Journal of Financial Economics 110 (2013) 110–138

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Pricing the term structure with linear regressions ☆

Tobias Adrian *, Richard K. Crump, Emanuel Moench

Federal Reserve Bank of New York, 33 Liberty Street, New York, NY 10045, United States

ARTICLE INFO

Article history:
Received 28 July 2011
Received in revised form
10 May 2012
Accepted 8 June 2012
Available online 22 April 2013

JEL classification:
G10
G12

Keywords:
Term structure of interest rates
Fama-MacBeth regressions
Dynamic asset pricing estimation
Empirical finance

ABSTRACT

We show how to price the time series and cross section of the term structure of interest rates using a three-step linear regression approach. Our method allows computationally fast estimation of term structure models with a large number of pricing factors. We present specification tests favoring a model using five principal components of yields as factors. We demonstrate that this model outperforms the Cochrane and Piazzesi (2008) four-factor specification in out-of-sample exercises but generates similar in-sample term premium dynamics. Our regression approach can also incorporate unspanned factors and allows estimation of term structure models without observing a zero-coupon yield curve.

1. Introduction

Affine models of the term structure of interest rates are a popular tool for the analysis of bond pricing. The models typically start with three assumptions: (1) the pricing kernel is exponentially affine in the shocks that drive the economy, (2) prices of risk are affine in the state variables, and (3) innovations to state variables and log yield observation errors are conditionally Gaussian (for examples, see Chen and Scott, 1993; Dai and Singleton, 2000; Collin-Dufresne and Goldstein, 2002; Duffee, 2002; Kim and Wright, 2005). These assumptions give rise to yields that are affine in the state variables and whose coefficients on the state variables are subject to constraints across maturities (for overviews, see Duffie and Kan, 1996; Piazzesi, 2003; Singleton, 2006). Empirically, the affine term structure literature has primarily used maximum likelihood (ML) methods to estimate coefficients and pricing factors, thus exploiting the distributional assumptions as well as the no-arbitrage constraints.

In this paper, we propose an alternative, regression-based approach to the pricing of interest rates. We start with observable pricing factors and develop a three-step ordinary least squares (OLS) estimator. In the first step, we decompose pricing factors into predictable components and factor innovations by regressing factors on their lagged levels. In the second step, we estimate exposures of Treasury returns with respect to lagged levels of pricing factors and contemporaneous pricing factor innovations.

☆ We would like to thank Greg Duffee, Darrell Duffie, Arturo Estrella, Eric Ghysels, Paul Glasserman, Scott Joslin, Don Kim, Monika Piazzesi, Joshua Rosenberg, Ken Singleton, Jonathan Wright, Hao Wu, and seminar participants at the Federal Reserve Bank of New York, the Stanford Institute for Theoretical Economics workshop at Stanford University, and New York University for helpful comments and discussions. Michael Abrahams provided outstanding research assistance. The views expressed in this paper are ours alone and do not necessarily represent those of the Federal Reserve Bank of New York or the Federal Reserve System.

* Corresponding author.
E-mail addresses: tobias.adrian@ny.frb.org (T. Adrian), richard.crump@ny.frb.org (R.K. Crump), emanuel.moench@ny.frb.org (E. Moench).
URLS: http://nyfedeconomists.org/adrian/ (T. Adrian), http://nyfedeconomists.org/crump/ (R.K. Crump), http://nyfedeconomists.org/moench/ (E. Moench).

0304-405X/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jfineco.2013.04.009

In the third step, we obtain the market price of risk parameters from a cross-sectional regression of the exposures of returns to the lagged pricing factors onto exposures to contemporaneous pricing factor innovations. We provide analytical standard errors that adjust for the generated regressor uncertainty. We also discuss the advantages of our method with respect to the recently suggested approaches by Joslin, Singleton, and Zhu (2011) and Hamilton and Wu (2012). In particular, we point out that the assumption of serially uncorrelated yield pricing errors underlying these likelihood-based methods imply excess return predictability not captured by the pricing factors. In contrast, because our approach is based on return regressions, we do not need to make assumptions about serial correlation in yield pricing errors.

We report a specification with five principal components of zero coupon yields as pricing factors. We show that models with fewer factors are rejected in specification tests. The pricing errors in the five-factor specification are remarkably small and return pricing errors do not exhibit autocorrelation. We further find that level risk is priced and that the time variation of level risk is best captured by the second (the slope factor) and fifth principal components. The five-factor specification exhibits substantial variation in risk premiums and at the same time gives reasonable maximal Sharpe ratios.

We next present a four-factor specification following Cochrane and Piazzesi (2008, CP) which includes the first three principal components of Treasury yields and a linear combination of forward rates designed to predict Treasury returns (the CP factor) as pricing factors. Unlike Cochrane and Piazzesi (2008), we allow for unconstrained prices of risk and find that the CP factor significantly prices all factors except slope. The magnitude and time pattern of the price of risk specification of the four-factor CP model is akin to that of the five-factor model, indicating that the two models capture term premium dynamics in a similar way. However, in-sample yield pricing errors are somewhat larger in the four-factor model.

To compare the four- and five-factor models, we perform two out-of-sample exercises. In the first, we use the model-implied term premiums to infer the future path of average short-term interest rates. In the second, we estimate the models using returns on bonds maturing in less than or equal to ten years and then impute the model-implied yields of bonds with longer maturities. In both of these exercises, the five-factor model outperforms the four-factor specification. Hence, we choose the five-factor model to be our preferred specification.

Our procedure can potentially be applied to any set of fixed income securities. In this paper, we use our approach to estimate an affine term structure model from returns on maturity-sorted portfolios of coupon-bearing Treasury securities. The availability of a zero coupon term structure is, therefore, not necessary to estimate the model. Yet, estimation from the returns on maturity sorted bond portfolios with pricing factors extracted from coupon bearing yields generates a zero coupon curve that is very similar to the Fama and Bliss discount bond yields.

We present a number of extensions. First, we show how to estimate the model in the presence of unspanned factors. Such factors do not improve the cross-sectional fit of yields but do affect the time variation of prices of risk through their predictive power for the yield curve factors. In contrast to the four- and five-factor specifications, incorporating an unspanned real activity factor produces a significant price of slope risk. Second, we show how to impose linear restrictions on risk exposures and market prices of risk in the estimation of the model. Third, we show that the maximal Sharpe ratios implied by the four- and five-factor specifications are very reasonable, with an average level below one and peaks below 2.5. Fourth, we explain how the model can be used to fit the yield curve at the daily frequency. Finally, we show that the implied principal component loadings from the term structure model are statistically indistinguishable from the actual principal component loadings.

Our paper is organized as follows. In Section 2, we present the model and our three-step estimator, and we show how to obtain model-implied yields from the estimated parameters. We further discuss the relation between our approach and other estimation methods in the literature. In Section 3, we present our main empirical findings. Section 4 discusses extensions and robustness checks. Section 5 concludes.

2. The model

In this section, we derive the data generating process for arbitrage-free excess holding period returns from a dynamic asset pricing model with an exponentially affine pricing kernel. We then show how to estimate the model parameters via three-step linear regressions, present their asymptotic distributions, and derive the no-arbitrage cross-equation constraints for bond yields.

2.1. State variables and expected returns

We assume that the dynamics of a $K \times 1$ vector of state variables X_t evolve according to the following vector autoregression (VAR):

$$X_{t+1} = \mu + \Phi X_t + v_{t+1}. \tag{1}$$

This specification of the dynamic evolution of the state variables can be interpreted as a discrete time analog to the intertemporal capital asset pricing model (ICAPM) state variable dynamics of Merton (1973) or the general equilibrium setup of Cox, Ingersoll, and Ross (1985). We assume that the shocks v_{t+1} conditionally follow a Gaussian distribution with variance–covariance matrix Σ:

$$v_{t+1} | \{X_s\}_{s=0}^{t} \sim \mathrm{N}(0, \Sigma), \tag{2}$$

where $\{X_s\}_{s=0}^{t}$ denotes the history of X_t. We denote $P_t^{(n)}$ the zero coupon Treasury bond price with maturity n at time t. The assumption of no-arbitrage implies (see Dybvig and Ross, 1987) that there exists a pricing kernel M_t such that

$$P_t^{(n)} = \mathrm{E}_t[M_{t+1} P_{t+1}^{(n-1)}]. \tag{3}$$

We assume that the pricing kernel M_{t+1} is exponentially affine:

$$M_{t+1} = \exp(-r_t - \tfrac{1}{2}\lambda'_t\lambda_t - \lambda'_t\Sigma^{-1/2} v_{t+1}), \tag{4}$$

where $r_t = \ln P_t^{(1)}$ is the continuously compounded risk-free rate. We further assume that market prices of risk are of the essentially affine form as suggested in Duffee (2002):

$$\lambda_t = \Sigma^{-1/2}(\lambda_0 + \lambda_1 X_t). \tag{5}$$

We denote $rx_{t+1}^{(n-1)}$ the log excess holding return of a bond maturing in n periods:

$$rx_{t+1}^{(n-1)} = \ln P_{t+1}^{(n-1)} - \ln P_t^{(n)} - r_t. \tag{6}$$

Using Eqs. (4) and (6) in Eq. (3) yields

$$1 = \mathrm{E}_t[\exp(rx_{t+1}^{(n-1)} - \tfrac{1}{2}\lambda'_t\lambda_t - \lambda'_t\Sigma^{-1/2}v_{t+1})]. \tag{7}$$

Assuming that $\{rx_{t+1}^{(n-1)}, v_{t+1}\}$ are jointly normally distributed, we then find

$$\mathrm{E}_t[rx_{t+1}^{(n-1)}] = Cov_t[rx_{t+1}^{(n-1)}, v'_{t+1}\Sigma^{-1/2}\lambda_t] - \tfrac{1}{2}Var_t[rx_{t+1}^{(n-1)}]. \tag{8}$$

We denote

$$\beta_t^{(n-1)'} = Cov_t[rx_{t+1}^{(n-1)}, v'_{t+1}]\Sigma^{-1}, \tag{9}$$

and using Eq. (5) we obtain

$$\mathrm{E}_t[rx_{t+1}^{(n-1)}] = \beta_t^{(n-1)'}[\lambda_0 + \lambda_1 X_t] - \tfrac{1}{2}Var_t[rx_{t+1}^{(n-1)}]. \tag{10}$$

We can decompose the unexpected excess return into a component that is correlated with v_{t+1} and another component that is conditionally orthogonal. Then we find

$$rx_{t+1}^{(n-1)} - \mathrm{E}_t[rx_{t+1}^{(n-1)}] = \gamma_t^{(n-1)'}v_{t+1} + e_{t+1}^{(n-1)}. \tag{11}$$

It is straightforward to show that $\gamma_t^{(n-1)} = \beta_t^{(n-1)}$ using Eq. (9). We assume that the return pricing errors $e_{t+1}^{(n-1)}$ are conditionally independently and identically distributed (i.i.d.) with variance σ^2.

In our baseline specifications, we use linear combinations of log yields (such as principal components) as observable factors X_t and estimate the model parameters using holding period returns based on the same set of yields. Per construction, this implies that $\beta_t = \beta\ \forall t$. We thus proceed with the assumption that β is constant.

The return generating process for log excess holding period returns is then

$$rx_{t+1}^{(n-1)} = \underbrace{\beta^{(n-1)'}(\lambda_0 + \lambda_1 X_t)}_{\text{Expected return}} \quad \underbrace{-\tfrac{1}{2}(\beta^{(n-1)'}\Sigma\beta^{(n-1)} + \sigma^2)}_{\text{Convexity adjustment}} \quad \underbrace{+\beta^{(n-1)'}v_{t+1}}_{\text{Priced return innovation}} \quad \underbrace{+e_{t+1}^{(n-1)}}_{\text{Return pricing error}} \tag{12}$$

Stacking this system across maturities and time periods, we rewrite it as

$$rx = \boldsymbol{\beta}'(\lambda_0\iota'_T + \lambda_1 X_-) - \tfrac{1}{2}(B^\star\,\mathrm{vec}(\Sigma) + \sigma^2\iota_N)\iota'_T + \boldsymbol{\beta}'V + E \tag{13}$$

where rx is an $N\times T$ matrix of excess returns, $\boldsymbol{\beta} = [\beta^{(1)}\ \beta^{(2)} \cdots \beta^{(N)}]$ is a $K\times\ N$ matrix of factor loadings, ι_T and ι_N are a $T\times 1$ and $N\times 1$ vectors of ones, $X_- = [X_0\ X_1 \cdots X_{T-1}]$ is a $K\times T$ matrix of lagged pricing factors, $B^\star = [\mathrm{vec}(\beta^{(1)}\beta^{(1)'}) \cdots \mathrm{vec}(\beta^{(N)}\beta^{(N)'})]'$ is an $N\times K^2$ matrix, V is a $K\times T$ matrix, and E is an $N\times T$ matrix.

2.2. Estimation

Based on Eq. (13), we propose the following simple three-step regression-based estimator for the parameters of the model.[1]

1. We begin by estimating Eq. (1) via ordinary least squares. This allows the decomposition of X_{t+1} into a predictable component and an estimate of the innovation $\hat{v}_{t+1}$. We stack these innovations into the matrix $\hat{V}$ and construct an estimator of the state variable variance–covariance matrix $\hat{\Sigma} = \hat{V}\hat{V}/T$.
2. Next we regress excess returns on a constant, lagged pricing factors and contemporaneous pricing factor innovations according to

$$rx = \mathbf{a}\iota'_T + \boldsymbol{\beta}'\hat{V} + \mathbf{c}X_- + E. \tag{14}$$

Collecting the regressors into the $(2K+1)\times T$ matrix $\tilde{Z} = [\iota_T\ \hat{V}'\ X'_-]'$, our estimators become

$$[\hat{\mathbf{a}}\ \hat{\beta}'\ \hat{\mathbf{c}}] = rx\tilde{Z}'(\tilde{Z}\tilde{Z}')^{-1}. \tag{15}$$

We collect the residuals from this regression into the $N\times T$ matrix $\hat{E}$. We then estimate $\hat{\sigma}^2 = \mathrm{tr}(\hat{E}\hat{E}')/NT$. We construct $\hat{B}^\star$ from $\hat{\boldsymbol{\beta}}$.
3. We estimate the price of risk parameters λ_0 and λ_1 via cross-sectional regression. We know from Eq. (13) that $\mathbf{a} = \boldsymbol{\beta}'\lambda_0 - \frac{1}{2}(B^\star\ \mathrm{vec}(\Sigma) + \sigma^2\iota_N)$ and $\mathbf{c} = \boldsymbol{\beta}'\lambda_1$. We use these expressions to obtain the following estimators for λ_0 and λ_1:

$$\hat{\lambda}_0 = (\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}')^{-1}\hat{\boldsymbol{\beta}}(\hat{\mathbf{a}} + \tfrac{1}{2}(\hat{B}^\star\mathrm{vec}(\hat{\Sigma}) + \hat{\sigma}^2\iota_N)) \tag{16}$$

and

$$\hat{\lambda}_1 = (\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}')^{-1}\hat{\boldsymbol{\beta}}\hat{c}. \tag{17}$$

2.3. Inference

Denoting the matrix $\Lambda = [\lambda_0\ \lambda_1]$, we can write the price of risk estimator in the single expression

$$\hat{\Lambda} = (\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}')^{-1}\hat{\boldsymbol{\beta}}[rx + \tfrac{1}{2}\hat{B}^\star\ \mathrm{vec}(\hat{\Sigma})\iota'_T + \tfrac{1}{2}\hat{\sigma}^2\iota_N\iota'_T]M_{\hat{V}}Z_-(Z_-M_{\hat{V}}Z'_-)^{-1}, \tag{18}$$

where $Z_- = [\iota_T\ X'_-]'$ and $M_{\hat{V}} = I_T - \hat{V}'(\hat{V}\hat{V}')^{-1}\hat{V}$. The equivalence of Eq. (18) with Eq. (16) and Eq. (17) follows from observing that $\iota'_T M_{\hat{V}}Z_-(Z_-M_{\hat{V}}Z'_-)^{-1} = \varrho'_1$, where ϱ_1 is a $(K+1)\times 1$ vector with first element equal to one and zeros elsewhere and because $[\hat{\mathbf{a}}\ \hat{\mathbf{c}}] = rxM_{\hat{V}}Z_-(Z_-M_{\hat{V}}Z'_-)^{-1}$.

We show in the Appendix that $\boldsymbol{\beta}$ and Λ have the joint limiting distribution:

$$\sqrt{T}\begin{bmatrix}\mathrm{vec}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ \mathrm{vec}(\hat{\Lambda} - \Lambda)\end{bmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{bmatrix}\mathcal{V}_\beta & \mathcal{C}_{\Lambda,\beta}{}' \\ \mathcal{C}_{\Lambda,\beta} & \mathcal{V}_\Lambda\end{bmatrix}\right), \tag{19}$$

[1] For a regression-based approach using a linear pricing kernel specification, see Adrian, Crump, and Moench (2012).

with

$$\mathcal{V}_\beta = \sigma^2(I \otimes \Sigma^{-1}). \tag{20}$$

We provide the analytical expressions for $\mathcal{V}_\Lambda$ and $\mathcal{C}_{\Lambda,\beta}$ in the Appendix. The asymptotic variance of the factor risk exposures $\boldsymbol{\beta}$ depends only on the variance of the return pricing errors in the regression of excess returns on lagged factors and factor innovations as well as on the variance of the VAR innovations of the pricing factors, but not on $\boldsymbol{\beta}$ or Λ. This implies that we could conduct inference about whether a given pricing factor is a significant explanatory variable for bond returns without estimating the full set of model parameters.

2.4. Affine yields

From the estimated model parameters, we can generate a zero coupon yield curve. Under the assumptions we have made so far, we can show that bond prices are exponentially affine in the vector of state variables:

$$\ln P_t^{(n)} = A_n + B'_n X_t + u_t^{(n)}. \tag{21}$$

By substituting Eq. (21) into Eq. (6), we see that

$$\begin{aligned} rx_{t+1}^{(n-1)} &= A_{n-1} + B'_{n-1}X_{t+1} + u_{t+1}^{(n-1)} - A_n - B'_n X_t - u_t^{(n)} \\ &\quad + A_1 + B'_1 X_t + u_t^{(1)}. \end{aligned} \tag{22}$$

Equating this expression for excess returns with the return generating expression in Eq. (12), we find

$$\begin{aligned} & A_{n-1} + B'_{n-1}(\mu + \Phi X_t + v_{t+1}) + u_{t+1}^{(n-1)} - A_n - B'_n X_t - u_t^{(n)} \\ & \quad + A_1 + B'_1 X_t + u_t^{(1)} \end{aligned} \tag{23}$$

$$= \beta^{(n-1)\prime}(\lambda_0 + \lambda_1 X_t + v_{t+1}) - \tfrac{1}{2}(\beta^{(n-1)\prime}\Sigma\beta^{(n-1)} + \sigma^2) + e_{t+1}^{(n-1)}. \tag{24}$$

This equation has to hold state by state. Let $A_1 = -\delta_0$ and $B_1 = -\delta_1$. Matching terms, we obtain the following system of recursive linear restrictions for the bond pricing parameters:

$$A_n = A_{n-1} + B'_{n-1}(\mu - \lambda_0) + \tfrac{1}{2}(B'_{n-1}\Sigma B_{n-1} + \sigma^2) - \delta_0, \tag{25}$$

$$B'_n = B'_{n-1}(\Phi - \lambda_1) - \delta'_1, \tag{26}$$

$$A_0 = 0, \quad B'_0 = 0, \tag{27}$$

and

$$\beta^{(n)} = B'_n. \tag{28}$$

We also obtain the following expression for the log bond pricing errors:

$$\underbrace{u_{t+1}^{(n-1)} - u_t^{(n)} + u_t^{(1)}}_{\text{Log yield pricing error}} = \underbrace{e_{t+1}^{(n-1)}}_{\text{return pricing error}} \tag{29}$$

Several remarks are in order. First, the derivation of log bond prices is exact, provided that $\beta^{(n)\prime} = B'_n$. Second, the recursions in Eqs. (25) and (26) are the standard linear difference equations for affine term structure models with homoskedastic shocks. The only difference with respect to the standard formulation is the appearance of the term $\frac{1}{2}\sigma^2$ in Eq. (25). This arises due to the fact that we allow for a maturity-specific return fitting error that is conditionally orthogonal to the state variable innovations. Our approach thus incorporates pricing errors explicitly into the no-arbitrage recursions.

Eq. (29) implies that if yield pricing errors are i.i.d., then return pricing errors are cross-sectionally and serially correlated. Likelihood-based estimation approaches for affine term structure models typically assume serially uncorrelated yield pricing errors and, thus, implicitly assume return pricing errors to be serially correlated.

Setting the price of risk parameters λ_0 and λ_1 to zero in the recursions in Eq. (25) and Eq. (26) generates the risk-adjusted bond pricing parameters A_n^{RF} and B_n^{RF}. These parameters have the property that $-(1/n)(A_n^{RF} + B_n^{RF\prime}X_t)$ equals the time t expectation of average future short rates over the next n periods. These risk-neutral yields are of independent economic interest, as the term premium can be calculated as the difference between the risk neutral yield and model-implied fitted yield.

2.5. Discussion of related literature

While Gaussian affine term structure models have historically been estimated using computationally intensive maximum likelihood techniques, Joslin, Singleton, and Zhu (2011, JSZ) and Hamilton and Wu (2012, HW) have recently proposed alternative, faster, estimation approaches. We now briefly compare these approaches with the method introduced here.

A crucial difference between our method and the approaches in JSZ and HW is that we do not impose the bond pricing recursions given in Section 2.4 in our estimation procedure. This allows us to derive simple regression-based estimators for all model parameters whereas both JSZ and HW require numerical optimization for a subset of the parameters in their models. We show in Section 3.3 that the risk exposures estimated via the regression approach satisfy these additional restrictions to a very high degree of precision without imposing them in the estimation. Moreover, because we do not impose these restrictions, our approach does not require the availability of a zero-coupon yield curve, but can instead be readily used in applications with returns on coupon-bearing bond portfolios as we discuss in Section 3.6. Another important difference between our approach and those in JSZ and HW is that we do not impose the constraint that principal components (or other linear combinations of yields) must be priced perfectly. As a consequence, there is a potential inconsistency between actual and model-implied principal components. However, we show in Section 4.5 that any such inconsistency is numerically negligible in our preferred five-factor specification.

