Accepted Manuscript

An asset pricing approach to testing general term structure models

Bent Jesper Christensen, Michel van der Wel

PII: S0304-405X(19)30064-9
DOI: https://doi.org/10.1016/j.jfineco.2019.03.010
Reference: FINEC 3038


Received date: 31 May 2016
Revised date: 1 March 2018
Accepted date: 30 March 2018

Please cite this article as: Bent Jesper Christensen, Michel van der Wel, An asset pricing approach to testing general term structure models, Journal of Financial Economics (2019), doi: https://doi.org/10.1016/j.jfineco.2019.03.010

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
An asset pricing approach to testing general term structure models

Bent Jesper Christensen\textsuperscript{a,b,c,*}, Michel van der Wel\textsuperscript{b,d,e,f}

\textsuperscript{a}Aarhus University, Department of Economics and Business Economics, Fuglesangs Alle 4, 8210 Aarhus V, Denmark
\textsuperscript{b}Center for Research in Econometric Analysis of Time Series, Aarhus University, Department of Economics and Business Economics, Fuglesangs Alle 4, 8210 Aarhus V, Denmark
\textsuperscript{c}Dale T. Mortensen Center, Aarhus University, Department of Economics and Business Economics, Fuglesangs Alle 4, 8210 Aarhus V, Denmark
\textsuperscript{d}Erasmus University Rotterdam, Erasmus School of Economics, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands
\textsuperscript{e}Tinbergen Institute, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands
\textsuperscript{f}Erasmus Research Institute of Management, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands

Abstract

We develop a new empirical approach to term structure analysis that allows testing for time-varying risk premiums and arbitrage opportunities in models with both unobservable factors and factors identified as the innovations to observed macroeconomic variables. Factors can play double roles as both covariance-generating common shocks driving yields and determinants of market prices of risk in cross-sectional pricing. The evidence favors time-varying risk prices significantly related to the second Stock-Watson principal component of macroeconomic variables and to changes in the industrial production index. Our preferred specification includes these two observable and two unobservable factors, with the no-arbitrage condition imposed.

Keywords: Bond aging effect, Macroeconomic conditioning variables, Nonlinear drift restriction, Time-varying risk premiums, Yield curve model

JEL classification: G12, C32

\textsuperscript{*}We are grateful to an anonymous referee, Bob Jarrow, and Peter Schotman for useful comments and to the Center for Research in Econometric Analysis of TimE Series (CREATES), funded by the Danish National Research Foundation (grant DNRF78), and the Danish Social Science Research Council for research support. Bent Jesper Christensen is also a Danish Finance Institute (DFI) research fellow. Michel van der Wel is grateful to Netherlands Organisation for Scientific Research (NWO) for a Veni grant.

\textsuperscript{*}Corresponding author. Tel.: +45 2964 5180; E-mail address: bjchristensen@econ.au.dk (B.J. Christensen).
1. Introduction

The Treasury market is among the largest financial markets in the world, of comparable order of magnitude even to the New York Stock Exchange in terms of total market capitalization and with most of the debt held at maturities of more than a year.\textsuperscript{1} Treasuries are held by many different financial institutions, including commercial and investment banks, pension funds, insurance companies, and hedge funds, as well as by private investors, for purposes of diversification, immunization, asset allocation, market timing, and risk management. The weight of government bonds in many portfolios has increased dramatically since the 2007-2009 financial crisis. In the relatively liquid Treasury market, where individual issues of comparable maturity and contractual terms are fairly close substitutes, market conditions at any given time are effectively summarized by the yield curve. Bonds, notes, and bills are always quoted in terms of their corresponding yields.

In this paper, we introduce a new asset pricing approach to test the relevant efficient market pricing conditions directly on the sequence of consecutive observed yield curves. As in the equity case, a zero intercept condition is tested. But, in addition to the standard bilinear term in factor loadings and market prices of risk, the relevant mean restriction in the term structure case involves a nonlinear (quadratic, or convexity) term in the loadings, and the test is applied to yield changes appropriately adjusted for both average slope (or yield spread) and local slope [the bond aging effect, cf. Litterman and Scheinkman (1991)] of the yield curve. These extensions to the classical asset pricing approach provide a link between intercepts and expected excess returns after adjustment for risk compensation, and the analysis can proceed in a fashion parallel to multivariate testing of equity pricing models. A tension between apparent nonzero intercepts (indicative of mispricing) and potentially omitted factors, similar to that common in the equity case, arises empirically in our term structure setting.

The idea that the yield curve captures current market conditions provides the foundation of the Ho and Lee (1986) and Heath, Jarrow, and Morton (1992; henceforth, HJM) approach to general interest rate modeling, in which the entire yield curve acts as the state variable in the dynamic term structure model. An assumption is made about the shapes of the volatility functions governing the stochastic evolution through time of the curve. To preclude arbitrage opportunities, a restriction linking drifts and volatilities of yields through market prices of risk is required. At this point, the distribution of all future interest rates is

\textsuperscript{1}The most recent bond market figure from Bank for International Settlements (BIS) debt securities statistics is from March 2017, with $17,037 billion worth of total government debt securities outstanding by the end of December 2016 and an average remaining maturity of 5.6 years. For the same month, from the World Federation of Exchanges, the NYSE total stock market capitalization is $19,573 billion.
specified, both under the physical and the risk-neutral measure, hence facilitating derivative pricing. Calibration to the entire current term structure is feasible, as the shape of the current yield curve is not restricted by parameters and state variables. The approach thus is overwhelmingly popular among practitioners, for pricing, trading, and hedging bonds and interest rate sensitive claims. The idea is to condition current prices on all information in the current yield curve, while reducing dependence on potentially obsolete parameter estimates based on past observations.

In spite of its importance and compelling logic, the HJM framework has rarely been analyzed econometrically. It has not been tested whether risk premiums are time-varying in this framework, whether any time-varying risk premiums can be related to aggregate observables, and whether the no-arbitrage drift condition is satisfied in practice in the market place.

We build an econometrically tractable dynamic term structure model consistent with the general HJM framework, and we also consider the potential reduction to well-known affine subclasses. The general model allows for an arbitrary number of latent factors, and a potentially time-varying risk price is associated with each of these. The model is formulated at the level of yields to maturity because this is how market prices are quoted and because factor loadings more realistically are stable through time for constant maturity slope-adjusted yield changes than for consecutive returns to an aging bond that by definition becomes shorter as maturity approaches. We develop an asset pricing approach to the empirical analysis of slope-adjusted yield changes, using likelihood-based methods and a state space implementation. We specify latent state variables such that market prices of risk are given as conditional expectations of these. In the sequel, we investigate whether some or all of the latent variables can be replaced by observable macroeconomic aggregates, each carrying a market price of risk that potentially differs from the conditional factor mean. Evidence exists that macroeconomic variables explain both bond market premiums (see, e.g., Wachter, 2006; Ludvigson and Ng, 2009; Joslin, Priebsch, and Singleton, 2014) and the factors underlying the bond market (see, e.g., Ang and Piazzesi, 2003; Diebold, Rudebusch, and Aruoba, 2006; Duffee, 2011b).

In the equity case (no convexity term, and using return data instead of slope-adjusted yield changes), and if further restricting attention to the special case of serially uncorrelated and purely latent state variables, the state space model reduces to the classical factor analysis applied by Roll and Ross (1980) when testing the arbitrage pricing theory (APT) of Ross (1976) in stock market data. Gultekin and Rogalski (1985) apply the classical factor analysis to the returns on a set of constant maturity bond portfolios and then regress returns cross-sectionally on estimated loadings from the factor analysis to estimate period-by-period.
risk premiums in an adaptation of the two-step approach of Fama and MacBeth (1973) to the factor analysis. They also consider multivariate APT tests by regressing returns on observable factors (stock market portfolios or estimated bond market risk prices) and testing for zero intercepts as in Gibbons, Ross, and Shanken (1989).

The classical factor analysis does not apply if state variables are correlated over time, and we use the Kalman filter to handle the generalization to a hidden Markov process for the latent state variables. To verify robustness to departures from distributional assumptions, we compare with results from an alternative two-step principal component analysis (PCA) based approach more closely related to that of Gultekin and Rogalski (1985). This delivers similar point estimates of risk prices in the specification without macroeconomic variables, but these are all insignificant, unlike in our approach. Furthermore, in the alternative approach, all time variation in risk prices stems from factor variation. Our generalized model with observable macroeconomic as well as unobservable factors effectively separates time variation in factors from that in conditional risk prices.

Using data on yields instead of bond portfolio returns, we adjust yield changes for average slope (spread) and local slope (aging) of the yield curve to approximate log excess returns to hypothetical discount bonds. Consequently, the arbitrage condition tested differs from the APT by a convexity term quadratic in loadings (the Jensen’s inequality or HJM drift restriction term). Information on risk premiums is obtained jointly from the cross section (as in the two-step approach) and from the time series pattern of yields, and the Kalman filter delivers the optimal relative weighting of information from these two complementary sources. Our analysis allows studying the appropriate number of observable and latent factors to be included to explain term structure movements in this framework, as well as testing whether risk premiums are in fact time-varying and whether the no-arbitrage drift restriction is satisfied in the data. Here, we consider the unsmoothed Fama-Bliss fixed maturity panel of monthly US zero-coupon Treasury yields of maturities ranging from three months to ten years, over the period 1985 through 2016.

Allowing for time-varying risk premiums is important when testing for the absence of arbitrage opportunities. If risk premiums were inappropriately constrained to be constant, volatility estimates would be inconsistent and, as the volatilities enter into the drift specification under the no-arbitrage condition, the test of this condition would be misspecified. Even if consistent volatility estimates were used, the test on the drift restriction would be misspecified if the wrong (constant in time) form of market prices of risk were entered into the condition tested. More generally, rejection of the arbitrage restriction can arise spuriously if the maintained model is misspecified. Another instance in which this occurs is when the number of factors included is insufficient. This reinforces the importance of our
general approach, allowing for an arbitrary number of factors and testing for time-varying risk premiums and no arbitrage for each given number of factors. What we are dealing with is nothing but the usual joint hypothesis problem of market efficiency tests, well known from the asset pricing literature using stock market data. Any rejection of the arbitrage restriction can be interpreted as an indication of either the presence of arbitrage opportunities or, if no arbitrage is a maintained hypothesis throughout, the need for a more general model (e.g., adding more factors or allowing more general movement of risk premiums through time). Either alternative would render the model under the null inappropriate for pricing purposes. Hence, it is crucial to have a procedure for testing the no-arbitrage condition.

The HJM framework for term structure modeling is general and so includes, e.g., the popular affine models and all standard short rate models. In the short rate approach, a stochastic process assumption is adopted for the short rate of interest. When combined with a rule for the measure change from the physical to the risk-neutral, e.g., based on a risk premium specification from equilibrium theory, this delivers a model for the shape and stochastic evolution through time of the entire term structure of interest rates of all maturities. The literature on short rate models is huge, and a complete survey is well beyond the scope of this paper, but some of the seminal contributions in the area are the Gaussian model of Vasicek (1977), the general equilibrium model of Cox, Ingersoll, and Ross (CIR) (1985a), the resulting square root model of CIR (1985b), the general affine model framework of Duffie and Kan (1996), and the associated empirical analysis in Dai and Singleton (2000). In all these models, the issue of whether or not risk premiums are time-varying is a relevant one, hence providing a link to our research. In most cases, risk premiums are set as given parametrized (e.g., affine) functions of state variables (see, e.g., Dai and Singleton, 2002; Duffee, 2002; Duarte, 2004). Stanton (1997) estimates this functional relation nonparametrically. Another empirical literature explores yield spreads in the context of the expectations hypothesis, see, e.g., Campbell and Shiller (1991). An early contribution on time-varying risk premiums in this area is Fama (1984), and Bams and Wolff (2003) is an early study using a state space framework.

In addition to time-varying risk premiums, many short rate models allow for time-varying volatility, starting with CIR (1985b). We focus instead on unrestricted shape in the maturity direction of the volatility function, while keeping it constant through time and letting dynamics enter through the risk premium transition equation. More general specifications could in principle be considered, because each short rate model can be extended within the HJM framework by letting the volatilities of yields of all maturities coincide with those in the short rate model in question and leaving the initial term structure shape unrestricted. This prescription highlights the generality of the HJM approach, although restrictions on
yield curve shapes are implied also in this framework, once a process specification for risk premiums is adopted.

We allow for time-varying market prices of risk that are dependent through time according to a Markov transition scheme. A similar assumption is adopted by Duffee (2002), with the key difference being that he also assumes that the short rate (equivalently, any bond yield) is linear in the same factors so, in this sense, our work is related to the literature on (essentially) affine models. Adrian, Crump, and Moench (2013) use risk premium estimates to determine the cross-sectional relation among yields in affine models. They consider only observable state variables. They suggest a three-step regression approach, first estimating Markov state transition parameters by regressing current on lagged observable states, then estimating volatilities or loadings by time series regression, and finally estimating risk prices by cross-sectional regression on estimated loadings, as in Fama and MacBeth (1973) and Gultekin and Rogalski (1985). Other regression-based approaches to affine models include Joslin, Singleton, and Zhu (2011) and Hamilton and Wu (2012). These papers assume that a set of principal components or another subset of (linear combinations of) yields is priced (and observed) according to the assumed factor model without error. This allows backing out the unobservable factors from the subset of yields, following Chen and Scott (1993) and Pearson and Sun (1994). Hamilton and Wu (2014) test the assumption that such a set of yields or principal components is priced without error and show that it implies serially correlated measurement errors. Adrian, Crump, and Moench (2013) find that formulating the analysis in terms of returns eliminates the serial correlation in errors. Golinski and Spencer (2017) also consider returns and assume that if the same linear transformation that when applied to yields produces the principal components is instead applied to returns, then the resulting transformed portfolio returns are observed without error, again allowing the backing out of unobservable factors. Their intercept tests are for whether yields conform to an affine term structure model, in contrast to our zero intercept tests of the no-arbitrage condition on risk-adjusted expected excess returns.

In this paper, we consider both observable and unobservable state variables, without assuming that any particular set of portfolios allows backing out the unobservable state variables and, hence, the state space formulation. The Kalman filter ensures optimal weighting of information from yields and state variables across maturities and calendar time. When all state variables are observed, our model reduces to a regression format, as in the literature on affine models with observable state variables. When all state variables are instead unobserved, the cross-sectional risk pricing parameters are not separately identified from the dynamic Markov transition parameters, and tests are essentially conducted on the factor returns structure instead of actual pricing. When one or more of the state variables
are observable, the associated pricing and transition parameters are separately identified. With the additional cross-dynamics of just one price, the risk premium assumption can also be verified with respect to whether the model accurately computes bond prices through the present value relation. Thus, both physical (dynamic) and risk-neutral (cross-sectional pricing) parameters are estimated and used in the construction of the no-arbitrage test. Our approach allows investigating the relative importance of observable and unobservable state variables and, in case of the former, determining which state variables primarily derive their importance in their capacity as conditioning variables in risk premiums and, hence, cross-sectional pricing and which are more important because their innovations serve as covariance-generating factors in the stochastic evolution of yields through calendar time.

Other papers on the HJM approach closely related to ours are few. Bliss and Ritchken (1996) and de Jong and Santa Clara (1999) consider a specific HJM-style model in a dynamic factor model setting, but they impose the no-arbitrage drift restriction throughout, instead of testing it, which is part of our main focus. Their model involves only a single market price of risk, and de Jong and Santa Clara (1999) restrict this to be proportional to the volatility of the short rate, as in CIR (1985b), whereas Bliss and Ritchken (1996) do not estimate it at all. We work with a vector of potentially time-varying risk premiums and, as a second part of our focus, investigate the appropriate number of observed versus unobserved factors and test for the special case that premiums are constant. Jeffrey, Kristensen, Linton, Nguyen and Phillips (2004) consider nonparametric estimation of the volatility function, but for a fixed number of factors, all of them observed, and they do not test the no-arbitrage drift condition or estimate any premiums.

Our empirical results show that time-varying risk premiums are preferred over constant premiums. We find that at least three or four factors are necessary to explain term structure movements, consistent with the bulk of the empirical literature, dating back to Litterman and Scheinkman (1991). Our preferred specifications include both unobservable factors and factors identified as the innovations to observed macroeconomic variables that in turn play separate roles as covariance-generating factors driving the distribution of yields and as determinants of market prices of risk in cross-sectional pricing. The evidence favors time-varying risk prices that are significantly related to the second Stock and Watson (2002) principal component of macroeconomic variables and to changes in the industrial production index. Our most preferred specification includes these two observable and two latent factors.

Regarding the drift condition, our results are broadly consistent with the absence of arbitrage opportunities, but only when observable macro factors are included. When up to seven factors, all latent, are included, the no-arbitrage restriction is strongly rejected at all conventional levels. When we start replacing some of these latent factors with macro factors,
much weaker evidence of arbitrage opportunities emerges. In our preferred specification with
two macro and two latent factors, the evidence is consistent with the no-arbitrage condition.
The difference in results is of interest in its own right. When researchers are working with too
few factors, or only with latent factors, and report apparent signs of arbitrage opportunities
in the market place, the proper understanding of their results can be that the model is
misspecified. This is reminiscent of the tension between nonzero Jensen’s alphas and omitted
factors in the asset pricing literature using stock market data.

Our work establishes the empirical importance of the nonlinear component of the no-
arbitrage condition and of suitable slope adjustment of the yield change data. We show that
if average and local slopes (yield spread and bond aging, respectively) and convexity are
not correctly accounted for, so that in effect the Ross (1976) APT condition applicable to
returns is tested instead of the appropriate no-arbitrage condition for slope-adjusted yield
changes, then this overturns the qualitative conclusions from the tests. In addition, we show
that reduction to any of a number of affine subclasses from the literature is rejected by the
data.

