State Variables Using Forecasting and Standard Loss Functions.
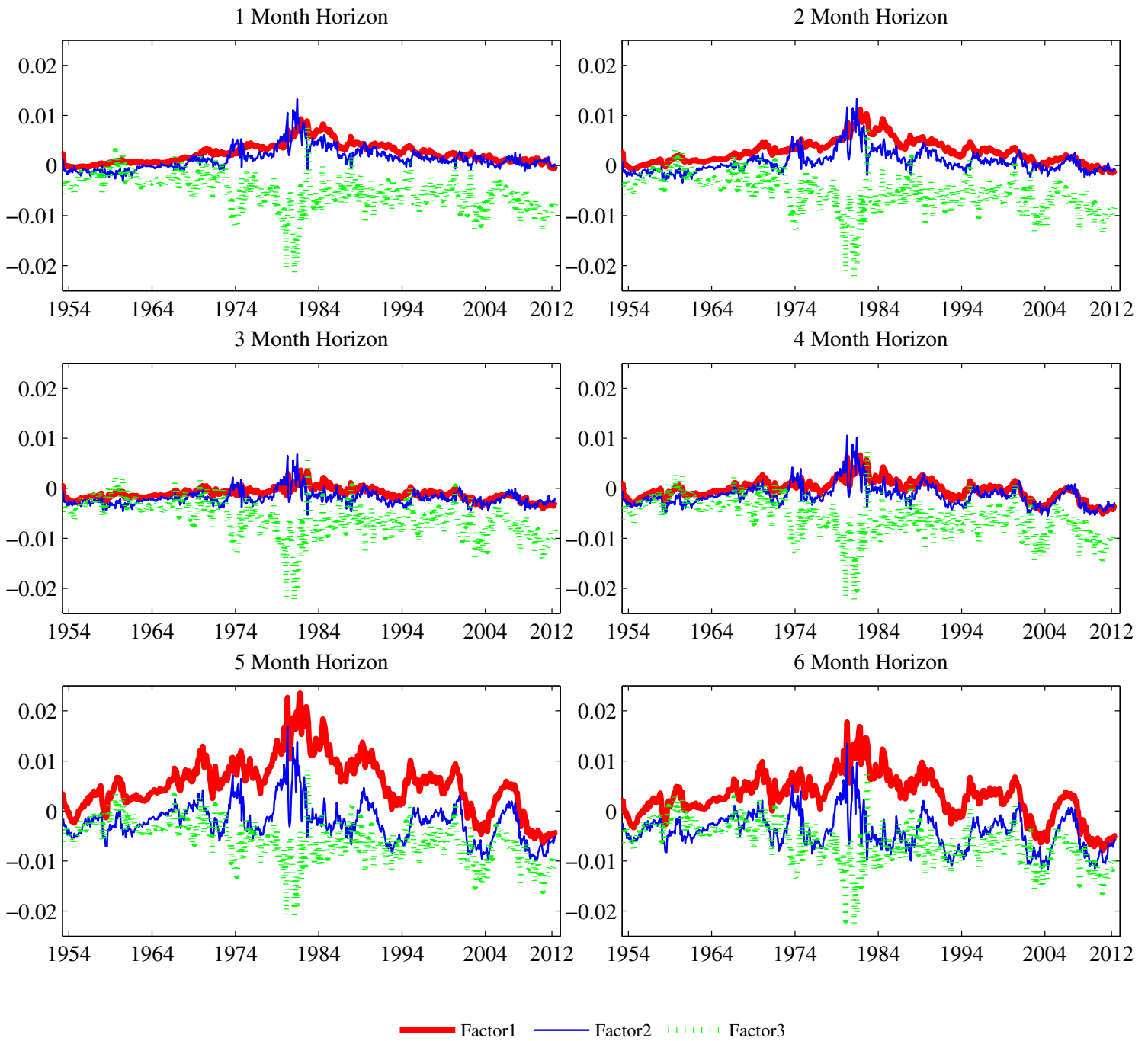
Notes to Figure: This figure presents the differences between the state variables from the JSZ canonical form with forecasting loss function and variable portfolio weights and the state variables from the JSZ canonical form with standard loss function. The state variables are obtained from in-sample estimation. These results are forecast-horizon specific, i.e. each panel represents the differences in the estimates for a given forecast horizon.

Figure 3: Differences Between Factor Loadings Using Forecasting and Standard Loss Functions.
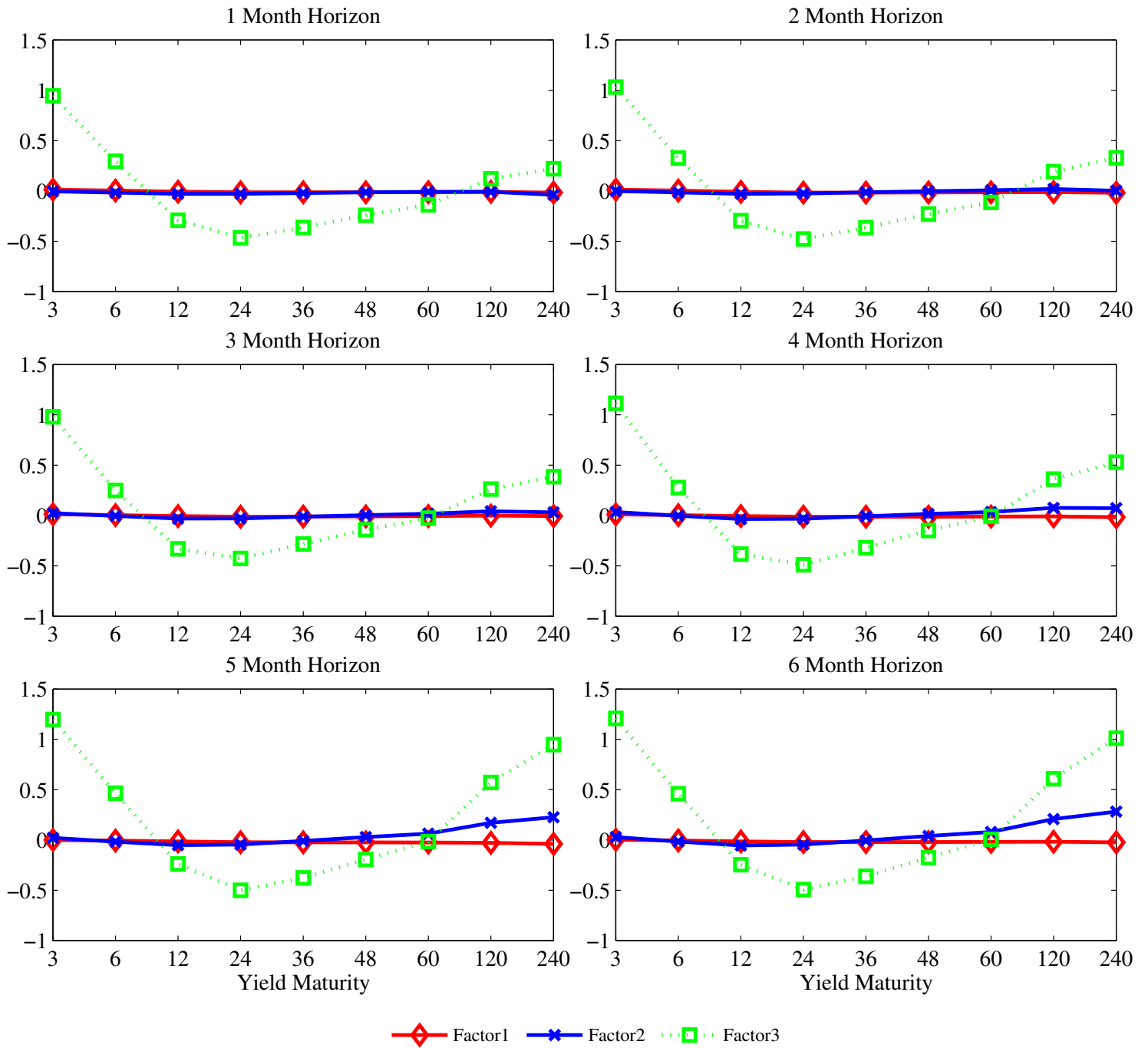
Notes to Figure: This figure presents the differences between the factor loadings from the JSZ canonical form with forecasting loss function and variable portfolio weights and the factor loadings from the JSZ canonical form with standard loss function. The factor loadings are obtained from in-sample estimation. These results are forecast-horizon specific, i.e. each panel represents the differences in the estimates for a given forecast horizon.

