ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Bauer, Michael D.

Working Paper

Restrictions on Risk Prices in Dynamic Term Structure Models

CESifo Working Paper, No. 5241

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Bauer, Michael D. (2015) : Restrictions on Risk Prices in Dynamic Term Structure Models, CESifo Working Paper, No. 5241, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/108791

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

CESifo Working Papers

www.cesifo.org/wp

Restrictions on Risk Prices in Dynamic Term Structure Models

Michael D. Bauer

CESIFO WORKING PAPER NO. 5241
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
MARCH 2015

An electronic version of the paper may be downloaded
- *from the SSRN website: www.SSRN.com*
- *from the RePEc website: www.RePEc.org*
- *from the CESifo website: www.CESifo-group.org/wp*

ISSN 2364-1428

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 5241*

Restrictions on Risk Prices in Dynamic Term Structure Models

Abstract

Restrictions on the risk-pricing in dynamic term structure models (DTSMs) can unleash the power of no-arbitrage by creating a tighter link between cross-sectional and time-series variation of interest rates. This paper presents a new econometric framework for estimation of affine Gaussian DTSMs under restrictions on risk prices, which addresses the issues of a large model space and of model uncertainty using a Bayesian approach. A simulation study demonstrates the good performance of the proposed method, both for model choice and for inference about the objects of interest. I obtain novel results for the U.S. Treasury yield curve. The data strongly favor tight restrictions on risk pricing: only level risk is priced, and only changes in the slope affect term premia. Incorporating the restrictions into an otherwise standard model substantially alters its conclusions. Interest rate persistence is significantly higher than in a maximally-flexible model, hence expectations of future short rates are more variable, and the role for term premia is somewhat diminished. Hence, restrictions on risk prices help resolve the puzzle of implausibly stable short-rate expectations which has plagued this literature. Restricted models attribute a larger share of the secular decline in long-term interest rates over the last twenty years to the expectations component, consistent with survey evidence on expectations of future interest rates and inflation.

JEL-Code: C520, E430, G120.

Keywords: no-arbitrage, prices of risk, Bayesian model selection, term premium.

Michael D. Bauer
Economic Research Department
Federal Reserve Bank of San Francisco
101 Market Street MS 1130
USA - 94105 San Francisco CA
michael.bauer@sf.frb.org

February 22, 2015
Previous versions of this paper were circulated under the title “Term Premia and the News.” The views in this paper do not necessarily reflect those of others in the Federal Reserve System.

1 Introduction

Policymakers and academic researchers are keenly interested in estimating term premia in long-term interest rates. To properly interpret the information in the yield curve, they need to separate the expectations and term premium components in long-term interest rates—the former reflects the market's projection for future monetary policy, while the latter captures the compensation that investors require for bearing interest rate risk. Dynamic term structure models (DTSMs), which impose absence of arbitrage, are widely used for this purpose, as well as in many other applications. However, the no-arbitrage assumption only becomes powerful if it creates a meaningful link between the cross-sectional variation of interest rates and their time-series variation, that is, if the allowed risk adjustment is restricted. Existing studies typically impose few or no restrictions on risk prices, and therefore effectively make little use of the no-arbitrage assumption. This paper provides an econometric framework for estimating DTSMs under restrictions on risk prices, which unleashes the power of no-arbitrage. The empirical results demonstrate that this substantially changes our conclusions about short-rate expectations and term premia.

For estimation of term premia, the key statistical problem is to accurately infer expectations of future short-term interest rates. Doing so with only time-series information is extremely difficult because of the very high persistence of interest rates.[1] Intuitively, the unconditional mean and the speed of mean reversion are hard to estimate because a very persistent time series does not revert to its mean very often. This leads to large statistical uncertainty, as well as to biased estimates of the speed of mean reversion. The no-arbitrage assumption in DTSMs can alleviate these problems. Because it requires that the cross section of interest rates reflects forecasts of future short rates, allowing for a risk adjustment, the cross-sectional information can potentially help to pin down the unobserved expectations. However, the risk adjustment loosens this connection between cross-sectional and dynamic properties. Absent any restrictions on the risk pricing, i.e., in a maximally-flexible model, the cross section provides *no information at all* for estimating the time-series parameters (Joslin et al., 2011). Notably, the vast majority of existing DTSMs impose only few or no risk-price restrictions, so that effectively only the time-series information is used for inference about expectations and term premia. Consequently, these DTSMs suffer from the same problems as time-series models, and typically imply implausibly fast mean reversion and puzzling term premium behavior (Bauer et al., 2012; Kim and Orphanides, 2012).

The potential of no-arbitrage in term structure models is only unleashed if the risk ad-

[1] Many studies have noted this problem, including Kozicki and Tinsley (2001), Rudebusch (2007), Cochrane and Piazzesi (2008), Duffee and Stanton (2012), and Kim and Orphanides (2012).

justment is restricted. This can be accomplished be imposing restrictions on the parameters that determine the behavior of risk prices. When choosing such restrictions, however, one faces at least two major challenges, which are likely the reasons that the DTSM literature has avoided this route. The first challenge is that the number of possible restrictions is large. Even in a small model with only three risk factors, and focusing only on zero restrictions on risk-price parameters, there are $2^{12} = 4096$ models to choose from. Brute-force estimation of this amount of highly nonlinear models is at the very least inconvenient and in most cases infeasible. The second challenge is that the choice of restrictions entails model uncertainty, which is particularly problematic in the DTSM context. Comparisons of apparently similar and equally plausible models, which differ only little in terms of risk-price restrictions, reveal that these models often have dramatically different economic implications, including expectations and term premia (Kim and Orphanides, 2012; Bauer and Neely, 2014).

This paper introduces a Bayesian econometric framework that overcomes these challenges. The framework relies on Markov chain Monte Carlo (MCMC) methods to estimate affine Gaussian DTSMs with risk price restrictions. Model selection does not require separate estimation of every single possible model specification because the samplers visit only the more plausible models and do not waste time in uninteresting areas of the model space. Model uncertainty is easily dealt with by means of Bayesian Model Averaging (BMA)—estimates of the objects of interest, such as short-rate expectations and term premia, are calculated as averages across specifications, using posterior model probabilities as weights. The key methodological contribution is to develop model-selection samplers for DTSM estimation. I use the insight that this model-selection problem resembles variable selection in multivariate regressions, and I adapt existing variable-selection samplers to the problem at hand. The paper provides sufficient details so that applied researchers can implement this framework on their own for yield-curve analysis.[2]

Bayesian estimation of DTSMs has often been found to be problematic, with most MCMC samplers displaying very slow convergence and requiring a lot of tuning—examples are Ang et al. (2007, 2011); see Chib and Ergashev (2009) for further discussion. The reason is that parameters enter the likelihood in a highly nonlinear fashion, which requires various Metropolis-Hastings (MH) steps that are often inefficient. A separate contribution of this paper is to substantially simplify Bayesian DTSM estimation. I show that risk price parameters—the majority of the model's parameters and those crucial for determining its economic implications—can in fact be sampled using a simple Gibbs step, because their conditional likelihood function corresponds to a restricted VAR. For the remaining parameters which need to be sampled using

[2]The estimation code is available from the author upon request.