The rest of the paper proceeds as follows. In Section 2, we describe the model and the
main hypotheses. In Section 3, we present in turn the data, empirical methodology, esti-
mation results, and hypothesis tests, all in the case of latent state variables and allowing
for time-varying risk premiums in the state space approach. In Section 4, we consider the
possibility that some or all of the latent state variables are replaced by observable macroeco-
nomic variables. The corresponding generalized state space approach is introduced, and the
empirical results presented. In a separate analysis in Section 5, we consider the sensitivity
of our general approach to a number of relevant issues arising in practice, such as possible
errors in model specification, data, distributional assumptions, etc. Section 6 concludes.
Technical details are deferred to Appendices A and B. Further empirical results and analysis
are available in an accompanying Online Appendix.

2. The model

The analysis is set in the Heath, Jarrow, and Morton (1992) framework for yields as
opposed to forward rates and uses the Brace and Musiela (1994) parametrization in which
term to maturity $\tau$ instead of the maturity date enters as a separate argument in the con-
tinuously compounded zero coupon yield to maturity $y(t, \tau)$ on calendar date $t$. The yield
curve dynamics are given by the infinite dimensional stochastic differential equation

$$dy(t, \tau) = \alpha(t, \tau) dt + \sigma(t, \tau)'dW_t,$$  \hspace{1cm} (1)
with drift $\alpha$ and yield volatility function $\sigma$. Writing $d$ for the dimension of the driving Wiener process $W_t$, the dimension of $\sigma(t, \tau)$ is $d \times 1$. Thus, the mapping $\tau \mapsto y(t, \cdot)$ gives the yield curve at date $t$, and Eq. (1) shows how this evolves through calendar time. The no-arbitrage drift condition in this setting is

$$\alpha(t, \tau) = \frac{1}{\tau} (y(t, \tau) - y(t, 0)) + \frac{\partial y}{\partial \tau}(t, \tau) + \sigma(t, \tau)' \lambda_t + \frac{\tau}{2} \sigma(t, \tau)' \sigma(t, \tau),$$

(2)

where $\lambda_t$ is the $d$-vector of market prices of risk. As HJM derived the relevant drift condition in an alternative parametrization with maturity date $t + \tau$ instead of term to maturity $\tau$ as a separate argument, and for instantaneous forward rates instead of the yield curve, we present a brief derivation of Eq. (2) in Appendix A. The first two terms in Eq. (2) are not present in the HJM version of the drift condition. The first term, $(y(t, \tau) - y(t, 0))/\tau$, is an average slope or yield spread and appears because we consider yields, not forward rates. The second term, $\partial y(t, \tau)/\partial \tau$, is a local slope, i.e., the yield curve differentiated in the maturity direction, and appears because we consider constant terms to maturity instead of constant maturity dates. Litterman and Scheinkman (1991) decompose bond returns into a locally deterministic bond aging effect and the returns to constant maturity zeroes, and the local slope reflects the former. The third term in Eq. (2) is the risk premium, given by the volatility functions multiplied by the relevant market prices of risk. For derivative pricing purposes, changing measure from the physical to the risk-neutral is equivalent to setting $\lambda_t$ equal to zero. The final term, linear in maturity and quadratic in volatility, is the Jensen’s inequality (Itô’s lemma) or convexity term, replacing a term involving an integral in the HJM representation.

Consider panel data on yields $y_t = (y_{t, \tau_0}, \ldots, y_{t, \tau_m})$, $t = 1, \ldots, T$. Thus, there are $m + 1$ observed yields in the cross section, corresponding to terms to maturity $\tau_0 < \ldots < \tau_m$, and $T$ is the number of time periods. By a discrete time (Euler) approximation to Eq. (1), allowing for idiosyncratic (e.g., measurement) error $\varepsilon_{t,i}$ in the $i$th yield, and imposing the drift condition Eq. (2), we have

$$y_{t, \tau_i} - y_{t-1, \tau_i} = \frac{1}{\tau_i - \tau_0} (y_{t-1, \tau_i} - y_{t-1, \tau_0}) + \frac{y_{t-1, \tau_i} - y_{t-1, \tau_{i-1}}}{\tau_i - \tau_{i-1}} + b_i \lambda_{t-1} + \frac{\tau_i}{2} b_i' b_i + b_i' w_t + \varepsilon_{t,i},$$

(3)

$i = 1, \ldots, m$, where $\sigma(t, \tau) = b_i$, i.e., we henceforth take the volatility function to be time-invariant. This is relevant because we use the parametrization with fixed term to maturity, not the HJM parametrization. The $d$-dimensional vector $w_t$ of driving factors corresponds to increments to the Brownian motions $W_t$. The shortest observed term to maturity $\tau_0$ in practice is positive, so $y(t, 0)$ from Eq. (2) is replaced by $y_{t, \tau_0}$, and one cross-section dimension
is lost. Thus, the shortest maturity $\tau_0$ is not considered on the left-hand side, but on the right-hand side it is used in calculation of the first term (the average slope or yield spread) for all longer maturities and also in the calculation of the second term (local slope or aging) for the second-shortest maturity $\tau_1$. Because $y_{t,\tau_0}$ enters these calculations, we measure it by the three-month T-bill yield, which should be a more market-based rate than, say, the one-month yield, and a good short rate proxy (see Chapman, Long, and Pearson, 1999).

It is natural to collect yield data on the left-hand side by defining the slope-adjusted yield changes

$$\tilde{y}_{t,\tau_i} = y_{t,\tau_i} - y_{t-1,\tau_i} - \frac{1}{\tau_i - \tau_0} (y_{t-1,\tau_i} - y_{t-1,\tau_0}) = \frac{y_{t-1,\tau_i} - y_{t-1,\tau_0}}{\tau_i - \tau_0},$$

i.e., the raw yield changes adjusted for both average and local slope of the yield curve. In the ideal case with complete data for all maturities, so that $\tau_i = \tau_{i-1} + 1$ and $\tau_0 = 0$, with the time unit given by the sampling frequency in the calendar time dimension, it follows from Eq. (4) that the one-period excess return to the discount bond with maturity $\tau_i$ at time $t$ is given by $-\tau_i \tilde{y}_{t,\tau_i}$. This facilitates an interpretation of our work using slope-adjusted yield changes. But, in realistic panel data sets, the difference between adjacent maturities is a multiple of the sampling interval (time unit) that varies along the yield curve. So, the condition $\tau_i = \tau_{i-1} + 1$ is violated, and the relation to excess returns is only an approximation.

Forming the $m$-vector $\tilde{y}_t = (\tilde{y}_{t,\tau_1}, ..., \tilde{y}_{t,\tau_m})$, the resulting model has factor structure

$$\tilde{y}_t = \mu_t + B w_t + \varepsilon_t,$$

where the $i^{th}$ row of the $m \times d$ matrix $B$ is given by $b'_i$, and $\varepsilon_t = (\varepsilon_{t,1}, ..., \varepsilon_{t,m})'$. Thus, $B$ gives the exposures to the common covariance-generating factors $w_t$, and $\mu_t = (\mu_{t1}, ..., \mu_{tm})'$ contains the conditional means of the slope-adjusted yield changes $\tilde{y}_t$, given information through $t - 1$. We assume there are more observed yield changes in the cross-section dimension than factors, $m > d$, and without loss of generality $\text{var}(w_t) = I_d$, because any covariance terms can be absorbed in $B$. In addition, we assume that the idiosyncratic errors are contemporaneously uncorrelated, i.e., $\text{var}(\varepsilon_t) = \Psi$ is a diagonal matrix, and that $w_t$ and

---

2To see this, write $p_{t,\tau} = \exp(-\tau \tilde{y}_{t,\tau})$ for the discount bond price, so that the log return is $r_{t,\tau} = \log p_{t,\tau} - \log p_{t-1,\tau+1} = -\tau y_{t,\tau} + (\tau + 1) y_{t-1,\tau+1}$. Because $p_{t,0} = 1$, the case $\tau = 0$ gives the short rate $r_{t,0} = y_{t-1,1}$, and the excess return is $r_{t,\tau} - r_{t,0} = -\tau (y_{t,\tau} - y_{t-1,\tau+1}) + y_{t-1,\tau+1} - y_{t-1,1}$. The Euler approximation used in Eq. (4) can as well use a right instead of a left difference for the last term (local slope), i.e., $(y_{t-1,\tau+1} - y_{t-1,\tau}) / (\tau_{i+1} - \tau_i)$, and in this and the rest of Eq. (4), the complete data case allows us to substitute $\tau_{i+1} = \tau_i + 1$ and $\tau_0 = 0$. The result after a cancellation (and using the spread up to $\tau_i$ instead of $\tau_i$ in the average slope term) is $-r_{x,t,\tau} / \tau_i$.
\( \varepsilon_t \) are independent of each other and across time (as \( w_t = \int_{t-1}^{t} dW_s \)).

In the system represented in Eq. (5), the no-arbitrage drift restriction Eq. (2) is recast as

\[
\mu_{it} = b_i' \lambda_{t-1} + \frac{\tau_i}{2} b_i' b_i,
\]

for \( i = 1, \ldots, m \). Note the similarity to the arbitrage pricing theory of Ross (1976). Following Roll and Ross (1980), a standard approach to testing the APT is to apply the classical factor analysis to excess stock returns, thus estimating the loadings, say, \( B \), and then testing the APT as a cross-sectional restriction on the mean excess returns \( \mu \), in particular, \( \mu = B \lambda \), for suitable market prices of risk \( \lambda \). Two important steps are involved in adapting this asset pricing approach to our term structure case. First, the factor analysis structure Eq. (5) applies to the appropriately slope-adjusted yield changes defined in Eq. (4), not to raw yields, spreads, or yield changes. This is natural, given the approximate relation between excess returns and slope-adjusted yield changes. Second, the theory restriction tested is not that means be linear in loadings, as in the APT, viz. \( \mu_{it} = b_i' \lambda_{t-1} \). Instead, the no-arbitrage term structure restrictions also include the terms \( \frac{\tau_i}{2} b_i' b_i \) in Eq. (6), which are nonlinear (quadratic) in loadings. These are the Itô adjustment terms from the continuous-time framework (see Appendix A) and, in case of lognormal returns, they are precisely the Jensen’s inequality or convexity terms turning expected log returns into expected raw returns. In this context, the original APT test simply leaves out the convexity term, whereas it is natural that we include it, because \( -\tau_i \tilde{y}_t \tau_i^{\top} \) approximates the logarithmic excess return.

The no-arbitrage condition Eq. (6) should be tested in term structure analysis. Unfortunately, the empirical literature has invariably left out this step. If the condition is rejected, then the model is simply wrong, as it admits arbitrage. The volatility function (loadings) or the price of risk specification is too simplistic. If a specification is determined in which the no-arbitrage condition is not rejected, then the framework could be used to explore the dynamic properties of the vector of risk prices \( \lambda_{t-1} \). Thus, risk prices must be estimated, both to explore their dynamics and to test the no-arbitrage restriction. Several alternative estimation approaches suggest themselves. Two-step procedures are widely used, in the first step estimating \( \mu \) by the sample means and the loadings \( B \) either by principal components or by the classical factor analysis. In the second step, risk prices are estimated by cross-sectional regression. Although time variation in risk premiums can to some extent be picked up in the second step by running the cross-sectional regressions period by period, information is lost by not allowing for time-varying premiums in the first-step estimation of loadings. Instead, we develop a one-step full information maximum likelihood procedure that allows introducing time-varying market prices of risk from the outset. In the special case of constant risk prices \( \lambda \), the procedure is equivalent to expanding the classical factor analysis likelihood function
by the mean specification Eq. (6). With time-varying risk prices, the state space form is used, and the prediction error decomposition of the log likelihood function is calculated based on the innovations from the Kalman filter. This description of both the one- and two-step approaches corresponds to the restricted model. The unrestricted alternative specified for testing purposes adds maturity-specific intercepts $\alpha_i$ to the drift specification to measure the deviation from the arbitrage restriction Eq. (6). Explicitly, writing

$$\mu_{it} = \alpha_i + b_i'\lambda_{t-1} + \frac{\tau_i}{2}b_i'b_i,$$

the no-arbitrage null hypothesis is $H_0 : \alpha = 0$, where $\alpha = (\alpha_1, ..., \alpha_m)'$ represents the expected excess log returns after adjustment for risk compensation to hypothetical zero coupon bonds within the constant maturity yield panel. Our main focus is on testing $H_0$, which can now proceed much as in multivariate testing of equity pricing models, as well as on determining the appropriate number of factors, $d$, and testing for time-varying market prices of risk $\lambda_{t-1}$.

In addition, we investigate whether some or all of the latent factors can be replaced by observable macro series.

3. Empirical results

In this section, we first describe the data used in the empirical analysis. We then turn to the estimation methodology and the results.

3.1. Data and summary statistics

The data set we use is the unsmoothed Fama-Bliss fixed maturity panel over the period 1985 through 2016. The data from January 1985 through December 2000 were constructed by Diebold and Li (2006), using end of month prices for US Treasuries from the Center for Research in Security Prices government files to obtain unsmoothed Fama and Bliss (1987) forward rates. With this procedure, the authors obtained a fixed maturity data set for a cross section of 17 maturities: three, six, nine, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months. Using the same procedure, we extend the data set through December 2016. Thus, in terms of the model, we have for the cross-sectional dimension of the yields $m + 1 = 17$, i.e., the dimension of the vector of slope-adjusted yield changes is $m = 16$.

Summary statistics appear on the left side of Table 1 for yields and on the right side for slope-adjusted yield changes, Eq. (4). The first column shows term to maturity in years. The yield portion of the table is similar to that in Diebold and Li (2006), who consider the subperiod from 1985 through 2000. Thus, on average, the term structure of interest rates is strictly upward-sloping, from 3.5% to 5.3% (column labeled “Mean”). The term structure
of volatilities is strictly upward-sloping until reaching a peak at 18 months ($\tau_i = 1.5$) and then is strictly downward-sloping, to below the short maturity level (“Sd”). This is similar to the tent-shaped volatility term structure in Cochrane and Piazzesi (2005). For each maturity, a considerable spread exists between the minimum and maximum observation, about 9 percentage points or more, indicating a great deal of variation over the period. Finally, strong positive autocorrelation is evident at the one- and 12-month lags. Even at the 30-month lag, the correlation is increasing in maturity and exceeds 50% for maturities of six months and above. For the slope-adjusted yield changes, the means are much smaller, consistent with their interpretation as approximations to excess returns in the constant maturity yield panel. The means are again increasing with maturity and the term structure of volatilities now is decreasing. Serial correlation is much less than for yields, close to zero at the 12- and 30-month lags, suggesting that the factor model approach is more appropriate for slope-adjusted yield changes.

Fig. 1 provides a three-dimensional view of our data on raw yields and slope-adjusted yield changes. In both cases, the curves appear somewhat flat in the term to maturity direction, consistent with level shifts being of greater order of magnitude than slope changes and curvature twists. The stronger autocorrelation in yields shows up as more systematic swings in the calendar date dimension for raw than for slope-adjusted yield changes. The vertical axes are different because of the smaller magnitude of the slope-adjusted yield changes.

3.2. Estimation methodology

The factor structure in Eq. (5) combined with the potential dynamics of the risk prices in Eq. (6) suggests a dynamic factor model approach for estimation and inference. To proceed further, the process for $\lambda_t$ must be specified. Following HJM, we assume that the risk price process is adapted to the filtration generated by the driving Wiener processes. In the discrete time formulation of the econometric model, we specify that risk prices be affine in a suitable vector of latent state variables, say, $x_t$, i.e.,

$$\lambda_t = a + Ax_t,$$  \hspace{1cm} (8)

where $a$ is $d \times 1$, $A$ is $d \times d$, and $x_t$ is measurable with respect to the $w_t$-process from Eq. (5). Thus, constant risk premiums are the special case $A = 0$. Smooth time variation in risk premiums is generated by ensuring this property for the state variable sequence, and analysis
under the risk-neutral measure (for derivative pricing) is facilitated by setting both $a$ and $A$ equal to zero. In the absence of arbitrage opportunities, i.e., in the system Eqs. (5)-(6), and aside from idiosyncratic noise $\varepsilon_t$, time variation in the slope-adjusted yield change $\tilde{y}_{t,\tau_i}$ is generated by $b_i'(\lambda_{t-1} + w_t)$. It is therefore natural to specify the latent state vector as

$$x_t = \lambda_{t-1} + w_t.$$  \hfill (9)

Thus, state variables are driven by the $w_t$ process and also reflect past risk prices. Inserting Eq. (8) into Eq. (9) produces

$$x_t = a + Ax_{t-1} + w_t,$$  \hfill (10)

showing that latent state variables are governed by a vector autoregression (VAR) system forming a hidden Markov process underlying the observed yield data. The state vector reflects the full shock $w_t = \int_{t-1}^{t} dW_s$ and, in addition, depending on the structure of the transition matrix $A$, a portion of the past state. In the special case $A = I_d$, the state variables $x_t$ are exactly tracking the driving Wiener processes $W_t = \sum_s w_s$. Presumably, more empirically relevant specifications involve less weight on past shocks, e.g., through eigenvalues below unity in $A$. From Eqs. (8) and (10), the market prices of risk are given by

$$\lambda_t = E_t(x_{t+1}),$$  \hfill (11)

i.e., the expected next period state variables, conditional on the sequence of states (or, equivalently, $w_s$) through $t$, thus allowing a natural economic interpretation of our latent $x_t$.