Figure 4: Differences Between Portfolio Weights Using Forecasting and Standard Loss Functions.
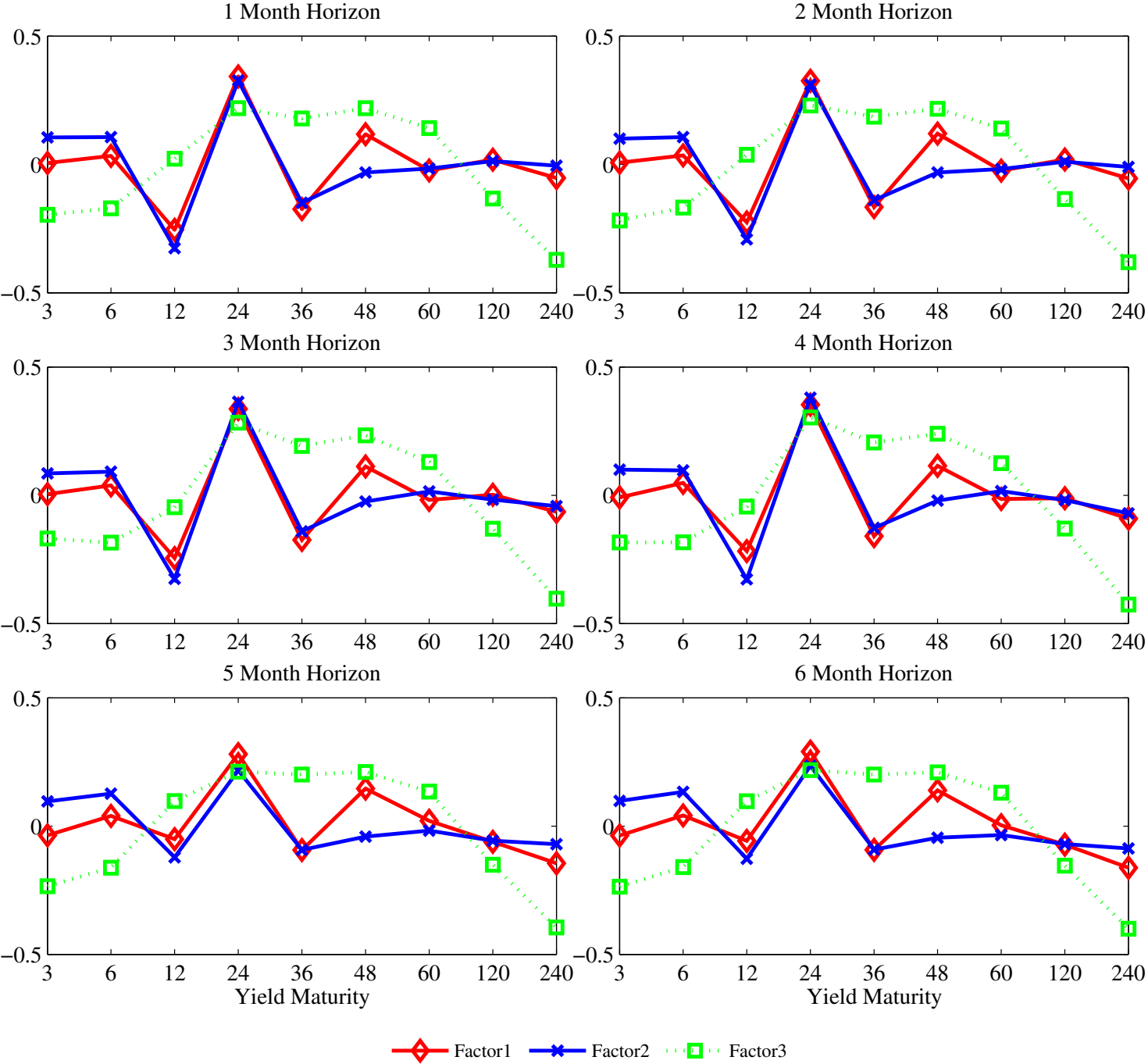
Notes to Figure: This figure presents the differences between the variable portfolio weights from the JSZ canonical form with forecasting loss function and the fixed portfolio weights from the JSZ canonical form with standard loss function. The portfolio weights are obtained from in-sample estimation. These results are forecast-horizon specific, i.e. each panel represents the differences in the estimates for a given forecast horizon.

Table 1: Summary Statistics

	Central Moments				Autocorrelation		
	Mean	St.Dev	Skewness	Kurtosis	Lag 1	Lag 12	Lag 30
3 month yield	0.0450	0.0290	0.8938	4.3247	0.9773	0.7944	0.5197
6 month yield	0.0479	0.0305	0.8717	4.2283	0.9850	0.8126	0.5359
1 year yield	0.0516	0.0306	0.6980	3.6594	0.9857	0.8317	0.5891
2 year yield	0.0536	0.0301	0.6734	3.4957	0.9878	0.8509	0.6485
3 year yield	0.0554	0.0294	0.6703	3.4460	0.9884	0.8611	0.6782
4 year yield	0.0569	0.0288	0.6903	3.4142	0.9882	0.8655	0.7000
5 year yield	0.0579	0.0282	0.7270	3.3717	0.9890	0.8739	0.7183
10 year yield	0.0617	0.0275	0.9148	3.5853	0.9890	0.8739	0.7183
20 year yield	0.0638	0.0265	0.9158	3.5373	0.9930	0.8936	0.7724

Notes to Table: We present summary statistics for the data used in estimation. We present the sample mean, standard deviation, skewness, kurtosis, and autocorrelations for each of the yields. The yields are continuously compounded monthly zero-coupon bond yields. The sample period is from 1953:04 to 2012:12.