MH, I show how to do so very efficiently using tailored MH steps based on Chib and Greenberg (1994) and Chib and Ergashev (2009). The sampling algorithms developed in this paper require little or no tuning, and display excellent convergence properties.[3]

The results of Bayesian model selection are generally sensitive to the priors, so a key question is how to choose the prior dispersion of the risk-price parameters. I construct a prior that is similar to Zellner's g-prior, so that only one hyperparameter, g, controls the prior dispersion.[4] Hence, we can intuitively understand the amount of prior information imposed on the estimation, and easily assess the sensitivity of the model selection results to changes in prior dispersion.

Because of the novelty of the methodology, it is important to verify that it works in simulated data. To this end, I apply the estimation method to data that is simulated from a known model.[5] The econometric framework performs well in recovering the zero restrictions on risk prices and the estimated risk-price parameters. In contrast, model selection based on credibility intervals of the parameters—similar to the common approach in the DTSM literature of choosing risk-price restrictions based on statistical significance—typically leads to the wrong model. Most importantly, the simulation study shows that estimation under risk-price restrictions accurately infers the persistence of interest rates and the volatility of short-rate expectations and term premia. In contrast, estimation of an unrestricted model leads to persistence that is too low, short-rate expectations that are too stable, and term premium estimates that are excessively volatile.

The framework is then applied to monthly U.S. Treasury yields over the period from 1990 to 2007. The model selection exercise reveals that the data call for tight restrictions on the market prices of risk, in stark contrast to most models in this literature. In fact, a model with only one unrestricted risk-price parameter (out of twelve) is strongly favored by the data. This is in some way a fortunate outcome, because it means that the connection between time-series variation and cross-sectional variation is relatively tight, and hence that the information in long rates can be used to pin down short-rate expectations and term premia. The model-selection results can be used to carry out posterior inference about the pricing of risks in Treasury markets. The results suggest that level risk is priced but slope and curvature risks are not.[6] Furthermore, there is strong evidence that level risk is time-varying, and that this

[3] Chib and Ergashev (2009) and Chib and Kang (2014) have also constructed efficient samplers for MCMC samplers for DTSM estimation. They take a different route by imposing strong prior restrictions.

[4] The prior also imposes conditional independence, which is a useful and common assumption in variable selection problems.

[5] My approach is similar to that of George et al. (2008), who also use repeated simulations to assess the accuracy of their Bayesian method for VAR estimation under parameter restrictions.

[6] Cochrane and Piazzesi (2008) also argue that only level risk is priced.

time variation is driven predominantly by changes in the slope of the yield curve. Notably, the only unrestricted parameter in the favored model is the one that captures the sensitivity of level risk to changes in the slope. The finding that variation in slope affects term premia goes back to Fama and Bliss (1987) and Campbell and Shiller (1991). My estimates are the first that provide direct evidence on risk pricing in DTSMs that is consistent with the long tradition of regression evidence on U.S. Treasury yields. While existing DTSMs allow for deviations from the expectations hypothesis, they do so in a much too general way. The data call for a very restricted form of time-varying risk prices.

Incorporating these restrictions into an otherwise standard DTSM substantially alters the economic implications. First and foremost, they change the VAR dynamics of the model, increasing its persistence relative to a maximally-flexible model. The explanation is that interest rates are extremely persistent under the risk-neutral ($\mathbb{Q}$) probability measure—far-ahead forward rates are not constant—and the risk-price restrictions tighten the link with the real-world ($\mathbb{P}$) probability measure, pulling up the persistence. As a consequence of the higher persistence and slower speed of mean reversion of model-implied interest rates, long-horizon expectations of future short-term interest rates are more variable in the restricted models. Estimated term premia, which are the difference between long-term rates and expected future short-term rates, are typically more stable. This finding is particularly noteworthy because the DTSM literature has long been puzzled by artificially stable expectations and overly variable term premia, see Kim and Orphanides (2012)[7] and others. My estimates show that when DTSMs are estimated under the appropriate risk-price restrictions, variability of short-rate expectations becomes much more plausible, and this long-standing puzzle is resolved.

Risk-price restrictions change our interpretation of the evolution of interest rates over certain historical episodes, in particular of the secular decline in long rates over the last two decades. Conventional DTSMs typically explain this by substantial declines in term premia and imply only a small role for short-rate expectations. In contrast, restricted models attribute a more important role to a declining expectations component. The finding that expectations of short-term interest rates have decreased over the 1990's and 2000's is consistent with the sizable declines in survey-based expectations of inflation and policy rates documented in Kozicki and Tinsley (2001), Kim and Orphanides (2012), and others. On the other hand the interpretation of the conundrum period of 2004 and 2005 is not materially changed when viewed through the lens of restricted DTSMs—the common finding that term premia dropped during an episode when monetary policy tightened appears robust.

[7] A working paper version of this paper, which notes the puzzling stability of far-ahead short-rate expectations, was first circulated in 2005.

Model uncertainty is an important but often-ignored factor when DTSMs are used to analyze the yield curve and to estimate term premia. My results show that there remains some uncertainty about which restrictions to impose on the model, and that small changes to risk-price restrictions can substantially alter the economic implications. Calculating the objects of interest using BMA is important, and it reveals that the uncertainty is quite significant. Hence, taking into account model uncertainty is crucial, and my results strongly caution against the common approach of conditioning on one specific DTSM when studying term premia.[8]

A key established fact in the yield curve literature is that excess bond returns are strongly predictable. This has been documented by many studies, and recent important examples include Cochrane and Piazzesi (2005) and Ludvigson and Ng (2009). Since the preferred model specifications in this paper imply tightly restricted risk pricing, the question arises whether these models can reproduce the strong predictability of bond returns. The paper shows that they can. While the model-implied return predictability in population is modest, the predictability in small samples is similar to that in actual interest rate data.

The method for DTSM estimation proposed here constitutes a significant advance over existing approaches in this literature. The vast majority of studies impose ad hoc restrictions on risk pricing by setting those risk-price parameters with large standard errors to zero.[9] This is unappealing because joint restrictions are imposed without considering joint significance, because the chosen significance level is necessarily arbitrary, and because model uncertainty is ignored. Only two studies impose restrictions on risk prices more systematically. In earlier work, Cochrane and Piazzesi (2008) derive the restrictions from a regression analysis of return predictability, and impose that all variation in term premia is driven by their tent-shaped return-forecasting factor. In contrast, I evaluate a large set of restrictions within a coherent statistical framework (including, in principle, the restrictions of Cochrane and Piazzesi), and explicitly address model uncertainty. In independent and parallel work, Joslin et al. (2014) estimate their macro-finance DTSM under all possible zero restrictions on risk-price parameters, and select a specification on the basis of information criteria. My approach has at least three distinct advantages over theirs: First, it does not require brute-force estimation of every possible model. Second, posterior model probabilities—in contrast to information criteria—can tell us in intuitive terms how plausible alternative models are, and allow us to perform inference about risk pricing. Third, instead of conditioning on a specific model, the approach here addresses the issue of model uncertainty, which is paramount in the context of DTSM

[8]Prominent examples of studies that use only one specific restricted DTSM for studying term premia are Cochrane and Piazzesi (2008) and Joslin et al. (2014).

[9]Prominent examples of studies that employ this approach include Dai and Singleton (2002), Duffee (2002), Ang and Piazzesi (2003), Kim and Wright (2005), and many others.

estimation.

The paper is structured as follows: Section 2 describes the econometric framework. Section 3 demonstrates that the method works well in simulated yield data. Section 4 shows the estimation results for U.S. Treasury yield data. Section 5 then turns to the economic implications, including interest rate persistence and volatilities, as well as estimated short-rate expectations and term premia. Section 6 concludes.