Joslin, Singleton, and Zhu (2011) show that Gaussian affine dynamic term structure models which impose these constraints imply additional restrictions across parameters. Under certain assumptions, linear combinations of yields can be used as observable pricing factors and the model parameters can be split into two subsets. The first subset summarizes the parameters governing the evolution of the pricing factors under the historical measure. JSZ show that for this subset, the ML estimator coincides with the OLS estimator of a VAR(1) of the pricing factors. Hence, they suggest a two step estimation approach. In the first step, they estimate the parameters of the VAR(1) via OLS. Then, in the second step, they estimate the remaining model parameters which govern the evolution of pricing factors under the risk-neutral measure, via numerical solution of the likelihood function taking as given the OLS estimates of the first step.

We see several advantages of our method with respect to that of JSZ. First, JSZ assume that the yield pricing errors are conditionally independent of lagged values of yield pricing errors. This assumption allows them to reduce the computational complexity of the estimation. Eq. (29) above shows that in affine models where bond prices are exponentially affine functions of the pricing factors, serially uncorrelated yield pricing errors give rise to autocorrelation of the return pricing errors. Serially correlated return pricing errors, in turn, generate excess return predictability not captured by the pricing factors. We think that this is an undesirable assumption and indeed our empirical results suggest that there is a strong level of serial correlation in yield fitting errors while there is little to no autocorrelation in return pricing errors. Second, the estimation approach suggested by JSZ is tailored to affine models that use linear combinations of yields (such as principal components) as pricing factors. In contrast, our approach does not require pricing factors to be linear combinations of yields but can also be readily applied to models that use observable (and economically interpretable) pricing factors such as macroeconomic variables. Third, while the approach in JSZ estimates a subset of parameters via OLS, it still requires estimation of a subset of parameters via numerical methods. These could be prone to issues related to the convergence to the global maximum of the likelihood function. By contrast, our approach relies exclusively on linear regressions which greatly alleviates such concerns.

Hamilton and Wu (2012) have recently proposed another multi-step estimation method for affine models which combines OLS regressions with numerical optimization. While their method does not require a particular normalization of the model, it is also based on the assumption that exactly K linear combinations of yields are observed and priced without error while the remaining yields are observed with error. Based on this assumption, HW first run an OLS regression of the vector of observed yields on their contemporaneous and lagged values. In further estimation steps, they then back out the remaining model parameters by numerically minimizing a chi-squared objective function which summarizes the deviations of the structural from the reduced form parameters, an approach which is asymptotically equivalent to full information maximum likelihood. As in JSZ, HW require that yield fitting errors are not serially correlated, which as we discuss above has strong implications for the properties of the return pricing errors. Indeed, Hamilton and Wu (forthcoming) show that popular affine term structure models feature serially correlated yield fitting errors. Moreover, it does not appear that the HW method can be readily applied to specifications which feature unspanned factors in the term structure of interest rates which we show is easy in our setup. Finally, the HW approach requires numerical optimization which our method avoids.

3. Empirical results

In this section, we provide estimation results from our regression approach for different specifications of affine term structure models. We start by testing for the number of factors necessary to explain the time series and cross section of Treasury returns. To demonstrate the robustness of these results with respect to the choice of data, we do so using different combinations of Treasury yield and return data that have previously been employed in the literature. The results of this analysis, discussed in Section 3.2, show that the first five principal components of Treasury yields are needed to explain Treasury returns. We, therefore, choose a $K=5$ factor specification as our baseline example and discuss its properties and implications in Section 3.3. We also estimate a model with $K=4$ factors in the style of Cochrane and Piazzesi (2008), which employs the first three principal components of Treasury yields and a return forecasting factor as pricing factors. This model is presented in Section 3.4. We compare the two model specifications in Section 3.5 and find that the five-factor model outperforms the four-factor model in economically important dimensions.

Both models use bond returns inferred from fitted zero coupon yield curves. However, a distinguishing feature of our estimation approach is that it does not require the availability of zero coupon yield data. Instead, given a risk-free short-term rate as well as a set of pricing factors that span the cross section of bond yields, we can obtain estimates of fitted zero coupon yield curves without observing these curves. To illustrate this capability, our third specification combines excess returns obtained from maturity-sorted portfolios of coupon bearing bonds with the first five principal components extracted from the set of Treasury yields published in the Federal Reserve Board's H.15 release. We discuss estimation results for this specification in Section 3.6.

3.1. Data

We estimate our baseline four- and five-factor specifications using the zero coupon yield data constructed by Gurkaynak, Sack, and Wright (2007, GSW).[2] These data are based on fitted Nelson-Siegel-Svensson curves, the parameters of which are published along with the estimated zero coupon curve. We use these parameters to back out the cross section of yields for maturities $n=3,\dots,120$ months. From these yields we extract principal components. Taking as the risk-free rate the $n=1$ month yield, we calculate excess returns for $n=6,12,18,24,\dots,60,84$, and 120 month zero coupon bonds directly from Eq. (6), giving a cross section of $N=12$ maturities.

Our third specification uses factors extracted from the constant maturity yields from the Federal Reserve Board's H.15 release and excess returns of Fama maturity-sorted bond portfolios and the Fama one-month risk-free rate,

[2] We thank the authors for making these data available for download, on the website http://www.federalreserve.gov/Pubs/feds/2006/200628/feds200628.xls.

available from the Center for Research in Security Prices (CRSP). The Fama bond portfolios are sorted into six-month maturity buckets with $n=1–6, 7–12, \ldots, 55–60$ months in addition to a portfolio grouping maturities 61–120 months, for a total of $N=11$ maturity groups. From the Fama risk-free rate R_t^f and the monthly portfolio returns $R_{t+1}^{(n-1)}$, we calculate log excess returns according to

$$rx_{t+1}^{(n-1)} = \log(1+R_{t+1}^{(n-1)}) - \log(1+R_t^f). \tag{30}$$

We estimate all models over the sample period 1987:01–2011:12, which provides a total of $T=300$ monthly observations. Taken as given the set of pricing factors, we estimate the parameters $(\Phi, \Sigma, \sigma, \beta, \lambda_0, \lambda_1)$ using our three-step estimation approach. We further obtain estimates of the short rate parameters δ_0 and δ_1 by regressing the one-month T-bill on the pricing factors. We then feed all parameter estimates into Eqs. (25) and (26) to obtain the recursive pricing parameters and use the latter to compute model-implied yields of the maturities of interest.

3.2. Model specification tests and identification

We start by testing for the number of pricing factors. While early research by Scheinkman and Litterman (1991) and Garbade (1996, Chapter 16) pointed out that three factors are enough to explain the cross-sectional variation of yields, more recent papers such as Cochrane and Piazzesi (2005, 2008) and Duffee (2011) emphasize the importance of additional factors to explain Treasury returns.

Our estimation approach uses return regressions to fit the cross section of yields and allows for direct tests of the number of pricing factors. Whether the pricing factors explain return variation is important for the identification of the price of risk parameters. This is easy to see in the three-step regression approach. Recall from Eqs. (16) and (17) that the market price of risk parameters are obtained via regressions onto the matrix of factor risk exposures. Identification of prices of risk thus requires that the stacked factor loadings $\boldsymbol{\beta}'$ are of full column rank.[3]

Because the asymptotic variance of $\boldsymbol{\beta}$ is of a simple OLS form [see Eq. (20)], we can apply standard tools to assess the rank of $\boldsymbol{\beta}$. We test for the possible rank deficiency using the Anderson (1951) canonical correlations test. We denote the sample partial canonical correlation between V and rx conditional on X_- by ρ_i for $i=1,\ldots,K$ factors. Then, under the null hypothesis that $rank(\boldsymbol{\beta}) \le r < K$, the Anderson test statistic is

$$rk_r = -T \sum_{i=r+1}^{K} \ln(1-\rho_i^2) \overset{a}{\sim} \chi^2((K-r)(N-r)). \tag{31}$$

We can also test for the presence of unspanned or useless factors by checking whether particular columns of $\boldsymbol{\beta}'$ are equal to zero. This is straightforward to do based on the asymptotic distribution of the factor risk exposures $\boldsymbol{\beta}$ from Eq. (20). Let β_i be the ith column of $\boldsymbol{\beta}'$. Then, under the null that $\beta_i = 0_{N\times 1}$, the Wald statistic is

$$W_{\beta^i} = \hat{\beta}'_i \hat{\mathcal{V}}_{\beta_i}^{-1} \hat{\beta}_i \overset{a}{\sim} \chi^2(N). \tag{32}$$

Table 1 displays the results of both test statistics for model specifications with three to five factors using various combinations of Treasury return and yield data. In particular, we report the Anderson test for the hypothesis that the rank of $\boldsymbol{\beta}$ in a K factor model is less than or equal to $K-1$ (denoted rk_{K-1}) and the Wald statistic for the null that the Kth column of $\boldsymbol{\beta}'$ is equal to zero. In Panel A of Table 1, we provide the test statistics for our benchmark specification, which uses both GSW returns and yields. Panel B reports tests for the number of factors using the same yield data as Duffee (2010), who merges Fama and Bliss yields of $n=3, 12, 24, \ldots, 60$ months with GSW yields of maturities $n=72$ and $n=120$ months. Panel C uses principal components extracted from the Federal Reserve H.15 release, selecting maturities $n=6$, 12, 24, 36, 60, 84, and 120 months.[4] Finally, the Panel D extracts factors from the CRSP constant maturity Treasury yield database using maturities $n=12$, 24, 60, 84, and 120 to which we append the three-month and six-month T-bill from the CRSP Fama and Bliss discount bond yield data. The specifications on the left-hand side are estimated using excess returns on zero coupon bonds calculated from the GSW data set. The specifications on the right hand side are estimated using excess returns on Fama maturity-sorted Treasury portfolios.

The tests unanimously support a five-factor specification across the different data sets. The Wald tests overwhelmingly reject that the last column of $\boldsymbol{\beta}'$ equals zero in the five-factor specification. The Anderson rank statistics also broadly support a five-factor model. We conclude from these results that a five-factor specification of an affine term structure model is likely to provide a better fit of bond returns and, hence, bond risk premiums than specifications with fewer factors. Importantly, this result is obtained for the different combinations of US Treasury yield and return data that have been used in the extant literature and is, therefore, not driven by the particular data set we use in our preferred five-factor specification.

3.3. Estimation of a five factor model

Following the evidence in favor of a five-factor model from the tests above, we use it as our baseline specification. We now show that this specification fits the yield curve close to perfectly and gives rise to substantial time variation in the prices of risk. Because traditional term structure models are estimated imposing nonlinear cross-equation restrictions, estimation of these models with more than three factors becomes computationally demanding. In contrast, adding factors to the regression-based approach comes at no computational cost.

Table 2 reports the time series properties of the yield pricing errors implied by the five-factor specification of the model. We set $\mu=0$ because principal components are

[3] See Kleibergen (2009) and Burnside (2012) for a related discussion in the context of static Fama and MacBeth equity pricing models.

[4] Joslin, Singleton, and Zhu (2011) also use H.15 yields but first extract the implied zero coupon yield curve. Our estimation procedure does not require this additional step.

Table 1
Identification tests.
This table reports the results of the identification tests described in Section 3.2. rk_{k-1} is the Anderson (1951) statistic for the hypothesis that the rank of $\boldsymbol{\beta}$ in a K factor model is $K-1$ or smaller. W_β is the Wald statistic for the hypothesis that the last row of $\boldsymbol{\beta}$ equals zero. These test statistics follow $\chi^2(N-K+1)$ and $\chi^2(N)$ distributions, respectively. Corresponding p-values are reported in parentheses. The four panels present alternative models estimated using different numbers of yield principal components as pricing factors. Panel A reports test results for specifications using Gurkaynak, Sack, and Wright (2007, GSW) yields to extract principal components. Panel B is based on a combination of Fama and Bliss and GSW yields, as in Duffee (2010). Panel C uses Federal Reserve H.15 yields. Panel D uses Center for Research in Security Prices (CRSP) constant-maturity Treasury yields. Specifications in the left column are estimated using excess returns from GSW zero coupon yields. Specifications in the right column are estimated using excess returns on Fama maturity-sorted portfolios.

	GSW returns		CRSP Fama returns	
Number of Factors	rk_{k-1}	W_β	rk_{k-1}	W_β
Panel A: GSW yields				
$K=3$	577.580	11216.622	333.092	2176.506
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)
$K=4$	629.077	4643.528	104.281	231.140
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)
$K=5$	885.261	21831.991	34.991	25.139
(p-value)	(0.000)	(0.000)	(0.000)	(0.009)
Panel B: Fama and Bliss yields + GSW six- and ten-year				
$K=3$	228.202	9105.729	103.634	2090.322
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)
$K=4$	155.195	6194.862	86.344	889.096
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)
$K=5$	34.348	166.716	32.952	86.717
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)
Panel C: H.15 yields				
$K=3$	242.768	384.954	189.816	454.848
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)
$K=4$	61.473	279.181	101.817	44.675
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)
$K=5$	18.845	55.866	63.757	68.442
(p-value)	(0.009)	(0.000)	(0.000)	(0.000)
Panel D: CRSP constant-maturity yields				
$K=3$	164.237	1154.031	68.376	1328.303
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)
$K=4$	152.737	462.968	61.836	728.340
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)
$K=5$	22.550	72.438	23.945	188.009
(p-value)	(0.002)	(0.000)	(0.001)	(0.000)

extracted from demeaned yields. The average yield pricing errors are very small, not exceeding 0.4 basis points in absolute value. Moreover, the standard deviation of the yield pricing errors is tiny, remaining below 1 basis point for all maturities. Finally, consistent with our decomposition of yield pricing errors in Eq. (29), we find evidence of strong serial correlation in yield pricing errors while the return pricing errors have essentially no autocorrelation.

The upper two panels of Fig. 1 show the time series of observed and fitted yields for the two- and ten-year Treasury notes. These plots show that the five-factor specification provides an extremely tight fit to yields. It is worth emphasizing that the model has been fitted to returns. All parameters have been obtained via linear regressions, without imposing the cross-equation recursions in Eqs. (25) and (26). These equations have been used only ex post, to allow the computation of all maturities of the yield curve. Our results, therefore, suggest that traditional estimation approaches in which latent pricing factors are jointly estimated with the model parameters by explicitly imposing the cross-equation constraints provide little or no gains in terms of fitting the yield curve.

The bottom two charts in Fig. 1 show the observed and fitted one-month excess holding returns on the same two maturities. The actual excess returns are shown as solid lines. The fitted excess returns are given by the expression $\hat{B}'_{n-1}(\hat{\lambda}_0+\hat{\lambda}_1 X_t)-\frac{1}{2}(\hat{B}'_{n-1}\hat{\Sigma}\hat{B}_{n-1}+\hat{\sigma}^2)+\hat{B}'_{n-1}\hat{v}_{t+1}$ and are shown as dashed lines. Both are almost indistinguishable from one another, suggesting that the dynamics of excess bond returns are close to perfectly captured by the five-factor model. We superimpose as dash-dotted lines the expected excess return component $\hat{B}'_{n-1}(\hat{\lambda}_0+\hat{\lambda}_1 X_t)-\frac{1}{2}(\hat{B}'_{n-1}\hat{\Sigma}\hat{B}_{n-1}+\hat{\sigma}^2)$, which captures the risk premium investors demand for holding a bond with n months to maturity for one month. The charts show that risk premiums in the five-factor specification exhibit substantial time variation.

We now investigate the role of each of the factors in pricing the various components of interest rate risk in the five-factor model. To that end, Table 3 reports the estimated elements of λ_0 and λ_1 as well as the corresponding t-statistics. Note that here and throughout, the standard errors are calculated under the assumption that μ is

Table 2
Five-factor model: fit diagnostics.

This table summarizes the time series properties of the pricing errors implied by the five-factor specification. The sample period is 1987:01–2011:12. Reported are the sample mean, standard deviation, skewness, and kurtosis of the errors; $\rho(1), \rho(6)$ denote their autocorrelation coefficients of order one and six. Panel A reports properties of the yield pricing errors $\hat{u}$, and Panel B reports properties of the return pricing errors $\hat{e}$. $n = 12, \ldots, 120$ denotes the maturity in months.

Summary Statistics	$n = 12$	$n = 24$	$n = 36$	$n = 60$	$n = 84$	$n = 120$
Panel A: Yield pricing errors						
Mean	−0.001	0.000	−0.001	−0.003	−0.003	−0.004
Standard deviation	0.004	0.006	0.006	0.004	0.004	0.008
Skewness	−0.342	0.420	−0.086	−0.069	0.305	−0.133
Kurtosis	3.245	2.932	2.414	2.205	2.545	2.726
$\rho(1)$	0.793	0.803	0.879	0.936	0.853	0.818
$\rho(6)$	0.565	0.526	0.745	0.718	0.657	0.442
Panel B: Return pricing errors						
Mean	0.000	−0.002	−0.007	−0.007	−0.003	−0.026
Standard deviation	0.081	0.092	0.151	0.132	0.179	0.617
Skewness	4.183	1.000	3.016	1.310	0.396	0.248
Kurtosis	46.489	9.708	40.674	12.555	5.685	7.981
$\rho(1)$	−0.001	0.175	−0.140	0.001	−0.162	−0.157
$\rho(6)$	0.104	0.098	0.083	−0.014	0.015	0.012

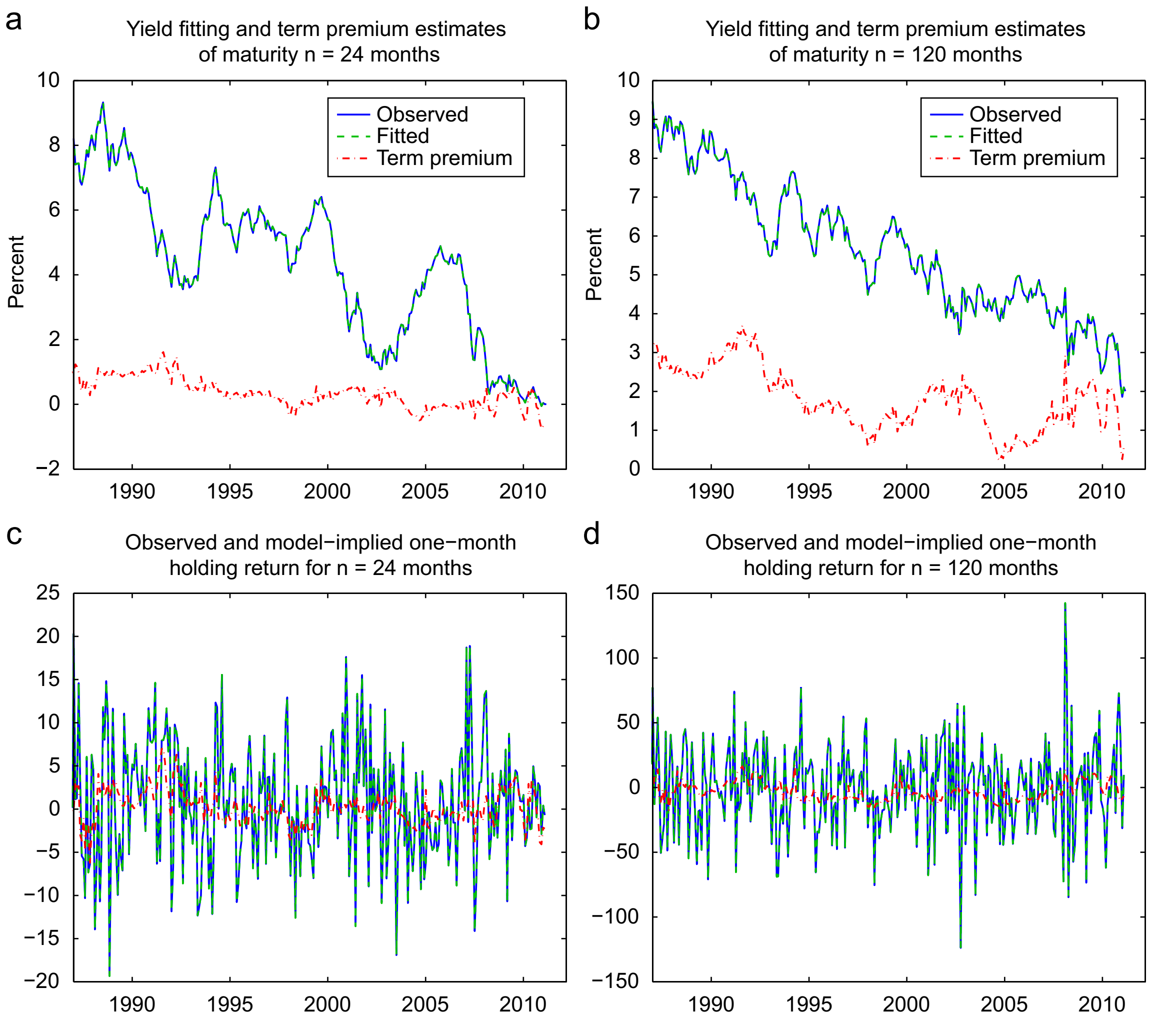

Fig. 1. Five-factor model: observed and model-implied time series. This figure provides plots of the yields and excess one-month holding returns for the two-year and ten-year maturities as observed and implied by the five factor specification. The observed yields and returns are plotted by solid lines, and dashed lines correspond to model-implied yields and returns. Dash-dotted lines show model-implied term premiums in the upper two charts and expected excess holding period returns in the lower two charts.

Table 3
Five-factor model: market prices of risk.

This table summarizes the estimates of the market price of risk parameters λ_0 and λ_1 for the five factor specification. t-Statistics are reported in parentheses. The standard errors have been computed according to the formulas from Section 2.3. Wald statistics for tests of the rows of Λ and of λ_1 being different from zero are reported along each row, with the corresponding p-values in parentheses below. PC1,…, PC5 denote the first through fifth principal components of Treasury yields. Bolded coefficients represent significance at the 5% level.