By considering the state variable $x_t$ in Eq. (9) instead of $w_t$ and $\lambda_{t-1}$ separately in the model, we are able to investigate time variation in the risk prices $\lambda_t$ without introducing additional noise terms. From Eq. (8), risk prices are adapted to the driving Wiener processes, as in HJM, with cross-effects (nondiagonal $A$) allowed. If $\lambda_t$ had involved separate sources of uncertainty not present in $w_t$, then the physical and risk-neutral ($\lambda_t = 0$) measures would not be equivalent.

That market prices of risk are affine in Markovian state variables follows Duffee (2002), who allows the cross-sectional pricing parameters in Eq. (8) to deviate from the dynamic transition parameters in Eq. (10). We consider this possibility in the case of observable state variables in Section 4.

With the above specifications, our model is

$$\tilde{y}_t = \alpha + \text{vec}_{1:m}\{b_i'b_i\tau_i/2\} + Bx_t + \varepsilon_t$$  \hfill (12)
and

\[ x_t = a + Ax_{t-1} + w_t, \quad (13) \]

with vec1:m{·} denoting the m × 1 vector with typical element given in curly braces. Eq. (12) corresponds to Eqs. (5), (7), and (9); and Eq. (13) to Eq. (10). In the restricted model imposing no arbitrage, \( \alpha \) is set equal to zero in Eq. (12).

The model is in state space form, and Eqs. (12) and (13) give the \( m \)-dimensional measurement equation and \( d \)-dimensional state transition equation. Assuming normal distributions for the shocks \( w_t \) and noise terms \( \varepsilon_t \), the Kalman filter applies [see Harvey (1989) and Durbin and Koopman (2012); Appendix B provides a brief summary of the Kalman filter recursions]. The filter provides estimates of the latent states \( x_t \) conditionally on data \( \tilde{y}_s \) through \( s = t - 1 \), i.e., predicted states \( E_{t-1}[x_t] \) and, conditionally on the expanded information set including current (time \( t \)) data \( \tilde{y}_t \), the filtered states \( E_t[x_t] \) (these expectations using square brackets are conditional on data, not on the true driving \( w_s \)). This allows calculating the innovations (prediction errors) \( \tilde{y}_t - E_{t-1}[\tilde{y}_t] \) in the data sequence and the (prediction error decomposition of the) corresponding conditional log likelihood function given the initial observation \( \tilde{y}_0 \). We maximize this log likelihood to obtain parameter estimates. With these and the filtered and predicted states in hand, we obtain estimates of our risk prices \( \lambda_{t-1} \) and covariance generating factors \( w_t \). As \( E_{t-1}[w_t] = 0 \), an estimate of \( \lambda_{t-1} \) is provided by the predicted state, \( \hat{\lambda}_{t-1} = E_{t-1}[x_t] \) [see Eq. (9)]. The covariance-generating factors are estimated off the expectation revisions from predicted to updated or filtered states, \( \hat{w}_t = E_t[x_t] - E_{t-1}[x_t] \).

For a given dimension of covariance-generating factors \( d = 1, 2, 3, 4 \), we estimate four variations of our model. The first distinction is between constant and time-varying risk prices. Constant risk prices correspond to the special case of \( A = 0 \), implying that the vector of market prices of risk is given by \( \lambda = a \). Time-varying risk prices is the case in which \( A \) is left free. A second distinction is between the presence or absence of arbitrage opportunities. Under the no-arbitrage restriction, \( \alpha = 0 \) is imposed on Eq. (12), and the parameters estimated are \((a, A, B, \Psi)\), except that \( A = 0 \) in the constant risk price case. In the unrestricted case, with \( \alpha \) free, \( a \) is set equal to zero for identification, as it enters the mean of the slope-adjusted yield changes through \( x_t \), and the parameters estimated are \((\alpha, A, B, \Psi)\), except again \( A = 0 \) for constant risk prices.

Our model specification implies that factors (state variables) are filtered from the history of (approximate excess) returns. Factors potentially matter both for the cross section of expected returns through their role as conditioning variables in market prices of risk via the pricing parameters \( A \) in Eq. (8) and as covariance-generating factors driving return dynamics through the loadings \( B \) [see Eq. (12)]. There can be no risk factor that matters for cross-sectional pricing without driving returns. To see this, in Eq. (5), for the, say, \( j^{th} \) factor not to
drive returns, the $j^{th}$ column in $B$ must vanish (since $-\tau_\text{i}\tilde{y}_{t,\tau_\text{i}}$ approximates the logarithmic excess return on a maturity $\tau_\text{i}$ bond). As is evident from the system in Eqs. (12) and (13), in this case nothing in the statistical approach ties the given factor ($j^{th}$ coordinate in $\boldsymbol{x}_\text{t}$) to the data. Thus, the $j^{th}$ row and column in $A$ and the $j^{th}$ entry in $a$ are unidentified and, by Eq. (8), so is the $j^{th}$ market price of risk. It follows that without factor loadings to identify it, a factor can enter neither risk prices nor time-varying expected returns (conditional expected values of $-\tau_\text{i}\tilde{y}_{t,\tau_\text{i}}$).

A formal statistical test of the null that the $j^{th}$ factor does not drive returns, i.e., the $j^{th}$ column in $B$ is zero, is problematic for the same reason, namely, that the $j^{th}$ row and column in $A$ and the $j^{th}$ entry in $a$ are unidentified under this null. Following Andrews and Ploberger (1994), the likelihood ratio (LR) test is not optimal in this case, in which a parameter (here, a component of the pricing parameters $A$ and $a$) is identified only under the alternative. Intuitively, the degrees of freedom in the test could not just be the number of parameters in the $j^{th}$ column in $B$, as parts of $A$ and $a$ are dropped, too. In practice, the joint LR test of both the $j^{th}$ column in $B$ and the relevant parts of $A$ and $a$ being zero could be implemented and the degrees of freedom set to the count of all these parameters, but again, the test is suboptimal, and the optimal (average exponential) test is rather complicated and dependent on correct specification. Similarly, if the $j^{th}$ factor is weak, i.e., the $j^{th}$ column in $B$ is close to zero and vanishing asymptotically (see Bryzgalova, 2016), then estimates of the relevant parts of $A$ and $a$ and their standard errors are unreliable. This situation is analogous to that in two-step linear asset pricing models (see Kleibergen, 2009).

While we have thus ruled out the possibility that a factor could matter for cross-sectional pricing without driving returns, the opposite possibility does exist, i.e., a factor can be important in driving returns, without conditioning market prices of risk. The phenomenon arises when the $j^{th}$ column in $B$ is nonzero and the $j^{th}$ column in $A$ is zero, i.e., the $j^{th}$ factor does not condition risk prices. A related but different hypothesis is that a given market price of risk is zero. In this case, the $j^{th}$ row in $A$ and the $j^{th}$ entry in $a$ are both zero, although the $j^{th}$ factor could still condition other risk prices, i.e., the $j^{th}$ column of $A$ need not be zero. For example, Adrian, Crump, and Moench (2013) find that, in both their four- and five-factor models, one of the risks (the slope risk) is unpriced. Our model allows testing both types of hypotheses. In this case, testing is not subject to the identification issue in Andrews and Ploberger (1994). But, again, if the $j^{th}$ factor is weak, standard inference on the relevant components of $A$ and $a$ remains unreliable.

Finally, we discuss the restrictions that we need to impose to ensure identifiability for all these models and the initialization of the Kalman filter. As Geweke and Zhou (1996) point out, a general identification issue in models with unobservable factors is that these
can be rotated (say, pre-multiplied by an orthogonal matrix $P$) and the factor loadings (the $B$ matrix in our case) adjusted for this (post-multiplied by $P'$) to yield an observationally equivalent model. A way to overcome this is to impose $d \times (d - 1)/2$ restrictions on the matrix $B$ such that this transformation is no longer possible. In practice, we do this by restricting the portion above the diagonal in the top square $d \times d$ part of $B$ to consist of zeros,

$$B = \begin{pmatrix}
  b_{1,1} & 0 & \ldots & 0 \\
  b_{2,1} & b_{2,2} & \ldots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  b_{d,1} & b_{d,2} & \ldots & b_{d,d} \\
  \vdots & \vdots & \ddots & \vdots \\
  b_{m,1} & b_{m,2} & \ldots & b_{m,d}
\end{pmatrix}, \quad (14)
$$

with $b_{i,j}$ the $(i,j)^{th}$ element of $B$ and $b_{i,i} \geq 0$. Next, to initialize the Kalman filter, we specify the distribution of the first unobserved state. In general, one can write

$$x_0 \sim N(\mu, \Sigma). \quad (15)$$

Different types of initializations can be implemented through choice of $\mu$ and $\Sigma$. When not much is known about the initial state, one can set its variance $\Sigma$ very large and take $\mu = 0$. As in our case the dynamics of $x_t$ are in fact a VAR system, we can use the properties of its unconditional distribution to initialize the system (see Appendix B). In this case, $\mu$ and $\Sigma$ are given functions of $a$ and $A$, and reversing the sign of a factor (state variable) is not simply counteracted by a reversal in sign of the associated column of $B$ but also impacts both $a$ and $A$ and, hence, the initialization of the system.

### 3.3. Estimates

We now present the results from applying the methodology from Subsection 3.2 to the data described in Subsection 3.1. We first consider the market prices of risk and their dynamics. We then turn to predictions, expectation revisions, pricing, and pricing errors. Finally, we present tests of time variation in premiums and arbitrage.

#### 3.3.1. Market prices of risk

In Table 2, we report the estimated market prices of risk and likelihood values for the restricted and unrestricted models with constant and time-varying risk premiums. This is done separately for $d = 1, 2, 3, \text{ and } 4$ factors. In the first model, the single-factor case, the restricted version has a negative market price of risk, estimated at about $-0.32$ in the
constant $\lambda$ case. In the time-varying case, the mean value of $\lambda_t$ given by $\mu = (I - A)^{-1}a$ is reported. This is slightly larger in magnitude, at $-0.38$, and both are strongly significant (asymptotic $t$-values below estimates). The difference in degrees of freedom between the restricted and unrestricted models is 15, corresponding to $m = 16$ parameters saved in $\alpha$ by imposing the no-arbitrage restriction, whereas $a$ is a new parameter introduced in the restricted model, both in the constant and time-varying $\lambda_t$ case.

In the two-factor model, $d = 2$, the first risk price is similar, but the second is smaller in magnitude. The total impact on the risk premium is the risk price times the relevant volatility or loading, so it is necessary to consider the risk price estimates in conjunction with the loadings. A graphical depiction of the estimated loadings $b_i$ for each factor as function of maturity is shown in Fig. 2. From the first exhibit in Panel A ($d = 1$), the volatility function in the one-factor model captures a downward-sloping term structure, in line with the decreasing volatilities of the slope-adjusted yield changes in the right-hand portion of Table 1. The second exhibit ($d = 2$) shows the volatility functions of the two-factor model. One is downward-sloping and similar to that in the one-factor model, and one is upward-sloping, with slightly more curvature. The volatility functions in the three-factor model in the third exhibit ($d = 3$) should in combination span the usual level, slope, and curvature features, cf. Litterman and Scheinkman (1991). One of the functions has a clear hump at $i = 4$, corresponding to maturity $\tau_i = 1.25$ years. In the estimation, as well as in the annotation to the figure, this corresponds to the second factor, and the upward-sloping or curvature volatility function now is associated with the third factor [the identifying restriction Eq. (14) on $B$ determines the ordering of the factors in each case]. The hump factor gets a risk price of $-0.33$ in Table 2 ($d = 3$, time-varying case, $t$-statistic of $-20$), and the factor corresponding to the upward-sloping or curvature volatility function gets a larger risk price than in the two-factor model.

The final exhibit in Panel A ($d = 4$) of Fig. 2 shows that allowing four factors leads to another upward-sloping volatility function similar to that already picked up, but with a twist and starting only for longer maturities [cf. Eq. (14)]. The market price of the risk factor associated with this volatility function has the opposite sign of the others, both in the constant and time-varying risk price case. As all four market prices of risk are strongly significant and of similar order of magnitude (Table 2, $d = 4$), reduction to three factors does not necessarily appear warranted. Further, only factors driving returns can matter for risk
prices, so we do not pursue the (nonstandard) formal testing of hypotheses about individual factors potentially not driving returns. If one of the factors were weak, then the associated risk price estimates would be unreliable. However, Table OA1 in the Online Appendix shows the results from estimation of $B$ for the four-factor model. As nearly all loadings in each column are strongly significant, factors do not appear weak, and inference on cross-sectional pricing parameters should be reliable.

3.3.2. Risk price dynamics

Throughout, average risk prices in the restricted models tend to be larger in the time-varying $\lambda_t$ than in the time-invariant specifications. This points to the adequacy of the time-varying risk price model. To investigate this, we show in Table 3 the estimated risk price dynamics, as given by the state transition matrix $A$. In the single-factor model, the autocorrelation coefficient is relatively high, at 0.47 in the unrestricted model and 0.51 with the no-arbitrage restriction imposed. The numbers are nearly as high or even higher for the first state variable (the upper left corner of $A$) in the higher-dimensional models. The next (curvature) state variable gets a marginally lower autocorrelation than the first, around 0.5, for $d = 2$, but higher than the first in the three- and four-factor models (recall from Fig. 2 that for $d = 3$ and 4 the factor associated with the curvature volatility function appearing first for $d = 2$ is the third factor, i.e., the ordering of the second and third factor is reversed, based on the shape of the volatility functions). The hump factor, appearing as the second in the three- and four-factor models, is strongly persistent, with estimated autocorrelation of 0.8 or higher, and so is the fourth (twist) factor, with autocorrelation in excess of 0.9 with no-arbitrage imposed.

The off-diagonal elements of $A$ show the dynamic dependence on lagged values of other state variables [see Eq. (10)] or, equivalently, the dependence of the price of a given risk factor on other factors [see Eq. (8)]. The first (slope) factor depends significantly and negatively on the lagged value of the curvature factor in the two-factor model. Some evidence also exists of the reverse causality, i.e., the lagged first state variable (first column of $A$) causing the second, but the coefficient is an order of magnitude smaller and significant only with no arbitrage imposed. Causality runs both ways between these two factors, slope and curvature, in the three- and four-factor models, where curvature, the second factor for $d = 2$, corresponds to the third for $d = 3$ and 4. The state variable associated with the hump factor (the second for $d = 3$ and 4) causes (helps pricing) all other factors in the three- and restricted four-factor models. It is not caused by other factors for $d = 3$, and not by slope for $d = 4$ either.
(t-statistic of 0.17), although it is caused by curvature and the fourth (twist) factor here. Twist itself exhibits significant causality running in both directions.

With the estimated risk pricing parameters $A$ and $a$ in hand, we explore the hypotheses that a given risk factor driving returns is unpriced or does not matter for cross-sectional pricing, in the sense that it does not condition risk prices. The first type of hypothesis is on a row of $A$ and $a$, and the second on a column of $A$. Evidently, from Table 3, for each $d$, every row and column contains strongly significant parameters, so factors driving returns are priced and matter for pricing of themselves (diagonals are significant) and other factors, too. We also carry out formal likelihood ratio tests of these hypotheses. Results of tests appear in Table 4 for the four-factor model. The first column shows LR statistics for testing each of the rows of $A$ and $a$ equal to zero, and they are all strongly significant (critical value of 11.1 at 5%), i.e., factors are priced. Similar (unreported) tests on rows of just $A$, not $a$, i.e., testing whether a given market price of risk is constant (but potentially nonzero), reject as well. The second column of the table shows LR statistics for testing each of the columns of $A$ equal to zero, and they are strongly significant, too (lowest value 48.7, for the slope factor, against a critical value of 9.5 at 5%), i.e., each factor matters for market prices of risk or, equivalently, for the prediction of future factors, cf. Eq. (11).

3.3.3. Predictions, expectation revisions, and pricing

Panel B of Fig. 2 shows the fitted time series of covariance-generating factors identified as the Kalman filter expectation revisions $\hat{w}_t = E_t[x_t] - E_{t-1}[x_t]$ from the restricted four-factor model, imposing the no-arbitrage condition and allowing for time-varying risk prices. Consistent with the model assumptions, the factors are all of the same order of magnitude, appear roughly serially uncorrelated, and move around a zero level. The corresponding estimated time-varying market prices of risk identified as the one-step-ahead Kalman filter predictions $\hat{\lambda}_t = E_t[x_{t+1}]$ [see Eq. (11)] appear in Fig. 3, with 95% confidence bands. Evidently, these are more strongly serially dependent, consistent with the results from Table 3. Each risk price moves around the stable nonzero level $\mu$ reported in Table 2. The levels and volatilities are different across the four risk prices, in contrast to the fitted factors from Panel B of Fig. 2. The first (slope) risk price is mostly negative, the second (on the persistent hump factor) is sometimes negative but also insignificant through extended subintervals, and the curvature and twist risk prices move in opposite directions and change signs during the period.
To interpret the results, note that the first factor is associated with a downward-sloping volatility function and gets a risk price that is negative. The upward slope or curvature factor, the third exhibit in Fig. 3, gets a risk price that is smaller in magnitude in the restricted two- and three-factor models (Table 2). The upshot is an upward-sloping (in maturity) risk premium, corresponding to the increasing means of slope-adjusted yield changes in Table 1. Fig. 4 shows in Panel A the combined effect on slope-adjusted yield changes of both loadings and risk prices, i.e., this is the contribution $b_{i,j}\mu_j$ to the total average risk premium $B\mu$ from each risk factor $j$, by maturity $i$. From the figure, at least in the one- through three-factor models, by far the majority of the risk premium stems from the pricing of the first (slope) factor. This dominates completely in the one- and two-factor models, and mostly so with three factors, especially for short maturities. The second (hump) factor provides a reinforcing contribution for maturities around 15 months ($i = 4$) and some counteracting contribution for long maturities. Thus, in the three- and four-factor models, the hump factor, although relatively unimportant for the level of the risk premium, contributes to the positive slope for maturities 15 months and longer. The slope and twist factors contribute to both the level and the positive slope of the risk premium, and the pricing of the curvature factor has a modest impact in the three-factor model and partly offsets the slope in the four-factor model.