2 Econometric framework

This section first lays out the model specification, explains the role of risk-price parameters, and then describes the methodology for estimation and model selection.

2.1 Model specification

The models used in this paper belong to the class of affine Gaussian DTSMs, a mainstay of modern finance.[10] Denote by X_t the $(N \times 1)$ vector of risk factors, which represents the new information that market participants obtain at time t. Assume that X_t follows a first-order Gaussian vector autoregression (VAR) under the physical (real-world) probability measure $\mathbb{P}$,

$$X_t = \mu + \Phi X_{t-1} + \Sigma \varepsilon_t, \tag{1}$$

with $\varepsilon_t \sim N(0, I_N)$, Σ lower triangular, and $E(\varepsilon_r \varepsilon_s{}') = 0$, $r \neq s$. The one-period interest rate r_t is an affine function of the factors,

$$r_t = \delta_0 + \delta_1' X_t. \tag{2}$$

Estimation will be carried out using monthly data, so r_t is the one-month interest rate. Assuming absence of arbitrage, there exists a risk-neutral probability measure, denoted by $\mathbb{Q}$, which prices all financial assets. The stochastic discount factor (SDF), which defines the change of probability measure between $\mathbb{P}$ and $\mathbb{Q}$, is specified as exponentially affine,

$$-\log M_{t+1} = r_t + \frac{1}{2}\lambda_t'\lambda_t + \lambda_t'\varepsilon_{t+1}, \tag{3}$$

[10] Extensions to other model classes, though beyond the scope of this paper, are easily possible. They are discussed in Section 6.

with the $(N \times 1)$ vector λ_t, the *market prices of risk*, being an affine function of the factors,

$$\lambda_t = \Sigma^{-1}(\lambda_0 + \lambda_1 X_t). \tag{4}$$

This is the essentially-affine risk-price specification of Duffee (2002). The risk prices λ_t measure the additional expected return required per unit of risk in each of the shocks in ε_t. Consider how the expected excess return of an n-period bond rate depends on risk prices:

$$E_t(rx_{t+1}^{(n)}) + \frac{1}{2}Var_t(rx_{t+1}^{(n)}) = \lambda_t' Cov(\varepsilon_{t+1}, rx_{t+1}). \tag{5}$$

The risk premium equals the prices of risk times the quantities of risk (the covariances). In a Gaussian model, the covariances are constant, and the only source of time-variation in term premia are changes in the market prices of risk.[11]

Under the above assumptions the risk-neutral dynamics (see Appendix A) are given by

$$X_t = \mu^{\mathbb{Q}} + \Phi^{\mathbb{Q}} X_{t-1} + \Sigma \varepsilon_t^{\mathbb{Q}}, \tag{6}$$

where $\varepsilon_t^{\mathbb{Q}} \overset{\mathbb{Q}}{\sim} N(0, I_k)$, $\mathrm{E}^{\mathbb{Q}}(\varepsilon_r^{\mathbb{Q}} \varepsilon_s^{\mathbb{Q}\prime}) = 0$, $r \neq s$, and the parameters describing the physical and risk-neutral dynamics are related in the following way:

$$\mu^{\mathbb{Q}} = \mu - \lambda_0, \quad \Phi^{\mathbb{Q}} = \Phi - \lambda_1. \tag{7}$$

In this model, yields are affine in the state variables. Denoting the J model-implied (fitted) yields by $\hat{Y}_t$, we have

$$\hat{Y}_t = A + BX_t, \tag{8}$$

where the J-vector A and the $J \times N$-matrix B contain the model-implied loadings of yields on risk factors, given in Appendix B. These are determined by the parameters δ_0, δ_1, $\mu^{\mathbb{Q}}$, $\Phi^{\mathbb{Q}}$, and Σ.

The observed bond yields used for estimation are $Y_t = \hat{Y}_t + e_t$, where e_t is a vector of measurement errors that is *iid* normal with diagonal covariance matrix R. Measurement errors are included because an N-dimensional factor model cannot perfectly price $J > N$ yields. For parsimony I impose $R = \sigma_e^2 I_J$, which implies that the model tries equally hard to fit the rates of the J instruments in a least-squares sense.

The risk factors X_t are taken as the first N principal components (PCs) of the observed yields, as in Joslin et al. (2013), Joslin et al. (2014), and many other recent studies. The

[11] See Appendix E for the details on model-implied excess bond returns and the derivation of equation (5).

assumption that risk factors are observable simplifies the estimation somewhat, but could easily be generalized.[12] The first three PCs of bond yields have loadings corresponding to level, slope, and curvature of the yield curve, which allows for an economic interpretation of the prices of risk. The empirical analysis will focus on the case with three risk factors, since these are sufficient to capture most of the variation in the yield curve (Litterman and Scheinkman, 1991).

Since not all parameters of the DTSM are identified, normalizing restrictions need to be imposed (Dai and Singleton, 2000). A DTSM is "maximally flexible" if it is exactly identified. Here I use the convenient canonical form for affine Gaussian DTSMs introduced by Joslin et al. (2011) (henceforth JSZ). Having observable risk factors implies that N linear combination of yields are priced exactly by the model. The weights are collected in the $N \times J$ matrix W, which in my case contains the loadings of the first N PCs of observed yields. Hence we have $X_t = WY_t = W\hat{Y}_t$.[13] The free parameters are $\theta = (\phi^{\mathbb{Q}}, k_{\infty}^{\mathbb{Q}}, \lambda_0, \lambda_1, \Sigma, \sigma_e^2)$. $\phi^{\mathbb{Q}}$ is an N-vector containing the eigenvalues of $\Phi^{\mathbb{Q}}$, assumed to be real, distinct, and less than unity. $k_{\infty}^{\mathbb{Q}}$ determines the unconditional mean of the short rate under $\mathbb{Q}$.[14] JSZ detail how $\mu^{\mathbb{Q}}$, $\Phi^{\mathbb{Q}}$, δ_0, and δ_1 are calculated from these parameters and W. The parameterization here includes λ_0 and λ_1, instead of μ and Φ as in JSZ, because the focus is on inference about these risk-price parameters.

2.2 The economics of risk-price restrictions

Absence of arbitrage requires the consistency of the time-series dynamics of interest rates with their cross-sectional behavior, allowing for a risk adjustment. The risk-price parameters λ_0 and λ_1 determine this risk adjustment and the behavior of term premia. Under the (weak) EH, λ_1 is zero, and term premia are constant. In a maximally-flexible model, all elements of λ_0 and λ_1 are unrestricted. The point of this paper is that the "truth" likely lies somewhere between these two extremes.

Restrictions on risk prices are desirable because in their absence, only the information in the time series of interest rates is used to estimate the VAR parameters (μ, Φ). The high persistence of interest rates, in combination with the small sample sizes used in practice,

[12] One can deal with latent either by conditioning on the factors as in Ang et al. (2007) or by marginalizing over the factors as in the approach of Chib and Ergashev (2009).

[13] Note that this implies $We_t = 0$ so that there are effectively only $J - N$ independent measurement errors.