Factor	λ_0	$\lambda_{1.1}$	$\lambda_{1.2}$	$\lambda_{1.3}$	$\lambda_{1.4}$	$\lambda_{1.5}$	W_Λ	W_{λ_1}
PC1	**−0.019**	−0.003	**−0.016**	−0.005	0.012	**0.030**	**30.367**	**23.705**
(t-statistic)	(−2.566)	(−0.443)	(−2.160)	(−0.648)	(1.605)	(3.987)	(0.000)	(0.000)
PC2	0.013	0.027	−0.011	−0.003	−0.011	0.015	7.097	6.167
(t-statistic)	(0.951)	(1.914)	(−0.818)	(−0.213)	(−0.792)	(1.077)	(0.312)	(0.290)
PC3	−0.030	**−0.077**	−0.001	**−0.093**	**−0.132**	−0.056	**37.170**	**36.277**
(t-statistic)	(−0.951)	(−2.466)	(−0.029)	(−2.987)	(−4.244)	(−1.783)	(0.000)	(0.000)
PC4	0.042	0.064	−0.007	0.015	−0.058	**−0.086**	10.444	9.374
(t-statistic)	(1.062)	(1.594)	(−0.189)	(0.367)	(−1.461)	(−2.147)	(0.107)	(0.095)
PC5	0.005	**−0.105**	0.012	−0.004	−0.073	**−0.324**	**45.691**	**45.688**
(t-statistic)	(0.097)	(−2.028)	(0.243)	(−0.070)	(−1.431)	(−6.287)	(0.000)	(0.000)

unknown to accommodate the sampling uncertainty from using demeaned yields. The price of level risk has a significant negative constant component (the top element in the vector λ_0). Excess return betas are negative multiples of the yield loadings $b_n = -(1/n)B_n$, implying that investors on average require a positive expected excess return for holding the level portfolio. In addition to level risk being nonzero unconditionally, we find that it varies significantly as a function of the slope and the fifth factor. The loading of level risk on the slope factor has a negative sign. The model, therefore, implies that a higher slope is associated with higher expected excess returns. As can be seen in the lower right-hand panel of Fig. 2, expected excess returns load positively on the slope factor with coefficients more or less linearly increasing in maturity. This is in line with prior evidence on the predictive power of yield spreads for bond returns as, e.g., in Campbell and Shiller (1991). We further find that the fifth principal component enters the price of level risk with a strongly significant positive coefficient. This underscores that factors with negligible contemporaneous effects on the yield curve can have strong predictive power for future excess returns, consistent with the findings of, e.g., Cochrane and Piazzesi (2008) and Duffee (2011). The estimated price of risk coefficients are also economically important. For example, the coefficient of the fifth principal component in the price of level risk is estimated to be 0.03. This implies that a 1 standard deviation decline in the fifth principal component increases the annualized expected excess return on the ten-year Treasury bond by about 6%.

We can conveniently summarize the pricing implications of the model by testing the null hypothesis that the different rows of Λ (which combines λ_0 and λ_1) are equal to zero. Given the asymptotic distribution of the estimator derived in Section 2.3, a Wald test can be used to that effect. In particular, let $\lambda'_{i\cdot}$ be the ith row of Λ. Then, under the null that $\lambda'_{i\cdot} = 0_{1\times(K+1)}$, the Wald statistic

$$W_{\Lambda_{i\cdot}} = \hat{\lambda}'_{i\cdot}\hat{\mathcal{V}}^{-1}_{\lambda_{i\cdot}}\hat{\lambda}_{i\cdot} \overset{a}{\sim} \chi^2(K+1) \tag{33}$$

has an asymptotic chi-square distribution with $K+1$ degrees of freedom. In a similar vein, we can test whether the price of risk associated with a given factor features significant time variation by testing if the corresponding row of λ_1 is jointly equal to zero. Let $\lambda'_{1i\cdot}$ be the ith row of λ_1. Then, under the null that $\lambda'_{1i\cdot} = 0_{1\times K}$, the Wald statistic

$$W_{\lambda_{1i\cdot}} = \hat{\lambda}'_{1i\cdot}\hat{\mathcal{V}}^{-1}_{\lambda_{1i\cdot}}\hat{\lambda}_{1i\cdot} \overset{a}{\sim} \chi^2(K) \tag{34}$$

has an asymptotic chi-square distribution with K degrees of freedom. The last two columns in Table 3 provide these Wald statistics as well as the corresponding p-values for the rows of Λ and λ_1, respectively. In line with the individual significance of the price of risk parameters, these Wald statistics show that level risk is priced significantly both unconditionally and conditionally.

Slope risk is not priced in the five-factor model. None of the individual elements of the second row of Λ is significantly different from zero, and both Wald statistics indicate that these coefficients are also jointly indistinguishable from zero. This result appears surprising given that slope captures the second largest share of the cross-sectional variation of yields. We return to this issue in Section 4.1.

While slope risk is not priced in the five-factor model, curvature risk, as measured by the exposure to the third principal component of Treasury yields, carries a significant price of risk. The level factor, the curvature factor itself, as well as the fourth principal component of yields all significantly affect the price of curvature risk over time. The coefficient of the price of curvature risk on the level factor is negative, indicating that expected excess returns on a portfolio that is long in short-term and long-term government bonds and short in intermediate maturities tend to be increasing in the level of rates. While the price of level risk is significantly determined by the fifth principal component, the price of curvature risk strongly varies with the fourth principal component, which again highlights the role of higher-order principal components for expected bond returns.

The upper two panels of Fig. 2 provide a plot of unconditional first and second moments of yields across maturities as observed and fitted by the five-factor model. The charts show that the specification fits both moments

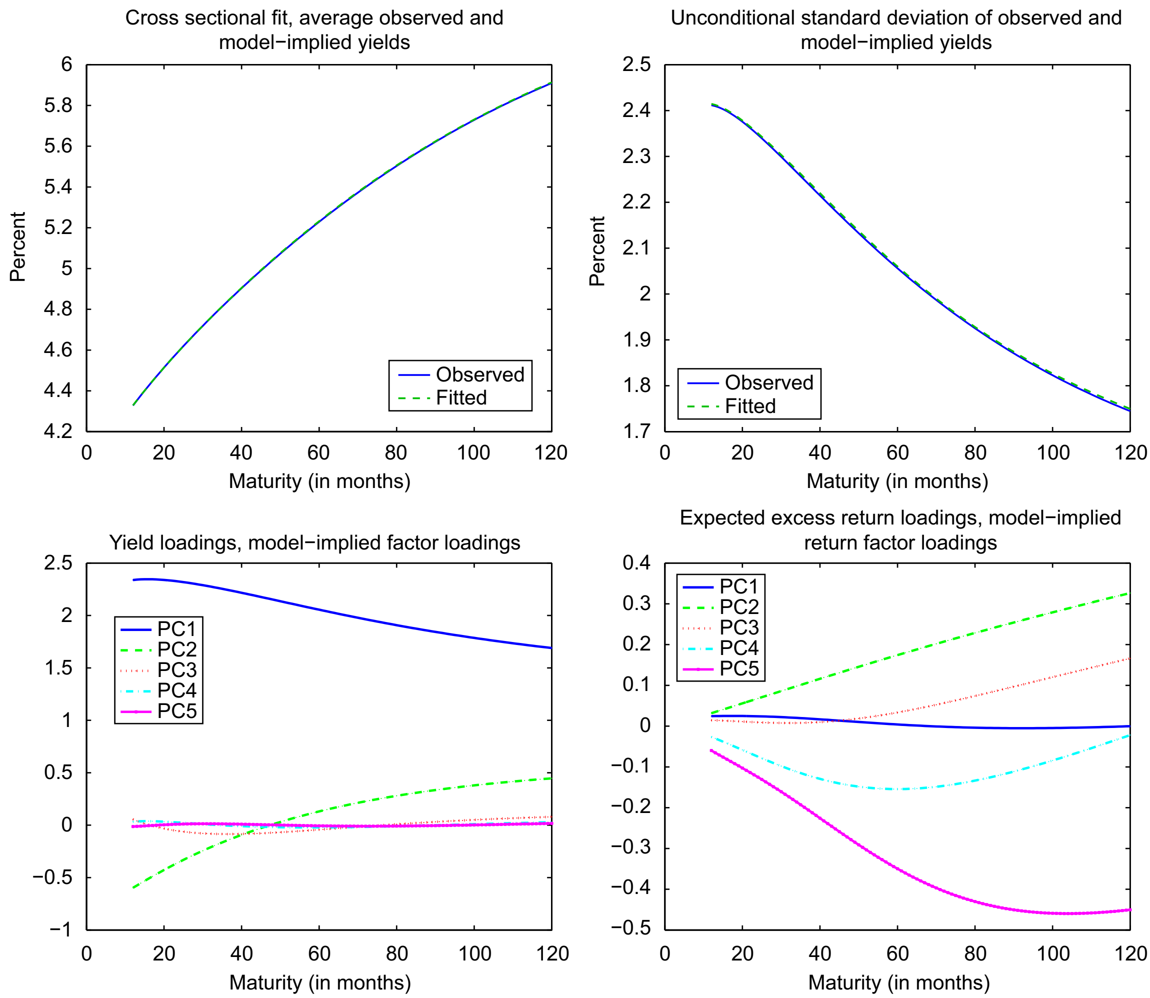

Fig. 2. Five-factor model: cross-sectional diagnostics. This figure provides graphs exhibiting the cross-sectional fit and interpretation of the factors in the five factor specification. The upper two panels plot unconditional means and standard deviations of yields against those implied by the model. The lower left panel plots the implied yield loadings $-(1/n)B_n$. These coefficients can be interpreted as the response of the n-month yield to a contemporaneous shock to the respective factor. The lower right panel plots the expected return loadings $B'_n\lambda_1$. These coefficients can be interpreted as the response of the expected one-month excess holding return on an n-month bond to a contemporaneous shock to the respective factor. PC1,…, PC5 denote the first through fifth principal components of Treasury yields.

very well. The lower left panel of Fig. 2 provides a plot of the estimated yield loadings b_n. While these graphs reinforce the common interpretation of the first three principal components of yields as level, slope, and curvature, they also highlight why higher order principal components of yields have been given much less attention in the literature. The lower left panel of Fig. 2 shows that the loadings of yields on the fourth and fifth principal components are very small. This is in sharp contrast to the lower right panel of Fig. 2, which shows the loadings $B'_n\lambda_1$ of expected excess returns on all model factors. The loadings on the fourth and fifth principal components exhibit strong variation across maturities. In fact, expected returns are explained nearly entirely by the second, fourth, and fifth principal components, while yields are explained almost exclusively by the first three principal components.

We can assess the economic significance of the various risk prices by computing their contribution to the variability of the pricing kernel. Recalling that

$$\ln M_{t+1} = -r_t - \tfrac{1}{2}\lambda'_t\lambda_t - \lambda'_t\Sigma^{-1/2}v_{t+1}, \tag{35}$$

it is straightforward to decompose the conditional volatility of the pricing kernel into the contributions due to each price of risk according to

$$\mathrm{Var}_t(\ln M_{t+1}) = \lambda'_t\lambda_t = \sum_{j=1}^{K}\lambda_{jt}^2. \tag{36}$$

Doing so, we find that risk prices of all five factors contribute to the time variation of the pricing kernel (see the upper left chart in Fig. 3). However, given that the exposure of long term bonds to risks other than level risk is fairly small, the statistical significance of their market prices of risk does not translate into a sizable impact on expected excess returns on those bonds. As the lower left chart in Fig. 3 shows, it is primarily, but not exclusively, the time variation of level risk

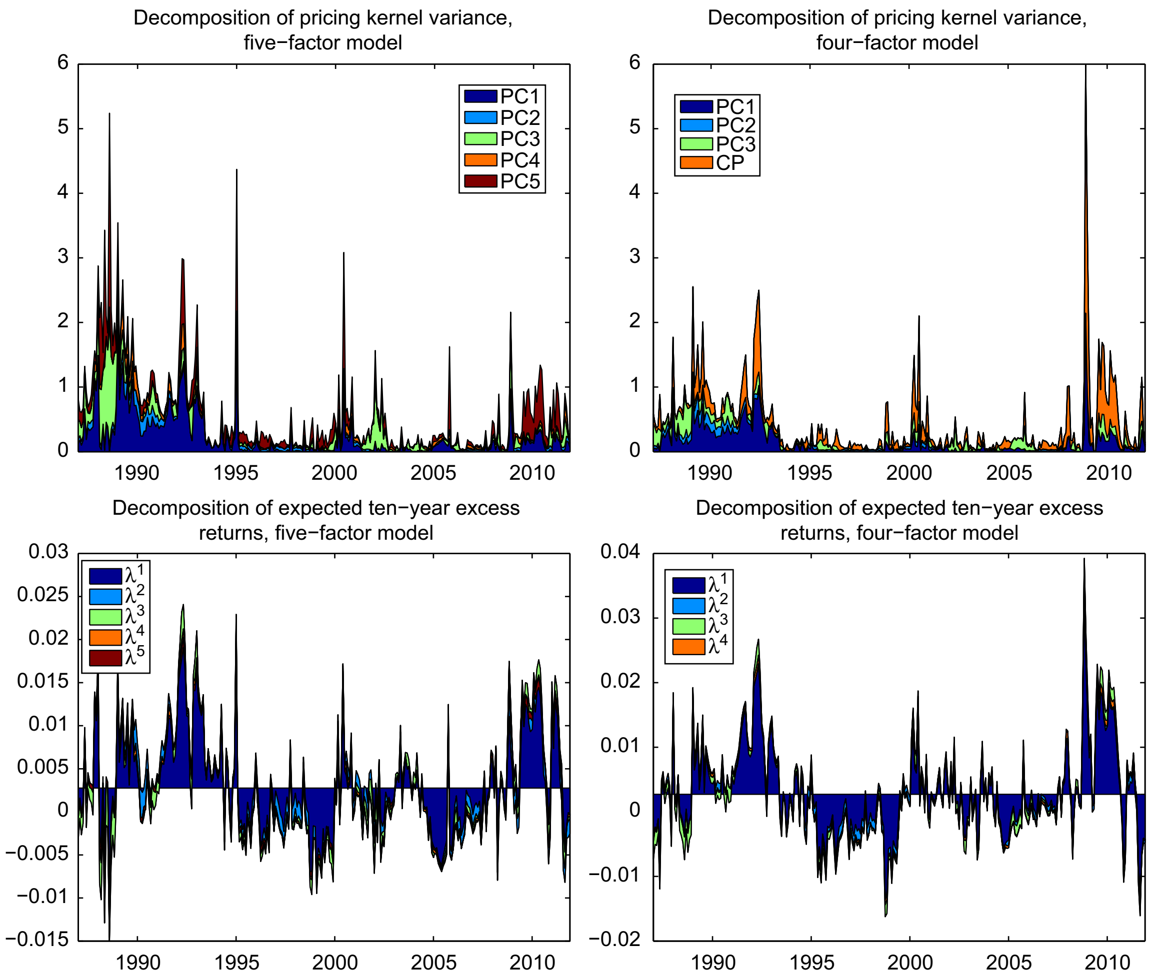

Fig. 3. Pricing kernel variance decomposition. This figure shows decompositions of the conditional volatility of the pricing kernel and the expected excess return in the four- and five-factor specification for the ten-year Treasury note. The upper two panels decompose $\lambda'_t\lambda_t$ into its K components. The lower panels decompose expected excess returns, $\beta'\Lambda Z_t=(\beta_{i1}\Lambda_1+\cdots+\beta_{iK}\Lambda_k)Z_t$, where β_{ij} is the jth element of β_i and Λ_k denotes the kth row of Λ. PC1,…,PC5 denote the first through fifth principal components of Treasury yields. CP denotes the Cochrane and Piazzesi (2008) return forecasting factor.

that contributes to time-varying expected excess returns on the ten-year Treasury bond. This time variation in turn is largely, but not exclusively, driven by movements in the slope and the fifth principal component, as shown in the lower right chart of Fig. 2.

We derive the recursive pricing formulas in Eqs. (25) and (26) by equating our return generating process in Eq. (12) with the definition of excess returns in Eq. (6) under the assumption that log bond prices are linear in the state variables X. The matching of terms implies that our recursive bond pricing parameters B_n, which are functions of the parameters $\Phi, \Sigma, \sigma, \delta_0, \delta_1, \lambda_0$, and λ_1, must be equal to the coefficients $\beta^{(n)}$ from the time series regressions of excess returns on the estimated factor innovations $\hat{v}$ for all maturities n. The restriction is not imposed in the estimation. The difference between the imputed recursive pricing parameters $\hat{B}_n$ and the regression coefficients $\hat{\beta}^{(n)}$ can serve as a diagnostic of how well the affine model can replicate Treasury return dynamics.

Fig. 4 provides plots of the two sets of coefficients. They are visually indistinguishable for all maturities for the first three factors. While they differ slightly for the fourth and fifth factors, that difference is economically negligible. Analytic standard error bands for $\boldsymbol{\beta}$ are plotted in gray. The tight intervals indicate that these elements are estimated very precisely. To summarize, the five-factor specification of our affine term structure model appears to be able to replicate the dynamics of Treasury yields and returns very well.

3.4. Four-factor specification

The results above show that the predominant share of the variation of expected excess bond returns in the five-factor specification is driven by the second, fourth, and fifth principal components of Treasury yields. However, Cochrane and Piazzesi (2008, CP) find that a single return forecasting factor is able to explain the bulk of the predictability in excess returns. They augment this factor with the first three principal components of orthogonalized yields and estimate an affine term structure model, positing a highly restricted form for the price of risk matrix

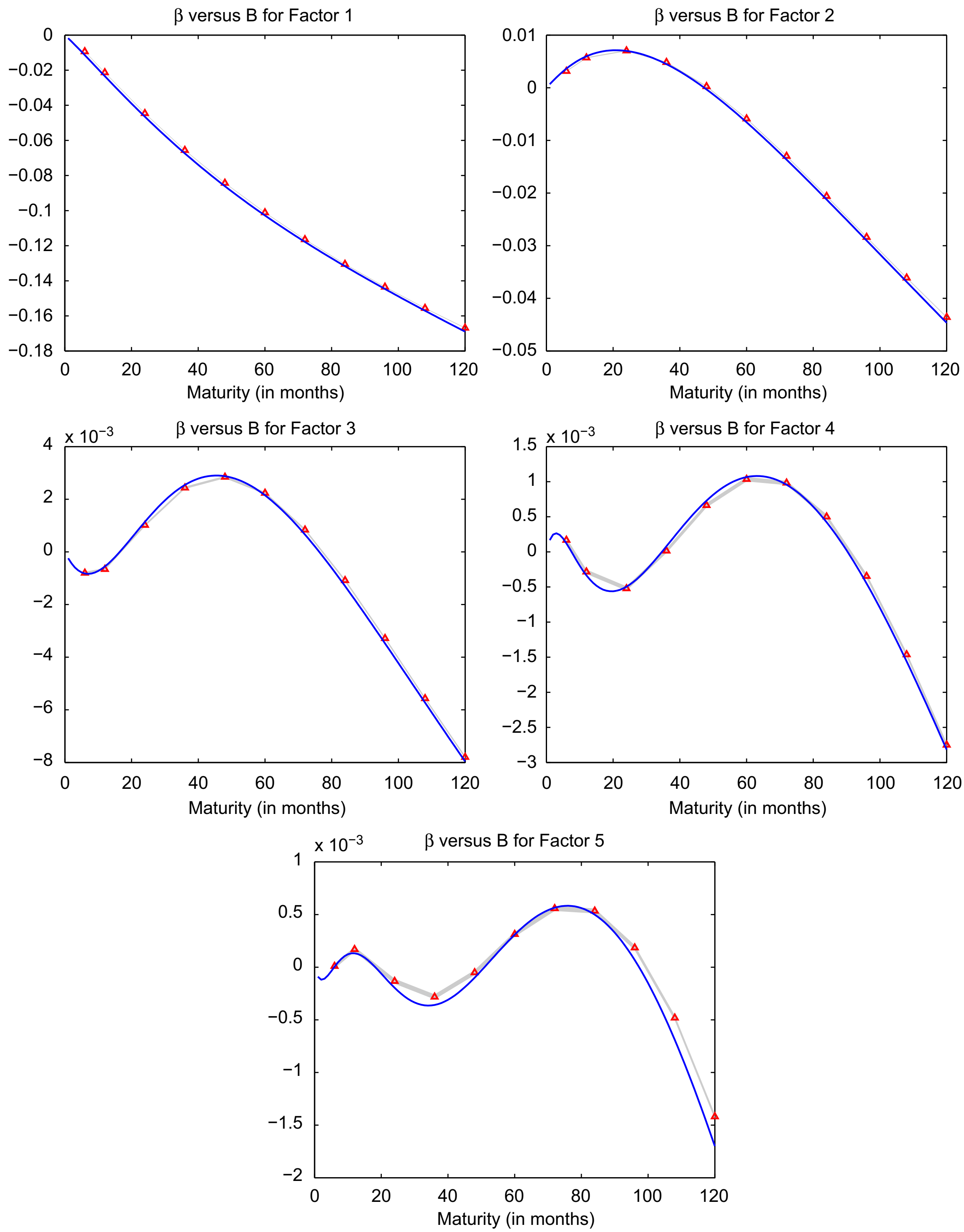

Fig. 4. Regression Coefficients $\beta^{(n)}$ versus recursive pricing parameters B_n. This figure provides plots of the coefficients $\hat{\beta}^{(n)}$ from the regression equation (15) of log excess holding period returns on the state variable innovations versus the recursive pricing parameters B_n used to generate fitted yields given in Eq. (26) for the five factor specification. The diamonds represent the former, and the solid lines correspond to the latter. The confidence intervals for $\hat{\beta}^{(n)}$ are plotted as shaded bands.

Λ. In their model, only the upper element of λ_0 and the loading of level risk on the return forecasting factor are allowed to be nonzero. These restrictions effectively imply that only level is allowed to be a priced risk factor and only the return forecasting factor drives time variation in bond risk premiums.

We estimate a four-factor model similar to the one in Cochrane and Piazzesi (2008) within our regression-based framework. This allows us to directly assess the statistical and economic evidence toward such a specification. We compare the results obtained from the four-factor model with that obtained from our baseline five-factor model and discuss the interpretation of the price of risk estimates.

To generate a factor that summarizes one-month excess return predictability optimally in-sample among linear combinations of bond yields, we calculate our version of the CP factor by regressing monthly GSW excess returns onto the vector of ten one month lagged GSW one-year forward rates F_t according to

$$rx_{t+1} = \gamma_0 + \Gamma F_t + \eta_{t+1}. \tag{37}$$

The average R^2 across these individual predictive return regressions is 7.5%. Following Cochrane and Piazzesi (2008), we define the CP factor x_t to be the first principal component of the fitted values $\hat{\Gamma} F_t$, standardized to zero mean and unit variance. The first principal component explains 94.65% of the cross-sectional variance of all fitted one-month excess returns. We estimate the model by augmenting this factor with the first three principal components extracted from the cross section of Treasury yields.

Our approach differs in a few regards from that employed by Cochrane and Piazzesi (2008). Most important, their factor is constructed by placing one-year excess holding period returns on the left hand side of the regression. However, because the specification in Eq. (12) holds exactly only for monthly excess returns, we prefer to extract a factor that best predicts returns at this frequency. Also in contrast to CP, we use ten annual GSW forward rates instead of five Fama and Bliss annual forward rates. This is because we find that the GSW data provide a better in-sample fit for excess returns at the monthly frequency. Finally, unlike CP, we do not orthogonalize yields with respect to the return forecasting factor before extracting principal components. Instead, we define level, slope, and curvature as the first three principal components of the Treasury yields in our data set. We do so to guarantee that the three factors are identical across the model specifications that we compare.