Table 2: Out-of-Sample RMSEs:
$A_0(3)$ with Latent Factors

Panel A: Forecasting Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	38.47	55.53	68.99	78.78	88.97	99.65
6 month yield	33.51	53.37	69.85	83.52	96.27	108.70
1 year yield	37.84	59.81	78.28	92.96	105.82	117.62
2 year yield	40.41	63.36	81.95	95.63	107.89	118.75
3 year yield	37.65	59.22	76.23	88.29	99.70	109.44
4 year yield	33.77	54.00	69.40	79.79	90.30	99.13
5 year yield	30.76	50.19	66.65	78.61	87.65	96.97
10 year yield	32.10	49.40	63.15	76.06	82.02	91.09
20 year yield	28.54	44.51	52.71	61.85	70.89	80.07

Panel B: Standard Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	42.22	57.51	73.06	87.73	103.87	120.70
6 month yield	36.79	55.87	70.27	85.84	101.77	119.88
1 year yield	38.75	60.28	80.46	95.58	108.62	121.45
2 year yield	41.07	65.36	82.99	96.92	110.11	127.41
3 year yield	38.60	59.48	76.86	92.61	109.65	125.54
4 year yield	35.22	58.49	72.90	88.85	104.24	118.98
5 year yield	33.92	52.85	70.46	85.83	100.68	114.68
10 year yield	33.24	50.40	63.15	77.06	89.02	101.09
20 year yield	28.44	44.51	58.71	71.85	84.89	97.07

Panel C: RMSE Ratio

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	0.91	0.97***	0.94***	0.90***	0.86***	0.83***
6 month yield	0.91	0.96***	0.99***	0.97***	0.95***	0.91***
1 year yield	0.98**	0.99***	0.97***	0.97***	0.97***	0.97***
2 year yield	0.98***	0.97***	0.99***	0.99***	0.98***	0.93***
3 year yield	0.98**	1.00***	0.99***	0.95***	0.91***	0.87***
4 year yield	0.96**	0.92***	0.95***	0.90***	0.87***	0.83***
5 year yield	0.91*	0.95***	0.95***	0.92***	0.87***	0.85***
10 year yield	0.97*	0.98***	1.00*	0.99***	0.92***	0.90***
20 year yield	1.00	1.00	0.90	0.86**	0.84**	0.82**

Notes to Table: We present the out-of-sample RMSEs for the $A_0(3)$ latent model with forecasting loss function and standard loss function. Panel A reports on the $A_0(3)$ model with forecasting loss function. At each time t and for each forecast horizon k, we estimate the specification through Kalman filter using data up to and including t, and forecast k periods ahead. Panel B reports on the $A_0(3)$ model with standard loss function. We estimate the model through Kalman filter using data up to and including t at each month t, and forecast one to six months ahead. We show the root mean square error (RMSE) of the forecast in basis points. In order to compare the forecasting performance of the latent model with two different loss functions, Panel C shows the ratios of the RMSEs in Panel A and Panel B. Gains in accuracy that are statistically different from zero are denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively, evaluated using the Diebold and Mariano (1995) t-statistics computed with a serial correlation-robust variance and the small-sample adjustment of Harvey, Leybourne, and Newbold (1997).

Table 3: Out-of-Sample RMSEs:
$A_1(3)$ with Latent Factors

Panel A: Forecasting Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	36.56	52.06	66.95	75.75	87.05	101.19
6 month yield	31.03	50.65	67.46	81.67	96.15	110.10
1 year yield	39.22	57.81	74.41	89.63	103.40	117.31
2 year yield	51.69	67.95	82.70	95.93	108.15	120.46
3 year yield	48.14	63.94	77.90	90.20	101.68	113.09
4 year yield	40.40	57.04	70.82	82.48	93.25	103.95
5 year yield	33.83	51.81	66.21	78.24	89.15	99.62
10 year yield	35.45	50.28	62.03	72.56	82.30	91.12
20 year yield	48.95	56.36	62.94	69.82	76.66	82.68

Panel B: Standard Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	39.87	57.78	73.87	85.81	99.21	113.15
6 month yield	34.58	55.15	74.23	90.77	106.90	122.76
1 year yield	38.47	61.77	83.00	101.43	118.47	134.27
2 year yield	41.62	67.91	90.03	109.03	125.89	141.42
3 year yield	40.00	65.47	86.63	104.83	120.89	135.61
4 year yield	36.69	60.82	80.83	97.93	113.12	127.15
5 year yield	34.80	58.21	77.61	94.37	109.23	122.87
10 year yield	37.20	57.48	74.02	89.00	102.47	114.87
20 year yield	28.23	45.76	60.71	74.48	87.01	98.73

Panel C: RMSE Ratio

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	0.92*	0.90***	0.91***	0.88***	0.88***	0.89***
6 month yield	0.90*	0.92***	0.91***	0.90***	0.90***	0.90***
1 year yield	1.02	0.94**	0.90***	0.88***	0.87***	0.87***
2 year yield	1.24***	1.00	0.92***	0.88***	0.86***	0.85***
3 year yield	1.20***	0.98	0.90***	0.86***	0.84***	0.83***
4 year yield	1.10***	0.94***	0.88***	0.84***	0.82***	0.82***
5 year yield	0.97	0.89***	0.85***	0.83***	0.82***	0.81***
10 year yield	0.95	0.87***	0.84***	0.82***	0.80***	0.79***
20 year yield	1.73***	1.23***	1.04	0.94	0.88***	0.84***

Notes to Table: We present the out-of-sample RMSEs for the $A_1(3)$ latent model with forecasting loss function and standard loss function. Panel A reports on the $A_1(3)$ model with forecasting loss function. At each time t and for each forecast horizon k, we estimate the specification through Kalman filter using data up to and including t, and forecast k periods ahead. Panel B reports on the $A_1(3)$ model with standard loss function. We estimate the model through Kalman filter using data up to and including t at each month t, and forecast one to six months ahead. We show the root mean square error (RMSE) of the forecast in basis points. In order to compare the forecasting performance of the latent model with two different loss functions, Panel C shows the ratios of the RMSEs in Panel A and Panel B. Gains in accuracy that are statistically different from zero are denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively, evaluated using the Diebold and Mariano (1995) t-statistics computed with a serial correlation-robust variance and the small-sample adjustment of Harvey, Leybourne, and Newbold (1997).

Table 4: Out-of-Sample RMSEs:
$A_2(3)$ with Latent Factors

Panel A: Forecasting Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	36.90	53.41	67.91	77.63	89.84	104.33
6 month yield	31.92	51.56	69.97	84.11	98.56	113.93
1 year yield	39.77	58.66	75.73	91.75	106.41	120.81
2 year yield	52.31	68.94	84.22	98.16	111.08	124.12
3 year yield	48.68	64.95	79.45	92.41	104.51	116.71
4 year yield	40.79	57.92	72.19	84.53	95.91	107.54
5 year yield	34.19	52.72	67.57	80.30	91.83	103.19
10 year yield	35.66	50.85	63.13	74.42	84.97	94.72
20 year yield	49.30	56.89	63.07	70.74	78.34	85.21

Panel B: Standard Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	40.24	58.80	74.93	87.55	100.86	115.43
6 month yield	35.09	56.10	75.74	93.03	109.91	126.48
1 year yield	38.81	62.69	84.71	104.04	121.97	138.72
2 year yield	42.25	69.32	92.32	112.17	130.04	146.67
3 year yield	40.67	66.97	88.95	107.90	124.94	140.78
4 year yield	37.29	62.27	83.05	100.77	116.84	132.08
5 year yield	35.52	59.77	79.93	97.26	112.93	127.73
10 year yield	38.17	59.30	76.42	91.84	106.27	119.99
20 year yield	29.21	47.75	63.12	77.33	90.76	103.78