[14] JSZ use the parameter $r_{\infty}^{\mathbb{Q}}$, the risk-neutral mean of the short rate. In online supplemental material they introduce an alternative parameterization using $k_{\infty}^{\mathbb{Q}}$, which is also valid when the system is non-stationary under $\mathbb{Q}$. Conveniently, the value of $k_{\infty}^{\mathbb{Q}}$ that maximizes the likelihood function for given values of the other parameters can be found analytically.

makes it extraordinarily difficult to accurately infer their dynamic behavior. The statistical uncertainty around μ and Φ is high (Kim and Orphanides, 2012), and small-sample bias is often severe (Bauer et al., 2012). Furthermore, the relatively flat likelihood surface makes it difficult to numerically find maximum likelihood (ML) estimates due to the existence of local maxima (Ang and Piazzesi, 2003; Hamilton and Wu, 2012; Duffee and Stanton, 2012). Without restrictions on risk pricing in a DTSM, the no-arbitrage assumption is not very restrictive at all. The only difference between a simple factor model for interest rates and a maximally-flexible DTSM is that under no-arbitrage, the yield loadings have to be consistent with some values for $\mu^{\mathbb{Q}}$ and $\Phi^{\mathbb{Q}}$. Hence, they live in a parameter space of smaller dimension than in a factor model. That this is of little practical relevance has been documented by Duffee (2011a) and Joslin et al. (2013).

The no-arbitrage assumption only has real punch under restrictions on the risk pricing. If λ_0 and λ_1 are restricted, then the information in the cross section of yields helps to pin down the time-series dynamics. Importantly, the data contain a lot of cross-sectional information, since we observe the entire yield curve at every point in time. Under tight restrictions on λ_0 and λ_1, as for example in Cochrane and Piazzesi (2008), "we are able to learn a lot about true-measure dynamics from the cross section" (p. 2). In such a setting, the power of the no-arbitrage restrictions is truly unleashed, and helps to overcome the above-mentioned statistical problems.

One can imagine various possible ways to restrict the risk pricing in a DTSM. With few exceptions (notably JSZ) the literature has focused on zero restrictions on λ_0 and λ_1, and these will be the focus of this paper. Many studies, including Dai and Singleton (2002), Ang and Piazzesi (2003), and Kim and Wright (2005), impose such restrictions in an ad hoc fashion, typically based on t-statistics from preliminary estimates that are not reported. This is unsatisfactory because joint restrictions are imposed without considering joint significance, and because the chosen significance level is necessarily arbitrary. Furthermore, slight differences in the restrictions can have major implications for term premia.

Here, I provide a systematic framework to select restrictions. Let γ be a vector of indicator variables, each of which corresponds to an element of $\lambda \equiv (\lambda_0', vec(\lambda_1)')'$. If an element of γ is equal to zero, the corresponding parameter is restricted to zero, and it is unrestricted otherwise. γ can take on 2^{N+N^2} different values—in a three-factor model, there are 4096 candidate specifications. The basic idea is to include γ in the set of parameters to be estimated, as in Bayesian variable selection. In this way, the most plausible restrictions can be found without having to estimate each model separately. I will lay out an algorithm to estimate the model parameters conditional on γ (Section 2.5) and then describe three different model-

selection algorithms to estimate model parameters jointly with γ (Section 2.6).

2.3 Likelihood

The conditional likelihood function of the data is

$$f(Y_t|Y_{t-1};\theta,\gamma) \quad = \quad f(Y_t|X_t;k_\infty^{\mathbb{Q}},\phi^{\mathbb{Q}},\Sigma,\sigma_e^2)\times f(X_t|X_{t-1};k_\infty^{\mathbb{Q}},\phi^{\mathbb{Q}},\lambda,\gamma,\Sigma).$$

The first factor—the "$\mathbb{Q}$-likelihood"—captures the cross-sectional dependence of yields on risk factors and its logarithm is given by

$$\log f(Y_t|X_t;\theta) = const - (J-N)\log(\sigma_e) - .5\|Y_t - A - BX_t\|/\sigma_e^2, \tag{9}$$

where $\|z\|$ denotes the Euclidean norm of vector z, and A and B depend on $(k_\infty^{\mathbb{Q}},\phi^{\mathbb{Q}})$ and Σ. The $\mathbb{Q}$-likelihood corresponds to a simple non-linear least-squares problem. The second factor—the "$\mathbb{P}$-likelihood"—captures the time-series dynamics of the risk factors, and corresponds to the likelihood of a Gaussian VAR:

$$\log f(X_t|X_{t-1};\theta,\gamma) = const - .5\log(|\Sigma\Sigma'|) - .5\|\Sigma^{-1}(X_t - \mu - \Phi X_{t-1})\|, \tag{10}$$

where μ and Φ depend on λ and γ, as well as on $(k_\infty^{\mathbb{Q}},\phi^{\mathbb{Q}},\Sigma)$, according to equation (7). The joint log-likelihood of the data $Y = (Y_1, Y_2, \ldots, Y_T)$ is given by

$$\log f(Y|\theta,\gamma) = \sum_{t=1}^{T}\log f(Y_t|X_t;\theta)\log f(X_t|X_{t-1};\theta,\gamma). \tag{11}$$

Due to the chosen parameterization, the risk prices λ affect only the $\mathbb{P}$-likelihood, through μ and Φ, which are determined by the linear restrictions in equation (7). Therefore, estimation of λ (for given $\mathbb{Q}$-parameters and restrictions γ) corresponds to estimation of a Restricted VAR (RVAR). Details on this are given in Appendix C.1. This fact substantially simplifies both ML and Bayesian estimation.

ML estimation of the model is further simplified by the fact that several parameters can be concentrated out of the likelihood function, in particular $k_\infty^{\mathbb{Q}}$, σ_e^2, and λ. That is, optimal values for these parameters, given the other parameters, can be obtained analytically. Hence numerical search is only necessary over $\phi^{\mathbb{Q}}$ and Σ. ML estimates for the unrestricted model are denoted by $\hat{\theta}^{ML}$.

2.4 Priors

The approach chosen here follows the objective Bayesian tradition in that little prior information is imposed, in order to let the data speak for itself. To this end, I employ largely uninformative and if possible semi-automatic prior formulations. This stands in contrast to the approach of Chib and Ergashev (2009), who use prior information to overcome the typical problems in DTSM estimation, such as flat and irregular likelihood surfaces. The idea here is that risk-price restrictions help in estimating DTSMs without the need for strong prior information, due to the power of the no-arbitrage assumption. Of course, the aim of imposing little prior information in a model selection problem leads to some tension, because model selection results can be quite sensitive to prior dispersion. This issue will be carefully addressed.

There are five blocks of parameters for which prior distributions are needed: (λ, γ), $k_\infty^{\mathbb{Q}}$, $\phi^{\mathbb{Q}}$, Σ, and σ_e^2. I assume prior independence across blocks. For $k_\infty^{\mathbb{Q}}$ the prior is taken to be normal with mean zero and standard deviation of one annualized percentage point (1/1200). The elements of $\phi^{\mathbb{Q}}$ are *a priori* independent and uniformly distributed over the interval from zero to one—this ensures that the model is stationary under $\mathbb{Q}$. The priors for Σ and σ_e^2 are taken to be completely diffuse.[15]

The key parameters of interest are λ and γ, which are taken to be *a priori* independent of the other model parameters. The joint prior for these parameters naturally factors into a parameter prior and a model prior: $P(\lambda, \gamma) = P(\lambda|\gamma)P(\gamma)$, which will be described below.