Table 4 reports summary statistics for the yield and return pricing errors of the four-factor model. These indicate that the in-sample fit is somewhat less precise in the four-factor model than in the five-factor model, but with average pricing errors no larger than 2.9 basis points in absolute value, and standard deviations of these errors of less than 5 basis points, the fit is good. The return pricing errors continue to display little serial correlation, whereas yield pricing errors are highly autocorrelated. The upper panels of Fig. 5 plot the observed and fitted values of the two- and ten-year Treasuries, showing that the fitting error is economically negligible. The lower panels of Fig. 5 plot the observed and fitted excess returns as well as the predicted excess returns. The predicted excess returns display significant time variation, comparable to the dynamics observed in the five-factor model.

Table 5 shows that the CP factor is a significant determinant of the prices of risk of all pricing factors except for slope. In fact, for level risk, the CP factor is the only significant pricing factor, driving out the significance of the slope factor that had been present in the five-factor specification. This is in line with the specification in Cochrane and Piazzesi (2008), who restrict this coefficient to be the only nonzero element in λ_1. As in the five-factor model, the estimated price of risk coefficients are economically important in the four-factor specification. In particular, the −0.043 coefficient of the CP factor in the price of level risk implies that a positive 1 standard deviation shock to the CP factor increases the annualized expected excess return on the ten-year Treasury bond by 8.7%. The slope factor is not significant for any of the price of risk parameters in the four-factor specification, also in line with CP. However, the level and curvature factors continue to remain significant determinants of the price of curvature risk. This implies that the CP factor does not entirely subsume the variation of expected bond returns that is captured by the principal components in the five-factor specification.

Table 4
Four-factor model: fit diagnostics.

This table summarizes the time series properties of the pricing errors implied by the four factor specification. The sample period is 1987:01–2011:12. Reported are the sample mean, standard deviation, skewness, and kurtosis of the errors; $\rho(1), \rho(6)$ denote their autocorrelation coefficients of order one and six. Panel A reports properties of the yield pricing errors $\hat{u}$, and Panel B reports properties of the return pricing errors $\hat{e}$. $n = 12, \ldots, 120$ denotes the maturity in months.

Summary Statistics	$n = 12$	$n = 24$	$n = 36$	$n = 60$	$n = 84$	$n = 120$
Panel A: Yield pricing errors						
Mean	0.028	0.010	−0.013	−0.029	−0.018	0.015
Standard deviation	0.048	0.030	0.026	0.029	0.021	0.028
Skewness	0.653	0.179	−0.358	−0.995	−0.399	0.662
Kurtosis	6.540	2.638	2.842	4.282	3.605	3.718
$\rho(1)$	0.769	0.843	0.912	0.858	0.877	0.748
$\rho(6)$	0.612	0.532	0.780	0.606	0.630	0.478
Panel B: Return pricing errors						
Mean	0.025	−0.041	−0.074	−0.037	0.024	0.142
Standard deviation	0.539	0.509	0.382	0.896	0.924	2.445
Skewness	2.042	0.384	0.566	0.126	0.592	−0.308
Kurtosis	18.013	5.534	5.386	4.411	5.144	6.829
$\rho(1)$	−0.073	−0.055	0.027	−0.147	−0.128	−0.337
$\rho(6)$	0.228	0.142	0.171	0.057	0.072	−0.017

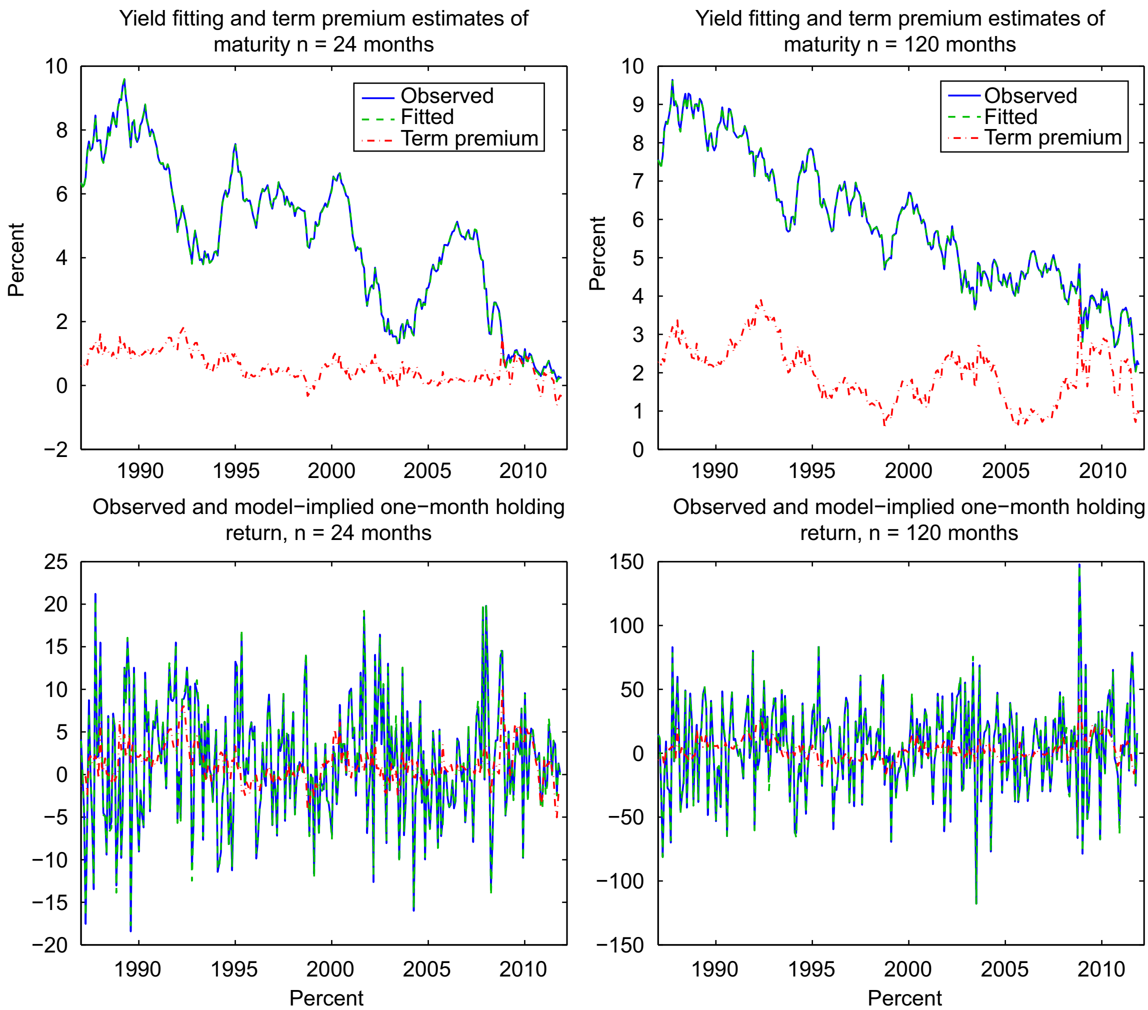

Fig. 5. Four-factor model: observed and model-implied time series. This figure provides plots of the yields and excess one-month holding returns for the two-year and ten-year zero coupon maturities as observed and implied by the four-factor specification. The observed yields and returns are plotted by solid lines, and dashed lines correspond to model-implied yields and returns. Dash-dotted lines show model-implied term premiums in the upper two charts and expected excess holding period returns in the lower two charts.

Table 5
Four-factor model: market prices of risk.

This table summarizes the estimates of the market price of risk parameters λ_0 and λ_1 for the four factor specification. t-Statistics are reported in parentheses. The standard errors have been computed according to the formulas from Section 2.3. Wald statistics for tests of the rows of Λ and of λ_1 being different from zero are reported along each row, with the corresponding p-values in parentheses below. PC1,…, PC5 denote the first through fifth principal components of Treasury yields. Bolded coefficients represent significance at the 5% level.

Factor	λ_0	$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{1,3}$	$\lambda_{1,4}$	W_Λ	W_{λ_1}
PC1	**−0.019**	−0.001	0.003	0.003	**−0.043**	**37.239**	**30.393**
(t-statistic)	(−2.597)	(−0.195)	(0.325)	(0.420)	(−5.002)	(0.000)	(0.000)
PC2	0.012	0.026	0.000	0.002	−0.025	6.746	5.981
(t-statistic)	(0.865)	(1.801)	(0.005)	(0.113)	(−1.524)	(0.240)	(0.201)
PC3	−0.015	−0.063	−0.051	**−0.109**	**0.111**	**18.326**	**18.110**
(t-statistic)	(−0.454)	(−1.862)	(−1.366)	(−3.186)	(2.861)	(0.003)	(0.001)
CP	−0.087	0.019	0.098	−0.010	**−0.291**	**27.851**	**25.546**
(t-statistic)	(−1.728)	(0.374)	(1.753)	(−0.188)	(−4.872)	(0.000)	(0.000)

These results suggest that the parameterization chosen by Cochrane and Piazzesi (2008) for the price of risk dynamics might be overly restrictive. Using a Wald statistic akin to Eq. (33), we can explicitly test for the joint significance of the elements of Λ that CP restrict to be zero. We overwhelmingly reject the null that these elements are jointly equal to zero with a p-value of 10^{-6}. Hence, we find that the tight restrictions imposed by Cochrane and

Piazzesi (2008) on the price of risk dynamics are clearly not supported in the model on statistical terms. Moreover, imposing the CP restriction on Λ using the minimum distance estimator discussed in Section 4.2, we find that the fit of the four-factor model deteriorates, with yield pricing errors increasing substantially across all maturities but especially at the long end of the yield curve.

That being said, the predictability in excess returns is clearly dominated by the CP factor within the four-factor specification. As can be seen in the lower right panel of Fig. 6, the CP factor predicts positive excess returns across the yield curve in a manner increasing in maturity. This closely matches the pattern shown by Cochrane and Piazzesi (2008) for one-year excess returns.

Given these findings and the results from the five-factor model, it is instructive to study the relation between the CP factor and the principal components of Treasury yields. Interestingly, we find that even five principal components fail to entirely span the CP factor: A regression of x_t on the five PCs from our benchmark model yields an R^2 of only 65.6%. This is especially striking given that 99.99% of variation in yields are explained by the first five principal components, considering that x_t is constructed as a linear combination of these yields. The pairwise correlations between the return forecasting factor and the first five principal components of Treasury yields in our data set are 5%, 45%, 19%, −17%, and −62%, respectively. This corroborates the findings of Section 3.3 where we show that the fifth principal component is a predominant driver of risk premiums in the five-factor specification, followed by the second, third, and fourth principal components.

Finally, the curvature and CP factors are both found to be priced in the four-factor specification, as indicated by the corresponding Wald statistics. This is in contrast to the restrictions imposed by Cochrane and Piazzesi (2008), who force these risk prices to be zero.

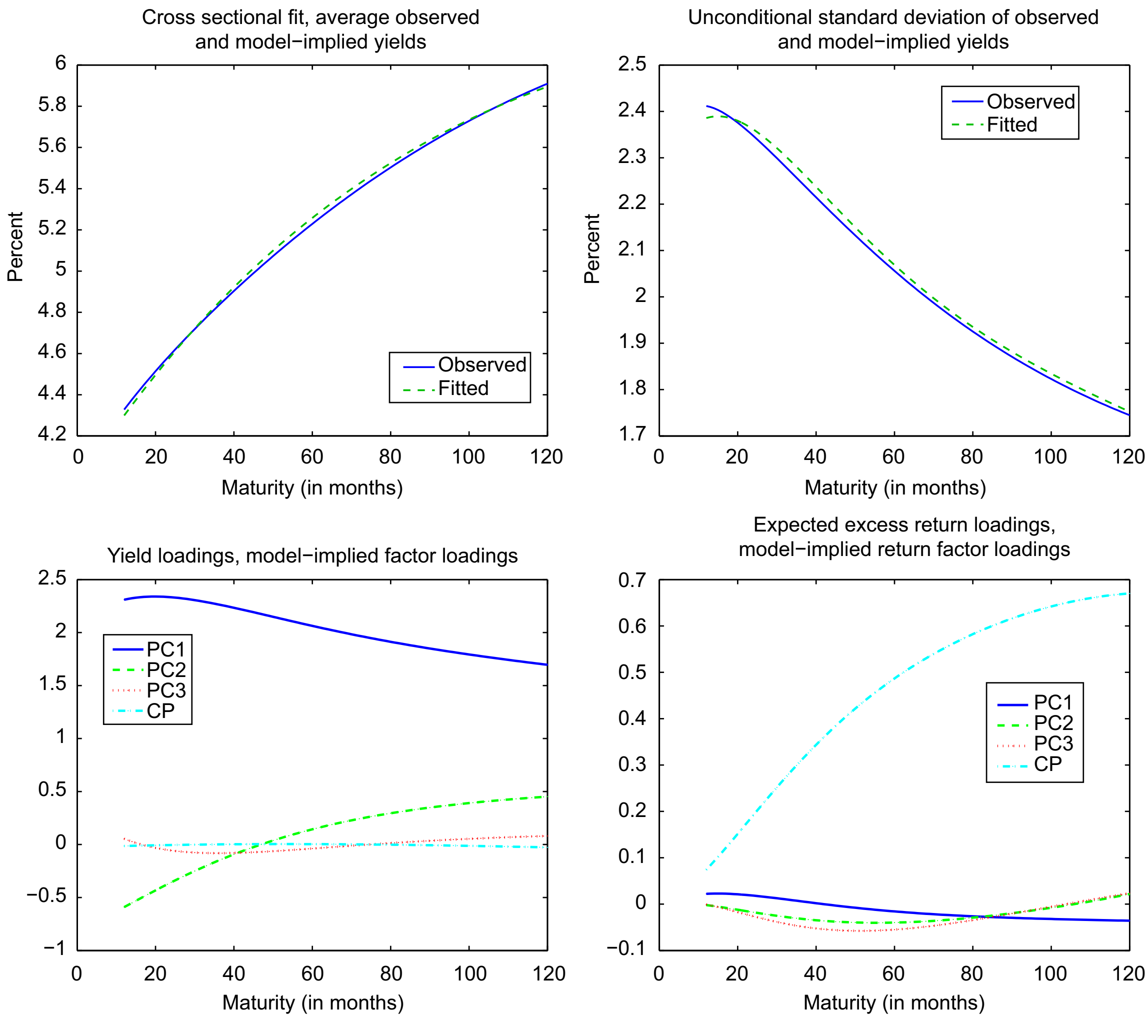

Fig. 6. Four-factor model: cross-sectional diagnostics. This figure provides graphs exhibiting the cross-sectional fit and interpretation of the factors in the four factor specification. The upper two panels plot the unconditional means and standard deviations of observed yields against those implied by the model. The lower left panel plots the implied yield loadings $-(1/n)B_n$. These coefficients can be interpreted as the response of the n-month yield to a contemporaneous shock to the respective factor. The lower right panel plots the expected return loadings $B'_n\lambda_1$. These coefficients can be interpreted as the response of the expected one-month excess holding return on an n-month bond to a contemporaneous shock to the respective factor. PC1,…, PC5 denote the first through fifth principal components of Treasury yields. CP denotes the Cochrane and Piazzesi (2008) return forecasting factor.

We again assess the economic significance of the various risk prices by computing their contribution to the variability of the pricing kernel. We find that risk prices of all four factors contribute to the volatility of the pricing kernel (see the upper right chart in Fig. 3). Yet, as in the five-factor model, only the exposure of long term bonds to level risk is sizable. Hence, the market prices of risk of the remaining factors have a modest impact, at best, on expected excess returns. As Fig. 3 shows, it is primarily the time variation of level risk that contributes to time varying expected excess returns on the ten-year Treasury bond. This time variation, in turn, is largely but not exclusively driven by movements in the CP factor, as highlighted in the lower right panel of Fig. 6.

To summarize, the results in this subsection have shown that strong support exists for the importance of a Cochrane and Piazzesi-type return forecasting factor in explaining Treasury return dynamics. However, the tight parametric restrictions that CP impose on risk price dynamics are statistically not supported by our estimates. Finally, comparing the term premium implied by the four-factor specification with that of the five-factor model, we conclude that the unrestricted four-factor CP specification captures similar term premium dynamics as the five principal component specification but implies somewhat larger in-sample yield fitting errors.

3.5. Comparing the five- and four-factor specifications

The evidence presented thus far does not appear sufficient to designate which of the two models provides a better representation of the data. In this subsection, we consider their out-of-sample performance both in the cross section and in the time series. We find that the five-factor model outperforms the four-factor model across both dimensions. To provide an additional benchmark, we report results for a four-factor model with *CP* treated as an unspanned factor and a five-factor model with *PC*4 and *PC*5 treated as unspanned factors (Section 4.1 explains how to impose the unspanning restriction in the estimation). The upper-left panel of Fig. 7 shows that the in-sample term premium dynamics are very similar across all specifications. In particular, the five-factor models and the four-factor models exhibit closely aligned term premiums. Term premiums are of economic interest because they allow inference of risk-adjusted expectations of the path of future short-term interest rates. Thus, a natural way to compare the different models is to investigate their ability to predict future short-term interest rates. To generate these forecasts, we begin by using the period 1987:01–1991:12 as a training sample and then proceed to reestimate each model at monthly intervals with an expanding estimation window and compare their average short-rate predictions up to five years ahead with the realized data. For comparison, we also include the average forecast error from a simple random walk of the one-month T-bill. In the upper-right panel of Fig. 7, we display the root mean square prediction error of average short rates at forecast horizons from one month through five years. The chart shows that the five-factor model and the model with *PC*4 and *PC*5 treated as unspanned factors outperform the random walk across all horizons. In contrast, the four-factor model and the model with *CP* treated as an unspanned factor display somewhat larger forecast errors, most notably at longer horizons. This short rate prediction exercise thus suggests that the five-factor model provides a better forecast performance in the time series than the four-factor model.

To assess the cross-sectional performance of the different model specifications we perform the following exercise. We continue to estimate each model using returns on bonds with maturities up to ten years. We then compare the model-implied yields on bonds with maturities up to 20 years relative to their actual values. In the lower-left panel of Fig. 7 we plot the results from this analysis. Although the four-factor model produces a good in-sample fit to maturities up to ten years, the chart shows that the model-implied yields for longer maturities are far below their realized values. This is not the case for the five-factor specification, which produces slightly higher average yields than in the realized data, but closer to the realized data than either of the specifications using unspanned factors. This suggests that the four-factor model strongly understates the degree of persistence of the pricing factors under the risk-neutral measure, whereas the five-factor model only slightly overstates it. Interestingly, Joslin, Singleton, and Zhu (2011), using their likelihood-based estimation approach, find that a model using all of the first five principal components of yields as pricing factors and estimated without observations on longer maturity bonds produces implausible average bond yields on very long maturity bonds. Here we show, using our methodology, that this is not a general feature of five-factor models. As a robustness check, the lower-right panel of Fig. 7 displays the in-sample fit when each model is estimated using returns on bonds with maturities up to 20 years. In this case, the five-factor model is very close to the observed yields, whereas the four-factor model exhibits more pronounced deviations.

In sum, the results presented in this subsection suggest that relative to the four-factor model with three principal components of yields and the CP factor, the five-factor model produces better out-of-sample forecasts of future short rates and is more robust to the choice of maturities used in estimation. While all of these specifications capture term premium dynamics in similar ways, the superior out-of-sample performance of the five-factor model leads us to designate it as our preferred specification.

3.6. Estimation using maturity-sorted portfolios

So far, we have seen that different specifications of multi-factor affine models estimated using our regression-based approach are able to tightly match data obtained from a smooth zero coupon yield curve. However, an important feature of our estimation approach is that the availability of a wide cross section of zero coupon bonds is not necessary for the estimation of the term structure. The only data required for our procedure are a panel of excess returns and a corresponding set of return spanning factors. Our third benchmark specification highlights this attractive feature of the regression-based approach. Here, we generate a no-arbitrage zero coupon yield curve from a cross section of

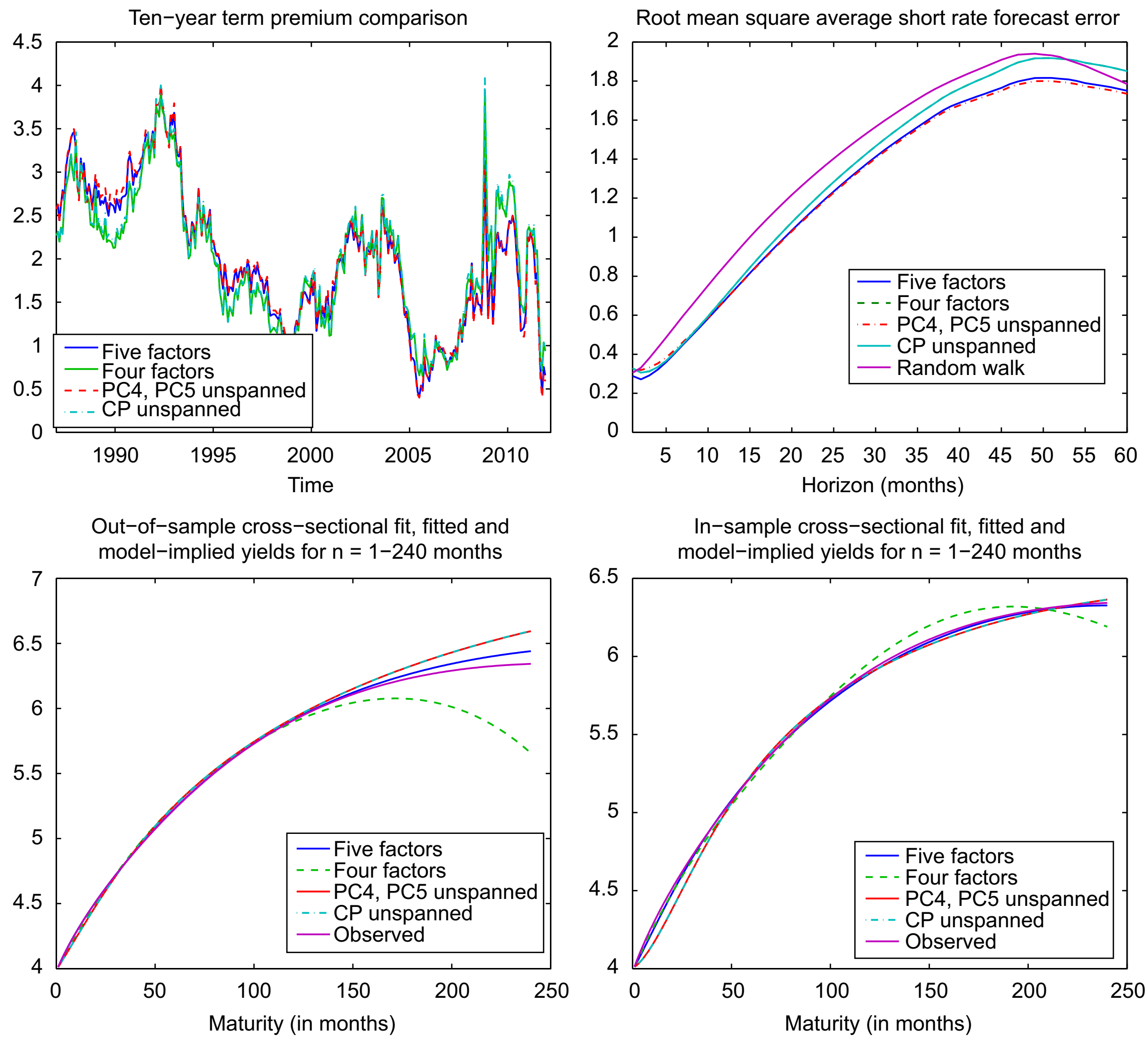

Fig. 7. Model comparison. This figure shows the model comparison diagnostics discussed in Section 3.5. The upper left panel plots the ten-year Treasury term premium implied by the five-factor model, the four-factor model, a five-factor model in which *PC*4, *PC*5 are restricted to be unspanned, and a four-factor model in which the *CP* factor is restricted to be unspanned. The upper right panel shows the out-of-sample root mean squared forecast error (RMSE) for average future short rates up to five years in the future as implied by the four model specifications. It also provides the RMSE for a simple random walk model of the short rate. The lower left panel shows the average yields implied by the four models for maturities up to 20 years, where the models have been estimated using only maturities up to ten years. The lower right chart shows the average yields implied by the four models when maturities up to 20 years have been used in the estimation. PC1, …, PC5 denote the first through fifth principal components of Treasury yields. CP denotes the Cochrane and Piazzesi (2008) return forecasting factor.