Panel C: RMSE Ratio

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	0.92	0.91***	0.91***	0.89***	0.89***	0.90***
6 month yield	0.91**	0.92***	0.92***	0.90***	0.90***	0.90***
1 year yield	1.02	0.94**	0.89***	0.88***	0.87***	0.87***
2 year yield	1.24***	0.99	0.91***	0.88***	0.85***	0.85***
3 year yield	1.20***	0.97*	0.89***	0.86***	0.84***	0.83***
4 year yield	1.09**	0.93***	0.87***	0.84***	0.82***	0.81***
5 year yield	0.96*	0.88***	0.85***	0.83***	0.81***	0.81***
10 year yield	0.93	0.86***	0.83***	0.81***	0.80***	0.79***
20 year yield	1.69***	1.19**	1.00	0.91*	0.86***	0.82***

Notes to Table: We present the out-of-sample RMSEs for the $A_2(3)$ latent model with forecasting loss function and standard loss function. Panel A reports on the $A_2(3)$ model with forecasting loss function. At each time t and for each forecast horizon k, we estimate the specification through Kalman filter using data up to and including t, and forecast k periods ahead. Panel B reports on the $A_2(3)$ model with standard loss function. We estimate the model through Kalman filter using data up to and including t at each month t, and forecast one to six months ahead. We show the root mean square error (RMSE) of the forecast in basis points. In order to compare the forecasting performance of the latent model with two different loss functions, Panel C shows the ratios of the RMSEs in Panel A and Panel B. Gains in accuracy that are statistically different from zero are denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively, evaluated using the Diebold and Mariano (1995) t-statistics computed with a serial correlation-robust variance and the small-sample adjustment of Harvey, Leybourne, and Newbold (1997).

Table 5: Out-of-Sample RMSEs:
$A_3(3)$ with Latent Factors

Panel A: Forecasting Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	37.33	55.82	70.73	78.92	92.61	105.56
6 month yield	33.03	53.25	71.10	85.70	101.64	117.49
1 year yield	40.01	59.22	76.68	93.14	108.27	123.17
2 year yield	52.57	69.53	85.18	99.55	112.93	126.46
3 year yield	48.94	65.52	80.40	93.77	106.33	119.03
4 year yield	41.01	58.43	73.07	85.82	97.66	109.79
5 year yield	34.44	53.26	68.47	81.61	93.60	105.46
10 year yield	35.82	51.27	63.90	75.59	86.60	96.85
20 year yield	50.87	57.08	63.30	71.33	79.36	86.74

Panel B: Standard Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	41.02	60.26	75.38	88.61	102.80	117.10
6 month yield	35.14	56.43	76.44	94.20	111.31	128.08
1 year yield	38.94	63.13	85.38	105.11	123.33	140.35
2 year yield	42.38	69.68	92.98	113.20	131.49	148.44
3 year yield	40.81	67.28	89.61	108.94	126.45	142.62
4 year yield	37.42	62.53	83.66	101.80	118.39	133.86
5 year yield	35.65	60.01	80.55	98.27	114.53	129.57
10 year yield	38.22	59.52	77.05	92.94	107.76	121.49
20 year yield	30.02	48.48	63.59	78.25	92.11	105.06

Panel C: RMSE Ratio

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	0.91	0.93***	0.94***	0.89***	0.90***	0.90***
6 month yield	0.94**	0.94**	0.93***	0.91***	0.91***	0.92***
1 year yield	1.03	0.94*	0.90***	0.89***	0.88***	0.88***
2 year yield	1.24***	1.00	0.92***	0.88***	0.86***	0.85***
3 year yield	1.20***	0.97	0.90***	0.86***	0.84***	0.83***
4 year yield	1.10***	0.93***	0.87***	0.84***	0.82***	0.82***
5 year yield	0.97	0.89***	0.85***	0.83***	0.82***	0.81***
10 year yield	0.94	0.86***	0.83***	0.81***	0.80***	0.80***
20 year yield	1.69***	1.18**	1.00	0.91*	0.86***	0.83***

Notes to Table: We present the out-of-sample RMSEs for the $A_3(3)$ latent model with forecasting loss function and standard loss function. Panel A reports on the $A_3(3)$ model with forecasting loss function. At each time t and for each forecast horizon k, we estimate the specification through Kalman filter using data up to and including t, and forecast k periods ahead. Panel B reports on the $A_3(3)$ model with standard loss function. We estimate the model through Kalman filter using data up to and including t at each month t, and forecast one to six months ahead. We show the root mean square error (RMSE) of the forecast in basis points. In order to compare the forecasting performance of the latent model with two different loss functions, Panel C shows the ratios of the RMSEs in Panel A and Panel B. Gains in accuracy that are statistically different from zero are denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively, evaluated using the Diebold and Mariano (1995) t-statistics computed with a serial correlation-robust variance and the small-sample adjustment of Harvey, Leybourne, and Newbold (1997).

Table 6: Out-of-Sample RMSEs:
JSZ Canonical Form with Fixed Portfolio Weights

Panel A: Forecasting Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	38.18	55.72	67.68	76.69	87.64	99.06
6 month yield	33.81	51.29	66.91	80.26	93.67	107.02
1 year yield	39.39	59.75	77.92	94.37	108.58	120.94
2 year yield	39.63	61.92	79.63	95.15	108.23	119.37
3 year yield	38.06	59.55	76.65	90.79	102.96	113.00
4 year yield	35.30	55.76	72.01	84.46	95.52	104.93
5 year yield	32.18	51.99	67.80	79.46	90.02	98.93
10 year yield	33.27	49.66	61.71	70.58	79.52	87.04
20 year yield	26.36	40.81	51.30	59.60	67.23	73.44

Panel B: Standard Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	38.11	55.06	69.89	80.61	91.84	103.21
6 month yield	33.68	52.32	69.43	83.91	97.71	110.97
1 year yield	38.68	60.60	79.92	96.19	110.51	123.39
2 year yield	39.13	62.34	81.09	96.46	109.60	121.24
3 year yield	37.45	59.98	77.79	92.29	104.59	115.25
4 year yield	34.71	56.23	73.01	86.40	97.79	107.71
5 year yield	31.62	52.50	68.68	81.67	92.68	102.08
10 year yield	33.02	49.93	62.56	73.33	82.63	90.45
20 year yield	26.24	40.85	51.63	60.85	68.80	75.40

Panel C: RMSE Ratio

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	1.00	1.01	0.97	0.95**	0.95**	0.96***
6 month yield	1.00	0.98**	0.96*	0.96**	0.96***	0.96***
1 year yield	1.02	0.99	0.97	0.98**	0.98***	0.98***
2 year yield	1.01	0.99	0.98*	0.99***	0.99***	0.98***
3 year yield	1.02*	0.99	0.99**	0.98***	0.98***	0.98***
4 year yield	1.02	0.99	0.99**	0.98***	0.98***	0.97***
5 year yield	1.02	0.99*	0.99**	0.97***	0.97***	0.97***
10 year yield	1.01	0.99**	0.99**	0.96***	0.96***	0.96***
20 year yield	1.00*	1.00	0.99	0.98	0.98	0.97

Notes to Table: We present the out-of-sample RMSEs for the JSZ canonical form with forecasting loss function and standard loss function. Panel A reports on the JSZ canonical form with forecasting loss function. At each time t and for each forecast horizon k, we estimate the specification using data up to and including t, and forecast k periods ahead. Panel B reports on the JSZ canonical form with standard loss function. We estimate the specification using data up to and including t at each month t, and forecast one to six months ahead. We show the root mean square error (RMSE) of the forecast in basis points. In order to compare the forecasting performance of the JSZ canonical form with two different loss functions, Panel C shows the ratios of the RMSEs in Panel A and Panel B. Gains in accuracy that are statistically different from zero are denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively, evaluated using the Diebold and Mariano (1995) t-statistics computed with a serial correlation-robust variance and the small-sample adjustment of Harvey, Leybourne, and Newbold (1997).