Panel B of Fig. 4 shows the evolution over time of the total risk premium $b'_i\lambda_t$ for four selected maturities $\tau_i$ (six, 12, 60, and 120 months). Due to the interaction between the cross-sectional effects (the different shapes of the volatility functions) and the time series effects (the risk prices), risk premiums move in different ways across maturities. Toward the end of the sample, they are decreasing over time for short maturities, mainly reflecting the behavior of the first risk price from Fig. 3, and increasing for long maturities, mainly due to the evolution of the fourth risk price.

3.3.4. Pricing errors

In Fig. 5, Panel A, we show the estimated intercepts $\alpha$ from the unrestricted models. They are strongly increasing in maturity for all four dimensions of the vector of factors and both for the constant and the time-varying risk price specifications. This pattern must be captured by $B\mu$ in the restricted models, which is achieved by adding the estimated risk premiums by factor in Fig. 4, where positive slopes are particularly useful in picking up the required pattern to preclude arbitrage opportunities. In Fig. 5, Panel A, we also show the resulting entries of $B\mu$ from the restricted models. With only one factor, the constant and
time-varying risk price restricted models overshoot the unrestricted α at short maturities
and undershoot at long maturities. The overshooting for short maturities continues with
two factors, and the match improves with three and, especially, four factors.

Finally, Panel B of Fig. 5 shows the estimated idiosyncratic variances in Ψ, by maturity.
Errors are again largest in the short end of the term structure, with a bump at the two-year
point (i = 7) and also at the one-year point (i = 3) in the three- and four-factor models.
Axes differ, and errors are smaller with more factors. In fact, they are similar in the three-
and four-factor cases.

3.3.5. Evidence of time-varying premiums and arbitrage

From Table 2, the likelihood values increase with the number of factors, at the expense
of an increase in number of parameters, but the Akaike information criterion (AIC) and the
Schwarz Bayesian information criterion (BIC) also improve, thus pointing to the four-factor
models. For each given number of factors d = 1, . . . , 4, the time-varying risk price model
achieves a better (lower) information criterion than the corresponding constant risk price
model, and the unrestricted model gets better information criteria than the corresponding
restricted (α = 0) model.

The corresponding results from formal LR tests of the no-arbitrage condition $H_0 : \alpha = 0$
in Eq. (12) for the one- through four-factor models appear in Table 5. The table shows tests
of the no-arbitrage condition separately for the constant and time-varying risk price specifi-
cations. Numbers under the test statistics are p-values for rejecting the no-arbitrage condition.
Consistently with the information criteria, the test rejects the no-arbitrage condition in all
cases.

The table also shows tests of the hypothesis of constant risk price (A = 0) against the
time-varying case, separately for the unrestricted and the restricted (no-arbitrage) models.
All tests strongly reject, which is again consistent with the information criteria.

Clearly, the evidence favors time-varying market prices of risk. Further, in spite of
the popularity of three-factor (typically level, slope, and hump or curvature) models in the
literature, our evidence points to at least four factors underlying yields. Even so, the evidence
of arbitrage opportunities suggests that we have still not found the final model. Therefore, we
also investigate models with even more factors included. Fig. OA1 in the Online Appendix
shows the LR tests of the null α = 0 in the time-varying risk price models as a function of
the number of factors \(d\), along with 95% and 99% critical values. The first four test statistics (all significant) are also in Table 5. The LR test turns insignificant (consistent with the absence of arbitrage opportunities) at the 5% level with eight factors, not before, and then again turns significant when even more factors are included. As shown in the panel below the figure, these models have very large numbers of parameters. Fig. OA2 in the Online Appendix shows the corresponding AIC and BIC for restricted and unrestricted models. The AIC decreases throughout and, except for \(d = 8\) or higher, is always better (lower) for the unrestricted model than for the restricted with the same number of factors. These facts together indicate a large number of factors. The BIC is better for the restricted models for \(d \geq 7\) and points to the restricted eight-factor model. As it is generally not expected that so many factors should be needed in fixed income markets, we also investigate the sensitivity to the time period, to see whether the lower-dimensional models perhaps were not rejected in previous time intervals. Panel A of Fig. OA3 in the Online Appendix shows the evolution of \(p\)-values from LR tests of \(\alpha = 0\) in the one- through four-factor time-varying risk price models over rolling ten-year time intervals. All tests reject in all time intervals for the one- through three-factor models. For the four-factor model, the LR test fails to reject at 5% in multiple of the ten-year intervals ending between 1996 and 2002. Panels B and C show the corresponding evolution of the rolling AIC and BIC. Both always favor more factors (up to the four considered in the figure). The unrestricted model is always favored over the corresponding restricted model in the same time interval by the AIC. The BIC selects the restricted model over the unrestricted model in the ten-year interval ending in 2009, in every subsequent interval with one through three factors, and in every interval ending before 2007 with four factors.

All in all, the results point to time-varying market prices of risk. Although some of the subperiod evidence is consistent with the number of factors not exceeding four, it is perhaps a bit surprising that our methodology, filtering factors from the history of (approximate excess) returns, points to at least four and potentially more factors driving bond returns in the full period analysis. Researchers frequently use level, slope, and curvature or other term structure factors to model yields, but it would seem excessive that up to eight distinct risk sources in the yield curve demand time-varying expected returns. Although Duffee (2011b) finds evidence of one hidden factor not explained by common yield factors, Barillas (2011) finds limited evidence of any strictly hidden factor. Perhaps some factors appear needed in our estimation because they matter for risk prices, possibly without driving returns, although this is ruled out by our specification (see Subsection 3.2).

Formally testing the null of redundancy of a factor in driving returns is nonstandard because associated risk price parameters are identified only under the alternative, so our
discussion of the number of factors has been focused on whether the no-arbitrage test rejects and on the information criteria, not on direct tests on loadings. Also, in case of weak factors, inference on risk price parameters $A$ and $a$ is unreliable and, although our main focus is on testing on the intercepts $\alpha$ in Eq. (12), weak or redundant factors could reduce the performance of our approach. This is particularly likely in high-dimensional cases, so we focus mostly on specifications with at most four factors.

Still, our specification could in part be driving the finding of relatively many factors, both through the exclusion of directly observable macroeconomic factors and by ruling out factors that matter for cross-sectional pricing without driving returns. For example, Joslin, Priebsch, and Singleton (2014) find evidence of five factors, with two of them from macroeconomic data. Therefore, we next introduce a generalized state space model for testing the arbitrage restriction, allowing for both macroeconomic factors and unobservable factors spanned by and filtered from the history of returns, to investigate whether more information can be gained by replacing some of the unobservable factors by observable ones.

4. Macroeconomic sources of time-varying risk prices

In this section, we generalize the model to account for information from macroeconomic aggregates on the relevant factors and associated market prices of risk. Doing so allows an investigation of how many and which of the unobservable factors can be replaced by observable ones, both as determinants of market prices of risk and as covariance-generating factors driving returns. Certainly, information would be lost if a factor for which useful macroeconomic data exist were handled as unobservable. Furthermore, including both observable and latent state variables in the model allows exploring whether the inclusion of observable macroeconomic series leads to an empirical specification that passes the requirement of absence of arbitrage opportunities.

In our generalized model, a given macroeconomic variable potentially takes on separate roles in explaining the covariance-generating factors driving returns and in determining the associated risk prices. Separate parameters are identified in the two places and, in contrast to Section 3, we can test whether a factor matters only for market prices of risk, without driving returns. We first use a regression approach to investigate which macroeconomic series have the most explanatory power for our time-varying market prices of risk and then use the results in the specification of the generalized state space model.

4.1. Regression evidence

We explore the potential macroeconomic underpinning of premiums in our model by regressing our estimated time-varying market prices of risk one by one on variables suggested
in the literature. We do not consider principal components computed from yield data, because the effect of these would be captured by the latent state variables already included in our state space model. The most explanatory power is achieved using the variables exhibited in Fig. OA4 of the Online Appendix as regressors, namely, the second and fourth of leading principal components extracted from the Stock and Watson (2002) balanced panel of 111 variables from the McCracken and Ng (2016) FRED-MD data set (appropriately transformed and detrended to stationarity, following the article) and the monthly change in industrial production (IP) and in nonfarm payroll employment. Principal components (PCs) from the Stock-Watson data were used by Ludvigson and Ng (2009) as well. In other experiments, we tried the other leading PCs up to the eighth, as well as the output gap, defined as deviations of industrial production from a nonlinear trend as in Cooper and Priestley (2009), the monthly change in the US Consumer Price Index inflation rate, the monthly relative change in nondurable consumption expenditures, the term spread (ten years less three-month Treasury yields), the credit spread (Baa less Aaa rates from Moody’s), the stock market risk price or Sharpe ratio, defined as the monthly Standard & Poor’s (S&P) 500 excess return above the three-month Treasury yield, relative to the sum of daily absolute returns over the month, and, finally, the short rate (three-month yield) itself.

Results from regressing our market prices of risk on the four macroeconomic variables appear in Table OA2 of the Online Appendix. Each (vertical) portion of the table shows results for one of the four risk prices from the estimated latent factor model. The four univariate regressions are reported in the first column of each portion of the table. For the second through fourth risk price, this is followed by the multivariate regression of the relevant, say, $k^{th}$ risk price on the first $k^{th}$ macroeconomic variables. The market price of the first (slope) risk factor is significantly related to the second principal component ($t$-statistic of $-4.6$) and that of the second (curvature) factor to this and IP. The hump risk price is significantly related to the fourth Stock-Watson PC; and the twist risk price, to employment and the second PC.

---

3 The selection of the four variables (IP, employment, and two PCs) results from the following iterative procedure. The predicted factor scores from the single-factor latent model were regressed on all the candidate macro variables in univariate specifications, and the most significant explanatory variable was chosen, namely, the second Stock-Watson PC. Each time a new latent factor was considered, say, the $k^{th}$, the predicted scores from the $k$-factor latent model were regressed on all remaining macro variables, both in univariate specifications and in multivariate specifications using the $k - 1$ macro variables already selected (second PC and so on) along with a macro variable not yet selected. If the most significant variable in the univariate regressions remained significant in the multivariate case, it was selected; otherwise, the second-most significant from the univariate was considered for significance in the multivariate, and so forth. Similar results were obtained from a non-iterative procedure, regressing the four factors from the latent four-factor model on all macro variables.

---
Overall, the results suggest that our estimated market prices of risk are significantly related to relevant macroeconomic variables. We conclude that our estimates represent true economic pricing mechanisms in the bond market.

4.2. Expanded state space model with macroeconomic variables

The regression evidence suggests that some of the state variables $x_t$ that condition the market prices of risk $\lambda_t$ in Eq. (8) are observed macroeconomic variables in the data set. This implies a loss of information from modeling $x_t$ as a purely latent state vector, as in Section 3. We now extend the state space model to accommodate both latent and observable state variables. It is natural to identify the covariance-generating factors in slope-adjusted yield changes, i.e., $w_t$ in Eq. (5), with the unanticipated shocks or innovations to state variables, as was done for the case of latent state variables in Eq. (10). This identification of factors with shocks to the states that condition risk prices corresponds to the HJM specification that the process for market prices of risk is adapted to the filtration generated by the driving Wiener processes $W_t$ in Eq. (1). The main difference arising when the state variables are observed is that no reason exists for the dynamics of the state variables, as given by the VAR transition matrix in Eq. (10), to coincide with the coefficient matrix $A$ showing how state variables determine market prices of risk in Eq. (8). For example, even a state variable that is not highly serially correlated could have a strong impact on risk prices.

Without loss of generality, let the $d$-vector of covariance-generating factors $w_t$ be decomposed as

$$ w_t = \begin{pmatrix} w_o^t \\ w_u^t \end{pmatrix}, $$

(16)

where $w_o^t$ is the $d^o$-vector of innovations to the observable macroeconomic state variables and $w_u^t$ is the $d^u$-vector of innovations to the unobservable or latent state variables, with $d = d^o + d^u$. Thus, the model for slope-adjusted yield changes generalizing Eq. (5) is

$$ \tilde{y}_t = \mu_t + B^o w_o^t + B^u w_u^t + \varepsilon_t, $$

(17)

where $B^o$ is the $m \times d^o$ matrix of loadings on macro factors and $B^u$ is the $m \times d^u$ matrix of loadings on latent factors. As before, the latent factors are normalized for identification purposes, so $\text{var}(w_u^t) = I_{d^u}$, and $B^u$ has zero loadings above the diagonal in the top square $d^u \times d^u$ matrix [compare Eq. (14)]. Because the macro factors are innovations to observed series, they are not normalized. So, we write $\text{var}(w_o^t) = \Sigma^o$, a general $d^o \times d^o$ variance-covariance matrix, and $B^o$ is completely unrestricted, to be thought of more as a matrix of regression coefficients.
We assume that macro and latent factors are uncorrelated, \( \text{cov}(w_o^t, w_u^t) = 0 \). This implies that, as regressors, the macro factors are exogenous, and allows recasting the null hypothesis of absence of arbitrage opportunities as the zero intercept restriction \( H_0: \alpha = 0 \) on the conditional means

\[
\mu_{it} = \alpha_i + (b_o^t)' \lambda_{i-1}^o + (b_u^t)' \lambda_{i-1}^u + \frac{T_i}{2} (b_o^y)' \Sigma^o (b_o^y) + \frac{T_i}{2} (b_u^y)' (b_u^y)
\]

that now replace Eq. (7). Here, \((b_o^y)'\) and \((b_u^y)'\) are the \( i \)th row in \( B^o \) and \( B^u \), respectively, \( \lambda_{i-1}^o \) and \( \lambda_{i-1}^u \) are the conformable vectors of market prices of macro and latent factor risk, and \( \Sigma^o \) appears in the no-arbitrage condition because the macro factors are allowed to correlate, which is empirically warranted.

Some structure is imposed, for identification and to facilitate interpretation. Thus, observable and unobservable state variables evolve through time according to separate VAR systems,

\[
x_o^t = c^o + C^o x_o^{t-1} + w_o^t
\]

and

\[
x_u^t = a^u + A^u x_u^{t-1} + w_u^t,
\]

where \( c^o \) is \( d^o \times 1 \), \( C^o \) is \( d^o \times d^o \), \( a^u \) is \( d^u \times 1 \), and \( A^u \) is \( d^u \times d^u \). Macro state variables depend on neither current nor past unobservable state variables, only on their own past and contemporaneous innovations, and vice versa. Furthermore, we assume that market prices of macro risk depend only on macro state variables and that market prices of latent factor risk depend only on latent state variables,

\[
\lambda_i^o = a^o + A^o x_i^o
\]

and

\[
\lambda_i^u = a^u + A^u x_i^u,
\]

where \( A^o \) is \( d^o \times d^o \) and \( a^o \) a conformable vector. Again, we have \( x_i^u = \lambda_{i-1}^u + w_i^u \) and thus

\[
\lambda_i^u = E_t (x_{i+1}^u),
\]

i.e., the prices of latent risks are given by the conditional expectations of the relevant state variables [compare Eqs. (9)-(11)]. A similar interpretation does not apply to the prices of macro risks, where possibly \( a^o \neq c^o \) and \( A^o \neq C^o \), because, with observable state variables, the cross-sectional pricing coefficients potentially differ from the dynamic transition coefficients, as in Duffee (2011a).
To derive the state space model, Eq. (18) is substituted in Eq. (17). Collecting stochastic terms multiplying \((b_i^o)'\) yields \(\lambda_{t-1}^o + w_t^o\), and the unobserved series \(x_t^o\) is substituted for this. This corresponds to the derivation of the model in Eqs. (12) and (13). The stochastic terms multiplying \((b_i^o)'\) are \(\lambda_{t-1}^o\) and \(w_t^o\), and they must be separately substituted, using the pricing equation from Eq. (21) for \(\lambda_{t-1}^o\) and the transition equation from Eq. (19) for \(w_t^o\). These steps produce a generalized state space model with unobservable state transition equation given by Eq. (20) and measurement equations for the two observed multivariate series \(x_t^o\) and \(\tilde{y}_t\) given by Eq. (19) for the former and

\[
\tilde{y}_t = \alpha + B^o (a_1^o - c^o) + \text{vec}_1m\{[(b_i^o)' \Sigma^o (b_i^o) + (b_i^u)' (b_i^u)] \tau_t/2\} + B^o [x_t^o + (A^o - C^o) x_{t-1}^o] + B^ux_t^u + \varepsilon_t
\]

for the latter. The measurement equation for \(\tilde{y}_t\) involves the unobservable state variables \(x_t^o\), implying that the model is in state space form. In this sense, the term \(B^o x_t^o\) in Eq. (24) corresponds to the term \(Bx_t\) in the simpler state space case [Eqs. (12) and (13)], where all of \(x_t\) was unobserved. In the generalized model, the measurement equation for \(x_t^o\) does not involve \(x_t^u\), so, if this had been the only measurement equation, the model would reduce to an autoregression (a trivial state space form). The term \(B^o [\cdot]\) in Eq. (24) involves current and lagged observables and, so, represents a tractable regression-style complication relative to the simpler state space model without macroeconomic variables. The presence of this term implies that the macro variables have predictive content for excess returns, which is one of the modeling criteria of Joslin, Priebsch, and Singleton (2014). The hypothetical maturity \(\tau_t\) bond has one-period conditional expected excess return given information through \(t - 1\) given by \(E_{t-1}(-\tau_t \tilde{y}_t, \tau_t)\). Using Eqs. (19) and (20) in Eq. (24), we have

\[
E_{t-1}(\tilde{y}_t) = \alpha + B^o (a_1^o + A^o x_{t-1}^o) + B^u (a_u^o + A^o x_{t-1}^u) + \text{vec}_1m\{[(b_i^o)' \Sigma^o (b_i^o) + (b_i^u)' (b_i^u)] \tau_t/2\}.
\]

Evidently, for the \(j\)th macro variable not to condition expected excess returns, we would require that the \(j\)th column of \(B^o A^o\) be zero, which generically requires that the \(j\)th column of \(A^o\) be zero, i.e., the factor does not condition market prices of risk, cf. Eq. (21), which is a testable restriction. Our approach similarly satisfies the second modeling criterion of Joslin, Priebsch, and Singleton (2014), that the macroeconomic risks are unspanned by bond yields. Our model does not impose that \(x_t^o\) can be backed out from yields or \(\tilde{y}_t\).4

\[4\text{Although risks can be unspanned, the spanned portion is priced in nominal yields, i.e., market prices of risk multiply on loadings that are spanned but can themselves depend on unspanned macro information, as emphasized by Joslin, Priebsch, and Singleton (2014) and accommodated in Eq. (21), where the effect can} \]
The parameters in \( c^o, C^o, \) and \( \Sigma^o \) are identified in the VAR measurement equation for \( x^o_t \), Eq. (19), and those in \( A^u \) are identified in the unobservable state transition equation for \( x^u_t \), Eq. (20). In the latter, \( a^u \) is identified under the null \( \alpha = 0 \) and is set to zero under the alternative. A simplified two-step estimation procedure is to run Eq. (19) as a simple vector autoregression in the first step. The VAR parameters are fixed in the second step, which estimates the remaining parameters using a state space framework involving the state transition equation for \( x^u_t \), Eq. (20), and the remaining measurement equation, namely, that for \( \tilde{y}_t \), given by Eq. (24). Together, the two steps deliver consistent but inefficient estimates of all parameters. Efficient estimates are obtained by using the two-step estimates as starting values in a one-step estimation of the full model, consisting of the state transition equation for \( x^u_t \), Eq. (20), and the measurement equations for both \( x^o_t \) and \( \tilde{y}_t \), Eqs. (19) and (24). This is the approach we pursue in our empirical work.