$K=5$ principal components extracted from the Federal Reserve's H.15 Treasury yield data and $N=11$ excess returns on Fama maturity sorted bond portfolios from CRSP. Both sets of data are constructed from coupon bearing bonds. This reduces estimation error with respect to the standard practice of fitting term structure models to zero coupon yields.

Given the excess holding period returns as well as the pricing factors, we estimate the model parameters $(\Phi, \Sigma, \sigma, \lambda_0, \lambda_1)$ using our three-step estimation approach. We further estimate δ_0 and δ_1 by regressing the log Fama one-month T-bill on the pricing factors. Based on these parameters, we can then use Eqs. (25) and (26) to compute the loadings A_n and B_n, which, along with the pricing factors X, allow us to impute zero coupon yields for bonds of any maturity.

The upper panels of Fig. 8 plot the two-year and five-year maturities implied by the model against the corresponding Fama and Bliss discount bond yields, displaying a visually close fit. It is important to note that this is not by construction, as the two sets of zero coupon yields have been estimated using different methodologies. The Fama and Bliss discount bond yields are fitted to the cross section of Treasury bonds period by period. Instead, our estimated zero curve is directly implied by a no-arbitrage term structure model, which imposes internal consistency between the time series and the cross section of yields. The lower panels of Fig. 8 show that the model-implied excess returns for zero coupon bonds closely match the observed excess returns of Fama maturity sorted portfolios. In sum, we see it as a particular strength of our

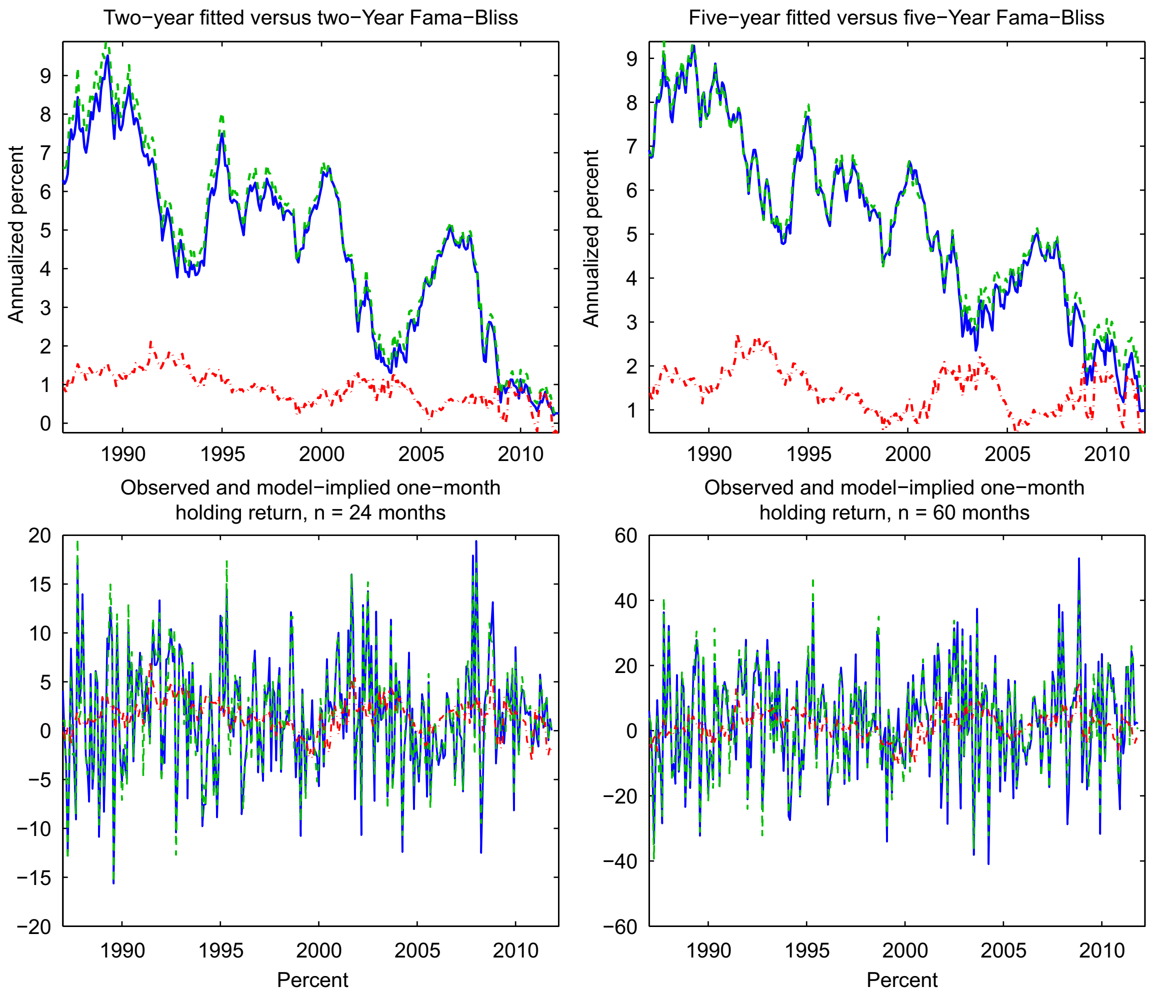

Fig. 8. Fama maturity-sorted portfolio model: observed and model-implied time series. This figure provides plots of the model-implied yields for the two-year and five-year maturities from the Fama maturity-sorted portfolio specification detailed in Section 3.6. Excess returns on maturity-sorted Fama bond portfolios are regressed on a set of spanning factors extracted from principal components of Federal Reserve H.15 bond yields. The top two panels plot the fitted yields from this specification against the Fama and Bliss zero coupon yields. The bottom two panels plot the observed excess returns on Fama maturity-sorted Treasury portfolios (solid lines) against the model-implied fitted returns (dashed lines). The dash-dotted lines show model-implied term premiums in the upper two charts and expected excess holding period returns in the lower two charts.

estimation approach to generate a model-based zero coupon yield curve from returns on coupon-bearing securities.

The use of excess returns from maturity-sorted portfolios could cause problems due to the changing composition of underlying securities. To address this concern, we first employ the test of parameter stability described by Andrews (1993). The generalized method of moments (GMM) framework of Andrews (1993) is easily applied to the OLS regression in Eq. (15). We obtain a value of 31.70 for the sup–*LM* statistic [we choose a standard calculation window of $\pi=(15\%, 85\%)$], which is well below the corresponding 90% critical value. In addition to assessing the stability of estimated betas for each portfolios' return, we approximate the portfolios' betas at every point by the weighted average of the betas corresponding to the individual cash flows of the underlying securities (principal and coupon payments). Doing so, we see that the betas at the portfolio level are very stable for maturities up to ten years but exhibit pronounced time variation for the portfolio containing bonds that mature in ten years or more. This latter result is not surprising as large differences exist in duration between the different securities in that portfolio and, hence, the interest rate sensitivity of that portfolio can vary considerably over time. Consequently, we use the maturity-sorted portfolios up to ten years and drop the Fama portfolio with maturities above ten years in the estimation.

4. Extensions

In this section we present a number of extensions and further results. First, we explain how to estimate the model in the presence of unspanned factors. Second, we show how to impose linear restrictions on risk exposures and market prices of risk in the estimation of the model.

Third, we show that the maximal Sharpe ratios implied by the four- and five-factor specifications are very reasonable, with an average level below one and peaks below 2.5. Fourth, we explain how the model can be used to fit the yield curve at the daily frequency. Finally, we show that the implied principal component loadings from the term structure model are statistically indistinguishable from the actual principal component loadings.

4.1. Unspanned factors

Recent models of the yield curve have featured unspanned factors that do not affect the dynamics of bonds under the pricing measure, but do affect them under the historical measure (see Joslin, Priebsch, and Singleton, 2012; Duffee, 2011; Wright, 2011). The predictability under the historical measure is consistent with previous work that finds that macroeconomic factors have forecasting power for the term structure (see, e.g., Moench, 2008; Ludvigson and Ng, 2009). The assumption that a given factor does not affect bond yields under the pricing measure can be implemented by imposing the restriction that the corresponding elements of $\{B_n, n=1,\dots,N\}$ be exactly equal to zero.

This restriction can be readily incorporated in our regression-based setting. Partition the factors into spanned factors X_t^s with nonzero risk exposures and unspanned factors X_t^u, which have zero risk exposures. The factors continue to follow a joint VAR process under the historical measure

$$\begin{bmatrix} X_t^s \\ X_t^u \end{bmatrix} = \mu + \Phi \begin{bmatrix} X_{t-1}^s \\ X_{t-1}^u \end{bmatrix} + \begin{bmatrix} v_t^s \\ v_t^u \end{bmatrix}, \tag{38}$$

where X_t^s is of dimension $K_s \times 1$, X_t^u is of dimension $K_u \times 1$, and μ and Φ are partitioned accordingly. The spanning restriction is that the risk exposures of the unspanned factors are equal to zero, i.e., $\beta^{(n)} = [\beta_s^{(n)'}\ 0]'$ for all $n=1,\dots,N$. Using Eqs. (26) and (28), this restriction implies $\delta'_1 = [\delta_1^{s'}\ 0]$, so that the short rate does not load on any unspanned factors. It further implies that the upper right $K_s \times K_u$ block of the risk-neutral transition matrix $\Phi^\star = (\Phi - \lambda_1)$ is zero, i.e.,

$$\Phi^\star = \begin{bmatrix} \Phi_{ss} - \lambda_1^{ss} & 0 \\ \Phi_{us} - \lambda_1^{us} & \Phi_{uu} - \lambda_1^{uu} \end{bmatrix}. \tag{39}$$

With this restriction, the return generating process can be rewritten as

$$rx_{t+1}^{(n-1)} = \beta^{(n-1)'}(\lambda_0 + \lambda_1 X_t + v_{t+1}) - \tfrac{1}{2}(\beta^{(n-1)'}\Sigma\beta^{(n-1)} + \sigma^2) + e_{t+1}^{(n-1)} \tag{40}$$

$$rx_{t+1}^{(n-1)} = \beta^{(n-1)'}(\lambda_0 + \lambda_1 X_t + X_{t+1} - \Phi X_t - \mu) - \tfrac{1}{2}(\beta^{(n-1)'}\Sigma\beta^{(n-1)} + \sigma^2) + e_{t+1}^{(n-1)} \tag{41}$$

$$rx_{t+1}^{(n-1)} = -\beta_s^{(n-1)'}(\mu_s^\star + \Phi_{ss}^\star X_t^s) - \tfrac{1}{2}(\beta_s^{(n-1)'}\Sigma_{ss}\beta_s^{(n-1)} + \sigma^2) + \beta_s^{(n-1)'}X_{t+1}^s + e_{t+1}^{(n-1)}, \tag{42}$$

where $\mu_s^\star$ denotes the upper $K_s \times 1$ subvector of the risk-neutral mean $\mu^\star = (\mu - \lambda_0)$, $\Phi_{ss}^\star$ denotes the upper left $K_s \times K_s$ block of $\Phi^\star$, and Σ_{ss} is equal to the upper left $K_s \times K_s$ block of Σ. Estimation of this model proceeds with only a slight modification of the three-step procedure. As before, we obtain estimates of the VAR parameters and innovations in Eq. (38) $\hat{v}_{t+1}$ using OLS. Denote X_-^s and X^s as the stacked lagged and contemporaneous values of spanned factors. We then estimate the regression

$$rx = \mathbf{a}_s \iota'_T + \mathbf{c}_s X_-^s + \boldsymbol{\beta}'_s X^s + E, \tag{43}$$

where $\mathbf{a}_s$ is an $N \times 1$ vector and $\boldsymbol{\beta}'_s$ and $\mathbf{c}_s$ are $N \times K_s$ matrices. We see from Eq. (40) that now $\mathbf{a}_s = -\boldsymbol{\beta}'_s \mu_s^\star - \frac{1}{2}(B^{\star s}\,\mathrm{vec}(\Sigma_{ss}) + \sigma^2 \iota_N)$ with $B^{\star s}$ the first K_s^2 columns of $B^\star$. We can also see that $\mathbf{c}_s = -\boldsymbol{\beta}'_s \Phi_{ss}^\star$. We can then estimate the parameters governing the risk-neutral dynamics of the pricing factors according to

$$\hat{\mu}_s^\star = -(\hat{\boldsymbol{\beta}}_s \hat{\boldsymbol{\beta}}'_s)^{-1}\hat{\boldsymbol{\beta}}_s(\hat{\mathbf{a}}_s + \tfrac{1}{2}(\hat{B}^{\star s}\,\mathrm{vec}(\hat{\Sigma}_{ss}) + \hat{\sigma}^2\iota_N)) \tag{44}$$

and

$$\hat{\Phi}_{ss}^\star = -(\hat{\boldsymbol{\beta}}_s \hat{\boldsymbol{\beta}}'_s)^{-1}\hat{\boldsymbol{\beta}}_s \hat{\mathbf{c}}_s. \tag{45}$$

Denoting $\hat{\Psi}_{ss}^\star = [\hat{\mu}_s^\star\ \hat{\Phi}_{ss}^\star]$, we can write these estimators in the single expression

$$\hat{\Psi}_{ss}^\star = -(\hat{\boldsymbol{\beta}}_s \hat{\boldsymbol{\beta}}'_s)^{-1}\hat{\boldsymbol{\beta}}_s(rx + \tfrac{1}{2}(\hat{B}^{\star s}\,\mathrm{vec}(\hat{\Sigma}_{ss}) + \hat{\sigma}^2\iota_N)\iota'_T)M_X^s Z_-^{s'}(Z_-^s M_X^s Z_-^{s'})^{-1}, \tag{46}$$

where $Z_-^s = [\iota_T\ X_-^{s'}]'$ and $M_x^s = I_T - X^{s'}(X^s X^{s'})^{-1}X^s$. We then obtain

$$\hat{\Lambda}_{ss} = \hat{\Psi}_{ss} - \hat{\Psi}_{ss}^\star \tag{47}$$

We further set

$$\hat{\lambda}_1^{su} = \hat{\Phi}_{su} \tag{48}$$

in accordance with the spanning restriction in Eq. (39). The parameters $\mu_u^\star$ and $\Phi_{u\cdot}^\star$ are not identified in this model. They do not matter for bond pricing, so we adopt the convention of setting the corresponding prices of risk $\hat{\lambda}_0^u$ and $\hat{\lambda}_1^{u\cdot}$ equal to zero. This is equivalent to setting the risk neutral parameters equal to the physical VAR estimates $\hat{\mu}_u$ and $\hat{\Phi}_u$.

We can see from Eqs. (47) and (48) above that the estimator of market prices of risk in the unspanned factor case is a linear combination of estimated parameters governing the historical and risk-neutral dynamics of pricing factors. We show that the two estimators have the joint limiting distribution

$$\sqrt{T}\begin{bmatrix} \mathrm{vec}(\hat{\Lambda}_{ss} - \Lambda_{ss}) \\ \mathrm{vec}(\hat{\lambda}_1^{su} - \lambda_1^{su}) \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \mathcal{V}_\Lambda^{ss} & \mathcal{C}_{\Lambda,\lambda_1}^{ss,su} \\ \mathcal{C}_{\Lambda,\lambda_1}^{ss,su'} & \mathcal{V}_{\lambda_1}^{su} \end{bmatrix} \right), \tag{49}$$

where $\mathcal{V}_\Lambda^{ss}$, $\mathcal{V}_{\lambda_1}^{su}$ and $\mathcal{C}_{\Lambda,\lambda_1}^{ss,su}$ are provided in the Appendix.

We illustrate estimation of an affine model with unspanned macroeconomic factors by using the specification in Joslin, Priebsch, and Singleton (2012). Their model features as pricing factors the first three principal components of Treasury yields as well as the first principal component of monthly core consumer price index inflation and monthly core personal consumption expenditures inflation and the Federal Reserve Bank of Chicago National Activity Index (CFNAI) as unspanned macroeconomic factors. According to a Wald test of the significance of columns of beta obtained by including the additional factors in the regression in Eq. (15), we do not reject the null that the inflation factor has zero betas with p=88.63.

Similarly, we do not reject that the betas for the exponentially smoothed CFNAI equal zero with $p = 96.81$. These findings, thus, justify the assumption of treating these macroeconomic factors as unspanned in a specification that also features the first three principal components of Treasury yields.

Unreported results show that the unspanned macro factor specification provides a somewhat poorer fit to the cross section of Treasury yields than the four- and five-factor specifications. Estimates of the market prices of risk and the corresponding test statistics are provided in Table 6. They differ in some ways from the ones already discussed. As in the five-factor specification, the second principal component significantly drives time variation in slope risk. Nonetheless, we can reject only that the first row of λ_1 is equal to zero at a 10% level. This is at odds with the implications of the five- and four-factor models which show substantial amounts of time variation in level risk that are largely due to movements in higher order principal components or the CP factor, respectively. This suggests that the specification proposed by Joslin, Priebsch, and Singleton (2012), which uses the first three principal components of yields and two macroeconomic variables as pricing factors, does not fully capture the dynamics of the price of level risk implied by the yield factor-based specifications. Furthermore, the coefficient of the price of slope risk on the CFNAI series is highly statistically significant and, as a consequence, slope risk is priced in this model specification. This is consistent with the results of Joslin, Priebsch, and Singleton (2012) but is in contrast to the pricing implications of the four- and five-factor models. This can be interpreted as indicating that the information contained in the yield curve is insufficient to completely characterize the time variation in the price of slope risk. Due to the ease of estimation of the term structure model in the regression-based approach, alternative specifications with other unspanned factors are straightforward to estimate.

4.2. Restricted market prices of risk

Numerous authors have considered affine term structure models in which certain elements of Λ are set equal to zero. These models could be easily estimated in our framework using a minimum distance procedure as follows. Let $\theta = (\mathrm{vec}(\Lambda)', \mathrm{vec}(\boldsymbol{\beta})', \mathrm{vec}(\Phi)', \sigma^2, \mathrm{vech}(\Sigma))'$ and $\hat{\theta}$ similarly, where $\mathrm{vech}(\cdot)$ is the vector-half operator. Let H be a $q \times p$ known matrix with full row rank, where $p = (K(K+1) + NK + K^2 + 1 + K(K+1)/2)$. This matrix contains the linear restrictions we would like to place on Λ. For example, if we wanted to impose a linear restriction only on the second element of λ_0, we would choose $H = [0\ 1\ 0 \cdots 0]$.

Under these restrictions, the minimum distance estimator solves

$$\min_{\theta} T \cdot (\hat{\theta} - \theta)' W_T (\hat{\theta} - \theta) \quad \text{s.t. } H\theta = 0, \tag{50}$$

where W_T is a $p \times p$ positive definite weighting matrix. Because the restrictions on θ are linear, the solution to this optimization problem is

$$\hat{\theta}_{\mathrm{md}} = \hat{\theta} - W_T^{-1} H' (H W_T^{-1} H')^{-1} H \hat{\theta}. \tag{51}$$

Then, by construction $H\hat{\theta}_{\mathrm{md}} = 0$, so we have new estimators for the parameters that satisfy the desired restrictions. The optimal choice of weighting matrix is a consistent estimator of the inverse of the asymptotic variance of the unconstrained estimator $\hat{\theta}$ (i.e., if $\mathcal{V}_\theta$ is the asymptotic variance of $\hat{\theta}$, then $W_T \rightarrow_p \mathcal{V}_\theta^{-1}$). In the Appendix we provide the elements of $\mathcal{V}_\theta$. Under this choice of weighting matrix,

$$\sqrt{T}(\hat{\theta}_{\mathrm{md}} - \theta) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_\theta - \mathcal{V}_\theta H' (H \mathcal{V}_\theta H')^{-1} H \mathcal{V}_\theta). \tag{52}$$

4.3. Sharpe ratios

From Eq. (8), the conditional Sharpe ratio can be expressed as

$$\frac{\mathrm{E}_t[rx_{t+1}^{(n-1)}] + \frac{1}{2}\mathrm{Var}_t[rx_{t+1}^{(n-1)}]}{\sqrt{\mathrm{Var}_t[rx_{t+1}^{(n-1)}]}} = \frac{Cov_t[rx_{t+1}^{(n-1)}, \ln M_{t+1}]}{\sqrt{\mathrm{Var}_t[rx_{t+1}^{(n-1)}]}}. \tag{53}$$

The Sharpe ratio in Eq. (53) is maximized for a hypothetical return to a portfolio of bonds that replicates the payoff to the log pricing kernel exactly. The model-implied maximal Sharpe ratio is, therefore,

$$\sqrt{\mathrm{Var}_t(\ln M_{t+1})} = \sqrt{\lambda'_t \lambda_t}. \tag{54}$$

The maximal Sharpe ratio provides a useful diagnostic of the validity of the stochastic discount factor. Duffee (2010) argues that five-factor affine models of the term structure can give rise to excessively high maximal Sharpe ratios due to overfitting. However, we do not find unreasonable maximal Sharpe ratios in our four- and five-factor specifications, as can be seen in Fig. 9. For the five-factor specification, the peak in the maximal Sharpe ratio is

Table 6
Macro factor model: market prices of risk.
This table summarizes the estimates of the market price of risk parameters λ_0 and λ_1 for the unspanned macro specification. t-Statistics are reported in parentheses. Standard errors have been computed according to the formulas from Section 4.1. Wald statistics for tests of the rows of Λ and of λ_1 being different from zero are reported along each row, with the corresponding p-values in parentheses. PC1,…, PC5 denote the first through fifth principal components of Treasury yields. Bolded coefficients represent significance at the 5% level.