Table 7: Out-of-Sample RMSEs:
JSZ Canonical Form with Variable Portfolio Weights

Panel A: Forecasting Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	36.60	49.69	60.41	69.09	79.40	90.58
6 month yield	29.97	45.55	59.59	73.09	86.01	98.95
1 year yield	35.06	54.58	69.97	84.23	97.20	109.72
2 year yield	40.70	61.01	76.67	89.55	101.50	113.23
3 year yield	39.19	57.80	73.04	84.54	95.55	106.31
4 year yield	35.54	53.10	67.71	77.95	87.92	97.94
5 year yield	32.16	49.41	63.80	73.82	83.81	93.32
10 year yield	39.47	50.97	60.34	69.03	77.73	84.47
20 year yield	41.32	50.58	58.34	64.61	73.16	80.14

Panel B: Standard Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	38.11	55.06	69.89	80.61	91.84	103.21
6 month yield	33.68	52.32	69.43	83.91	97.71	110.97
1 year yield	38.68	60.60	79.92	96.19	110.51	123.39
2 year yield	39.13	62.34	81.09	96.46	109.60	121.24
3 year yield	37.45	59.98	77.79	92.29	104.59	115.25
4 year yield	34.71	56.23	73.01	86.40	97.79	107.71
5 year yield	31.62	52.50	68.68	81.67	92.68	102.08
10 year yield	33.02	49.93	62.56	73.33	82.63	90.45
20 year yield	26.24	40.85	51.63	60.85	68.80	75.40

Panel C: RMSE Ratio

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	0.96	0.90	0.86*	0.86***	0.86***	0.88***
6 month yield	0.89	0.87***	0.86**	0.87***	0.88***	0.89***
1 year yield	0.91	0.90**	0.88**	0.88***	0.88***	0.89***
2 year yield	1.04	0.98	0.95**	0.93***	0.93***	0.93***
3 year yield	1.05**	0.96	0.94***	0.92***	0.91***	0.92***
4 year yield	1.02*	0.94*	0.93***	0.90***	0.90***	0.91***
5 year yield	1.02	0.94**	0.93***	0.90***	0.90***	0.91***
10 year yield	1.20	1.02***	0.96***	0.94***	0.94***	0.93***
20 year yield	1.57**	1.24	1.13	1.06	1.06	1.06

Notes to Table: We present the out-of-sample RMSEs for the JSZ canonical form with forecasting loss function and variable portfolio weights, and the JSZ canonical form with standard loss function and fixed portfolio weights. Panel A reports on the JSZ canonical form with forecasting loss function and variable portfolio weights. At each time t and for each forecast horizon k, we estimate the specification using data up to and including t, and forecast k periods ahead. Panel B reports on the JSZ canonical form with standard loss function and fixed portfolio weights. We estimate the specification using data up to and including t at each month t, and forecast one to six months ahead. We show the root mean square error (RMSE) of the forecast in basis points. In order to compare the forecast performance between the two specifications, Panel C shows the ratios of the RMSEs in Panel A and Panel B. Gains in accuracy that are statistically different from zero are denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively, evaluated using the Diebold and Mariano (1995) t-statistics computed with a serial correlation-robust variance and the small-sample adjustment of Harvey, Leybourne, and Newbold (1997).

Table 8: In-Sample RMSEs: JSZ Canonical Form

Panel A: Forecasting Loss Function with Variable Portfolio Weights

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	10.10	10.15	10.65	10.80	10.84	10.81
6 month yield	13.43	13.93	14.84	13.37	13.64	14.89
1 year yield	12.88	13.35	12.42	13.02	14.33	14.60
2 year yield	15.34	16.02	17.07	17.62	16.92	19.12
3 year yield	16.95	12.63	15.54	17.72	19.49	21.27
4 year yield	17.95	15.41	16.51	16.82	17.68	17.53
5 year yield	16.38	15.49	15.47	16.74	17.62	17.35
10 year yield	28.09	29.20	28.85	28.29	28.86	27.96
20 year yield	33.08	34.99	30.94	30.48	32.25	28.70

Panel B: Forecasting Loss Function with Fixed Portfolio Weights

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month
3 month yield	15.45	15.30	15.35	15.55	15.37	15.62
6 month yield	14.48	14.52	14.58	14.97	14.78	14.90
1 year yield	15.32	15.34	15.42	15.76	16.06	15.92
2 year yield	9.71	9.79	9.88	10.00	10.00	10.98
3 year yield	8.10	8.13	8.18	8.32	8.15	8.31
4 year yield	11.93	11.94	11.98	11.95	11.96	12.07
5 year yield	12.73	12.73	12.75	12.75	12.78	12.77
10 year yield	14.24	14.24	14.26	14.31	14.39	14.39
20 year yield	13.52	13.54	13.57	13.76	13.99	13.92

Panel C: Standard Loss Function with Fixed Portfolio Weights

3 month yield	15.18
6 month yield	13.99
1 year yield	15.04
2 year yield	8.55
3 year yield	6.87
4 year yield	9.86
5 year yield	10.68
10 year yield	11.62
20 year yield	11.54

Notes to Table: We present the in-sample RMSEs for the JSZ canonical form with forecasting loss function and variable portfolio weights (Panel A), the JSZ canonical form with forecasting loss function and fixed portfolio weights (Panel B), and the JSZ canonical form with standard loss function and fixed portfolio weights (Panel C). The estimates of the specifications with forecasting loss functions are forecast-horizon specific, so we report the in-sample RMSEs for each forecast horizon. All the RMSEs are reported in basis points.