An additional restriction on the parameter space is that the largest eigenvalue of Φ does not exceed unity, ensuring that the model is also stationary under the $\mathbb{P}$-measure.[16]

2.4.1 Model prior

I assume a uniform prior distribution over models: each element of γ is independently Bernoulli distributed with success probability 0.5, i.e., each model has equal prior probability of $0.5^{N(N+1)}$. Such a uniform prior is a common choice in the variable selection literature. It should be noted that this prior does not imply a uniform distribution over model size, but instead a Binomial distribution, with prior expectation that half of the elements of λ are unrestricted. This could be adjusted by changing the prior probability of inclusion from 0.5 to a higher or lower value, in order to a priori favor less or more restricted models (Chipman et al., 2001; Clyde and George, 2004), but in the present case there is no reason for doing so.

[15] An alternative is to use proper distributions that are rather dispersed, such as inverse-Wishart for $\Sigma\Sigma'$ and inverse-gamma for σ_e^2. This has no material impact on the estimation results.

[16] This induces a very small amount of prior dependence, which I will ignore in the notation for priors and posteriors.

2.4.2 Parameter prior for λ

The parameter prior $P(\lambda|\gamma)$ is of central importance, because the results of Bayesian model selection tend to be very sensitive to the choice of parameter priors (Kass and Raftery, 1995). In particular, overly disperse parameter priors will tilt the results in favor highly restricted models—the Lindley-Bartlett paradox (Bartlett, 1957). On the other hand, overly informative parameter priors would unduly influence posterior inference. This trade-off presents a challenge, which will be dealt with in the following way: The prior for λ will be parameterized using a single, easily interpretable hyperparameter, g. The main results are obtained for a value of g that strikes a reasonable balance between a prior that is too informative and too disperse. Then, a sensitivity analysis is carried out to assess the robustness of my main results to the prior dispersion, by varying the value of g by orders of magnitude.

Conditional prior independence of the elements of λ will be assumed throughout. This assumption is only restrictive when the parameter prior is very informative. It is commonly made in studies that use similar variable selection methods, including Dellaportas et al. (2002), George et al. (2008), and many others. Beyond the reason that there is no obvious choice for prior correlations, an additional advantage for prior conditional independence is that it substantially simplifies the model selection problem.

The alternative model selection approaches, described in Section 2.6 and in Appendix D, differ in how $P(\lambda|\gamma)$ is specified. In particular, they differ in how the elements of λ which are *excluded* from the model are distributed. However, all estimation approaches will share the same prior specification for the elements of λ that are *included* in the model.

The priors for the included risk-price parameters are assumed to be independent normal distributions centered around zero: $\lambda_i|\gamma_i = 1 \sim N(0, v_i)$. The choice of the normal family leads to normal conditional posterior distributions for λ. I parameterize the prior dispersion in terms of a single hyperparameter, using a variant of Zellner's g-prior approach. A g-prior is a Normal distribution with covariance matrix proportional to that of the least-squares estimator, for given error variance. In Appendix C.3 I show what this prior looks like in the context of an RVAR. To obtain prior conditional independence, I will use an "orthogonalized g-prior" for λ: I calculate the g-prior matrix numerically, by using an estimated covariance matrix of λ, and set off-diagonal elements equal to zero. To obtain estimates of the covariance matrix, I set all model parameters other than λ to their ML estimates and γ to a vector of ones, and calculate the least-squares estimates of λ. Denote the covariance matrix of these estimates by $\hat{V}$ and the diagonal elements of this matrix by $\hat{\sigma}^2_{\lambda_i}$. The prior variances are taken to be $v_i = g\hat{\sigma}^2_{\lambda_i}$.

With this semi-automatic prior specification, only the hyperparameter g has to be chosen. Very large values of g are to be avoided since such choices would induce the Lindley-Bartlett

paradox, i.e., they would tend to strongly favor the most restrictive model (Clyde and George, 2004). I use a moderate value of $g = 100$ to obtain my main results, so that prior dispersion is ten times as large as the standard errors of the ML estimates for the unrestricted model. Then, I will assess the sensitivity of these results to the choice of g.

2.5 Estimation with given risk price specification

I now describe an MCMC sampler to estimate the model for a given risk price specification γ. The sampler provides draws for the parameters which are approximately from their joint posterior distribution,

$$P(\theta|Y,\gamma) \propto P(Y|\theta,\gamma)P(\theta|\gamma).$$

The algorithm is a block-wise Metropolis-Hastings (MH) sampler, which iteratively draws the parameter blocks λ, $k_\infty^{\mathbb{Q}}$, $\phi^{\mathbb{Q}}$, Σ, and σ_e^2. The chain is initialized at $\hat{\theta}^{ML}$. Each block is sampled from its conditional posterior distribution, given the current values of the remaining blocks. Because the stationarity of the VAR depends on both λ and $\phi^{\mathbb{Q}}$, this prior restriction is checked at the end of each iteration, and in the case of explosive roots for Φ, all parameters are reverted to the previous iteration's values.[17] Iterating on the block-wise algorithm, the first B observations (the burn-in sample) are discarded so that the effect of the starting values becomes negligible, and the draws from the next M iterations (the MCMC sample size) are retained for posterior analysis.

2.5.1 Drawing λ

Conditional on all other parameters, estimation of the risk-price parameters in a DTSM simply amounts to estimating the parameters of an RVAR. This is very convenient, because the conditional posterior for λ is available in closed form, and λ can be sampled using a Gibbs step. The posterior is normal with mean and covariance matrix given in Appendix C.2. It is a key advantage of the MCMC sampler proposed here that the majority of parameters can be sampled very efficiently using a straightforward Gibbs step.

2.5.2 Drawing $k_\infty^{\mathbb{Q}}$

To draw $k_\infty^{\mathbb{Q}}$ I use a tailored proposal density, based on Chib and Greenberg (1994) and Chib and Ergashev (2009). In particular, the proposal is a Student-t distribution with five degrees of freedom, denoted by $q(k_\infty^{\mathbb{Q}})$. Its mean equal is set to the mode of the conditional

[17] This prior constraint of stationarity is not made explicit in the notation.

posterior of $k_\infty^{\mathbb{Q}}$, given all other parameters. Conveniently, since $k_\infty^{\mathbb{Q}}$ can be concentrated out of the likelihood function, this mode can be found analytically. The variance of the proposal distribution is the inverse of the second derivative of the conditional posterior, which I find numerically. The acceptance probability is

$$\alpha(k_\infty^{\mathbb{Q}\,(j-1)}, k_\infty^{\mathbb{Q}\,*}) = \min\left\{\frac{P(Y|k_\infty^{\mathbb{Q}\,*}, \theta_-)P(k_\infty^{\mathbb{Q}\,*}, \theta_-)q(k_\infty^{\mathbb{Q}\,(j-1)})}{P(Y|k_\infty^{\mathbb{Q}\,(j-1)}, \theta_-)P(k_\infty^{\mathbb{Q}\,(j-1)}, \theta_-)q(k_\infty^{\mathbb{Q}\,*})}, 1\right\},$$

where $k_\infty^{\mathbb{Q}\,*}$ is the proposed value and θ_- denotes all parameters other than $k_\infty^{\mathbb{Q}}$.