Factor	λ_0	$\lambda_{1.1}$	$\lambda_{1.2}$	$\lambda_{1.3}$	$\lambda_{1.4}$	$\lambda_{1.5}$	W_Λ	W_{λ_1}
PC1	**−0.019**	−0.028	**−0.018**	−0.005	−0.022	0.017	**15.746**	9.423
(t-statistic)	(−2.492)	(−1.719)	(−2.062)	(−0.541)	(−1.360)	(1.835)	(0.015)	(0.093)
PC2	0.010	0.025	−0.028	−0.030	−0.028	**−0.057**	**18.764**	**18.264**
(t-statistic)	(0.700)	(0.850)	(−1.837)	(−1.853)	(−0.975)	(−3.412)	(0.005)	(0.003)
PC3	0.000	−0.093	−0.019	**−0.099**	−0.057	−0.030	9.972	9.972
(t-statistic)	(0.000)	(−1.373)	(−0.528)	(−2.621)	(−0.865)	(−0.797)	(0.126)	(0.076)

2.30, with a sample average maximal Sharpe ratio below one.

As Eq. (54) shows, the maximal Sharpe ratio implied by an affine term structure model depends on the inner product of the market prices of risk and, hence, on the price of risk parameters λ_0 and λ_1. As discussed in Section 4.1, a subset of the latter is not identified in the presence of unspanned factors. We know that the fourth and fifth principal components are weakly spanned only by bond yields because level, slope, and curvature explain almost all of the cross-sectional variation of yields. Accordingly, the rows of λ_0 and λ_1 corresponding to the fourth and fifth principal component in the five-factor specification of the model could only be weakly identified, thus potentially resulting in somewhat higher maximal conditional Sharpe ratios than what one can perceive as reasonable.

To assess whether this weak spanning of the fourth and fifth principal component is a potential concern in the model, we perform the following robustness check. We make the extreme assumption that the fourth and fifth principal components are entirely unspanned and, hence, the corresponding market price of risk parameters is unidentified. We then reestimate the five-factor specification of the model by imposing the spanning restrictions outlined above to the fourth and fifth principal components, effectively setting to zero the lower two rows of λ_0 and λ_1 and fixing the upper right block of λ_1 to be equal to the respective block of $\hat{\Phi}$. As can be seen in the lower left panel of Fig. 9, the resulting Sharpe ratio is estimated at slightly lower values. However, as in Section 3.5, the term premium implied by the restricted five-factor model is very similar when the fourth and fifth principal components are treated as unspanned. In unreported results, we also computed the maximal Sharpe ratio of the five-factor model in which we impose the restriction that the second and fourth principal components are not priced. Consistent with our finding that the prices of risk of these two factors are statistically indistinguishable from zero in the

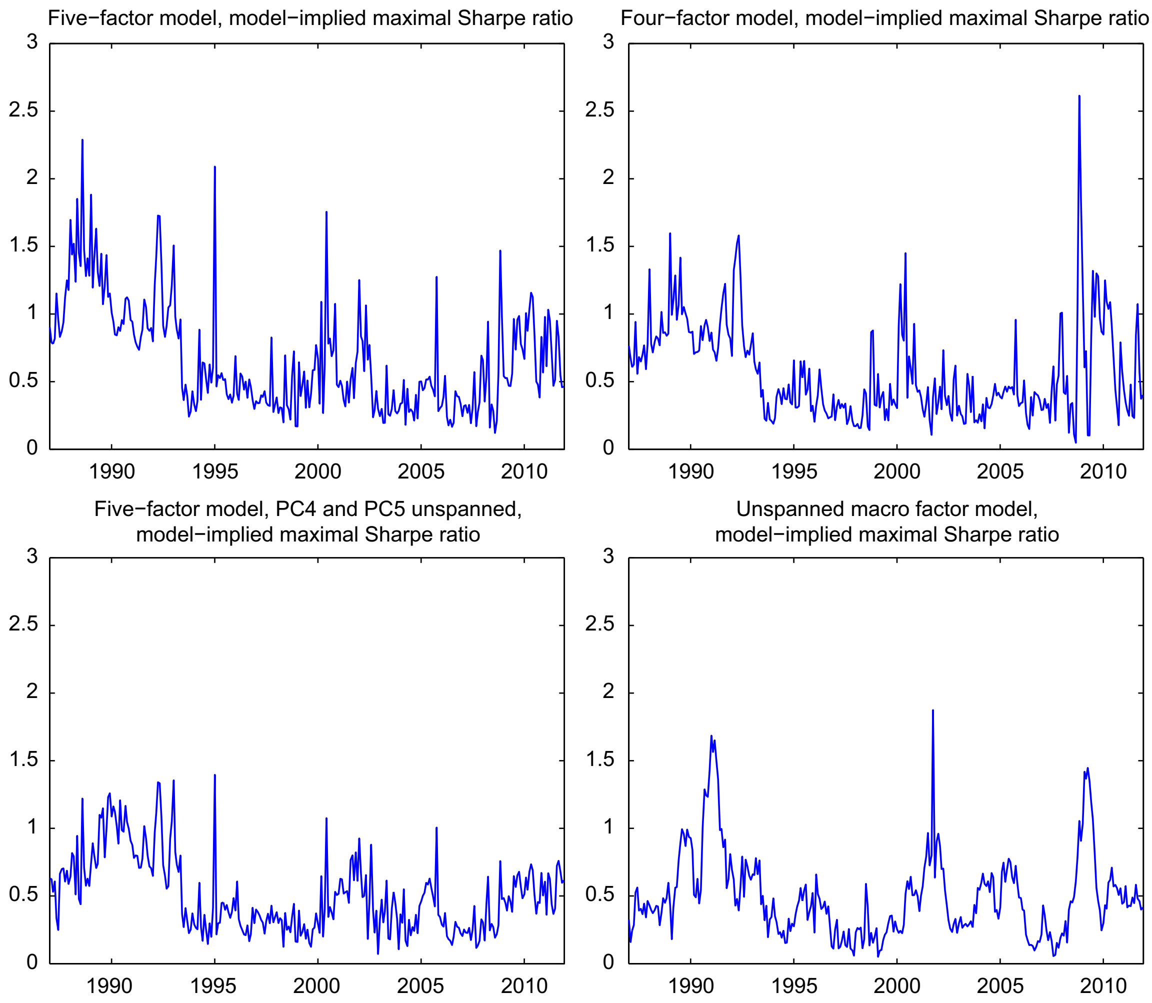

Fig. 9. Model-implied maximal Sharpe ratios. This figure provides plots of the time series of the Sharpe ratio diagnostic $\sqrt{\lambda'_t \lambda_t}$ as described in Section 4.3. The upper left panel plots the Sharpe ratio for the standard five factor specification and the upper right panel plots the Sharpe ratio for the four factor specification. The lower left panel plots the Sharpe ratio for a specification using five principal components as factors but with the fourth and fifth principal components estimated as unspanned factors. The lower right panel plots the Sharpe ratio for the unspanned macro factor specification. PC1,…, PC5 denote the first through fifth principal components of Treasury yields. CP denotes the Cochrane and Piazzesi (2008) return forecasting factor.

full five-factor model, this specification produces a very similar time series of the maximal Sharpe ratio as in the upper left panel of Fig. 9.

We also show the Sharpe ratio of the four-factor CP specification in the upper right panel of Fig. 9. While it is comparable in magnitude to that of the five-factor model, the time series patterns of the two Sharpe ratios do at times differ substantially. For example, during the financial crisis, the Sharpe ratio implied by the four-factor model tends to be higher than that of the five-factor model, while that of the five-factor model is higher during the late 1980s. The lower right-hand panel of Fig. 9 shows the Sharpe ratios of the unspanned macro factor model. This time series of maximal Sharpe ratios is somewhat smoother than the ones implied by the five- and four-factor specifications but otherwise shows similar dynamics.

4.4. Term structure estimation in real time

As we fit the model using simple linear regressions, estimation is extremely fast. This is in sharp contrast to traditional likelihood-based estimation of such models subject to nonlinear cross-equation restrictions, which typically require a long numerical optimization process. The simplicity and the speed at which it is estimated makes our approach particularly appealing for the real time analysis of term structure dynamics. In this subsection, we show how we fit an affine model to the US Treasury yield curve at the daily frequency. This allows us to interpret yield curve movements in terms of risk-neutral yield versus term premiums dynamics in real time.

Estimation at the daily frequency requires a slight modification of the empirical approach outlined so far. In the daily application of the model, we use the daily yields of maturities from $n = 3, \ldots, 120$ months obtained from Gurkaynak, Sack, and Wright (2007). We aggregate the daily yields to the monthly frequency by selecting end-of-month values and extract principal components from these monthly yields. We then apply the weights from the monthly principal components to the daily yields to obtain daily estimates of our pricing factors.[5]

We continue to compute log excess holding period returns at the monthly frequency and obtain estimates of the parameters of the model as outlined in Sections 2.2 and 2.4. Finally, we use the estimated parameters and the daily yield factors to impute model-implied yields and term premiums at the daily frequency. To illustrate the ability of the model to fit daily term structures, Fig. 10 shows for the $K=5$ factor specification the daily observed and model-implied yields for the two-year, five-year, and ten-year Treasury notes since August 2009. The model prices all three maturities precisely.

Perhaps more interesting is an analysis of term premiums over this sample period. As discussed in Section 2.4, we can use the model to achieve a decomposition of interest rates into risk neutral yields and term premiums. The lower right panel of Fig. 10 displays the estimated term premium for the ten-year Treasury along with the one-month Merrill Lynch Option Volatility Estimate (MOVE) index. The latter is a measure of implied volatilities from options on Treasury futures. The plot shows that our term premium estimate and the MOVE index exhibit a strong correlation. This is striking because the model is estimated without using any option data. We interpret the correlation between the two time series as evidence that our term premium estimate reflects the risk of holding Treasury securities.

4.5. Consistency of principal components as observable factors

Traditional estimation methods for affine term structure models typically treat pricing factors as latent variables that are backed out from observed yields using filtering or observation equation inversion techniques. Our estimation approach is different in that it requires the pricing factors X to be observed. We treat principal components extracted from yields as observed, thus ignoring the fitting error associated with the principal component extraction. This potentially gives rise to an inconsistency between actual and model-implied principal components. However, this inconsistency is negligible.

Let X be the $K \times T$ matrix of principal components of yields y. We can write

$$X = PyM_\iota. \tag{55}$$

where $M_\iota = I_{T\times T} - (1/T)\iota_T \iota'_T$ is the $T \times T$ time series demeaning operator and P is the $K \times N$ matrix of principal component loadings. Actual observed yields can be decomposed into fitted yields and fitting errors: $y_t^{(n)} = \hat{y}_t^{(n)} + \hat{\varepsilon}_t^{(n)}$. Then, because model-implied yields are assumed to be affine in the pricing factors X, we have $y_t^{(n)} = \hat{a}_n + \hat{b}'_n X_t + \hat{\varepsilon}_t^{(n)} = -(1/n)(\hat{A}_n + \hat{B}'_n X_t + \hat{u}_t^{(n)})$. We can thus expand Eq. (55) to find

$$X = PyM_\iota = P(\hat{a} \cdot \iota'_T + \hat{b}'X + \hat{\varepsilon})M_\iota = P\hat{b}'X + P\hat{\varepsilon}M_\iota, \tag{56}$$

where we have used the fact that $\iota'_T M_\iota = 0$ and that X is mean zero by construction. For equality to hold over arbitrary values of X, the model must thus satisfy the consistency conditions

$$P\hat{b}' = I_K \quad \text{and} \quad P\hat{\varepsilon}M_\iota = 0. \tag{57}$$

We find that $P\hat{\varepsilon}M_\iota$ is extremely close to zero because the yield fitting error $\hat{\varepsilon}$ is tiny, as shown in Table 2. Moreover, the condition $P\hat{b}' = I_K$ is satisfied to a high degree of precision in the five-factor specification of the model.

We reach this conclusion using the following Monte Carlo procedure. Because their standard errors do not have a simple closed form, we compute bootstrapped distributions of the recursive pricing parameters B_n. We construct the bootstrap by saving the residuals $\hat{v}$ and $\hat{e}$ from the three-step regression estimation of the model parameters as well as the residuals from the regression of the one-month T-bill on the pricing factors. We resample all three sets of residuals using the same random time indexation to obtain an artificial sample of pricing factors and log excess holding

[5] Because we extract the principal components from demeaned monthly yields, we need to make an adjustment to the daily factors. We apply the monthly principal components weights to the sample average of the monthly yields and then subtract this vector from the daily factors obtained as described before.

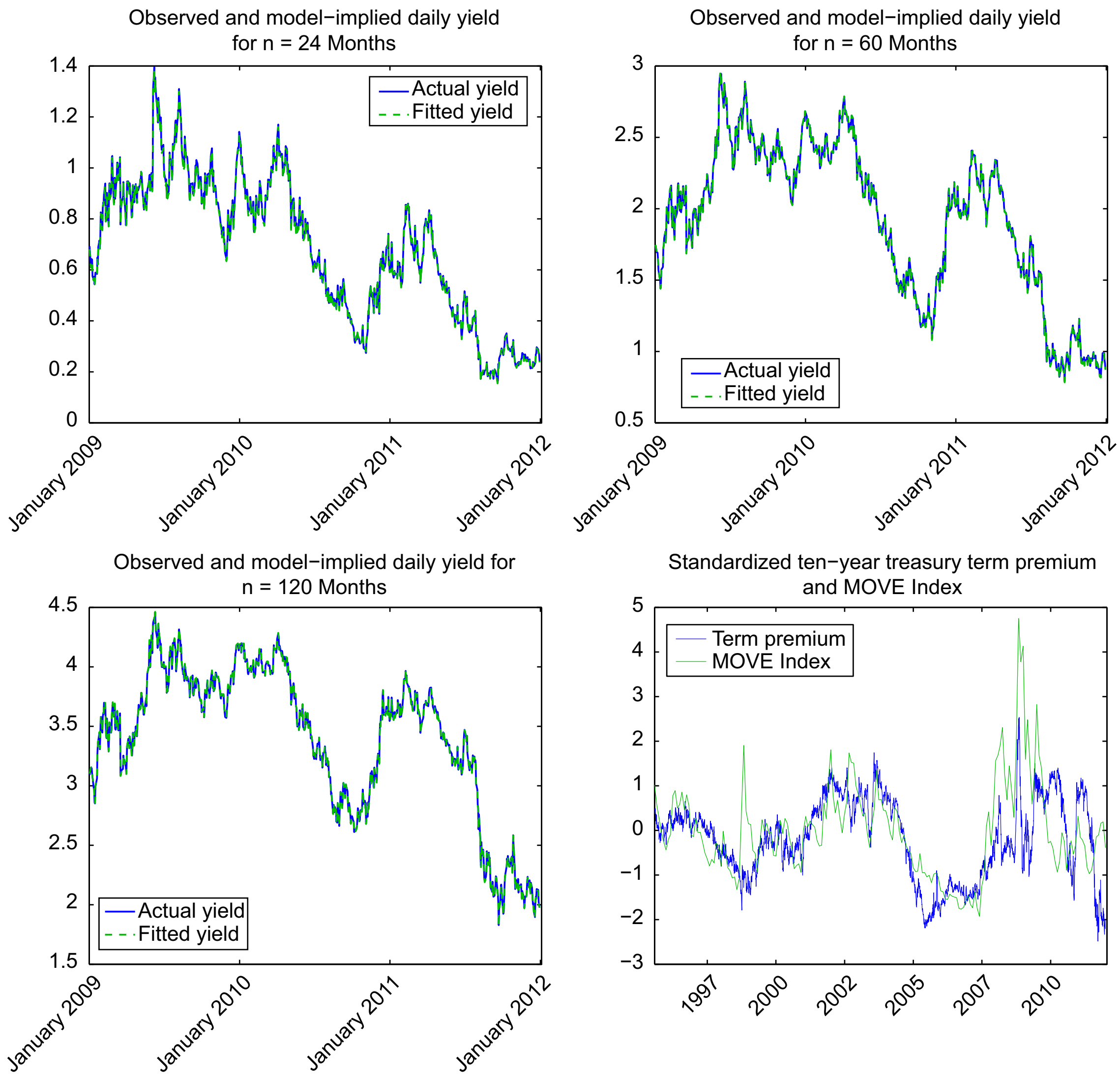

Fig. 10. Five-factor model: observed and model-implied daily yields. This figure provides plots of daily yields for the two-, five- and ten-year maturities as outlined in Section 4.4 fitted by the five-factor specification. Loadings from monthly principal components are applied to daily yields to obtain daily factors. Solid lines plot observed yields, and dashed lines plot yields as implied by the model. The lower right panel compares the standardized five-factor model-implied ten-year term premium to the standardized one-month Merrill lynch Option Volatility Estimate (MOVE) index.

period returns. Based on these, we reestimate the model parameters using the three-step regression approach and compute recursive pricing parameters B_n from Eq. (26). We repeat this procedure one thousand times. We find that the identity matrix is well within the 1 standard deviation confidence interval around the mean of the empirical distribution of $P\hat{b}'$. To illustrate this graphically, Fig. 11 plots the five principal components used as pricing factors along with the bootstrapped 95% confidence interval of the quantity $P\hat{b}'X$. As the plots show, these distributions are extremely tight. The principal components are visually almost indistinguishable from their model-implied 95% confidence intervals. The sampling error of the principal components is comparable in magnitude to the error that arises due to fitting zero coupon curves to actual bond data. We, therefore, believe that the inconsistency that is implied by treating pricing factors as observed is numerically negligible. These results suggest that deviations from Eq. (56) in the model are fully explained by observation error in yields and excess returns.

5. Conclusion

We outline an empirical approach to the estimation of dynamic term structure models. Our approach is computationally fast, gives rise to small pricing errors, and provides asymptotic standard errors for the model parameters of interest. Our method can be used for applications with observable factors and allows for unspanned factors.

Our empirical analysis uncovers a number of new results and revisits certain controversies. First of all, we show in specification tests that the first three principal

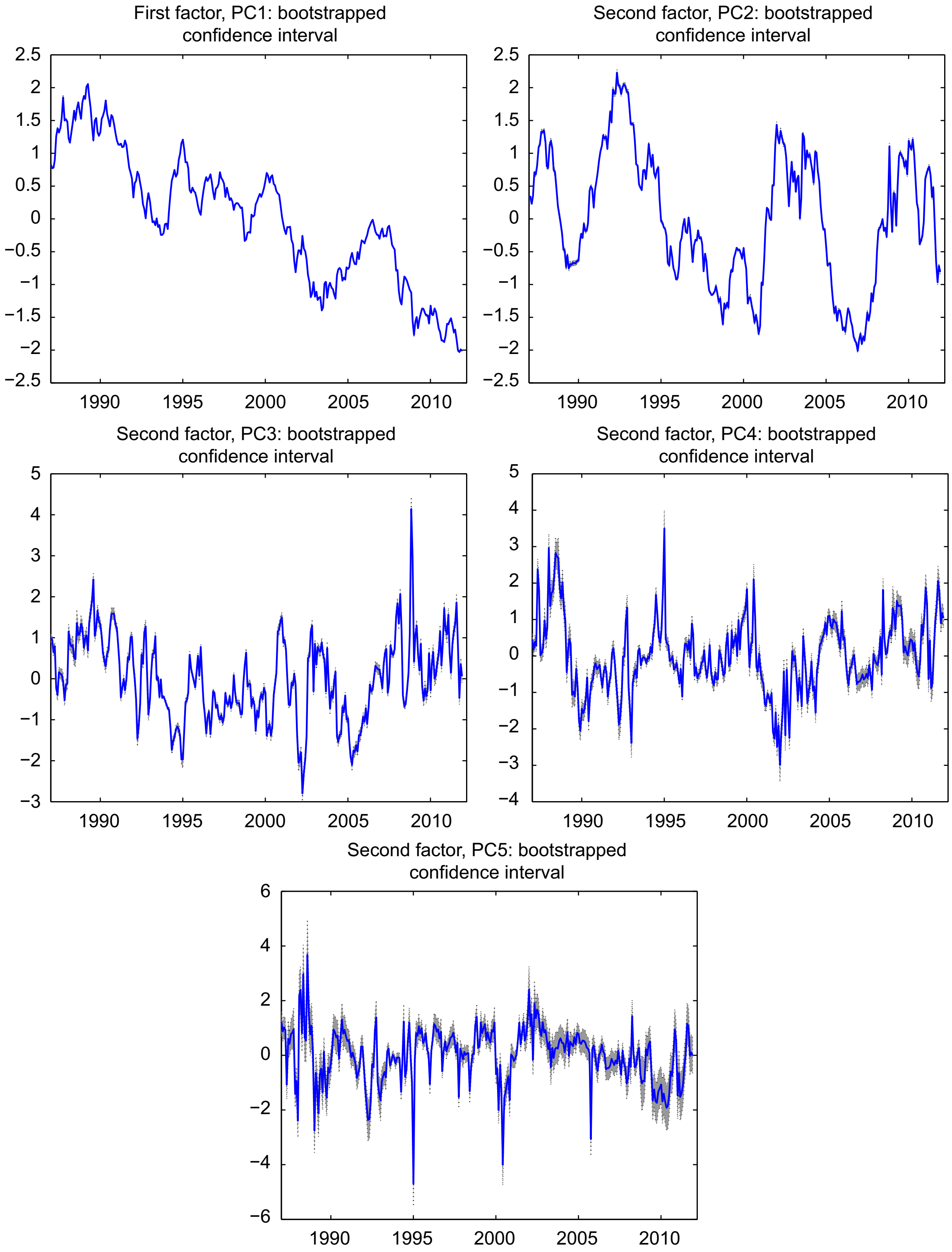

Fig. 11. Consistency of observed factors. This figure provides plots of bootstrapped 95% confidence intervals around the yield principal components. Principal components from the fitted yields are constructed according to the bootstrap procedure in Section 4.5. PC1,…, PC5 denote the first through fifth principal components of Treasury yields.

components of Treasury yields are not sufficient to span the cross section of Treasury returns. We, therefore, study a baseline specification that uses the first five principal components as pricing factors. Second, we show that the five-factor model gives rise to similar risk premiums and pricing kernel dynamics as a specification with three principal components and the Cochrane and Piazzesi (2008) forecasting factor. In both specifications, we find that practically all of the time variation in risk premiums is associated with level risk. However, the dynamics of the risk premium are mainly explained by the second and fifth principal component in the five-factor model and by the CP factor in the four-factor model. Nevertheless, we reject the restriction that the CP factor influences only the price of risk of the level shock and find that it significantly prices other sources of risk as well. For both the five-factor and the four-factor specifications, we find that slope risk is not priced. When comparing the two models based on out-of-sample predictions, we find that the five-factor model outperforms the four-factor model. We, therefore, designate it as our preferred specification.