Table 9: Parameter Estimates: JSZ Canonical Form with Variable Portfolio Weights

Panel A: Forecasting Loss Function

Forecast Horizon	K_0^P	K_1^P			Eigenvalues	K_0^Q	K_1^Q			Eigenvalues
1 month										
	-0.0016	0.9993	0.0631	0.6416	0.9938	0.0004	0.9982	0.0949	-0.6988	0.9991
	0.0005	0.0059	0.9339	0.3664	0.9259	-0.0004	-0.0006	0.9492	0.6744	0.9593
	0.0005	-0.0024	-0.0032	0.7770	0.7906	0.0002	0.0006	0.0031	0.8143	0.8034
2 month										
	-0.0017	1.0000	0.0668	0.7077	0.9942	0.0004	0.9976	0.0970	-0.7441	0.9992
	0.0004	0.0061	0.9383	0.3870	0.9287	-0.0003	-0.0002	0.9524	0.6930	0.9608
	0.0005	-0.0022	-0.0036	0.7833	0.7986	0.0001	0.0005	0.0028	0.8082	0.7981
3 month										
	-0.0014	1.0003	0.0668	0.6341	0.9940	0.0005	0.9987	0.0880	-0.7480	0.9993
	0.0004	0.0062	0.9423	0.4072	0.9324	-0.0004	-0.0009	0.9606	0.7453	0.9605
	0.0005	-0.0029	-0.0045	0.7507	0.7669	0.0002	0.0006	0.0008	0.7760	0.7755
4 month										
	-0.0014	0.9988	0.0757	0.7156	0.9942	0.0005	0.9993	0.0849	-0.8361	0.9992
	0.0004	0.0059	0.9451	0.4420	0.9303	-0.0004	-0.0018	0.9691	0.7976	0.9616
	0.0004	-0.0023	-0.0061	0.7428	0.7621	0.0002	0.0009	-0.0008	0.7641	0.7717
5 month										
	-0.0021	0.9974	0.1140	0.8687	0.9966	0.0004	0.9999	0.0919	-0.7866	0.9993
	0.0003	0.0043	0.9549	0.3965	0.9322	-0.0003	-0.0027	0.9765	0.6517	0.9691
	0.0004	-0.0011	-0.0089	0.7626	0.7861	0.0001	0.0011	-0.0003	0.8030	0.8110
6 month										
	-0.0021	1.0001	0.1242	0.8707	0.9969	0.0004	0.9991	0.0884	-0.7991	0.9995
	0.0003	0.0047	0.9555	0.3872	0.9306	-0.0003	-0.0016	0.9806	0.6556	0.9694
	0.0004	-0.0017	-0.0107	0.7559	0.7840	0.0001	0.0008	-0.0017	0.7951	0.8059

Panel B: Standard Loss Function

|	K_0^P	K_1^P			Eigenvalues	K_0^Q	K_1^Q			Eigenvalues
	-0.0021	0.9940	0.0549	0.3129	0.9948	0.0004	1.0052	0.1039	-0.2569	1.0000
	0.0004	0.0017	0.9337	0.1538	0.9274	-0.0003	-0.0073	0.9370	0.2717	0.9648
	0.0012	-0.0002	-0.0042	0.8084	0.8139	0.0003	0.0042	0.0136	0.8685	0.8458

Notes to Table: We present the estimated P- and Q-parameters governing the state variables in the JSZ canonical form with forecasting loss function and variable portfolio weights, and the JSZ canonical form with standard loss function. The estimates are based on in-sample estimation with the entire sample. Panel A reports on the JSZ canonical form with forecasting loss function and variable portfolio weights. The estimates are forecast-horizon specific. Panel B reports on the JSZ canonical form with standard loss function.

Table 10: Parameter Estimates: JSZ Canonical Form with Fixed Portfolio Weights

Panel A: Forecasting Loss Function

Forecast Horizon	K_0^P	K_1^P			Eigenvalues	K_0^Q	K_1^Q			Eigenvalues
1 month										
	-0.0022	0.9950	0.0547	0.3116	0.9956	0.0004	1.0051	0.1033	-0.2582	0.9995
	0.0004	0.0017	0.9274	0.1542	0.9196	-0.0003	-0.0073	0.9353	0.2699	0.9634
	0.0012	-0.0002	-0.0041	0.8302	0.8374	0.0003	0.0042	0.0137	0.8647	0.8423
2 month										
	-0.0021	0.9945	0.0544	0.3023	0.9951	0.0004	1.0051	0.1023	-0.2581	0.9995
	0.0004	0.0017	0.9324	0.1543	0.9243	-0.0003	-0.0073	0.9354	0.2703	0.9636
	0.0012	-0.0002	-0.0041	0.8376	0.8451	0.0003	0.0042	0.0138	0.8643	0.8417
3 month										
	-0.0021	0.9942	0.0546	0.2793	0.9950	0.0004	1.0051	0.1016	-0.2590	0.9994
	0.0004	0.0017	0.9315	0.1540	0.9248	-0.0003	-0.0073	0.9352	0.2705	0.9637
	0.0012	-0.0002	-0.0042	0.8145	0.8205	0.0003	0.0042	0.0138	0.8636	0.8409
4 month										
	-0.0019	0.9947	0.0537	0.2595	0.9955	0.0004	1.0051	0.1007	-0.2553	0.9995
	0.0004	0.0018	0.9306	0.1547	0.9243	-0.0003	-0.0073	0.9351	0.2706	0.9634
	0.0012	-0.0002	-0.0042	0.8019	0.8074	0.0003	0.0043	0.0138	0.8622	0.8396
5 month										
	-0.0020	0.9954	0.0512	0.2727	0.9962	0.0004	1.0052	0.0994	-0.2581	0.9995
	0.0004	0.0017	0.9325	0.1494	0.9275	-0.0003	-0.0072	0.9350	0.2701	0.9639
	0.0014	-0.0002	-0.0041	0.7731	0.7772	0.0003	0.0043	0.0142	0.8618	0.8386
6 month										
	-0.0018	0.9947	0.0523	0.2703	0.9955	0.0004	1.0052	0.1002	-0.2573	0.9994
	0.0004	0.0017	0.9303	0.1534	0.9252	-0.0003	-0.0073	0.9348	0.2702	0.9631
	0.0014	-0.0002	-0.0042	0.7684	0.7727	0.0003	0.0042	0.0138	0.8616	0.8392

Panel B: Standard Loss Function

|	K_0^P	K_1^P			Eigenvalues	K_0^Q	K_1^Q			Eigenvalues
	-0.0021	0.9940	0.0549	0.3129	0.9948	0.0004	1.0052	0.1039	-0.2569	1.0000
	0.0004	0.0017	0.9337	0.1538	0.9274	-0.0003	-0.0073	0.9370	0.2717	0.9648
	0.0012	-0.0002	-0.0042	0.8084	0.8139	0.0003	0.0042	0.0136	0.8685	0.8458

Notes to Table: We present the estimated P- and Q-parameters governing the state variables in the JSZ canonical form with forecasting loss function and fixed portfolio weights, and the JSZ canonical form with standard loss function. The estimates are based on in-sample estimation with the entire sample. Panel A reports on the JSZ canonical form with forecasting loss function and fixed portfolio weights. The estimates are forecast-horizon specific. Panel B reports on the JSZ canonical form with standard loss function.