2.5.3 Drawing $\phi^{\mathbb{Q}}$

A Random Walk (RW) step is used to draw $\phi^{\mathbb{Q}}$, the eigenvalues of $\Phi^{\mathbb{Q}}$. Because it is useful to have these eigenvalues sorted (in descending order) and to have similar scaling of parameters that are drawn jointly, I reparameterize them as $\chi = (\phi_1^{\mathbb{Q}} - 1, \phi_2^{\mathbb{Q}} - \phi_1^{\mathbb{Q}}, \phi_3^{\mathbb{Q}} - \phi_2^{\mathbb{Q}})$. The elements of χ are required to lie between -1 and 0. The proposed values are obtained as $\chi^* = \chi^{(i-1)} + t_5$, where t_5 is drawn from a multivariate t-distribution with five degrees of freedom, and covariance matrix set to an estimate of the negative inverse of the Hessian matrix of the conditional posterior density at the current values of the parameters.[18] This automatic approach for choosing the dispersion of the innovations in the RW step, which is borrowed from the tailored proposal approach, avoids the need to tune scaling parameters for RW proposals and automatically leads to acceptance probabilities in the 20-50% range that is recommended in the MCMC literature (Gamerman and Lopes, 2006, p. 196). The acceptance probability for this RW step is is calculated as

$$\alpha(\chi^{(j-1)}, \chi^*) = \min\left\{\frac{P(Y|\chi^*, \theta_-^{(j-1)}, \gamma)P(\chi^*, \theta_-)}{P(Y|\chi^{(j-1)}, \theta_-^{(j-1)}, \gamma)P(\chi^{(j-1)}, \theta_-^{(j-1)})}, 1\right\},$$

where χ^* is the proposed draw and θ_- denotes all parameters other than $\phi^{\mathbb{Q}}$.

2.5.4 Drawing Σ

For drawing Σ, one could again use a tailored proposal step, but finding the mode and Hessian of the target distribution numerically is computationally costly. An efficient alternative is to

[18] The Hessian is approximated numerically. To save on computation time, I update the estimate only every 100 iterations. If the Hessian is not positive definite, then it is replaced by a positive definite matrix that is close to it, using an eigenvalue decomposition and replacing the negative eigenvalue(s) with small positive eigenvalues.

use a proposal for the shock covariance matrix, $\Sigma\Sigma'$, that is Inverse-Wishart (IW) with mean equal to the value of the previous draw and scale adjusted to tune the acceptance probability. This is a proposal akin to that of a RW step, but without being symmetric and incorporating the requirement of a positive-definite covariance matrix. Since the prior for Σ is taken to be uninformative, the acceptance probability is

$$\alpha(\Sigma\Sigma'^{(j-1)}, \Sigma\Sigma'^{*}) = \min\left\{\frac{P(Y|\Sigma^{*}, \theta_{-})q(\Sigma\Sigma'^{*}, \Sigma\Sigma'^{(j-1)})}{P(Y|\Sigma^{(j-1)}, \theta_{-})q(\Sigma\Sigma'^{(j-1)}, \Sigma\Sigma'^{*})}, 1\right\}.$$

Here $q(A, B)$ denotes the proposal density, an IW distribution with mean A.

2.5.5 Drawing σ_e^2

The measurement error variance can be drawn using a Gibbs step, because conditional on the other parameters and the data, the measurement errors can be viewed as regression residuals. For an inverse-gamma (IG) prior with shape parameter $\alpha_0/2$ and scale parameter $\delta_0/2$, the conditional posterior is IG with shape and scale determined by $\alpha_1 = \alpha_0 + n$ and $\delta_1 = \delta_0 + ssr$, where n is the number of observations and ssr is the sum of squared residuals. Since the variance is the same across all measurement equations, the measurement errors are pooled. We have $n = T(J-N)$, because at each point in time there are $J-N$ independent measurement errors, and $ssr = \sum_{t=1}^{T} ||Y_t - \hat{Y}_t||$.[19] Furthermore, $\alpha_0 = \delta_0 = 0$ because the prior is taken to be completely diffuse.

2.6 Selection of risk-price restrictions

The general principle of Bayesian model selection is to compare the posterior model probabilities of alternative specifications. Calculating these requires knowledge of the marginal likelihood of all models. However, this is not practical when the number of candidate models is large. MCMC algorithms that sample jointly across models and parameters can identify a smaller set of plausible model specifications without the need to estimate all candidate models. The most interesting models, namely the ones with high posterior probability, will naturally be visited more frequently by such samplers.[20] The task of choosing zero restrictions on risk-price parameters in a DTSM closely parallels the problem of selecting variables in multivariate regressions. The two differences are, first, that the parameters of interest, λ, are those of an

[19] See footnote 13.

[20] For a recent overview of methods available for joint model-parameter sampling, see Godsill (2001) and Sisson (2005). For a review of MCMC methods for variable selection, see O'Hara and Sillanpää (2009).

RVAR and not of a multivariate regression, and second, that there are additional parameter blocks. Despite these differences, existing approaches to variable selection can be adapted to the present context. The MCMC chain is initialized at $\gamma = (1, 1, \ldots, 1)$ and $\theta = \hat{\theta}^{ML}$. In each iteration, first (λ, γ) is drawn as specified by the model-selection algorithm, and then $k_\infty^{\mathbb{Q}}$, $\phi^{\mathbb{Q}}$, Σ, and σ_e^2 are drawn as described above in Section 2.5.

The approach I describe here is based on Gibbs variable selection (GVS). Two alternative approaches which I will also use to ensure the reliability and robustness of the model selection—namely, Stochastic Search Variable Selection (SSVS) and Reversible-Jump Markov chain Monte Carlo—are described in Appendix D. GVS, which was developed by Dellaportas et al. (2002), is a special case of the product-space sampling of Carlin and Chib (1995). In product-space sampling, the parameter space is the product-space of all models under consideration, which can become very large. For GVS, models are treated as nested, so that the product-space is simply the parameter space of the unrestricted model. The parameter prior for GVS is

$$P(\lambda|\gamma) = P(\lambda_\gamma|\gamma)P(\lambda_{\setminus\gamma}|\lambda_\gamma, \gamma)$$

where $P(\lambda_\gamma|\gamma)$ is the prior for those elements of λ that are included, and $P(\lambda_{\setminus\gamma}|\lambda_\gamma, \gamma)$ denotes the prior for elements that are not currently included, the so-called "pseudo-prior" or "linking density" in the parlance of Carlin and Chib (1995). It is generally recommended to use a pseudo-prior similar to the posterior distribution from the unrestricted model. To this end, I use independent normal distributions with mean and variance equal to the conditional posterior moments of λ given the ML estimates of all other parameters.[21]

In each iteration of the GVS sampler, first λ_γ is drawn in a Gibbs step from its conditional posterior distribution. Due to prior conditional independence, this only depends on the prior for λ_γ, on γ, on the remaining parameters, and on the data. The conditional posterior distribution is given in Appendix C.2. For those elements of λ that are not included in the current model, $\lambda_{\setminus\gamma}$, the data is not informative and the values are drawn from the pseudo-prior.

The success probability for the Bernoulli conditional posterior of $\gamma_i^{(i)}$ is determined by

$$\begin{aligned}
&\frac{P(\gamma_i^{(j)} = 1|\lambda^{(j)}, \theta_-^{(j-1)}, \gamma_{-i}^{(j)}, Y)}{P(\gamma_i^{(j)} = 0|\lambda^{(j)}, \theta_-^{(j-1)}, \gamma_{-i}^{(j)}, Y)} \\
&\quad = \frac{P(Y|\gamma_i^{(j)} = 1, \lambda^{(j)}, \theta_-^{(j-1)}, \gamma_{-i}^{(j)})}{P(Y|\gamma_i^{(j)} = 0, \lambda^{(j)}, \theta_-^{(j-1)}, \gamma_{-i}^{(j)})} \cdot \frac{P(\lambda_i^{(j)}|\gamma_i^{(j)} = 1)}{P(\lambda_i^{(j)}|\gamma_i^{(j)} = 0)} \cdot \frac{P(\gamma_i^{(j)} = 1, \gamma_{-i}^{(j)})}{P(\gamma_i^{(j)} = 0, \gamma_{-i}^{(j)})}.
\end{aligned}$$

[21] This is a convenient alternative to the common practice of doing a pilot run for the unrestricted model, taking advantage of the closed-form availability of conditional posterior moments for λ.