Adrian, Crump, and Moench (2013) consider the special case with no unobservable state variables. The state space model reduces to a nonlinear regression in the second step, because the term \( B^u x^u_t \) is not present in Eq. (24). As in Fama and MacBeth (1973), the nonlinear regression is decomposed into two consecutive linear regressions, the first estimating loadings \( B^o \) [and \( (A^o - C^o) \)] by regressing \( \tilde{y}_t \) on current (and lagged) observed state variables over time [alternatively, Adrian, Crump and Moench (2013) estimate \( B^o \) and \( A^o \) by regressing instead on the fitted factors \( \tilde{w}^o_t = x^o_t - \tilde{C}^o x^o_{t-1} \) from the first step and lagged state variables \( x^u_{t-1} \)] and the second estimating \( (a^o - c^o) \) by cross-sectional regression of the \( m \) unconstrained time series intercepts less the vec(m{·})-terms from Eq. (24) on \( \tilde{B}^o \). The observable state variables considered by Adrian, Crump, and Moench (2013) include principal components computed from yield data. In our empirical work, we focus on observable macroeconomic variables instead of PCs, because our state space approach generates unobservable state variables \( x^u_t \) that would likely subsume the information content in such yield-based PCs.

Whether two-step or full model estimation is used, identification of the remaining parameters in Eqs. (19), (20), (21), (22), and (24) deserves some attention. First off, \( B^u \) is identified [under the condition of zero loadings in the upper right corner, as in Eq. (14)] as the loading matrix on the factor scores or latent state variables \( x^u_t \) in the state space framework; \( A^u \), as the transition matrix for these latent factors; and \( B^o \), as the regression coefficients on observed \( x^o_t \). Given identification of \( B^o \), the matrix \( (A^o - C^o) \) is identified by the variation in lagged macro variables \( x^o_{t-1} \). Because \( C^o \) is already identified in the VAR for \( x^o_t \) in Eq. (19), the upshot is that the pricing parameters \( A^o \) for the macro variables are identified. Finally, under the null of \( \alpha = 0 \), the restricted intercepts in Eq. (24) take the form \( B^o (a^o - c^o) \) and, be tested directly as a zero restriction on \( A^o \).
with $B^o$ already identified, the $d^o$-vector $(a^o - c^o)$ is identified, too. Because $c^o$ is identified in the VAR for $x^o_t$ in Eq. (19), all parameters under the null are identified, including the macro pricing intercepts $a^o$ and the latent factor intercepts $a^u$. Under the alternative, an unrestricted $\alpha$ is estimated, $a^u$ is set to zero as in the model without macro variables, and $a^o$ is unidentified.

As discussed in Subsection 3.2, our framework allows investigating whether a factor matters for pricing without driving returns, and vice versa. By similar arguments as in the model without macro variables, if the $j^{th}$ latent factor does not drive returns, i.e., the $j^{th}$ column of $B^u$ is zero, then this factor cannot matter for pricing, because in this case, the $j^{th}$ row and column of $A^u$ and the $j^{th}$ entry in $a^u$ are unidentified. However, this is different for observable factors. Even if the $j^{th}$ macro factor does not drive returns, i.e., the $j^{th}$ column of $B^o$ is zero, the presence of the term $(A^o - C^o) x^o_{t-1}$ in Eq. (24) provides for a potentially important pricing role for the state variable in question. If the $j^{th}$ column of $(A^o - C^o)$ is nonzero and there is one or more nonzero off-diagonal entries down the $j^{th}$ column, then the lagged value of the $j^{th}$ state variable is not wiped out by the zero $j^{th}$ column of $B^o$. The $j^{th}$ coordinate of $x^o_{t-1}$ matters in the model for $\tilde{y}_t$, even though the $j^{th}$ coordinate of current $x^o_t$ does not. The effect is through the state variable’s role in pricing, not through the covariance-generating factors driving returns. This situation arises when the $j^{th}$ state variable conditions the market prices of one or more of the risks associated with other state variables, i.e., through elements $(i,j)$ in the pricing matrix in Eq. (21), $A^o$, for $i \neq j$, and when these elements do not coincide with the corresponding $(i,j)$ entries in the transition matrix in Eq. (19), $C^o$. This simply means that the state variable in question does not condition the market prices of risks associated with other state variables in the same manner in which it conditions the future distribution of those other state variables. This is essentially to say that the future state variables are not just the current market prices of risk plus noise, and there is no reason they should be. This type of relation was adopted for simplicity for the unobservable state variables, in Eqs. (11) and (23), given the lesser information from data in that case.

Thus, in contrast to the model with latent factors only, the presence of macro variables implies that a state variable can matter for pricing without driving returns, but testing for this phenomenon is again nonstandard. The reason is that under the null that the $j^{th}$ column of $B^o$ is zero, the $j^{th}$ row of $A^o$ and $a^o$ are unidentified.

Conversely, it can be the case that a factor does not matter for pricing, even though it is driving returns, i.e., although the $j^{th}$ column of $B^o$ is nonzero, the $j^{th}$ row of $A^o$ and $a^o$ can be zero (the factor is unpriced) or the $j^{th}$ column of $A^o$ can be zero (the factor does not condition market prices of risk), or similarly for $A^u$, $a^u$, and $B^u$. Testing for this is
unreliable, though, if the factor in question is weak.

Important similarities and differences exist compared with the case of testing the APT in equity data. In the APT case, if the factors are traded portfolios, then they have risk prices given by their mean returns. If they are not spanned by stock prices, then the intercepts can be nonzero even under the null, where instead they take the form of the factor loading matrix times a vector of market prices of risk that potentially deviates from the factor means and is unidentified under the alternative. Comparing this with a static version of our model of yields, the APT market prices of unspanned risk correspond to our pricing intercepts $\alpha^o$ for observable factors; the APT factor means, to our $c^o$.

It could appear also that our average risk prices for unobservable factors play a role similar to that of the traded factor risk prices in the APT case, i.e., they are estimated by the factor means $\mu^u = (I - A^u)^{-1} a^u$ (or just $a^u$ in the static case). Yields are not traded assets, and our state variables need not be either. In our case, the association between factor means and risk prices arises not because factors are traded, but for identification purposes. Furthermore, although the case of observable factors in a static version of our yield model does have similarities to the non-traded factor version of the APT test, in that we identify $\alpha^o$ and $c^o$ under the null, we also identify the pricing matrix $A^o$ and the transition matrix $C^o$ separately in our dynamic version, and this is both under the null and the alternative. Finally, our model also involves the regression-style generalization $B^{o \cdot}$ in Eq. (24), stemming from the dynamic specification.

The analysis shows that adding observations on state variables allows separating the cross-sectional pricing from the state variable dynamics underlying slope-adjusted yield changes. The difference between the two is given in the model by $\alpha^o - c^o$ and $A^o - C^o$ appearing in Eq. (24). The first is identified under the null (as are $\alpha^o$ and $c^o$ individually), and the second is in addition identified under the alternative (as are $A^o$ and $C^o$ individually).

4.3. Results from expanded model

Table 6 shows results from the expanded state space model with time-varying risk premiums and using the macroeconomic aggregates both for covariance-generating factors and for conditioning risk prices. In the table, columns correspond to the cases $d = 1, \ldots, 4$ factors, of which $d^o = 0, 1, \ldots, d$ (corresponding to the row dimension of the table) are based on macroeconomic observables, with the remaining $d^u = d - d^o$ factors in each model treated as unobservable. For each combination of $d$ and $d^o$ considered, both an unrestricted model and a restricted version with the no-arbitrage condition imposed are estimated. The resulting maximized log likelihood value, the number of parameters, and the AIC and BIC are reported. Based on the results from Table OA2 of the Online Appendix, the macroeconomic
variables are entered in a certain sequence. Thus, when \( d^o = 1 \), the second Stock-Watson principal component is used as the sole macro variable in the model, given that this appeared as the most powerful predictor of the first risk price in the regression analysis. When \( d^o = 2 \), IP is included as well. When \( d^o = 3 \), the fourth principal component is added; for \( d^o = 4 \), nonfarm payroll employment.

From Table 6, likelihood values increase when more factors are left unobservable in the one- and two-factor cases. With three factors, likelihoods increase as unobservable factors replace observable ones, until a maximum is reached with \( d^o = 1 \) of the factors observed. In the \( d = 4 \) case, maximum likelihood is reached with two of the four factors observed, and the other two unobserved. Although the unobservable factors should be able to take on either the same values as the observed, or different values if this would increase likelihood, two reasons exist that likelihood can nevertheless be increased by replacing unobservable factors by observed factors. First, the likelihood function includes the contributions from both measurement equations, i.e., for the slope-adjusted yield changes [see Eq. (24)] as well as for the macroeconomic variables [see Eq. (19)], and this leaves the likelihoods not directly comparable when the number of macro variables differs. Second, when an unobserved factor is replaced by an observed one, the restriction that the dynamic transition parameters coincide with the cross-sectional pricing parameters is relaxed, which leads to an increase in likelihood, at the expense of an increase in number of parameters, as also reflected in the table.

The results in Table 6 suggest that some of the unobservable factors can be replaced by observable ones. The three- and four-factor specifications are preferred, based on the previous analysis and, here, the new results indicate that one or two of these factors should be selected from the set of available macroeconomic variables. For each combination of number of observable factors \( d^o \) and total number of factors \( d \), and for each of AIC and BIC, numbers set in boldface indicate whether the information criterion favors the unrestricted model or the restricted version with the no-arbitrage condition imposed. The BIC favors the restricted over the unrestricted model whenever at least one macroeconomic variable is included \((d^o \geq 1)\). Also, both the AIC and the BIC (and the likelihood ratio test; see Table 7) favor the restricted over the unrestricted specifications when there are two or more factors and they are all observable \( (d^o = d \geq 2) \).
Throughout Table 6, the best information criteria are achieved by the specification with four factors, of which two are observable and two are unobservable. Within this specification \((d^o = 2, d = 4)\), the BIC favors the restricted version with the no-arbitrage condition imposed over the unrestricted, and the difference between the AIC is smaller than when \(d^o\) or \(d\) is reduced \((-61,789\) for the unrestricted versus \(-61,783\) for the no-arbitrage model). If one unobservable factor is dropped (so, \(d^o = 2, d = 3\)), the BIC still favors the restricted over the unrestricted model. However, for \(d = 3\) factors, both the AIC and the BIC improves by taking only one of the factors observable \((d^o = 1, d = 3)\). Here, the BIC continues to favor the restricted over the unrestricted model, but both criteria are considerably improved by moving to the model with four factors, of which two are observed. Again, in this model \((d^o = 2, d = 4)\), the BIC selects the restricted version over the unrestricted version \((\text{BIC} = -61,096)\). From Table 7, the LR test of \(\alpha = 0\) does reject \((\text{statistic of 29.7, on 12 degrees of freedom})\) for the full period. From Panel A of Fig. OA3 in the Online Appendix, this is not typical for this model. With rolling ten-year windows, the LR test fails to reject the null of \(\alpha = 0\) at 5% over the windows ending in the majority of years. From Panel C, the BIC selects the restricted no-arbitrage version over the unrestricted in every time interval and, from Panel B, even the AIC selects the restricted specification in almost every ten-year interval.

Overall, this evidence suggests that four factors in interest rates suffice for an arbitrage-free model and are preferred over three factors. Four factors \((d^o = 2, d = 4)\) are favored by the conservative information criteria that would penalize an excessive number of parameters. Furthermore, for each model with \(d \leq 3\) in the table, adding an unobservable factor (moving one model horizontally to the right in the table) or adding an observable factor (moving one model right and one down) improves information criteria, except for a few cases along the diagonal \((d^o = d)\), where adding one more macro variable need not be an improvement. Finally, the evidence supports the no-arbitrage restriction when at least one macro variable is included. In this sense, both statistical and financial considerations point to four instead of just three (or fewer) factors and reinforce the value of using both macro and latent factors.

Estimates of loadings on unobserved and observed factors in the preferred model appear in Table OA3 of the Online Appendix. As they are all significant, all four factors are driving returns. Table 8 shows the results of LR tests of whether factors are priced and whether they matter for risk prices. All tests are strongly significant. The statistics for the macro factors not to be priced are just in excess of 40 for both, against a critical value of 7.8 at 5%. Statistics for unspanned macroeconomic information not to condition risk prices are about 26 and 31, against a critical value of 6 at 5% on two degrees of freedom. By Eq. (25), this implies that the same unspanned macroeconomic information conditions expected
returns, too. The table also reports LR tests (again by row and column) of whether the cross-sectional pricing parameters $a^o$ and $A^o$ from Eq. (21) are equal to the dynamic state transition parameters $c^o$ and $C^o$ for the macro variables in Eq. (19). They are not. The pricing (or risk-neutral, or martingale) measure differs significantly from the physical.

[insert Table 8 near here]

All in all, the results show that replacing some of the unobservable factors with observed macroeconomic aggregates is useful and leads to new important insights. In contrast to the analysis using unobservable factors only, in Section 3, we now find that just four factors suffice to get a well-specified model that precludes arbitrage opportunities.

5. Sensitivity analysis

In this section, we investigate the sensitivity of our general approach to a number of relevant issues arising in practice, such as errors in model specification, data, distributional assumptions, etc. In Subsection 5.1, we examine the effect on test statistics of omitting the slope adjustment of yield changes and the convexity term. In Subsection 5.2, we explore a number of affine subclasses from the literature. In additional diagnostic checks, for robustness to alternative distributional assumptions, reported in Section B of the Online Appendix, we estimate two model variations in which we use a principal components analysis in the first step of a two-step procedure similar to that of Gultekin and Rogalski (1985). In the second step, risk prices are determined by cross-sectional regression on loadings from the first step. We show that this does not deliver genuine time-varying risk prices, whereas our full state space approach does. In addition to the cross-sectional information on risk prices, the state space model uses the time series pattern of yields, and the Kalman filter determines the optimal combination of the two sources of information.

5.1. Yield changes and convexity

Our approach is to test for no arbitrage at the level of yields, which is how traded instruments are quoted. In this case, the relevant no-arbitrage condition is $a_i = 0$ imposed on Eq. (7) in the unobservable factor case and on Eq. (18) in the general case with observable and unobservable factors. In both cases, the test is on the conditional means of slope-adjusted yield changes. An alternative analysis at the level of returns to the heterogeneous bonds, bills, and notes that compose the market instead could focus on the standard APT condition omitting the quadratic form in loadings in Eqs. (7) and (18). Further, the slope-adjusted yield changes [Eq. (4)] are related to raw yield changes $\Delta y_{t,\tau} = y_{t,\tau} - y_{t-1,\tau}$. The differences are the average slope or yield spread $(y_{t-1,\tau} - y_{t-1,\tau_0})/\tau_i - \tau_0)$ and local slope or bond aging.
\((y_{t-1,\tau_i} - y_{t-1,\tau_{i-1}})/(\tau_i - \tau_{i-1})\) terms, which empirically tend to be decreasing in \(i\) as the step length between the maturities increases. We examine, as a first diagnostic, the behavior of the test when it is applied directly to raw yield changes \(\Delta y_{t,\tau_i}\), without slope adjustment, and with the quadratic form in factor loadings left out. In Eq. (24), the convexity term \(\left[\left(b_0^\tau\right)^\prime \Sigma_o \left(b_0^\tau\right) + \left(b_u^\tau\right)^\prime \left(b_u^\tau\right)\right] \tau_i/2\) is omitted, and raw instead of slope-adjusted yield changes are used as data on the left-hand side. The rest of the analysis is identical to that in Section 4.