Figure A1: $A_1(3)$ Model with Latent Factors
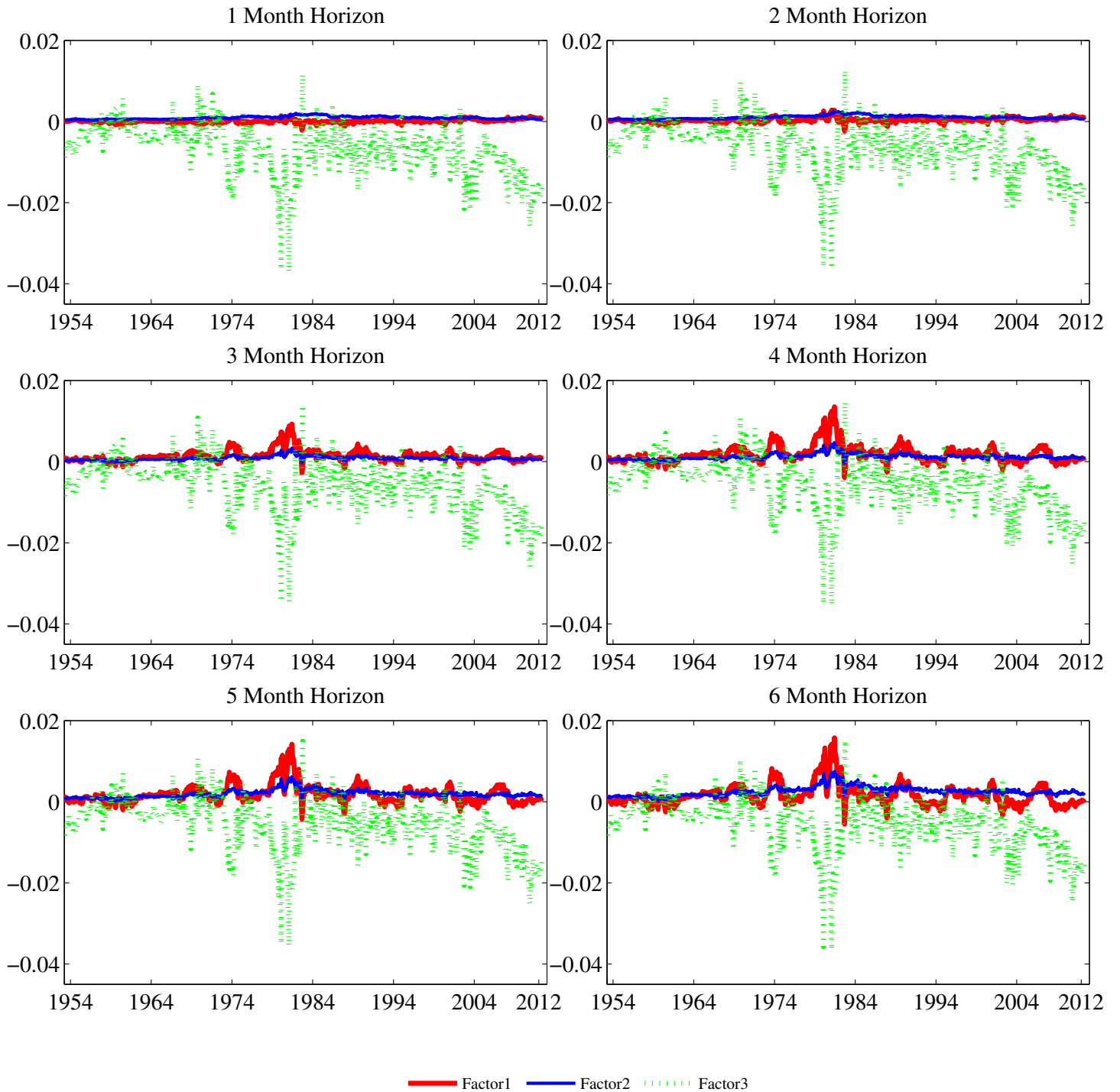Differences Between State Variables Using Forecasting and Standard Loss Functions.
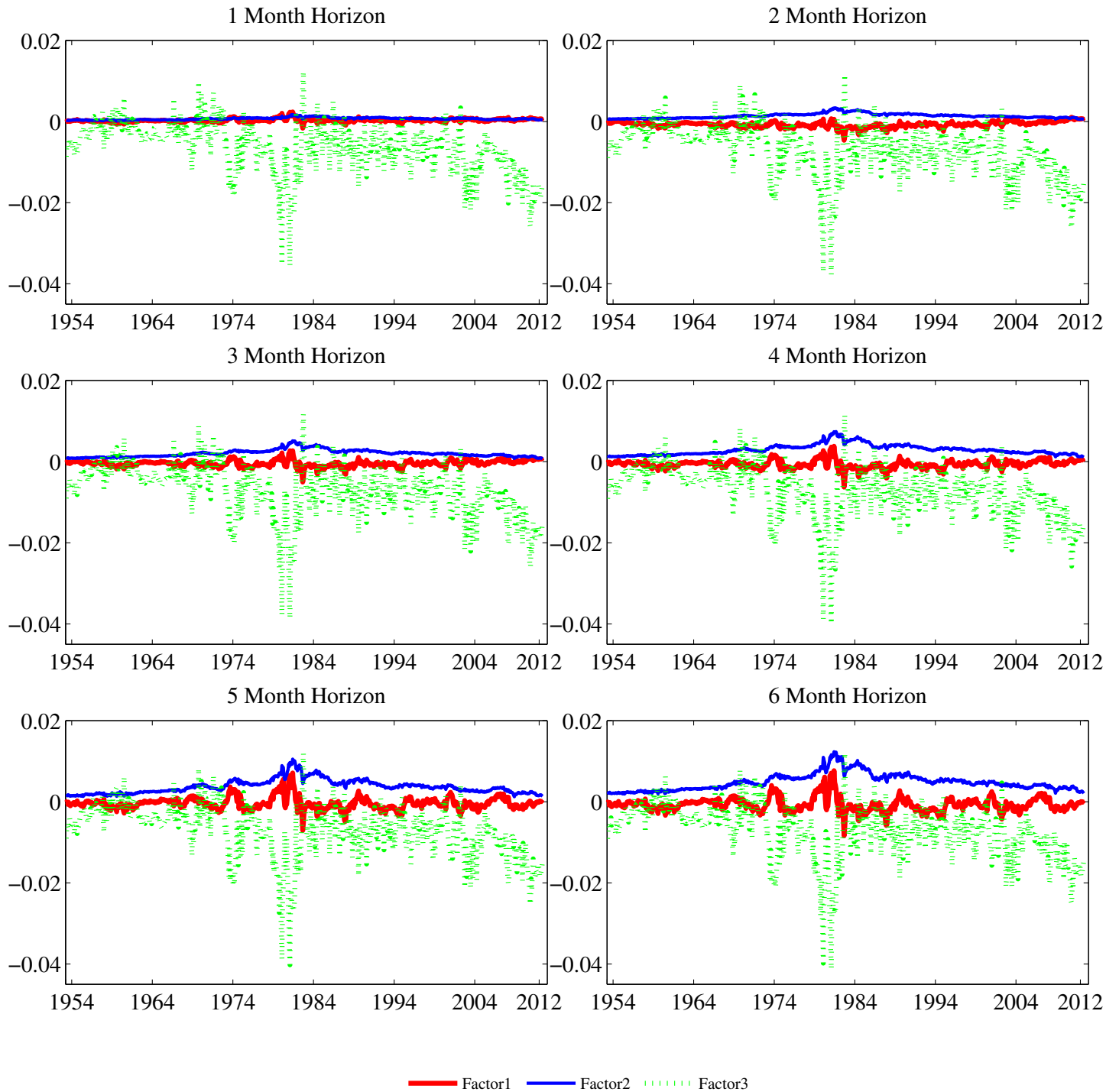
Notes to Figure: This figure presents the differences between the state variables from the $A_1(3)$ model with forecasting loss function and standard loss function. The state variables are obtained from in-sample estimation. These results are forecast-horizon specific, i.e. each panel represents the differences in the estimates for a given forecast horizon.