To calculate the first term, the ratio of likelihoods, only the $\mathbb{P}$-likelihoods need to be evaluated, since the parameters of the $\mathbb{Q}$-likelihood remain unchanged and this term cancels out. For the second term, the numerator is the density of the prior, i.e., a normal distribution with mean zero and variance v_i, and the denominator is the density of the pseudo-prior. The third term cancels out due to equal prior model probabilities.

2.7 Bayesian Model Averaging

The approach described above provides a way to infer plausible risk-price restrictions. But it is important not to restrict attention to one specific model, and instead to entertain the possibility that different models may be plausible but imply different results. This can be accomplished by means of Bayesian Model Averaging (BMA). In BMA, estimates of model parameters and any objects of interest—such as, in the present context, interest rate persistence, volatilities, short-rate expectations, and term premia—are calculated as averages across specifications, using posterior model probabilities as weights. Conveniently, BMA estimates can be obtained in a straightforward fashion from the output of the model-selection sampler: Ignoring the draws for the model indicator γ, the MCMC sample reflects the posterior distribution of the parameters across models. The BMA distribution of any object of interest can then be obtained by simply calculating it for every draw of these parameters. BMA allows me to take into account the statistical uncertainty over models in the empirical analysis, in order to avoid the sense of false confidence by conditioning on one specific restricted model.

3 Simulation study

Because of the novelty of the econometric framework, it is important to assess its reliability and effectiveness in a simulation study. To this end, I simulate yield data from a given DTSM, and assess the performance of the samplers in recovering the true parameters and restrictions, as well as in estimating the persistence and volatility of interest rates.

The data-generating process (DGP) is a two-factor DTSM with restrictions and parameters specified based on ML estimates on the actual data. The two factors correspond to the level and the slope of the yield curve. The DGP is constructed as follows: First, I obtain the MLE for the unrestricted model. Second, I choose the restrictions on risk-prices based on which elements of λ are found to be significantly different from zero at the 5% level. Third, I re-

estimate the model under these restrictions. This results in the following risk-price parameters:

$$\lambda_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \lambda_1 = \begin{pmatrix} 0 & -0.0236 \\ 0 & 0 \end{pmatrix}.$$

The remaining parameters are $k_\infty^{\mathbb{Q}} = 6.6 \cdot 10^{-5}$, $\phi^{\mathbb{Q}} = (0.9899, 0.9789)$, $\sigma_e = 2/120000$ (2 basis points), and $\Sigma = ((2 \cdot 10^{-4}, 1 \cdot 10^{-4})', (0, 2 \cdot 10^{-4})')$. Using this DGP, I generate 50 samples of size $T = 300$. The sample length is chosen to be similar to the sample sizes we encounter in practice when using monthly observations of post-1980 interest rate data, as many studies do. For each sample I estimate the maximally-flexible model using the MCMC sampler described in Section 2.5, and then run the model-selection samplers SSVS, GVS, and RJMCMC, in order to identify plausible restrictions and to estimate the model using BMA. The burn-in period is 5,000 for all samplers. The model-specific MCMC sampler is run for 10,000 iterations, whereas the model-selection samplers are each run for 40,000 iterations.

How well do the samplers fare in recovering the true model and in estimating risk-price parameters? Table 1 reports the estimation results for γ and λ. The first two rows show the specification of the data-generating process (DGP). The following two rows show the results for estimation of the unrestricted model using MCMC, first the average for λ across all samples and iterations, and below the frequency of "significant" estimates—those for which the credibility intervals do not straddle zero—among the 50 samples. For the one parameter that is non-zero in the DGP, the average estimate is close to the true value, but it is significant in only 22% of the samples. For the parameters which are zero in the DGP, estimates from the simulated data are not centered on zero, and they are often found to be significant. There appears to be some small-sample bias, i.e., the posterior distributions are, on average across samples, not centered around the DGP parameters. If, for any given sample, we select a model based on which parameter estimates are significant, then the model is correctly identified in only 6% of the cases. Overall, MCMC estimation of the unrestricted model leads to unsatisfactory results—neither does it accurately recover the risk-price parameters nor does it correctly identify those that should be non-zero. This indicates that the common approach in the DTSM literature of choosing risk-price restrictions based on statistical significance will typically lead to the wrong model.

The following rows in Table 1 show the results for the model-selection samplers. The first row reports average value of γ—the posterior probability of inclusion—across samples and replications, the second row shows the average value of λ, and the third row shows the average value of λ conditional on the sampler visiting the DGP model. All three model-selection samplers exhibit similar performance in estimating λ and choosing among models.

The posterior probability of inclusion is largest for that parameter which is non-zero in the DGP. For this parameter, the inclusion probability is above 50%, whereas for the others it is below 50%—the algorithms correctly identify the parameters which should be non-zero. Averaged across the samples in which the parameter is correctly identified to be unrestricted, the estimated value is close to the DGP value. The percentage of samples in which the modal model (the model with the highest posterior probability) corresponds to the DGP model, reported in the last column, is higher than 50%. The posterior model probability of the DGP model estimated across all samples and iterations, which is given in squared brackets in the last column, is over 30%. Overall, the model-selection samplers show a high accuracy in determining the correct model, and its performance should be viewed as a significant success, given that we have a rather small sample and estimation of risk prices is otherwise notoriously difficult.[22]

Small-sample bias is evident in the simulation results. For those elements of λ which are restricted to zero in the DGP, the MCMC sampler leads to estimates which are quite different from zero, and the model-selection samplers in some cases find significant evidence that these parameters may be non-zero. The small-sample bias for estimates of λ parallels the bias in estimates of VAR parameters in the DTSM context (Bauer et al., 2012). It stems from the high persistence of interest rates in combination with the small sample size. Due to the parameterization of the model, and the fact that $\mathbb{Q}$-parameters are estimated very precisely, the bias in μ and Φ translates into bias in λ_0 and λ_1.[23] The simulation results demonstrate that in spite of the small-sample bias, risk-price restrictions can be recovered quite accurately. Furthermore, imposing risk-price restriction generally reduces this bias: for parameters that are zero in the DGP, the model-selection samplers generally deliver estimates that are much closer to zero than those for the unrestricted model.

The typical purpose of estimating a DTSM is to recover short-rate expectations and term premia. Is the approach proposed in this paper more accurate in recovering the key objects of interest than estimation of a maximally-flexible DTSM? Table 2 shows the estimated interest rate persistence and the implied volatilities of expectations and term premia, in comparison to the values implied by the DGP. The persistence is measured by the largest eigenvalue of Φ, and the impulse-response function for the level factor in response to level shocks at the five year horizon. The DGP parameters imply highly persistent VAR dynamics. The MCMC estimation

[22]Note that the sampler is much more accurate in determining the correct model than the comparable approach of George et al. (2008) in the VAR context.

[23]Naturally, this issue disappears for larger sample sizes. As a consequence, the model-selection samplers become more accurate in recovering the DGP specification when longer samples are simulated (results not shown).

leads to persistence that is considerably lower, reflecting the usual downward bias in estimated persistence. In contrast, estimation under risk-price restrictions leads to substantially higher persistence, close to the values implied by the DGP.