LR tests of the no-arbitrage condition \(\alpha = 0\) in specifications with \(d = 1, \ldots, 4\) factors, of which \(d^o = 0, 1, \ldots, 4\) observable, are reported in Panel A of Table 9, which is directly comparable to Table 7. The results are entirely different. Whereas in Table 7 the null is rejected at 1% unless two or more factors are included, and all of them observable, it is simply never rejected in Panel A of Table 9. This verifies the importance of correct specification, using slope-adjusted instead of raw yield changes, and accounting for convexity.

Some further details are available in the Online Appendix, where Panel A of Table OA4 is directly comparable to Table 6. In both tables, the information criteria point to the \((d, d^o) = (4, 2)\) specification, within which the BIC selects the no-arbitrage version. In Panel A of Table OA4, both information criteria favor the no-arbitrage restricted model over the corresponding unrestricted model, for every combination of observable and unobservable factors, consistent with little power to reject in Table 9, Panel A. With slope adjustment and convexity (Tables 6 and 7), the AIC favors the restricted over the corresponding unrestricted model only when the LR test in Table 7, Panel A, fails to reject; the BIC, only when observable factors are included, such as in the preferred model with two macro and two unobservable factors.

While precluding arbitrage opportunities in equities requires that mean excess returns be spanned by the volatilities through a linear function (the market prices of risk), in the term structure case it is the means of appropriately slope-adjusted yield changes less a nonlinear function of the volatilities that must be spanned in this manner. Ignoring this could lead to a loss in power to detect arbitrage opportunities. This point is particularly important when the absence of arbitrage opportunities is included as a requirement in specification searches, because ignoring it could lead to acceptance of a too small model.

5.2. **Affine subclasses**

Well-known affine term structure models arise by appropriately restricting the general volatility functions. For additional evidence on the sensitivity to model misspecification, we consider restrictions generating the affine models of Ho and Lee (1986) and Hull and White (1990), as well as more general volatility restrictions.
The volatility functions in Panel A of Fig. 2 behave like factor loadings in the setup of the general model. In combination, the estimated functions capture standard level, slope, and curvature or hump shapes. This suggests the possibility of more parsimonious parametrizations of the volatility or factor loading matrix $B$. In the case of a single covariance generating factor, $d = 1$, the model could, for example, be rigged to generate the forward rate volatility function $\sigma_f(\cdot)$ from the Ho-Lee setting: $\sigma_f(\tau) = \sigma_1$, i.e., flat forward rate volatility. Integrating and dividing by the time to maturity provides a yield volatility of exactly $\sigma_1$ in this case for all maturities: $\sigma(\tau) = \sigma_1$. Our volatility matrix $B$ would be a vector: $\sigma_1 \times 1_m$, i.e., the first factor is a level factor. For $d = 2$, we could take the first volatility factor as in the Ho-Lee case and the second from the Hull-White model: $\sigma_f(\tau) = \sigma_2 \exp(-\gamma_1 \tau)$, exponentially declining forward rate volatility, with $\gamma_1$ a parameter to be estimated. Similar derivations as before produce the yield volatility $\sigma^2(\tau) = \sigma_2 (1 - \exp(-\gamma(\tau)))/\gamma_1 \tau$ in this case, i.e., the second factor is a slope factor, and the level and slope loadings or volatility functions only use up three degrees of freedom through $(\sigma_1, \sigma_2, \gamma_1)$. Given our empirical findings of the relevance of a hump factor and the similarity of the first two volatility functions with the Nelson and Siegel (1987) model, we take a similar hump-shaped yield volatility as in that model for the next dimension. For the full four-factor case $d = 4$, we set

$$
\begin{pmatrix}
\sigma_1 \\
\sigma_2 \frac{1 - e^{-\gamma_1 \tau}}{\gamma_1 \tau} \\
\sigma_3 \frac{1}{2} \left[ 1 + (1 + 2 \gamma_2 \tau) e^{-\gamma_2 \tau} \right] \\
\sigma_4 \frac{1 - e^{-2 \gamma_3 \tau}}{2 \gamma_3 \tau}
\end{pmatrix},
$$

and, in case $d < 4$, we use the first $d$ elements of this vector. The specification of the fourth factor as a second slope factor is motivated by our empirical findings (see Fig. 2). In addition, following Björk and Christensen (1999), an arbitrage-free yield curve shape consistent with the Hull and White (1990) extended Vasicek (1977) yield volatility function requires in addition to the three Nelson-Siegel terms a fourth term, namely, a second slope term with double slope, in the sense $\gamma_3 = \gamma_1$ in Eq. (26), a condition that therefore can be tested.

Panel B of Table 9, again directly comparable with Table 7, provides the LR test results when we estimate the time-varying risk price HJM model with the restricted volatility functions [Eq. (26)] and different combinations of observable and unobservable factors. We consider an ordering with the macro factors (if any) first and then the unobservable factors (if any). Here, the no-arbitrage restrictions are rejected at 1% in all cases. The restricted volatility functions could imply a rotation that does not go well with our sequencing of the macro variables, but even with all factors unobservable, the $d^p = 0$ portion of Table 9,
Panel B, all tests reject strongly. From Panel B of Table OA4 in the Online Appendix, the information criteria again point to the \((d, d^o) = (4, 2)\) specification, but the AIC no longer selects the no-arbitrage version of this. It selects the unrestricted \(\alpha\) version of every \((d, d^o)\) combination, and the BIC now requires at least two factors, and all of them macro, before pointing to no arbitrage. Thus, restricting the volatility functions does change inference.

For each \((d, d^o)\) combination, likelihood values and information criteria can be compared across Table 6 (general volatility functions) and Panel B of Table OA4. In spite of the vastly reduced number of parameters with restricted volatility functions in \(B\), both the AIC and the BIC favor our general volatility function specifications (Table 6) in all cases.

Thus, on the methodological side, the analysis shows the importance of using our approach to testing the no-arbitrage restrictions in the general HJM framework, without parametric restrictions imposed on the factor loadings. On the substantive side, the results provide evidence against the affine subclasses considered.

6. Conclusion

Our general dynamic factor model approach facilitates analysis of several relevant issues in term structure analysis, such as the number of factors, the shapes of the volatility functions, the no-arbitrage drift restriction, the dynamics of risk premiums, and their relation to macroeconomic conditioning variables. On the methodological side, the implementation clearly demonstrates the tractability of our likelihood-based state space approach. Formally, Gaussian distributional assumptions are adopted, but we demonstrate consistency of our results with those obtained using robust alternative procedures. Furthermore, even without Gaussianity, the Kalman filter generates minimum mean squared error linear predictions of the latent state variables and the market prices of risk. On the substantive side, our results show the importance of time-varying risk premiums in the Heath, Jarrow, and Morton (1992) framework. We show that some of the unobservable factors can be replaced with observed macroeconomic aggregates. In our generalized state space model, the observed macroeconomic variables can play different roles as covariance-generating common factors (through the innovations to the macro variables) determining the dynamic evolution of yields and as determinants of risk prices in the cross-sectional pricing mechanism. The most successful specifications include both observed and unobserved factors. The evidence favors a model with four factors, of which two are unobservable and two are given by observed macroeconomic variables, namely, the second Stock-Watson principal component and the change in industrial production. For this specification, little evidence exists of any arbitrage opportunity. Thus, our results reinforce those of Joslin, Priebsch, and Singleton (2014),
i.e., important information on market prices of risk and the prediction of excess returns is contained in unspanned macro factors over and beyond what is available in yields only.

Our work establishes the empirical importance of the nonlinear (quadratic in loadings) term in the no-arbitrage condition and of proper treatment of bond aging and yield spread through adjustment of yield changes for local and average slope of the yield curve. Judging from our results, derivative pricing in the HJM framework should not be based on model specifications with only one, two, or three factors and should not restrict the shape of the volatility function to the standard parametrized forms associated with the affine subclasses considered, even though both simplifications are extremely popular in practical pricing applications. The indication is that the framework should not be adopted for derivative pricing based on a volatility function and an initial yield curve only, without a preceding analysis of the appropriate structure involving dynamic risk premium specification and examining the model under the physical measure for the presence of arbitrage opportunities.

Appendix A. The no-arbitrage drift condition

We derive the appropriate no-arbitrage drift condition for the yield dynamics with fixed term to maturity. Our setting and that in HJM have two main differences: (1) we study yields to maturity, while HJM consider instantaneous forward rates; and (2) we use the parametrization from Brace and Musiela (1994) with fixed term \( \tau \) to maturity, whereas HJM use a fixed maturity date \( T \) in their notation.

The HJM specification of the dynamics of instantaneous forward rates takes the form

\[
df(t, T) = \alpha_f(t, T)\, dt + \sigma_f(t, T)\, dW_t, \tag{A.1}
\]

i.e., an infinite dimensional stochastic differential equation for the forward curve \( f(t, \cdot) \), where \( \alpha_f \) and \( \sigma_f \) are the forward rate drift and volatility functions, \( t \) is calendar time, and \( T \) is the fixed maturity date. Equivalently, we assume that under the physical measure the yield curve dynamics can be written as

\[
dy(t, \tau) = \alpha(t, \tau)\, dt + \sigma(t, \tau)'\, dW_t, \tag{A.2}
\]

which is Eq. (1). Here, \( \tau \geq 0 \) indicates term to maturity. For the traded bond with maturity date \( T \), write \( T = t + \tau \), i.e., term to maturity \( \tau \) shrinks as calendar date \( t \) increases, which is the bond aging effect. Write \( p(t, T) = \exp(-\,(T - t)\, y(t, T - t)) \) for the zero coupon bond
price. By Itô’s lemma, using $T = t + \tau$, bond price dynamics are

\[
\frac{dp(t, T)}{p(t, T)} = -\tau dy(t, \tau) + \frac{1}{2} \tau^2 \sigma(t, \tau)' \sigma(t, \tau) \, dt + y(t, \tau) \, dt + \tau \frac{\partial y}{\partial \tau}(t, \tau) \, dt
\]

(A.3)

\[
= \left( -\tau \alpha(t, \tau) + \frac{1}{2} \tau^2 \sigma(t, \tau)' \sigma(t, \tau) + y(t, \tau) + \tau \frac{\partial y}{\partial \tau}(t, \tau) \right) dt - \tau \sigma(t, \tau)' dW_t
\]

where $\alpha_p$ and $\sigma_p$ are the expected return and return volatility of the bond. By matching volatility coefficients, we have $\sigma_p(t, \tau) = -\tau \sigma(t, \tau)$. The absence of arbitrage opportunities requires the existence of a market price of risk process $\lambda_t$ such that

\[
\alpha_p(t, \tau) = r_t + \sigma_p(t, \tau)' \lambda_t,
\]

(A.4)

where $r_t = y(t, 0)$ is the short rate at time $t$. Inserting for the bond drift and volatility from Eq. (A.3), we get the yield drift condition under the physical measure,

\[
\alpha(t, \tau) = \frac{1}{\tau} (y(t, \tau) - y(t, 0)) + \frac{\partial y}{\partial \tau}(t, \tau) + \frac{1}{2} \tau^2 \sigma(t, \tau)' \sigma(t, \tau) + \sigma(t, \tau)' \lambda_t,
\]

(A.5)

which is Eq. (2). The average slope (or spread) $(y(t, \tau) - y(t, 0))/\tau$ enters because we inherit an excess return condition, whereas the local slope $\partial y(t, \tau)/\partial \tau$ reflects the change in term to maturity $\tau$ (bond aging).

**Appendix B. The generalized state space model**

We briefly describe the state space approach and the Kalman filter, following Harvey (1989) and Durbin and Koopman (2012), as we apply it to our most general model from Eqs. (19), (20) and (24), restated as

\[
y_t = \alpha + B^o (a^o - c^o) + \text{vec}_{1:m} \left\{ \left[ (b^{o'})' \Sigma^o (b^o) + (b^{u'})' (b^u) \right] \tau_i/2 \right\} + B^o x^o + \varepsilon_t,
\]

(B.1)

with

\[
x^o_t = c^o + C^o x^o_{t-1} + w^o_t
\]

(B.2)

and

\[
x^u_t = a^u + A^u x^u_{t-1} + w^u_t.
\]

(B.3)

The model is in state space form, with measurement equation given by Eqs. (B.1) and (B.2), and state transition equation given by Eq. (B.3).
In the special case of a static model, \( C^o = 0 \) and \( A^u = 0 \), we consider \( c^o \) and \( \Sigma^o \) known (use the sample mean and variance-covariance matrix of the observable state variables \( x_i^o \)), and the model collapses to
\[
\tilde{y}_t = \alpha + B^o (a^o - c^o) + \text{vec}_{1:m} \left\{ \left[ (b_i^o)' \Sigma^o (b_i^o)' + (b_i^o)' (b_i^o) \right] \tau_i / 2 \right\} + B^o [x_i^o + A^o x_{t-1}^o] + B^u a^u + B^u w_t^o + \varepsilon_t,
\]
which is similar to the classical factor analysis model with factors \( w_t^o \) but with an extension to time-varying conditional means, depending on \( x_i^o \) and \( x_{t-1}^o \). If no arbitrage is not imposed, \( \alpha \) is set to match the sample averages of the left- and right-hand sides, and the risk prices \( \lambda = (\lambda^o, \lambda^u)' = (a^o, a^u)' \) are unidentified. A simple but inefficient set of estimates, e.g., to get useful starting values, can be obtained by estimating \( B^o \) and \( B^o A^o \) row by row in time series regression of \( \tilde{y}_{it} \) on current and lagged observed state variables, fixing these parameters at the estimated values, and applying the classical factor analysis to \( \tilde{y}_t = \tilde{y}_t - B^o [x_i^o + A^o x_{t-1}^o] \) to get \( B^u \). Now \( \alpha \) is set, too. In the restricted case \( (\alpha = 0) \), risk prices \( \lambda \) are obtained by running a cross-sectional generalized least squares (GLS) regression of the time series average of \( \tilde{y}_t + B^o c^o - \text{vec}_{1:m} \left\{ \left[ (b_i^o)' \Sigma^o (b_i^o)' + (b_i^o)' (b_i^o) \right] \tau_i / 2 \right\} \) on the columns of \( B = (B^o, B^u) \). The relevant variance-covariance matrix for GLS takes the same form as in the classical factor analysis, \( B^u B^u + \Psi \). This generalizes the two-step procedure of Gultekin and Rogalski (1985) to both unobservable (as in their case) and observable state variables. An estimator that is more efficient under the null can be obtained by iterating the procedure. In each iteration, the initial time series regressions use the restricted intercepts based on the risk prices from the previous iteration. The efficient one-step estimator is obtained by inserting the expressions for the restricted time series and cross-sectional regression estimators of \( B^o, B^o A^o \), and \( \lambda \) into the factor analysis likelihood function and optimizing the resulting concentrated likelihood with respect to \( B^u \) and \( \Psi \).

For the dynamic case with unrestricted state transition \( C^o \) and \( A^u \), we use the Kalman filter. We briefly describe the filter for the case of latent factors only, suppressing the superscript \( u \) for notational convenience. The Kalman filter recursions provide estimates of the unobserved state \( x_t \) given the data, with a distinction between conditioning on two different information sets. In each recursive step, the estimate of the unobserved state at time \( t \) is first made based on the information up to the previous time point, that is, \( \tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_{t-1} \). We denote this estimate as \( x_{it-1} = E[x_t | \tilde{y}_1, \ldots, \tilde{y}_{t-1}] = E_{t-1}[x_t] \) and refer to it as the predicted state. The uncertainty around this estimate is measured by the covariance matrix \( P_{it-1} := \text{var}_{t-1}[x_t] \). When the new observation \( \tilde{y}_t \) becomes available, we can update our information set to \( \tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_{t-1}, \tilde{y}_t \) and obtain an updated estimate of the unobserved state. We denote this updated estimate as \( x_{it} = E[x_t | \tilde{y}_1, \ldots, \tilde{y}_t] = E_t[x_t] \) and refer to it as...
the filtered state. The uncertainty around this estimate is measured by $P_{t|t} := \text{var}_t[x_t]$.

Due to the properties of the Gaussian $\varepsilon_t$ and $w_t$, the predicted and filtered estimates of the unobserved state, and the uncertainty around these, can be calculated in a simple recursive manner. Given that for time $t - 1$ a filtered estimate $x_{t-1|t-1}$ (and error variance $P_{t-1|t-1}$) of the state has been obtained, we get a predicted estimate and error variance for the state at the next time $t$ through

$$x_{t|t-1} = a + A x_{t-1|t-1}$$  \hspace{1cm} (B.5)

and

$$P_{t|t-1} = A P_{t-1|t-1} A' + \text{var}(w_t) = A P_{t-1|t-1} A' + I_d,$$  \hspace{1cm} (B.6)

with $I_d$ the $d$-dimensional identity matrix. After the new observation $\tilde{y}_t$ comes in, we can update these estimates and obtain the filtered estimate and error variance for the state using

$$x_{t|t} = x_{t|t-1} + P_{t|t-1} B' F_t^{-1} v_t$$  \hspace{1cm} (B.7)

and

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} B' F_t^{-1} B P_{t|t-1},$$  \hspace{1cm} (B.8)

where we define the prediction error (or innovation) and its variance as

$$v_t = \tilde{y}_t - \tilde{y}_{t|t-1} = \tilde{y}_t - (\alpha + \text{vec}_{1:m}\{b_i' \tau_i / 2\} + B x_{t|t-1})$$  \hspace{1cm} (B.9)

and

$$F_t = \text{var}_{t-1}(v_t) = BP_{t|t-1} B' + \text{var}(\varepsilon_t) = BP_{t|t-1} B' + \Psi.$$  \hspace{1cm} (B.10)

To initialize the Kalman filter, we use the properties of the VAR process for the latent state $x_t$ from Eq. (B.3) to get the unconditional distribution (see also the discussion in Subsection 3.2). The unconditional mean satisfies $\mu = a + A \mu$, i.e., we use $\mu = (I - A)^{-1} a$ for the initialization of the unconditional mean [see Eq. (15)]. We can write the dynamics of the state as

$$x_t = (I - A)(I - A)^{-1} a + A x_{t-1} + w_t \iff x_t = (I - A)\mu + A x_{t-1} + w_t \iff x_t - \mu = A(x_{t-1} - \mu) + w_t,$$  \hspace{1cm} (B.11)

so for the unconditional variance of $x_t$, denoted $\Sigma$, we get

$$\Sigma = A \Sigma A' + I_d,$$  \hspace{1cm} (B.12)
which we can solve using the properties of the vectorization operator $vec$ to give

$$vec(\Sigma) = [I_d^2 - (A \otimes A)]^{-1}vec(I_d^2),$$

(B.13)

with $\otimes$ the Kronecker product. Thus, the filter is initialized by setting $x_{0|0} = \mu$ and $P_{0|0} = \Sigma$.