We also allow for certain factors to be unspanned and provide asymptotic standard errors in this case. Once we add unspanned macroeconomic factors to a model with the first three principal components as pricing factors, we find that slope risk is significantly time varying as a function of a real activity indicator. This result suggests that there is time variation in the pricing kernel that is not spanned by the yield curve.

Our estimation method can be easily adapted and extended, as it solely relies on linear regressions. We present several examples of such extensions in the paper. First, we demonstrate that affine term structure models can be estimated even if zero coupon yields are not available. We estimate such a model by using five principal components of coupon bearing yields to price the cross section of maturity sorted returns. The resulting zero coupon yield curve is very similar to the Fama and Bliss discount curve, even though the estimation methods are vastly different. We furthermore present estimation results at the daily frequency, which are readily computed due to the ease of our estimation method. We leave it to future work to adapt the model to further applications, such as the estimation of inflation risk premiums, or to credit risk models.

Appendix A

A.1. Asymptotic variance: no unspanned factors

Here we derive explicit expressions for $\mathcal{V}_\Lambda$, $\mathcal{V}_\beta$ and $\mathcal{C}_{\Lambda,\beta}$ given in Eq. (19) and the remaining variances and covariances required for the minimum distance estimator of Section 4.2. Before proceeding, we require some definitions. Let $\mathbf{f}=[\mathbf{a}\ \mathbf{c}]$ and define the $NK\times N$ matrix A_β as $A_\beta=\mathrm{diag}(\beta^{[1]},\ldots,\beta^{(N)})$. Also, define $\Upsilon_{zz}=\mathrm{plim}_{T\to\infty}(Z_-Z'_-/T)$ and $\Upsilon_{xx}=\mathrm{plim}_{T\to\infty}(X_-X'_-/T)$. For an $m\times n$ matrix A the $mn\times mn$ commutation matrix $\kappa_{m,n}$ satisfies $\mathrm{vec}(A')=\kappa_{m,n}\ \mathrm{vec}(A)$. For a symmetric $n\times n$ matrix A, the $n^2\times(n(n+1)/2)$ duplication matrix G_n satisfies $\mathrm{vec}(A)=\ G_n\ \mathrm{vech}(A)$. Finally, denote the Moore-Penrose inverse of G_n by G_n^+.

We first focus on $\mathcal{V}_\Lambda$. We could decompose the estimator as

$$\sqrt{T}\hat{\Lambda}=\mathcal{T}_1+\mathcal{T}_2+\mathcal{T}_3, \tag{58}$$

where

$$\mathcal{T}_1=\sqrt{T}(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}')^{-1}\hat{\boldsymbol{\beta}}[rx+\tfrac{1}{2}(B^\star\ \mathrm{vec}(\Sigma)+\sigma^2\iota_N)\iota'_T]M_{\hat{V}}Z'_-(Z_-M_{\hat{V}}Z'_-)^{-1}, \tag{59}$$

$$\mathcal{T}_2=\tfrac{1}{2}\cdot(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}')^{-1}\hat{\boldsymbol{\beta}}[\sqrt{T}(\hat{B}^\star\ \mathrm{vec}(\hat{\Sigma})-B^\star\mathrm{vec}(\Sigma))]\varrho'_1, \tag{60}$$

and

$$\mathcal{T}_3=\tfrac{1}{2}\cdot(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}')^{-1}\hat{\boldsymbol{\beta}}\left[\sqrt{T}(\hat{\sigma}^2-\sigma^2)\right]\iota_N\varrho'_1. \tag{61}$$

Let us first consider $\mathcal{T}_1$. We could further decompose $\mathcal{T}_1$ as

$$\mathcal{T}_1=\mathcal{T}_{1,1}+\mathcal{T}_{1,2}+\mathcal{T}_{1,3}+\mathcal{T}_{1,4}+\mathcal{T}_{1,5}, \tag{62}$$

where

$$\mathcal{T}_{1,1}=\sqrt{T}\Lambda, \tag{63}$$

$$\mathcal{T}_{1,2}=\sqrt{T}VM_{\hat{V}}Z'_-(Z_-M_{\hat{V}}Z'_-)^{-1}, \tag{64}$$

$$\mathcal{T}_{1,3}=\sqrt{T}(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}')^{-1}\hat{\boldsymbol{\beta}}EM_{\hat{V}}Z'_-(Z_-M_{\hat{V}}Z'_-)^{-1}, \tag{65}$$

$$\mathcal{T}_{1,4}=-\sqrt{T}(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}')^{-1}\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\Lambda, \tag{66}$$

and

$$\mathcal{T}_{1,5}=-\sqrt{T}(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}')^{-1}\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'VM_{\hat{V}}Z'_-(Z_-M_{\hat{V}}Z'_-)^{-1}. \tag{67}$$

$\mathcal{T}_{1,1}$ is already simplified. The properties of $\mathcal{T}_{1,2}$ depend on whether the assumption that $\mu=0$ is imposed. When the assumption is not imposed, then $M_{\hat{V}}Z'_-=Z'_-$ and

$$\mathcal{T}_{1,2,\mu}=(VZ'_-/\sqrt{T})\Upsilon_{zz}^{-1}+o_p(1). \tag{68}$$

When the assumption that $\mu=0$ is imposed, we have

$$\mathcal{T}_{1,2,0}=[0_{K\times 1}\ \left(VX'_-/\sqrt{T}\right)]\Upsilon_{zz}^{-1}+o_p(1). \tag{69}$$

This follows because

$$\left(VM_{\hat{V}}Z'_-/\sqrt{T}\right)=\left(VZ'_-/\sqrt{T}\right)-(V\hat{V}'/T)(\hat{V}\hat{V}'/T)^{-1}(\hat{V}Z'_-/\sqrt{T}) \tag{70}$$

$$\left(VM_{\hat{V}}Z'_-/\sqrt{T}\right)=[0_{K\times 1}\ \ (VX'_-/\sqrt{T})]+o_p(1) \tag{71}$$

and

$$(Z_-M_{\hat{V}}Z'_-/T)=(ZZ'_-/T)-(Z\hat{V}'/T)(\hat{V}\hat{V}'/T)^{-1}(\hat{V}Z'_-/T) \tag{72}$$

$$(Z_-M_{\hat{V}}Z'_-/T)=\Upsilon_{zz}+o_p(1). \tag{73}$$

Next,

$$\mathcal{T}_{1,3}=(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}')^{-1}\hat{\boldsymbol{\beta}}\left(EM_{\hat{V}}Z'_-/\sqrt{T}\right)(Z_-M_{\hat{V}}Z'_-/T)^{-1} = (\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}[\sqrt{T}(\hat{\mathbf{f}}-\mathbf{f})]+o_p(1), \tag{74}$$

by our assumptions. Similarly,

$$\mathcal{T}_{1,4}=-\sqrt{T}(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}')^{-1}\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\Lambda = -(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}[\sqrt{T}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]'\Lambda+o_p(1). \tag{75}$$

Finally, it is straightforward to show that $\mathcal{T}_{1,5}=o_p(1)$.

We now consider $\mathcal{T}_2$. First, note that

$$\sqrt{T}(\hat{B}^{\star}\,\mathrm{vec}(\hat{\Sigma}) - B^{\star}\,\mathrm{vec}(\Sigma)) \tag{76}$$

$$= \sqrt{T}(\hat{B}^{\star}\,\mathrm{vec}(\hat{\Sigma}) - \hat{B}^{\star}\,\mathrm{vec}(\Sigma) + \hat{B}^{\star}\,\mathrm{vec}(\Sigma) - B^{\star}\,\mathrm{vec}(\Sigma)), \tag{77}$$

$$= B^{\star}\left(\sqrt{T}(\mathrm{vec}(VV'/T) - \mathrm{vec}(\Sigma))\right) + \sqrt{T}(\hat{B}^{\star} - B^{\star})\mathrm{vec}(\Sigma) + o_p(1), \tag{78}$$

because $\hat{V}\hat{V}'/T = VV'/T + O_p(T^{-1})$ under our assumptions. Thus,

$$\mathcal{T}_2 = \tfrac{1}{2}\cdot(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}')^{-1}\hat{\boldsymbol{\beta}}[\sqrt{T}(\hat{B}^{\star}\,\mathrm{vec}(\hat{\Sigma}) - B^{\star}\,\mathrm{vec}(\Sigma))]\varrho'_1, \tag{79}$$

$$\mathcal{T}_2 = \tfrac{1}{2}\cdot(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}[\sqrt{T}(\hat{B}^{\star}\,\mathrm{vec}(\hat{\Sigma}) - B^{\star}\,\mathrm{vec}(\Sigma))]\varrho'_1 + o_p(1), \tag{80}$$

$$\mathcal{T}_2 = \mathcal{T}_{2,1} + \mathcal{T}_{2,2} + o_p(1), \tag{81}$$

where

$$\mathcal{T}_{2,1} = \tfrac{1}{2}\cdot(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}[\sqrt{T}(\hat{B}^{\star} - B^{\star})\mathrm{vec}(\Sigma)]\varrho'_1, \tag{82}$$

and

$$\mathcal{T}_{2,2} = \tfrac{1}{2}\cdot(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}B^{\star}(\sqrt{T}(\mathrm{vec}(VV'/T) - \mathrm{vec}(\Sigma)))\varrho'_1. \tag{83}$$

The ith element of the interior matrix of $\mathcal{T}_{2,1}$ is

$$\sqrt{T}(\hat{\beta}^{(i)'}\hat{\Sigma}\hat{\beta}^{(i)} - \beta^{(i)'}\Sigma\beta^{(i)}) = 2\sqrt{T}\beta^{(i)'}\Sigma(\hat{\beta}^{(i)} - \beta^{(i)}) + o_p(1). \tag{84}$$

Thus,

$$\sqrt{T}(\hat{B}^{\star} - B^{\star})\mathrm{vec}(\Sigma) = 2A'_{\beta}(I_N\otimes\Sigma)\mathrm{vec}(\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) + o_p(1) \tag{85}$$

and

$$\mathcal{T}_{2,1} = (\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}A'_{\beta}(I_N\otimes\Sigma)\mathrm{vec}(\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))\varrho'_1 + o_p(1). \tag{86}$$

$\mathcal{T}_{2,2}$ does not require any further simplification. Finally, consider $\mathcal{T}_3$. Because $\hat{E}\hat{E}'/T = EE'/T + O_p(T^{-1})$,

$$\mathcal{T}_3 = \tfrac{1}{2}\cdot(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}')^{-1}\hat{\boldsymbol{\beta}}[\sqrt{T}(\hat{\sigma}^2 - \sigma^2)]\iota_N\varrho'_1 \tag{87}$$

$$\mathcal{T}_3 = \tfrac{1}{2}\cdot(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}[\sqrt{T}(\hat{\sigma}^2 - \sigma^2)]\iota_N\varrho'_1 + o_p(1) \tag{88}$$

$$\mathcal{T}_3 = \tfrac{1}{2}\cdot(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}[\sqrt{T}(\mathrm{tr}(EE')/NT - \sigma^2)]\iota_N\varrho'_1 + o_p(1). \tag{89}$$

Combining these results, we have

$$\sqrt{T}(\mathrm{vec}(\hat{\Lambda}) - \mathrm{vec}(\Lambda)) = \mathrm{vec}(\mathcal{T}_{1,2} + \mathcal{T}_{1,3} + \mathcal{T}_{1,4} + \mathcal{T}_{2,1} + \mathcal{T}_{2,2} + \mathcal{T}_3) + o_p(1), \tag{90}$$

where

$$\mathrm{vec}(\mathcal{T}_{1,2,\mu}) = (\Upsilon_{zz}^{-1}\otimes I_K)\mathrm{vec}(VZ'_-/\sqrt{T}) + o_p(1), \tag{91}$$

$$\mathrm{vec}(\mathcal{T}_{1,2,0}) = (\Upsilon_{zz}^{-1}\otimes I_K)\mathrm{vec}([0_{K\times 1}(VX'_-/\sqrt{T})]) + o_p(1), \tag{92}$$

$$\mathrm{vec}(\mathcal{T}_{1,3}) = (I_{K+1}\otimes(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta})\,\mathrm{vec}(\sqrt{T}(\hat{\mathbf{f}} - \mathbf{f})) + o_p(1), \tag{93}$$

$$\mathrm{vec}(\mathcal{T}_{1,4}) = -(\Lambda'\otimes(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta})\kappa_{K,N}\mathrm{vec}(\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) + o_p(1), \tag{94}$$

$$\mathrm{vec}(\mathcal{T}_{2,1}) = (\varrho_1\otimes(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}A'_{\beta}(I_N\otimes\Sigma))\mathrm{vec}(\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) + o_p(1), \tag{95}$$

$$\mathrm{vec}(\mathcal{T}_{2,2}) = \tfrac{1}{2}\cdot(\varrho_1\otimes(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}B^{\star})[\sqrt{T}\,\mathrm{vec}((VV'/T) - \Sigma)] + o_p(1), \tag{96}$$

and

$$\mathrm{vec}(\mathcal{T}_3) = \tfrac{1}{2}(\varrho_1\otimes(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta})\iota_N\cdot\sqrt{T}(\mathrm{tr}(EE')/NT - \sigma^2) + o_p(1). \tag{97}$$

Note next that under our assumptions $\mathrm{vec}(VZ'_-/\sqrt{T}) \to_d \mathcal{N}(0, \Upsilon_{zz}\otimes\Sigma)$, $\sqrt{T}\,\mathrm{vec}((VV'/T) - \Sigma) \to_d \mathcal{N}(0, (I_{K^2} + \kappa_{K,K})(\Sigma\otimes\Sigma))$, and $\sqrt{T}(\mathrm{tr}(EE')/NT - \sigma^2) \to_d \mathcal{N}(0, 2\sigma^4)$. These are all asymptotically independent of

$$\begin{bmatrix}\mathrm{vec}(\sqrt{T}(\hat{f} - \mathbf{f}))\\ \mathrm{vec}(\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})')\end{bmatrix} \to_d \mathcal{N}\left(0, \sigma^2\cdot\left(\begin{bmatrix}\Upsilon_{zz}^{-1} & 0\\ 0 & \Sigma^{-1}\end{bmatrix}\otimes I_N\right)\right), \tag{98}$$

because

$$\operatorname*{plim}_{T\to\infty}\left(\frac{\tilde{Z}\tilde{Z}'}{T}\right) = \begin{bmatrix}\Upsilon_{zz} & 0\\ 0 & \Sigma\end{bmatrix}. \tag{99}$$

Thus, the only asymptotic covariance term we need to consider is between $\mathcal{T}_{1,4}$ and $\mathcal{T}_{2,1}$, which implies that $\sqrt{T}\,\mathrm{vec}((\hat{\Lambda} - \Lambda)) \to_d \mathcal{N}(0, \mathcal{V}_{\Lambda})$, where $\mathcal{V}_{\Lambda} = \mathcal{V}_{\Lambda,\mathcal{T}} + \mathcal{C}_{\Lambda,\mathcal{T}} + \mathcal{C}_{\Lambda,\mathcal{T}}'$,

$$\mathcal{C}_{\Lambda,\mathcal{T}} = \mathrm{asycov}_{T\to\infty}(\mathcal{T}_{1,4}, \mathcal{T}_{2,1}), \tag{100}$$

$$\mathcal{C}_{\Lambda,\mathcal{T}} = -(\Lambda'\otimes(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta})\kappa_{K,N}[\sigma^2\cdot(I_N\otimes\Sigma^{-1})](\varrho_1\otimes(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}A'_{\beta}(I_N\otimes\Sigma))', \tag{101}$$

and $\mathcal{V}_{\Lambda,\mathcal{T}} = \mathcal{V}_{\Lambda,\mathcal{T},1} + \mathcal{V}_{\Lambda,\mathcal{T},2} + \mathcal{V}_{\Lambda,\mathcal{T},3} + \mathcal{V}_{\Lambda,\mathcal{T},4} + \mathcal{V}_{\Lambda,\mathcal{T},5} + \mathcal{V}_{\Lambda,\mathcal{T},6}$. If $\mu = 0$ is imposed, then

$$\mathcal{V}_{\Lambda,\mathcal{T},1} = \mathrm{asyvar}_{T\to\infty}(\mathcal{T}_{1,2,0}) = (\Upsilon_{zz}^{-1}\otimes I_K)\times\left(\begin{bmatrix}0 & 0_{1\times K}\\ 0_{K\times 1} & \Upsilon_{xx}\end{bmatrix}\otimes\Sigma\right)(\Upsilon_{zz}^{-1}\otimes I_K)' \tag{102}$$

and otherwise

$$\mathcal{V}_{\Lambda,\mathcal{T},1} = \mathrm{asyvar}_{T\to\infty}(\mathcal{T}_{1,2,\mu}) = (\Upsilon_{zz}^{-1}\otimes I_K)(\Upsilon_{zz}\otimes\Sigma)(\Upsilon_{zz}^{-1}\otimes I_K)' = (\Upsilon_{zz}^{-1}\otimes\Sigma). \tag{103}$$

The other terms are

$$\mathcal{V}_{\Lambda,\mathcal{T},2} = \mathrm{asyvar}_{T\to\infty}(\mathcal{T}_{1,3}) = \sigma^2\cdot(\Upsilon_{zz}^{-1}\otimes(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}), \tag{104}$$

$$\mathcal{V}_{\Lambda,\mathcal{T},3} = \mathrm{asyvar}_{T\to\infty}(\mathcal{T}_{1,4}) = \sigma^2\cdot(\Lambda'\Sigma^{-1}\Lambda\otimes(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}), \tag{105}$$

$$\mathcal{V}_{\Lambda,\mathcal{T},4} = \mathrm{asyvar}_{T\to\infty}(\mathcal{T}_{2,1}) = \sigma^2\cdot(\varrho_1\varrho'_1\otimes(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}A'_{\beta}(I_N\otimes\Sigma)A_{\beta}\boldsymbol{\beta}'(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}), \tag{106}$$

$$\mathcal{V}_{\Lambda,\mathcal{T},5} = \mathrm{asyvar}_{T\to\infty}(\mathcal{T}_{2,2}) \tag{107}$$

$$\mathcal{V}_{\Lambda,\mathcal{T},5} = \tfrac{1}{4}\cdot(\varrho_1\varrho'_1\otimes(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}B^{\star}(I_{K^2} + \kappa_{K,K})\times(\Sigma\otimes\Sigma)B^{\star'}\boldsymbol{\beta}'(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}), \tag{108}$$

and

$$\mathcal{V}_{\Lambda,\mathcal{T},6} = \mathrm{asyvar}_{T\to\infty}(\mathcal{T}_3) = \frac{\sigma^4}{2}\cdot(\varrho_1\varrho'_1\otimes(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}\iota_N\iota'_N\boldsymbol{\beta}'(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}). \tag{109}$$

It is clear from Eq. (98) that $\mathcal{V}_{\beta} = \sigma^2\cdot(I_N\otimes\Sigma^{-1})$. Finally, we need only calculate $\mathcal{C}_{\Lambda,\beta}$. From our above results we have that the only asymptotic covariance term between $\hat{\Lambda}$ and $\hat{\boldsymbol{\beta}}$ comes from $\mathcal{T}_{1,4}$ and $\mathcal{T}_{2,1}$. Thus,

$$\mathrm{asycov}_{T\to\infty}(\mathrm{vec}(\mathcal{T}_{1,4}), \mathrm{vec}(\hat{\boldsymbol{\beta}})) = -(\Lambda'\otimes(\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta})\kappa_{K,N}[\sigma^2\cdot(I_N\otimes\Sigma^{-1})], \tag{110}$$

$$\text{asycov}_{T\to\infty}(\text{vec}(\mathcal{T}_{1,4}), \text{vec}(\hat{\boldsymbol{\beta}})) \\ = -\sigma^2 \cdot \kappa_{K+1,K}((\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta} \otimes \Lambda'\Sigma^{-1}), \tag{111}$$

and

$$\text{asycov}_{T\to\infty}(\text{vec}(\mathcal{T}_{2,1}), \text{vec}(\hat{\boldsymbol{\beta}})) \\ = (\varrho_1 \otimes (\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}A'_{\beta}(I_N \otimes \Sigma))[\sigma^2 \cdot (I_N \otimes \Sigma^{-1})], \tag{112}$$

$$\text{asycov}_{T\to\infty}(\text{vec}(\mathcal{T}_{2,1}), \text{vec}(\hat{\boldsymbol{\beta}})) = \sigma^2 \cdot (\varrho_1 \otimes (\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}A'_{\beta}). \tag{113}$$

Thus,

$$\mathcal{C}_{\Lambda,\beta} = \text{asycov}_{T\to\infty}(\text{vec}(\hat{\Lambda}), \text{vec}(\hat{\boldsymbol{\beta}})), \tag{114}$$

$$\mathcal{C}_{\Lambda,\beta} = -\sigma^2 \cdot \kappa_{K+1,K}((\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta} \otimes \Lambda'\Sigma^{-1}) \\ +\sigma^2 \cdot (\varrho_1 \otimes (\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}A'_{\beta}). \tag{115}$$

Finally, we provide the remaining asymptotic variances and covariances necessary to construct $\mathcal{V}_\theta$ from Section 4.2. Let $\mathcal{C}_{\Lambda,\sigma^2}$ and $\mathcal{V}_{\sigma^2}$ be the asymptotic covariance between $\hat{\Lambda}$ and $\hat{\sigma}^2$ and the asymptotic variance of $\hat{\sigma}^2$, respectively, and similarly for the other parameters. $\mathcal{C}_{\beta,\Psi}, \mathcal{C}_{\beta,\sigma^2}, \mathcal{C}_{\beta,\Sigma}, \mathcal{C}_{\Psi,\sigma^2}, \mathcal{C}_{\Psi,\Sigma}$, and $\mathcal{C}_{\sigma^2,\Sigma}$ are all equal to the zero matrix of the appropriate dimensions. Because we impose the assumption that $\mu=0$ when estimating the VAR but not when we calculate the asymptotic variance of $\hat{\Lambda}$, then

$$\sqrt{T}\,\text{vec}(\hat{\Phi}-\Phi) = (\Upsilon_{xx}^{-1} \otimes I_K)\text{vec}(VX'_{-}/\sqrt{T}) + o_p(1) \tag{116}$$

and

$$\mathcal{C}_{\Lambda,\Phi} = \text{asycov}_{T\to\infty}(\text{vec}(\hat{\Lambda}), \text{vec}(\hat{\Phi})) \\ = (\Upsilon_{zz}^{-1} \otimes I_K)[(\Upsilon_{zz} \otimes \Sigma)'_2(\Upsilon_{xx}^{-1} \otimes I_K), \tag{117}$$

where $[(\Upsilon_{zz} \otimes \Sigma)]_2$ is the $K(K+1)\times K^2$ sub-matrix formed by extracting the last K^2 columns from the matrix $(\Upsilon_{zz} \otimes \Sigma)$. Next, we have that

$$\mathcal{C}_{\Lambda,\sigma^2} = \text{asycov}_{T\to\infty}(\text{vec}(\hat{\Lambda}), \text{vec}(\hat{\sigma}^2)) = \sigma^4(\varrho_1 \otimes (\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta})\iota_N, \tag{118}$$

and

$$\mathcal{C}_{\Lambda,\Sigma} = \text{asycov}_{T\to\infty}(\text{vec}(\hat{\Lambda}), \text{vech}(\hat{\Sigma})) \tag{119}$$

$$\mathcal{C}_{\Lambda,\Sigma} = \tfrac{1}{2} \cdot (\varrho_1 \otimes (\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}B^{\star})G_K G_K^{+}(I_{K^2} + \kappa_{K,K})(\Sigma \otimes \Sigma)(G_K^{+})' \tag{120}$$

$$\mathcal{C}_{\Lambda,\Sigma} = (\varrho_1 \otimes (\boldsymbol{\beta}\boldsymbol{\beta}')^{-1}\boldsymbol{\beta}B^{\star})G_K G_K^{+}(\Sigma \otimes \Sigma)(G_K^{+})', \tag{121}$$

by properties of the duplication matrix. Because the assumption that $\mu=0$ is imposed when we estimate the VAR, then, $\mathcal{V}_{\Phi,0} = (\Upsilon_{xx}^{-1} \otimes \Sigma)$. As given above, $\mathcal{V}_{\sigma^2} = 2\sigma^4$ and by properties of the duplication matrix, $\mathcal{V}_{\Sigma} = 2G_K^{+}(\Sigma \otimes \Sigma)(G_K^{+})'$.