Figure A2: $A_2(3)$ Model with Latent Factors
Differences Between State Variables Using Forecasting and Standard Loss Functions.
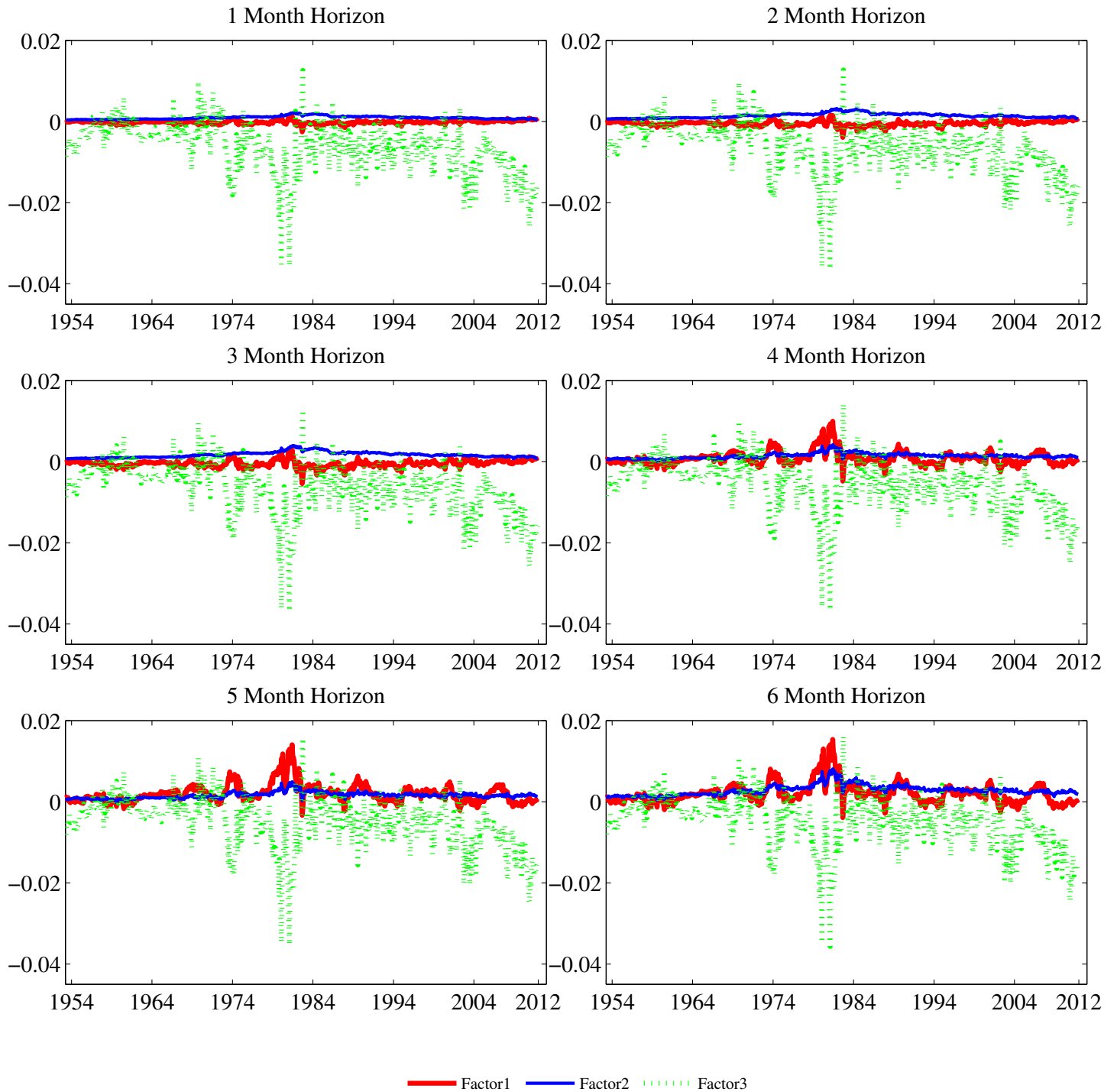
Notes to Figure: This figure presents the differences between the state variables from the $A_2(3)$ model with forecasting loss function and standard loss function. The state variables are obtained from in-sample estimation. These results are forecast-horizon specific, i.e. each panel represents the differences in the estimates for a given forecast horizon.

Figure A3: $A_3(3)$ Model with Latent Factors
Differences Between State Variables Using Forecasting and Standard Loss Functions.

Notes to Figure: This figure presents the differences between the state variables from the $A_3(3)$ model with forecasting loss function and standard loss function. The state variables are obtained from in-sample estimation. These results are forecast-horizon specific, i.e. each panel represents the differences in the estimates for a given forecast horizon.

Table A1: In-Sample RMSEs

Panel A: $A_1(3)$ with Latent Factors

Forecast Horizon k	Forecasting Loss Function						Standard Loss Function
	1 month	2 month	3 month	4 month	5 month	6 month	
3 month yield	28.96	28.52	28.63	28.95	29.93	28.41	18.21
6 month yield	16.76	16.13	16.67	17.49	17.95	21.66	15.66
1 year yield	22.88	23.78	25.39	25.10	26.48	27.35	17.26
2 year yield	20.96	22.41	23.03	24.43	23.81	22.73	11.01
3 year yield	19.60	20.48	20.89	22.04	21.56	20.58	7.83
4 year yield	15.84	15.00	15.46	16.63	16.10	17.21	10.43
5 year yield	16.49	18.43	20.12	23.37	24.02	28.18	12.12
10 year yield	22.90	22.82	24.57	23.51	26.18	28.43	12.82
20 year yield	21.55	21.06	23.98	24.42	24.79	25.81	15.34

Panel B: $A_2(3)$ with Latent Factors

Forecast Horizon k	Forecasting Loss Function						Standard Loss Function
	1 month	2 month	3 month	4 month	5 month	6 month	
3 month yield	28.51	29.81	29.74	29.70	31.18	29.12	18.18
6 month yield	17.00	15.99	16.64	17.87	18.83	23.12	16.77
1 year yield	23.03	24.82	26.39	26.18	27.48	28.16	18.51
2 year yield	21.03	22.44	23.12	25.69	23.96	22.84	14.07
3 year yield	19.65	20.51	21.98	22.31	21.68	22.66	12.57
4 year yield	17.14	15.06	16.54	16.81	16.08	18.12	11.79
5 year yield	16.55	18.45	21.11	23.53	24.22	28.30	10.77
10 year yield	23.42	24.12	26.86	24.68	25.32	28.68	11.74
20 year yield	21.86	21.33	24.04	24.77	25.67	27.89	16.42

Panel C: $A_3(3)$ with Latent Factors

Forecast Horizon k	Forecasting Loss Function						Standard Loss Function
	1 month	2 month	3 month	4 month	5 month	6 month	
3 month yield	31.23	30.54	30.02	30.26	31.72	30.01	19.29
6 month yield	18.21	16.04	16.81	18.71	19.03	23.44	17.79
1 year yield	25.05	25.57	26.46	26.78	29.30	28.43	19.51
2 year yield	21.58	22.76	23.96	26.66	24.92	23.78	14.92
3 year yield	20.73	21.58	23.03	23.26	22.44	25.40	13.34
4 year yield	18.38	17.34	18.81	19.00	17.90	19.87	12.50
5 year yield	17.58	19.19	21.62	24.09	24.09	29.03	11.42
10 year yield	25.66	25.10	27.79	26.65	27.93	29.47	11.39
20 year yield	23.14	22.58	25.42	26.41	26.90	28.76	18.99

Notes to Table: We present the in-sample RMSEs for the $A_1(3)$ (Panel A), $A_2(3)$ (Panel B) and $A_3(3)$ (Panel C) models with forecasting loss function and standard loss function. For the models with forecasting loss function, the estimates are forecast-horizon specific, so we report the in-sample RMSEs for each forecast horizon. All the RMSEs are reported in basis points.