The vectors $\alpha$ and $\alpha$ and the matrices $B$, $\Psi$ and $A$ contain the parameters that need to be estimated. Given normality of the $\varepsilon_t$ and $w_t$, we can use the output from the above Kalman recursions. In general, the logarithm of the likelihood for a model with conditional probability density function $p(\tilde{y}_t|\tilde{y}_1, \ldots, \tilde{y}_{t-1})$ is given by

$$\log L(\tilde{y}_1, \ldots, \tilde{y}_T) = \sum_{t=1}^{T} \log p(\tilde{y}_t|\tilde{y}_1, \ldots, \tilde{y}_{t-1}),$$

(B.14)

where $p(\tilde{y}_1|\tilde{y}_0)$ is according to the initialization of the system (see above) and $T$ is the number of observations. Due to normality of all distributions and the additivity in the system, the latent state $x_t$ conditional on $\tilde{y}_1, \ldots, \tilde{y}_{t-1}$ is normal with mean $x_{t|t-1}$ and variance $P_{t|t-1}$. For $\tilde{y}_t$, we can write

$$\tilde{y}_t = \alpha + \text{vec}_{1:m}\{b'_i \tau_i / 2 \} + B(x_t - x_{t|t-1}) + Bx_{t|t-1} + \varepsilon_t,$$

(B.15)

from which we get that the distribution of $\tilde{y}_t$ conditional on $\tilde{y}_1, \ldots, \tilde{y}_{t-1}$ is normal with mean $\alpha + \text{vec}_{1:m}\{b'_i \tau_i / 2 \} + Bx_{t|t-1}$ and variance matrix $F_t$. Thus, with $v_t$ from Eq. (B.9), we can write the prediction error decomposition of the log-likelihood for the state space model as

$$\log L(\tilde{y}_1, \ldots, \tilde{y}_T) = -Tm/2 - \sum_{t=1}^{T} \log |F_t|/2 - \sum_{t=1}^{T} v_t'F_t^{-1}v_t/2.$$  

(B.16)

This is maximized numerically, rerunning the filter at each trial value of the parameter vector. Standard errors are estimated off the square-roots of the diagonal elements of the negative inverse Hessian at the optimum.

With both observable and unobservable state variables, the generalized Kalman filter recursions are similar, but using Eq. (B.1) for $\tilde{y}_t$ in the definition of $v_t$ in (B.9), again using $x^n_{t|t-1}$ in place of $x^n_t$, and the likelihood includes that of the VAR model for $x^n_t$, Eq. (B.2).

References

Andrews, D., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62 (6), 1383–1414.


44


Table 1
Summary statistics of yields and slope-adjusted yield changes.
The table reports summary statistics for the yields and slope-adjusted yield changes from the unsmoothed Fama-Bliss data set for 17 maturities over the period January 1985 through December 2016. The table shows term to maturity in years ($\tau_i$), mean, standard deviation ("Sd"), minimum ("Min"), maximum ("Max"), and three autocorrelation coefficients: one month [$\hat{\rho}(1)$], one year [$\hat{\rho}(12)$], and 30 months [$\hat{\rho}(30)$]. For readability, yields and slope-adjusted yield changes are in percent (for yields, $y_{t,\tau_i} \times 100$; for slope-adjusted yield changes, $\tilde{y}_{t,\tau_i} \times 100$).

<table>
<thead>
<tr>
<th>$\tau_i$</th>
<th>Yield (in percent, $y_{t,\tau_i} \times 100$)</th>
<th>Slope-adjusted yield changes (in percent, $\tilde{y}_{t,\tau_i} \times 100$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>Mean 3.502, Sd 2.649, Min -0.020, Max 9.133, $\hat{\rho}(1)$ 0.990, $\hat{\rho}(12)$ 0.811, $\hat{\rho}(30)$ 0.491</td>
<td>Mean -1.051, Sd 1.175, Min -6.523, Max 2.274, $\hat{\rho}(1)$ 0.715, $\hat{\rho}(12)$ 0.045, $\hat{\rho}(30)$ 0.015</td>
</tr>
<tr>
<td>0.5</td>
<td>3.631, 2.672, 0.030, 9.325, 0.990, 0.813, 0.505</td>
<td>-0.805, 0.918, -4.542, 1.481, 0.723, 0.059, 0.078</td>
</tr>
<tr>
<td>0.75</td>
<td>3.719, 2.696, 0.045, 9.343, 0.990, 0.817, 0.527</td>
<td>-0.883, 0.777, -4.177, 0.829, 0.640, 0.162, 0.091</td>
</tr>
<tr>
<td>1</td>
<td>3.827, 2.720, 0.096, 9.561, 0.989, 0.820, 0.541</td>
<td>-0.880, 0.756, -3.665, 1.161, 0.679, 0.294, 0.262</td>
</tr>
<tr>
<td>1.25</td>
<td>3.933, 2.758, 0.121, 9.956, 0.989, 0.823, 0.561</td>
<td>0.712, 0.639, -2.964, 1.360, 0.697, 0.214, 0.071</td>
</tr>
<tr>
<td>1.5</td>
<td>4.005, 2.759, 0.161, 10.150, 0.989, 0.824, 0.579</td>
<td>-0.665, 0.597, -2.683, 1.015, 0.680, 0.213, -0.036</td>
</tr>
<tr>
<td>1.75</td>
<td>4.071, 2.748, 0.188, 10.231, 0.988, 0.825, 0.594</td>
<td>-0.560, 0.601, -2.391, 1.696, 0.647, 0.166, -0.113</td>
</tr>
<tr>
<td>2</td>
<td>4.118, 2.728, 0.205, 10.371, 0.988, 0.824, 0.608</td>
<td>-0.616, 0.525, -2.549, 0.751, 0.609, 0.258, -0.089</td>
</tr>
<tr>
<td>2.5</td>
<td>4.249, 2.718, 0.251, 10.660, 0.987, 0.825, 0.629</td>
<td>-0.569, 0.491, -1.984, 0.905, 0.588, 0.267, -0.191</td>
</tr>
<tr>
<td>3</td>
<td>4.365, 2.678, 0.289, 10.717, 0.987, 0.826, 0.646</td>
<td>-0.543, 0.477, -2.114, 0.867, 0.513, 0.269, -0.137</td>
</tr>
<tr>
<td>4</td>
<td>4.595, 2.615, 0.423, 11.127, 0.985, 0.821, 0.670</td>
<td>-0.449, 0.431, -2.002, 0.799, 0.496, 0.276, -0.125</td>
</tr>
<tr>
<td>5</td>
<td>4.758, 2.529, 0.589, 11.196, 0.985, 0.817, 0.685</td>
<td>-0.434, 0.402, -1.836, 0.577, 0.422, 0.232, -0.097</td>
</tr>
<tr>
<td>6</td>
<td>4.923, 2.482, 0.778, 11.450, 0.984, 0.810, 0.693</td>
<td>-0.382, 0.383, -1.821, 0.694, 0.415, 0.228, -0.066</td>
</tr>
<tr>
<td>7</td>
<td>5.053, 2.404, 0.962, 11.570, 0.983, 0.806, 0.700</td>
<td>-0.334, 0.356, -1.931, 0.591, 0.300, 0.152, -0.083</td>
</tr>
<tr>
<td>8</td>
<td>5.152, 2.372, 1.088, 11.425, 0.983, 0.806, 0.706</td>
<td>-0.306, 0.343, -1.606, 0.691, 0.329, 0.162, -0.038</td>
</tr>
<tr>
<td>9</td>
<td>5.238, 2.329, 1.277, 11.562, 0.983, 0.799, 0.704</td>
<td>-0.239, 0.350, -2.036, 0.673, 0.282, 0.174, 0.014</td>
</tr>
<tr>
<td>10</td>
<td>5.274, 2.270, 1.396, 11.531, 0.982, 0.794, 0.705</td>
<td>-0.209, 0.355, -2.096, 0.663, 0.272, 0.170, 0.006</td>
</tr>
</tbody>
</table>
Table 2
Estimates of time-varying risk premiums.

This table reports the estimated risk prices. For each number of covariance-generating factors $d = 1, 2, 3, 4$, four models are estimated: the model with a constant market price of risk $\lambda$ (“Constant $\lambda$”) and with time-varying $\lambda_t$ (“TV $\lambda$”), and for each of these variations an unrestricted version (“Unrestr”) and a model imposing the no-arbitrage drift restriction (“Restr”). In the case with constant risk premiums, the estimate $\lambda = a$ is reported; for time-varying risk premiums, the estimated mean of $\lambda_t$ [given by $\mu = (I - A)^{-1}a$]. Asymptotic $t$-statistics are given below the estimates in parentheses, and * and ** indicate significance at the 5% and 1% level, respectively. In addition, the table reports the log likelihood (“Likelihood”), number of parameters, Akaike information criterion (“AIC”) and Schwarz Bayesian information criterion (“BIC”).

<table>
<thead>
<tr>
<th>$d$</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
<th>Likelihood</th>
<th>No. of Parameters</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Constant $\lambda$</td>
<td>Unrestr</td>
<td>-0.323**</td>
<td></td>
<td>27,365</td>
<td>48</td>
<td>-54,634</td>
<td>-54,312</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Restr</td>
<td></td>
<td></td>
<td>27,236</td>
<td>33</td>
<td>-54,405</td>
<td>-54,184</td>
</tr>
<tr>
<td></td>
<td>TV $\lambda$</td>
<td>Unrestr</td>
<td>-0.380**</td>
<td></td>
<td>27,411</td>
<td>49</td>
<td>-54,723</td>
<td>-54,394</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Restr</td>
<td></td>
<td></td>
<td>27,279</td>
<td>34</td>
<td>-54,490</td>
<td>-54,262</td>
</tr>
<tr>
<td>2</td>
<td>Constant $\lambda$</td>
<td>Unrestr</td>
<td>-0.320**</td>
<td>-0.208**</td>
<td>29,167</td>
<td>63</td>
<td>-58,209</td>
<td>-57,785</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Restr</td>
<td></td>
<td></td>
<td>29,066</td>
<td>49</td>
<td>-58,033</td>
<td>-57,704</td>
</tr>
<tr>
<td></td>
<td>TV $\lambda$</td>
<td>Unrestr</td>
<td>-0.611**</td>
<td>-0.0654**</td>
<td>29,432</td>
<td>67</td>
<td>-58,730</td>
<td>-58,280</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Restr</td>
<td></td>
<td></td>
<td>29,330</td>
<td>53</td>
<td>-58,553</td>
<td>-58,197</td>
</tr>
<tr>
<td>3</td>
<td>Constant $\lambda$</td>
<td>Unrestr</td>
<td>-0.271**</td>
<td>-0.152**</td>
<td>-0.229**</td>
<td>29,603</td>
<td>77</td>
<td>-59,052</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Restr</td>
<td></td>
<td></td>
<td>29,487</td>
<td>64</td>
<td>-58,846</td>
<td>-58,415</td>
</tr>
<tr>
<td></td>
<td>TV $\lambda$</td>
<td>Unrestr</td>
<td>-0.499**</td>
<td>-0.333**</td>
<td>-0.15**</td>
<td>30,060</td>
<td>86</td>
<td>-59,948</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Restr</td>
<td></td>
<td></td>
<td>29,951</td>
<td>73</td>
<td>-59,756</td>
<td>-59,266</td>
</tr>
<tr>
<td>4</td>
<td>Constant $\lambda$</td>
<td>Unrestr</td>
<td>-0.273**</td>
<td>-0.0882**</td>
<td>-0.269**</td>
<td>29,782</td>
<td>90</td>
<td>-59,385</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Restr</td>
<td></td>
<td></td>
<td>29,676</td>
<td>78</td>
<td>-59,197</td>
<td>-58,673</td>
</tr>
<tr>
<td></td>
<td>TV $\lambda$</td>
<td>Unrestr</td>
<td>-0.486**</td>
<td>-0.282**</td>
<td>-0.543**</td>
<td>30,405</td>
<td>106</td>
<td>-60,598</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Restr</td>
<td></td>
<td></td>
<td>30,310</td>
<td>94</td>
<td>-60,432</td>
<td>-59,800</td>
</tr>
</tbody>
</table>
Table 3
Estimated risk price dynamics.
This table reports the estimated risk price dynamics. For each number of covariance-generating factors $d = 1, 2, 3, 4$, two models are estimated: an unrestricted version and a model imposing the no-arbitrage drift restriction. The table provides the estimated $4 \times 4$ matrix for the state transitions. Asymptotic $t$-statistics are given below the estimates in parentheses, and * and ** denote significance at the 5% and 1% level, respectively. VAR = vector autoregression.

<table>
<thead>
<tr>
<th>No. of Risk Factors</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: VAR estimates $d = 1$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unrestricted</td>
<td>0.467**</td>
<td>0.131**</td>
<td>-0.111**</td>
<td>0.911**</td>
</tr>
<tr>
<td></td>
<td>(33.9)</td>
<td>(10.0)</td>
<td>(-7.30)</td>
<td>(189)</td>
</tr>
<tr>
<td>Restricted</td>
<td>0.509**</td>
<td>0.030**</td>
<td>-0.113**</td>
<td>0.911**</td>
</tr>
<tr>
<td></td>
<td>(38.6)</td>
<td>(-3.30)</td>
<td>(32.1)</td>
<td>(176)</td>
</tr>
<tr>
<td><strong>Panel B: VAR estimates $d = 2$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unrestricted</td>
<td>0.769**</td>
<td>0.893**</td>
<td>-0.435**</td>
<td>0.911**</td>
</tr>
<tr>
<td></td>
<td>(13.9)</td>
<td>(16.6)</td>
<td>(-20.1)</td>
<td>(128)</td>
</tr>
<tr>
<td>Factor 2</td>
<td>-0.0429</td>
<td>0.475**</td>
<td>0.0766**</td>
<td>0.911**</td>
</tr>
<tr>
<td></td>
<td>(-0.690)</td>
<td>(16.8)</td>
<td>(6.92)</td>
<td>(128)</td>
</tr>
<tr>
<td>Restricted</td>
<td>0.703**</td>
<td>-0.113**</td>
<td>0.603**</td>
<td>0.911**</td>
</tr>
<tr>
<td></td>
<td>(13.8)</td>
<td>(-3.30)</td>
<td>(32.1)</td>
<td>(176)</td>
</tr>
<tr>
<td>Factor 2</td>
<td>-0.0738**</td>
<td>0.541**</td>
<td>-0.435**</td>
<td>0.911**</td>
</tr>
<tr>
<td></td>
<td>(-15.8)</td>
<td>(176)</td>
<td>(-20.1)</td>
<td>(128)</td>
</tr>
<tr>
<td><strong>Panel C: VAR estimates $d = 3$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unrestricted</td>
<td>0.518**</td>
<td>0.320**</td>
<td>-0.435**</td>
<td>0.911**</td>
</tr>
<tr>
<td></td>
<td>(34.8)</td>
<td>(7.84)</td>
<td>(-20.1)</td>
<td>(128)</td>
</tr>
<tr>
<td>Factor 2</td>
<td>-0.0903</td>
<td>0.893**</td>
<td>0.0766**</td>
<td>0.911**</td>
</tr>
<tr>
<td></td>
<td>(-1.20)</td>
<td>(16.6)</td>
<td>(6.92)</td>
<td>(128)</td>
</tr>
<tr>
<td>Factor 3</td>
<td>0.169**</td>
<td>-0.113**</td>
<td>0.603**</td>
<td>0.911**</td>
</tr>
<tr>
<td></td>
<td>(3.60)</td>
<td>(-3.30)</td>
<td>(32.1)</td>
<td>(176)</td>
</tr>
<tr>
<td>Restricted</td>
<td>0.455**</td>
<td>0.320**</td>
<td>-0.435**</td>
<td>0.911**</td>
</tr>
<tr>
<td></td>
<td>(14.2)</td>
<td>(7.84)</td>
<td>(-20.1)</td>
<td>(128)</td>
</tr>
<tr>
<td>Factor 2</td>
<td>-0.0240</td>
<td>0.842**</td>
<td>0.0153</td>
<td>0.911**</td>
</tr>
<tr>
<td></td>
<td>(-0.299)</td>
<td>(7.89)</td>
<td>(0.302)</td>
<td>(176)</td>
</tr>
<tr>
<td>Factor 3</td>
<td>-0.185**</td>
<td>0.204**</td>
<td>0.714**</td>
<td>0.911**</td>
</tr>
<tr>
<td></td>
<td>(-3.15)</td>
<td>(4.16)</td>
<td>(41.8)</td>
<td>(176)</td>
</tr>
<tr>
<td><strong>Panel C: VAR estimates $d = 4$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unrestricted</td>
<td>0.589**</td>
<td>-0.0133</td>
<td>-0.388**</td>
<td>-0.385**</td>
</tr>
<tr>
<td></td>
<td>(36.9)</td>
<td>(-0.530)</td>
<td>(-16.0)</td>
<td>(-30.4)</td>
</tr>
<tr>
<td>Factor 2</td>
<td>-0.105**</td>
<td>0.881**</td>
<td>-0.0928**</td>
<td>-0.114**</td>
</tr>
<tr>
<td></td>
<td>(-5.07)</td>
<td>(95.4)</td>
<td>(-4.80)</td>
<td>(-8.62)</td>
</tr>
<tr>
<td>Factor 3</td>
<td>-0.144**</td>
<td>0.0484</td>
<td>0.665**</td>
<td>-0.261**</td>
</tr>
<tr>
<td></td>
<td>(-7.95)</td>
<td>(1.79)</td>
<td>(24.9)</td>
<td>(-14.1)</td>
</tr>
<tr>
<td>Factor 4</td>
<td>-0.0895**</td>
<td>0.0605*</td>
<td>-0.255**</td>
<td>0.743**</td>
</tr>
<tr>
<td></td>
<td>(-3.56)</td>
<td>(2.26)</td>
<td>(-9.41)</td>
<td>(48.3)</td>
</tr>
<tr>
<td>Restricted</td>
<td>0.474**</td>
<td>0.361**</td>
<td>-0.403**</td>
<td>-0.232**</td>
</tr>
<tr>
<td></td>
<td>(57.3)</td>
<td>(46.8)</td>
<td>(-36.7)</td>
<td>(-54.5)</td>
</tr>
<tr>
<td>Factor 2</td>
<td>0.00203</td>
<td>0.800**</td>
<td>0.0437**</td>
<td>0.0157**</td>
</tr>
<tr>
<td></td>
<td>(0.167)</td>
<td>(186)</td>
<td>(3.66)</td>
<td>(3.03)</td>
</tr>
<tr>
<td>Factor 3</td>
<td>-0.191**</td>
<td>0.223**</td>
<td>0.711**</td>
<td>-0.140**</td>
</tr>
<tr>
<td></td>
<td>(-19.3)</td>
<td>(47.1)</td>
<td>(47.2)</td>
<td>(-26.0)</td>
</tr>
<tr>
<td>Factor 4</td>
<td>-0.0701**</td>
<td>0.131**</td>
<td>-0.111**</td>
<td>0.911**</td>
</tr>
<tr>
<td></td>
<td>(-5.44)</td>
<td>(10.9)</td>
<td>(-7.30)</td>
<td>(189)</td>
</tr>
</tbody>
</table>
Table 4
Testing \( a \) and \( A \).