A.2. Asymptotic variance: unspanned factors

Here we will derive explicit expressions for $\mathcal{V}_{\Lambda}^{ss}$, $\mathcal{V}_{\lambda_1}^{su}$ and $\mathcal{C}_{\Lambda,\lambda_1}^{ss,su}$ given in Eq. (49). In the unspanned case, we can write the equation as

$$rx = \mathbf{f}_s Z_{-}^{s} + \boldsymbol{\beta}'_s X^s + E, \quad \mathbf{f}_s = [\mathbf{a}_s \;\; \mathbf{c}_s], \tag{122}$$

where $Z_{-}^{s} = [\iota_T \; X_{-}^{s'}]'$. Under our assumptions, the OLS estimators of $\mathbf{f}_s$ and $\boldsymbol{\beta}'_s$ satisfy

$$\sqrt{T}\begin{bmatrix} \text{vec}(\hat{\mathbf{f}}_s - \mathbf{f}_s) \\ \text{vec}((\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)') \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_{\mathbf{f}\beta}^{s}), \tag{123}$$

where

$$\mathcal{V}_{\mathbf{f}\beta}^{s} = \sigma^2 \cdot \left(\text{p}\lim_{T\to\infty}(\tilde{Z}^s\tilde{Z}^{s'}/T)^{-1} \otimes I_N\right), \quad \tilde{Z}^s = [Z_{-}^{s'} \;\; X^{s'}]'. \tag{124}$$

Our estimator is

$$\hat{\Psi}_{ss}^{\star} = -(\hat{\boldsymbol{\beta}}_s\hat{\boldsymbol{\beta}}'_s)^{-1}\hat{\boldsymbol{\beta}}_s(rx + \tfrac{1}{2}(\hat{B}^{\star s}\text{vec}(\hat{\Sigma}_{ss}) \\ +\hat{\sigma}^2\iota_N)\iota'_T)M_X^s Z_{-}^{s'}(Z_{-}^{s} M_X^s Z_{-}^{s'})^{-1}, \tag{125}$$

which could be decomposed as

$$\sqrt{T}\hat{\Psi}_{ss}^{\star} = \mathcal{T}_1^s + \mathcal{T}_2^s, \tag{126}$$

where

$$\mathcal{T}_1^s = -\sqrt{T}(\hat{\boldsymbol{\beta}}_s\hat{\boldsymbol{\beta}}'_s)^{-1}\hat{\boldsymbol{\beta}}_s(rx + \tfrac{1}{2}(B^{\star s}\,\text{vec}(\Sigma_{ss}) \\ +\sigma^2\iota_N)\iota'_T)M_X^s Z_{-}^{s'}(Z_{-}^{s} M_X^s Z_{-}^{s'})^{-1} \tag{127}$$

and

$$\mathcal{T}_2^s = -\sqrt{T}(\hat{\boldsymbol{\beta}}_s\hat{\boldsymbol{\beta}}'_s)^{-1}\hat{\boldsymbol{\beta}}_s(\tfrac{1}{2}(\hat{B}^{\star s}\,\text{vec}(\hat{\Sigma}_{ss}) - B^{\star s}\,\text{vec}(\Sigma_{ss}) \\ +(\hat{\sigma}^2 - \sigma^2)\iota_N))\varrho_1^{s'}, \tag{128}$$

where $\varrho_1^s = (1,0,\ldots,0)'$ is $(K_s+1)\times 1$. Then, $\mathcal{T}_1^s$ could be further decomposed as

$$\mathcal{T}_1^s = \mathcal{T}_{1,1}^s + \mathcal{T}_{1,2}^s + \mathcal{T}_{1,3}^s, \tag{129}$$

where

$$\mathcal{T}_{1,1}^s = \sqrt{T}\Psi_{ss}^{\star}, \tag{130}$$

$$\mathcal{T}_{1,2}^s = -\sqrt{T}(\hat{\boldsymbol{\beta}}_s\hat{\boldsymbol{\beta}}'_s)^{-1}\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}'_s)\Psi_{ss}^{\star}, \tag{131}$$

and

$$\mathcal{T}_{1,3}^s = -\sqrt{T}(\hat{\boldsymbol{\beta}}_s\hat{\boldsymbol{\beta}}'_s)^{-1}\hat{\boldsymbol{\beta}}_s E M_X^s Z_{-}^{s'}(Z_{-}^{s} M_X^s Z_{-}^{s'})^{-1}. \tag{132}$$

$\mathcal{T}_{1,1}^s$ requires no further simplification. Next,

$$\mathcal{T}_{1,2}^s = -\sqrt{T}(\hat{\boldsymbol{\beta}}_s\hat{\boldsymbol{\beta}}'_s)^{-1}\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)'\Psi_{ss}^{\star} \\ = -(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s[\sqrt{T}(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)]'\Psi_{ss}^{\star} + o_p(1), \tag{133}$$

and

$$\mathcal{T}_{1,3}^s = -\sqrt{T}(\hat{\boldsymbol{\beta}}_s\hat{\boldsymbol{\beta}}'_s)^{-1}\hat{\boldsymbol{\beta}}_s E M_X^s Z_{-}^{s'}(Z_{-}^{s} M_X^s Z_{-}^{s'})^{-1} \\ = -(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s[\sqrt{T}(\hat{\mathbf{f}}_s - \mathbf{f}_s)] + o_p(1). \tag{134}$$

$\mathcal{T}_2^s$ could be decomposed as

$$\mathcal{T}_2^s = \mathcal{T}_{2,1}^s + \mathcal{T}_{2,2}^s + o_p(1), \tag{135}$$

where

$$\mathcal{T}_{2,1}^s = -\tfrac{1}{2} \cdot (\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s(\sqrt{T}(\hat{B}^{\star s}\,\text{vec}(\hat{\Sigma}_{ss}) - B^{\star s}\,\text{vec}(\Sigma_{ss})))\varrho_1^{s'} \tag{136}$$

and

$$\mathcal{T}_{2,2}^s = -\tfrac{1}{2} \cdot (\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s(\sqrt{T}(\hat{\sigma}^2 - \sigma^2)\iota_N)\varrho_1^{s'}. \tag{137}$$

Similar to the derivation in Appendix A.1, we can show that $\mathcal{T}_{2,1}^s = \mathcal{T}_{2,1,1}^s + \mathcal{T}_{2,1,2}^s + o_p(1)$, where

$$\mathcal{T}_{2,1,1}^s = (\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s A'_{\beta_s}(I_N \otimes \Sigma_{ss})\text{vec}(\sqrt{T}(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s))\varrho_1^{s'}, \tag{138}$$

$$\mathcal{T}_{2,1,2}^s = -\tfrac{1}{2} \cdot (\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s B^{\star s}[\sqrt{T}(\text{vec}(V^s V^{s'}/T) - \text{vec}(\Sigma_{ss}))]\varrho_1^{s'}, \tag{139}$$

and

$$\mathcal{T}^s_{2,2} = -\tfrac{1}{2}\cdot(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s(\sqrt{T}(\mathrm{tr}(EE'/(NT))-\sigma^2))\varrho_1^{s'} + o_p(1). \tag{140}$$

Putting all these results together, we have

$$\sqrt{T}\,\mathrm{vec}(\hat{\Psi}^\star_{ss}-\Psi^\star_{ss}) = \mathrm{vec}(\mathcal{T}^s_{1,2}+\mathcal{T}^s_{1,3}+\mathcal{T}^s_{2,1,1}+\mathcal{T}^s_{2,1,2}+\mathcal{T}^s_{2,2}), \tag{141}$$

where

$$\mathrm{vec}(\mathcal{T}^s_{1,2}) = -(\Psi^{\star'}_{ss}\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s)\kappa_{K_s,N}\,\mathrm{vec}(\sqrt{T}(\hat{\boldsymbol{\beta}}_s-\boldsymbol{\beta}_s))+o_p(1), \tag{142}$$

$$\mathrm{vec}(\mathcal{T}^s_{1,3}) = -(I_{(K_s+1)}\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s)\mathrm{vec}(\sqrt{T}(\hat{\mathbf{f}}_s-\mathbf{f}_s))+o_p(1), \tag{143}$$

$$\mathrm{vec}(\mathcal{T}^s_{2,1,1}) = -(\varrho_1^s\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s A'_{\beta_s}(I_N\otimes\Sigma_{ss}))\mathrm{vec}(\sqrt{T}(\hat{\boldsymbol{\beta}}_s-\boldsymbol{\beta}_s)) + o_p(1), \tag{144}$$

$$\mathrm{vec}(\mathcal{T}^s_{2,1,2}) = -\tfrac{1}{2}\cdot(\varrho_1^s\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s B^{\star s}) \times[\sqrt{T}(\mathrm{vec}(V^sV^{s'}/T)-\mathrm{vec}(\Sigma_{ss}))]+o_p(1), \tag{145}$$

and

$$\mathrm{vec}(\mathcal{T}^s_{2,2}) = -\tfrac{1}{2}\cdot(\varrho_1^s\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s)\iota_N\cdot\sqrt{T}(\mathrm{tr}(EE'/(NT))-\sigma^2)+o_p(1). \tag{146}$$

Then,

$$\sqrt{T}\,\mathrm{vec}(\hat{\Psi}^\star_{ss}-\Psi^\star_{ss}) \to_d \mathcal{N}(0,\mathcal{V}_{\Psi^\star_{ss}}), \tag{147}$$

where

$$\mathcal{V}_{\Psi^\star_{ss}} = \sum_{j=1}^{5}\mathcal{V}_{\Psi^\star_{ss},\mathcal{T},j} + \mathcal{C}_{\Psi^\star_{ss},\mathcal{T},1} + (\mathcal{C}_{\Psi^\star_{ss},\mathcal{T},1})' + \mathcal{C}_{\Psi^\star_{ss},\mathcal{T},2} + (\mathcal{C}_{\Psi^\star_{ss},\mathcal{T},2})' + \mathcal{C}_{\Psi^\star_{ss},\mathcal{T},3} + (\mathcal{C}_{\Psi^\star_{ss},\mathcal{T},3})'. \tag{148}$$

Let us partition $\mathcal{V}^s_{\mathbf{f}\beta}$ conformably

$$\mathcal{V}^s_{\mathbf{f}\beta} = \begin{bmatrix} \mathcal{V}^s_{\mathbf{f}\beta,11} & \mathcal{V}^s_{\mathbf{f}\beta,12} \\ \mathcal{V}^{s'}_{\mathbf{f}\beta,12} & \mathcal{V}^s_{\mathbf{f}\beta,22} \end{bmatrix}, \tag{149}$$

and

$$\mathcal{V}_{\Psi^\star_{ss},\mathcal{T},1} = \mathrm{asyvar}_{T\to\infty}(\mathcal{T}^s_{1,2}) = (\Psi^{\star'}_{ss}\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s)\mathcal{V}^s_{\mathbf{f}\beta,22}(\Psi^{\star'}_{ss}\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s)', \tag{150}$$

$$\mathcal{V}_{\Psi^\star_{ss},\mathcal{T},2} = \mathrm{asyvar}_{T\to\infty}(\mathcal{T}^s_{1,3}), \tag{151}$$

$$\mathcal{V}_{\Psi^\star_{ss},\mathcal{T},2} = (I_{(K_s+1)}\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s)\mathcal{V}^s_{\mathbf{f}\beta,11}(I_{(K_s+1)}\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}'_s), \tag{152}$$

$$\mathcal{V}_{\Psi^\star_{ss},\mathcal{T},3} = \mathrm{asyvar}_{T\to\infty}(\mathcal{T}^s_{2,1,1}), \tag{153}$$

$$\mathcal{V}_{\Psi^\star_{ss},\mathcal{T},3} = (\varrho_1^s\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s A'_{\beta_s}(I_N\otimes\Sigma_{ss}))\kappa_{N,K_s}\mathcal{V}^s_{\mathbf{f}\beta,22}\kappa'_{N,K_s} \tag{154}$$

$$\times(\varrho_1^s\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s A'_{\beta_s}(I_N\otimes\Sigma_{ss}))', \tag{155}$$

$$\mathcal{V}_{\Psi^\star_{ss},\mathcal{T},4} = \mathrm{asyvar}_{T\to\infty}(\mathcal{T}^s_{2,1,2}), \tag{156}$$

$$\mathcal{V}_{\Psi^\star_{ss},\mathcal{T},4} = \tfrac{1}{4}\cdot(\varrho_1^s\varrho_1^{s'}\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s B^{\star s}(I_{K_s^2}+\kappa_{K_s,K_s}) \times(\Sigma_{ss}\otimes\Sigma_{ss})B^{\star s'}\boldsymbol{\beta}'_s(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}), \tag{157}$$

$$\mathcal{V}_{\Psi^\star_{ss},\mathcal{T},5} = \mathrm{asyvar}_{T\to\infty}(\mathcal{T}^s_{2,2}) = \frac{\sigma^4}{2}\cdot(\varrho_1^s\varrho_1^{s'}\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s\iota_N\iota'_N\boldsymbol{\beta}'_s(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}), \tag{158}$$

and

$$\mathcal{C}_{\Psi^\star_{ss},\mathcal{T},1} = \mathrm{asycov}_{T\to\infty}(\mathcal{T}^s_{1,2},\mathcal{T}^s_{1,3}), \tag{159}$$

$$\mathcal{C}_{\Psi^\star_{ss},\mathcal{T},1} = (\Psi^{\star'}_{ss}\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s)\mathcal{V}^{s'}_{\mathbf{f}\beta,12}(I_{(K_s+1)}\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s)', \tag{160}$$

$$\mathcal{C}_{\Psi^\star_{ss},\mathcal{T},2} = \mathrm{asycov}_{T\to\infty}(\mathcal{T}^s_{1,2},\mathcal{T}^s_{2,1,1}), \tag{161}$$

$$\mathcal{C}_{\Psi^\star_{ss},\mathcal{T},2} = (\Psi^{\star'}_{ss}\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s)\mathcal{V}^s_{\mathbf{f}\beta,22}\kappa'_{N,K_s}(\varrho_1^s\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s A'_{\beta_s}(I_N\otimes\Sigma_{ss}))', \tag{162}$$

$$\mathcal{C}_{\Psi^\star_{ss},\mathcal{T},3} = \mathrm{asycov}_{T\to\infty}(\mathcal{T}^s_{1,3},\mathcal{T}^s_{2,1,1}), \tag{163}$$

$$\mathcal{C}_{\Psi^\star_{ss},\mathcal{T},3} = (I_{(K_s+1)}\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s)\mathcal{V}^s_{\mathbf{f}\beta,12}\kappa'_{N,K_s}(\varrho_1^s\otimes(\boldsymbol{\beta}_s\boldsymbol{\beta}'_s)^{-1}\boldsymbol{\beta}_s A'_{\beta_s}(I_N\otimes\Sigma_{ss}))'. \tag{164}$$

Recall that our estimators are $\hat{\Lambda}_{ss}=\hat{\Psi}_{ss}-\hat{\Psi}^\star_{ss}$ and $\hat{\lambda}_1^{su}=\hat{\Phi}_{su}$. Under our assumptions, $\hat{\Psi}^\star_{ss}$ and $\hat{\Psi}$ are asymptotically independent and so

$$\mathrm{asyvar}_{T\to\infty}\begin{pmatrix}\mathrm{vec}(\hat{\Lambda}_{ss})\\ \mathrm{vec}(\hat{\lambda}_1^{su})\end{pmatrix} \tag{165}$$

$$= \begin{bmatrix} \mathcal{V}_{\Psi^\star_{ss}}+\mathrm{asyvar}_{T\to\infty}\mathrm{vec}(\hat{\Psi}_{ss}) & -\mathrm{asycov}_{T\to\infty}(\mathrm{vec}(\hat{\Psi}_{ss}),\mathrm{vec}(\hat{\Phi}_{su})) \\ -\mathrm{asycov}_{T\to\infty}(\mathrm{vec}(\hat{\Phi}_{su}),\mathrm{vec}(\hat{\Psi}_{ss})) & \mathrm{asyvar}_{T\to\infty}\mathrm{vec}(\hat{\Phi}_{su}) \end{bmatrix}. \tag{166}$$

Thus we require only $\mathcal{V}_\Psi=\mathrm{asyvar}_{T\to\infty}(\hat{\Psi})=(\Upsilon_{zz}^{-1}\otimes\Sigma)$. When $\mu=0$ is imposed, then $(\hat{\Lambda}_{ss},\hat{\lambda}_1^{su})$ should be adjusted accordingly and $\mathcal{V}_{\Phi,0}=(\Upsilon_{xx}^{-1}\otimes\Sigma)$ should be used for the analogous result to Eq. (166).

References

Adrian, T., Crump, R.K., Moench, E., 2012. Efficient, regression-based estimation of dynamic asset pricing models. Staff reports 493. Federal Reserve Bank of New York, NY.

Anderson, T.W., 1951. Estimating linear restrictions on regression coefficients for multivariate normal distributions. Annals of Mathematical Statistics 22, 327–351.

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.

Burnside, C., 2012. The cross section of foreign currency risk premia and consumption growth risk: comment. American Economic Review 101, 3456–3476.

Campbell, J.Y., Shiller, R.J., 1991. Yield spreads and interest rate movements: a bird's eye view. Review of Economic Studies 58, 495–514.

Chen, R.R., Scott, L., 1993. Maximum likelihood estimation for a multi-factor equilibrium model of the term structure of interest rates. Journal of Fixed Income 3, 14–31.

Cochrane, J., Piazzesi, M., 2005. Bond risk premia. American Economic Review 95, 138–160.

Cochrane, J., Piazzesi, M., 2008. Decomposing the yield curve. Unpublished working paper. University of Chicago, Chicago, IL.

Collin-Dufresne, P., Goldstein, R.S., 2002. Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility. Journal of Finance 57, 1685–1730.

Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–407.

Dai, Q., Singleton, K.J., 2000. Specification analysis of affine term structure models. Journal of Finance 55, 1943–1978.

Duffee, G., 2002. Term premia and interest rate forecasts in affine models. Journal of Finance 57, 405–443.

Duffee, G., 2010. Sharpe ratios in term structure models. Unpublished Working Paper 575. Johns Hopkins University, Baltimore, MD.

Duffee, G.R., 2011. Information in (and not in) the term structure. Review of Financial Studies 24, 2895–2934.
Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. Mathematical Finance 6, 379–406.
Dybvig, P.H., Ross, S.A., 1987. Arbitrage. In: Eatwell, J., Milgate, M., Newman, P. (Eds.), The New Palgrave: A Dictionary of Economics, Palgrave Macmillan, New York, NY, pp. 100–106.
Garbade, K., 1996. Fixed Income Analytics. MIT Press, Cambridge, MA.
Gurkaynak, R.S., Sack, B., Wright, J.H., 2007. The US treasury yield curve: 1961 to the present. Journal of Monetary Economics 54, 2291–2304.
Hamilton, J.D., Wu, C. Testable implications of affine term structure models. Journal of Econometrics, forthcoming.
Hamilton, J.D., Wu, C., 2012. Identification and estimation of affine-term-structure models. Journal of Econometrics 168, 315–331.
Joslin, S., Priebsch, M., Singleton, K.J., 2012. Risk premiums in dynamic term structure models with unspanned macro risks. Unpublished working paper.
Joslin, S., Singleton, K.J., Zhu, H., 2011. A new perspective on Gaussian dynamic term structure models. Review of Financial Studies 24, 926–970.
Kim, D.H., Wright, J.H., 2005. An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates. Finance and Economics Discussion Series 2005-33. Board of Governors of the Federal Reserve System, Washington, DC.
Kleibergen, F., 2009. Tests of risk premia in linear factor models. Journal of Econometrics 149, 149–173.
Ludvigson, S.C., Ng, S., 2009. Macro factors in bond risk premia. Review of Financial Studies 22, 5027–5067.
Merton, R.C., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867–887.
Moench, E., 2008. Forecasting the yield curve in a data-rich environment: a no-arbitrage factor-augmented var approach. Journal of Econometrics 146, 26–43.
Piazzesi, M., 2003. Affine term structure models. In: Ait-Sahalia, Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics, vol. 1. , Elsevier, Amsterdam, Netherlands, pp. 691–758. (Chapter 12).
Scheinkman, J.A., Litterman, R., 1991. Common factors affecting bond returns. Journal of Fixed Income 1, 54–61.
Singleton, K.J., 2006. Empirical Dynamic Asset Pricing. Princeton University Press, Princeton, NJ.
Wright, J.H., 2011. Term premia and inflation uncertainty: empirical evidence from an international panel dataset. American Economic Review 101, 1514–1534.