This table reports likelihood ratio test statistics for restrictions on the pricing matrix. The tests are on the rows (first column of table) and columns (second column of table) for each of the four factors in the fully latent \( d = 4 \) model. Tests on rows are joint on the \( j^{th} \) row of the \( A \) matrix and the \( j^{th} \) element of the \( a \) vector.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Row</th>
<th>Column</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>578.5</td>
<td>48.7</td>
</tr>
<tr>
<td>2</td>
<td>125.8</td>
<td>154.3</td>
</tr>
<tr>
<td>3</td>
<td>111.4</td>
<td>149.0</td>
</tr>
<tr>
<td>4</td>
<td>108.8</td>
<td>188.0</td>
</tr>
</tbody>
</table>
Table 5
Likelihood ratio tests of no-arbitrage restriction and time-varying risk premiums.

This table reports results of tests of the no-arbitrage drift restriction and time-varying (“TV”) risk prices. For each number of covariance-generating factors \( d = 1, 2, 3, 4 \), four models are estimated: models with constant and time-varying \( \lambda_t \), and for each of these variations an unrestricted version and a model with the no-arbitrage drift restriction imposed. The restricted versus unrestricted results provide LR tests of the no-arbitrage drift restriction, for both constant and time-varying risk prices. The constant versus time-varying results provide LR tests of the restriction that \( \lambda \) is constant instead of time-varying, with and without the no-arbitrage drift restriction imposed. The \( p \)-value of rejecting the null hypothesis of the test is reported below the test statistic in parenthesis, and * and ** denote significance at the 5% and 1% level, respectively.

<table>
<thead>
<tr>
<th>Number of risk factors</th>
<th>( d = 1 )</th>
<th>( d = 2 )</th>
<th>( d = 3 )</th>
<th>( d = 4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Restricted versus unrestricted</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant ( \lambda )</td>
<td>259**</td>
<td>204**</td>
<td>232**</td>
<td>212**</td>
</tr>
<tr>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
</tr>
<tr>
<td>TV ( \lambda )</td>
<td>263**</td>
<td>205**</td>
<td>218**</td>
<td>190**</td>
</tr>
<tr>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
</tr>
<tr>
<td><strong>Constant versus time-varying</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unrestricted</td>
<td>90.8**</td>
<td>530**</td>
<td>914**</td>
<td>1,246**</td>
</tr>
<tr>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
</tr>
<tr>
<td>Restricted</td>
<td>86.6**</td>
<td>528**</td>
<td>929**</td>
<td>1,267**</td>
</tr>
<tr>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.00)</td>
</tr>
</tbody>
</table>
Table 6
Model performance overview, including macro factors.
This table reports the model performance of the time-varying risk premium models including macro factors. For each number of covariance-generating factors \( d = 1, 2, 3, 4 \), models with varying number of macro factors \( d^o = 0, 1, \ldots, d \) are estimated. The macro factors are the second principal component from a large set of macroeconomic variables (PC2), industrial production (IP), the fourth principal component from the set of macro variables, and nonfarm payroll employment (of which for the \( d^o = 1 \) case only PC2 is included; in the \( d^o = 2 \) case, PC2 and IP; etc.). The table focuses on the case with time-varying risk premiums, and considers both an unrestricted version (“Unr”) and a model imposing the no-arbitrage drift restriction (“Restr”). The table reports the log likelihood (“Likelihood”) and the number of parameters for each of the models and provides the Akaike and the Schwarz Bayesian information criteria (“AIC” and “BIC”), with the best value for each model highlighted in boldface. As the loadings on the second and third macro factors are similar we restrict \( a_3 = 0 \) and \( A_{31} = A_{32} = 0 \) in the \( d^o = 3 \) case.

<table>
<thead>
<tr>
<th>Number of macro factors</th>
<th>( d = 1 )</th>
<th>( d = 2 )</th>
<th>( d = 3 )</th>
<th>( d = 4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d^o = 0 )</td>
<td>Unr</td>
<td>Restr</td>
<td>Unr</td>
<td>Restr</td>
</tr>
<tr>
<td>Likelihood</td>
<td>27,411</td>
<td>27,279</td>
<td>29,432</td>
<td>29,330</td>
</tr>
<tr>
<td>No. of Pars</td>
<td>49</td>
<td>34</td>
<td>67</td>
<td>53</td>
</tr>
<tr>
<td>AIC</td>
<td>-54,723</td>
<td>-54,490</td>
<td>-58,730</td>
<td>-58,563</td>
</tr>
<tr>
<td>BIC</td>
<td>-54,394</td>
<td>-54,262</td>
<td>-58,280</td>
<td>-58,197</td>
</tr>
<tr>
<td>( d^o = 1 )</td>
<td>Unr</td>
<td>Restr</td>
<td>Unr</td>
<td>Restr</td>
</tr>
<tr>
<td>Likelihood</td>
<td>23,327</td>
<td>23,281</td>
<td>28,233</td>
<td>28,175</td>
</tr>
<tr>
<td>No. of Pars</td>
<td>52</td>
<td>37</td>
<td>69</td>
<td>55</td>
</tr>
<tr>
<td>( d^o = 2 )</td>
<td>Unr</td>
<td>Restr</td>
<td>Unr</td>
<td>Restr</td>
</tr>
<tr>
<td>Likelihood</td>
<td>23,193</td>
<td>23,191</td>
<td>29,077</td>
<td>29,060</td>
</tr>
<tr>
<td>No. of Pars</td>
<td>76</td>
<td>62</td>
<td>93</td>
<td>80</td>
</tr>
<tr>
<td>AIC</td>
<td>-46,234</td>
<td>-46,258</td>
<td>-57,968</td>
<td>-57,959</td>
</tr>
<tr>
<td>BIC</td>
<td>-45,714</td>
<td>-45,834</td>
<td>-57,322</td>
<td>-57,404</td>
</tr>
<tr>
<td>( d^o = 3 )</td>
<td>Unr</td>
<td>Restr</td>
<td>Unr</td>
<td>Restr</td>
</tr>
<tr>
<td>Likelihood</td>
<td>23,405</td>
<td>23,404</td>
<td>30,310</td>
<td>30,295</td>
</tr>
<tr>
<td>No. of Pars</td>
<td>102</td>
<td>88</td>
<td>119</td>
<td>106</td>
</tr>
<tr>
<td>AIC</td>
<td>-46,606</td>
<td>-46,631</td>
<td>-60,381</td>
<td>-60,377</td>
</tr>
<tr>
<td>BIC</td>
<td>-45,903</td>
<td>-46,025</td>
<td>-59,544</td>
<td>-59,631</td>
</tr>
<tr>
<td>( d^o = 4 )</td>
<td>Unr</td>
<td>Restr</td>
<td>Unr</td>
<td>Restr</td>
</tr>
<tr>
<td>Likelihood</td>
<td>23,506</td>
<td>23,504</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of Pars</td>
<td>133</td>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>-46,747</td>
<td>-46,769</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>-45,823</td>
<td>-45,935</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7
Likelihood ratio tests of no-arbitrage restriction, including macro factors.

This table reports results of tests of the no-arbitrage drift restriction when including macro factors. For each number of covariance-generating factors \( d = 1, 2, 3, 4 \), models with varying number of macro factors \( d^o = 0, 1, \ldots, d \) are estimated. The macro factors are the second principal component from a large set of macroeconomic variables (PC2), industrial production (IP), the fourth principal component from the set of macro variables, and nonfarm payroll employment (of which for the \( d^o = 1 \) case only PC2 is included; in the \( d^o = 2 \) case, PC2 and IP, etc.). The table focuses on the case with time-varying risk premiums and considers both an unrestricted version and a model imposing the no-arbitrage drift restriction. The table reports LR tests of the no-arbitrage drift restriction. The \( p \)-value of rejecting the null hypothesis of the test is reported below the test statistic in parenthesis, and * and ** denote significance at the 5% and 1% level, respectively.

<table>
<thead>
<tr>
<th>Number of</th>
<th>Number of risk factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of</td>
<td>( d = 1 )</td>
</tr>
<tr>
<td>macro factors</td>
<td>( d^o = 0 )</td>
</tr>
<tr>
<td>( d^o = 0 )</td>
<td></td>
</tr>
<tr>
<td>( d^o = 1 )</td>
<td></td>
</tr>
<tr>
<td>( d^o = 2 )</td>
<td></td>
</tr>
<tr>
<td>( d^o = 3 )</td>
<td></td>
</tr>
<tr>
<td>( d^o = 4 )</td>
<td></td>
</tr>
</tbody>
</table>
Table 8
Testing $a^u$ and $a^o$ vectors and $A^u$ and $A^o$ matrices.

This table reports likelihood ratio test statistics for restrictions on the rows and columns of the pricing and dynamics matrices of the $d = 4$ factor model with $d^o = 2$ macro factors in which risk premiums are time-varying and the no-arbitrage restriction imposed. The tests are for zero restrictions on $a^u$ and $A^u$, on $a^o$ and $A^o$, and for entries in $a^o$ and $A^o$ constrained equal to those in $c^o$ and $C^o$. Tests on rows are joint on the $j^{th}$ row of the $A$ ($C$) matrix and the $j^{th}$ element of the $a^u$($c$) vector.

<table>
<thead>
<tr>
<th>Factor</th>
<th>$a^u, A^u$ elements at zero</th>
<th>$a^o, A^o$ elements at zero</th>
<th>$a^o, A^o$ restricted to $c^o, C^o$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Row</td>
<td>Column</td>
<td>Row</td>
</tr>
<tr>
<td>Factor 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>216.8</td>
<td>59.9</td>
<td>42.8</td>
</tr>
<tr>
<td>Factor 2</td>
<td>35.4</td>
<td>89.0</td>
<td>41.7</td>
</tr>
</tbody>
</table>

53
Table 9
Sensitivity analysis of likelihood ratio tests of no-arbitrage restriction, including macro factors.

This table provides two sensitivity analyses of the LR tests of the no-arbitrage drift restriction when including macro factors. Panel A reports the tests for the case in which the nonlinear term is omitted and raw yield changes are used instead of slope-adjusted yield changes. Panel B reports the tests when restricting $B$ to be of the Nelson-Siegel functional form. In all cases, models are estimated with varying number of covariance-generating factors $d = 1, 2, 3, 4$ and varying number of macro factors $d_o = 0, 1, \ldots, d$. In the case with restricted $B$, the ordering is such that the first factors are macro factors (if any are included) and the last factors (if any) are the unobserved factors. The table focuses on the case with time-varying risk premiums and provides LR tests of the no-arbitrage drift restriction. The $p$-value of rejecting the null hypothesis of the test is reported below the test statistic in parenthesis, and * and ** denote significance at the 5% and 1% level, respectively. APT = arbitrage pricing theory.

<table>
<thead>
<tr>
<th>Number of macro factors</th>
<th>Number of risk factors</th>
<th>$d = 1$</th>
<th>$d = 2$</th>
<th>$d = 3$</th>
<th>$d = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: LR tests APT on yield changes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d^o = 0$</td>
<td></td>
<td>1.96 (1.00)</td>
<td>3.06 (0.00)</td>
<td>0.939 (1.00)</td>
<td>0.817 (1.00)</td>
</tr>
<tr>
<td>$d^o = 1$</td>
<td></td>
<td>1.75 (1.00)</td>
<td>1.94 (1.00)</td>
<td>1.99 (1.00)</td>
<td>0.893 (1.00)</td>
</tr>
<tr>
<td>$d^o = 2$</td>
<td></td>
<td>0.176 (&lt;0.00)</td>
<td>0.488 (&lt;0.00)</td>
<td>0.542 (&lt;0.00)</td>
<td>0.542 (&lt;0.00)</td>
</tr>
<tr>
<td>$d^o = 3$</td>
<td></td>
<td>0.0536 (&lt;0.00)</td>
<td>0.556 (&lt;0.00)</td>
<td>1.21 (&lt;0.00)</td>
<td></td>
</tr>
<tr>
<td>$d^o = 4$</td>
<td></td>
<td>1.21 (&lt;0.00)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel B: LR tests restricted $B$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d^o = 0$</td>
<td></td>
<td>586** (1.00)</td>
<td>444** (1.00)</td>
<td>486** (1.00)</td>
<td>242** (1.00)</td>
</tr>
<tr>
<td>$d^o = 1$</td>
<td></td>
<td>4379** (1.00)</td>
<td>408** (1.00)</td>
<td>282** (1.00)</td>
<td>420** (1.00)</td>
</tr>
<tr>
<td>$d^o = 2$</td>
<td></td>
<td>53.7** (1.00)</td>
<td>615** (1.00)</td>
<td>313** (1.00)</td>
<td></td>
</tr>
<tr>
<td>$d^o = 3$</td>
<td></td>
<td>55.7** (1.00)</td>
<td>196** (1.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d^o = 4$</td>
<td></td>
<td>72.4** (1.00)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 1. This figure provides a three-dimensional plot of both the yields and slope-adjusted yield changes from the unsmoothed Fama-Bliss data set for 17 maturities over the period January 1985 through December 2016.
Fig. 2. This figure shows estimates of the volatility functions and a time series plot of the estimated covariance-generating factors for the case with time-varying risk premiums and the no-arbitrage restriction imposed. Panel A shows the estimated $B$ matrix for $d = 1, 2, 3, 4$ factors. The time series plots in Panel B show the estimated $w_t$ time series for each of the factors in the four factor case $d = 4$. 
Fig. 3. This figure shows time series plots of the estimated risk prices for the $d = 4$ factor model with time-varying risk premiums and the no-arbitrage restriction imposed, together with 95% confidence boundaries.
Panel A: Loading matrix $B$ times mean risk price $\mu = (I - A)^{-1}a$ for $d = 1, 2, 3, 4$

Panel B: Loading matrix $B$ times risk price $\lambda_t$ for $d = 4$, selected maturities

Fig. 4. This figure shows how the estimated risk prices and loadings interact to form risk premiums for different maturities in the model with time-varying risk premiums and the no-arbitrage restriction imposed. Panel A shows the estimated mean risk prices $\mu = (I - A)^{-1}a$ pre-multiplied by the factor loading matrix $B$ by maturity for each case $d = 1, 2, 3, 4$. Panel B shows time series plots of the estimated risk prices $\lambda_t$ pre-multiplied by the loading matrix $B$ from the $d = 4$ factor model, for four maturities (six months, and one, five and ten years).
Panel A: Unrestricted intercept ($\alpha$) and risk premium ($B\mu$)

Panel B: Idiosyncratic variance ($\Psi$)

Fig. 5. This figure shows estimates of the mean and measurement error variance of the HJM model by maturity. For each number of covariance-generating factors $d = 1, 2, 3, 4$, four models are estimated: models with constant and time-varying (“TV”) risk prices $\lambda_t$, and for each of these variations an unrestricted version (“unrestr”) and a model imposing the no-arbitrage drift restriction (“restr”). Panel A shows intercepts $\alpha$ from the unrestricted models, along with its spanned analogue $B\mu$ from the restricted models. Panel B shows the idiosyncratic variance $\Psi$. 