Table 2 also reports model-implied volatilities of changes in five-to-ten-year forward rates, risk-neutral forward rates (i.e., short-rate expectations), and forward term premia. In the DGP, forward term premia are less volatile than risk-neutral forward rates. Estimation of the unrestricted model gets this wrong, implying forward term premia that are substantially more volatile than risk-neutral forward rates, which are very stable. In contrast, estimation under restrictions on risk pricing much more accurately recovers the true volatilities. In particular, this approach ensures that the expectations component is correctly identified as being more volatile than the term premium component.

In sum, the results of the simulation study shows that the model selection samplers appear to be working quite well for recovering restrictions on risk pricing. Furthermore, estimation under plausible risk-price restrictions significantly improves the accuracy in recovering the persistence of interest rates and the volatilities of short-rate expectations and term premia. This stands in stark contrast to the common approach of estimating maximally-flexible models, which in this simulation setting leads to seriously distorted results about the objects of interest.

4 Estimation results

The econometric framework is now applied to real-world data. I use monthly observations of nominal zero-coupon U.S. Treasury yields, with maturities of one through five, seven, and ten years.[24] The sample period starts in January 1990 and ends in December 2007 (as in, for example, Joslin et al., 2011), which yields $T = 216$ monthly observations. The start of the sample is chosen to avoid the structural break in interest rate dynamics that was likely caused by changing monetary policy procedures in the 1980s. The sample excludes the recent period of near-zero interest rates because affine Gaussian models are ill-suited to deal with the zero lower bound Bauer and Rudebusch (2013a).[25]

[24] The yields are unsmoothed Fama-Bliss yields, bootstrapped from bond prices in the same way as the Fama-Bliss data commonly used. The data was generously made available by Anh Le.

[25] The estimation approach proposed here to could be extended to models that restrict yields to remain positive, such as, for example, shadow-rate DTSMs. This and other extensions are discussed in Section 6.

4.1 Maximally-flexible model

Estimates of the unrestricted, maximally-flexible DTSM, denoted henceforth as model $\mathcal{M}_0$, will serve as a benchmark against which to compare subsequent results. This comparison will reveal how risk-price restrictions change the economic implications of a typical affine Gaussian DTSMs. Choosing $\lambda = (1, 1, \ldots, 1)$, I run the sampler described in Section 2.5 for 50,000 iterations, and discard the first 10,000 iterations as a burn-in sample. Table 3 reports the prior and posterior means and standard deviations for all model parameters. Also shown are average acceptance probabilities for those blocks not sampled using MH steps, and inefficiency factors, calculated as in, for example, Chib and Ergashev (2009). The parameter estimates reported in the table are very similar to those obtained using ML estimation (not shown), because the priors are not very informative.

The efficiency of the sampler is satisfactory. While estimation of the risk-price parameters λ_0 and λ_1 has been found difficult in the DTSM literature due to the flat likelihood surface, sampling of this block is very efficient here. The reason, of course, is that it is carried out using a Gibbs step, which is an important advantage of my estimation methodology. For the blocks sampled using MH steps, the sampling efficiency is acceptable: For Σ and $\phi^{\mathbb{Q}}$, the acceptance probabilities are within the recommended range of 20-50%, while for $k_\infty^{\mathbb{Q}}$ it is on the high side, around 90%. The inefficiency factors are slightly high for $k_\infty^{\mathbb{Q}}$ and $\phi^{\mathbb{Q}}$, which warrants running the sampler for a higher number of iterations, as is done here. In any case, since estimation of $\mathbb{Q}$-parameters is very precise, some inefficiency of the sampler here is of little concern.

The risk-price parameters are estimated very imprecisely. This is due to the high persistence of interest rates. As a consequence, the dynamic properties of interest rates, determined by μ and Φ, are hard to pin down, which in my parameterization translates into high uncertainty about the λ_0 and λ_1. Only for two of the twelve risk-price parameters do the 95%-CIs intervals not straddle zero.[26] This is a first indication that the data support a parsimonious specification, with many risk-price parameters set to zero.

4.2 Restrictions on risk prices

To obtain the model-selection results, the GVS, SSVS, and RJMCMC samplers are each run for 200,000 iterations, and the first 10,000 iterations are discarded. Based on the output of these samplers, we can carry out posterior analysis of the risk-price restrictions and model specifications. Table 4 shows summary statistics for each MCMC sample of γ: posterior

[26] The standard deviations for elements of λ under the prior and posterior distributions differ by a factor very close to ten. This is no coincidence, but due to the fact that the prior variance is set to $g = 100$ times the variance of the ML estimates.

means (the posterior probabilities of inclusion for the corresponding elements of λ), Monte Carlo standard errors, and inefficiency factors. Table 5 shows posterior model probabilities, calculated as the relative frequencies of each specification in the sampled chain for γ, for the 15 most frequently visited models, sorted using the output from the SSVS sampler. I also report posterior odds (the ratio of posterior probabilities) for each model relative to the modal model, which correspond to Bayes factors because the prior over models is flat.

The results for all three samplers are closely in line with each other. In Table 4, differences in estimated inclusion probabilities are generally within the tolerance indicated by the numerical standard errors. In Table 5, the posterior model probabilities and posterior odds are very similar across the three sampling algorithms, and the model rankings are largely identical, certainly near the top. Table 5 also reports how many models were visited, in total, by each of the algorithms. None of the algorithms visits more than about one fourth of all possible models. This demonstrates the advantage of joint-model-parameter samplers: they focus on exploring models with sufficiently high posterior probability, and do not waste time in the uninteresting area of the model space. While the model-selection results are consistent across the three samplers, they differ in terms of their convergence properties, as evident from the inefficiency factors in Table 4. The GVS sampler is the most efficient sampler, followed closely by the SSVS sampler. The RJMCMC sampler, on the other hand, has significantly higher inefficiency factors. The GVS algorithm emerges as the favored model-selection sampler, because it converges quickly and, as opposed to the SSVS sampler, restricts the excluded parameters to be exactly zero. From now on, I will therefore focus on the GVS results.

The evidence strongly favors tightly restricted models, with only very few free risk-price parameters. In Table 4, only one element of λ has a high posterior probability for inclusion, which is above 95%. For all other parameters, the inclusion probabilities are below 50%, and for most of the parameters they are near zero. Table 5 shows that all of the 15 most plausible models leave only one to three risk-price parameters unrestricted. The evidence for tight restrictions is clearest with a prior-posterior analysis for the number of unrestricted parameters, which has a prior distribution that is Binomial $(12, 0.5)$. While the prior mean number of unrestricted risk-price parameters is 6, the posterior mean is only 2.1. And while the prior probability of at most two unrestricted risk-price parameters is below 1%, the corresponding posterior probability is 67%. The posterior distribution puts substantially more weight on models with a low number of parameters than the prior distribution.

The evidence is also quite clear on which restrictions are favored by the data. By far the most important risk-price parameter is the $(1, 2)$ element of the λ_1 matrix, which determines the sensitivity of the price of level risk to variation in the slope factor. The evidence strongly

favors the model with only this one element of λ being non-zero. This model will be denoted as $\mathcal{M}_1$. There is also some evidence, albeit weaker, in favor of two other models which have one or two additional unrestricted risk-price parameters. These models in the second and third row of Table 5 will be denoted by $\mathcal{M}_2$ and $\mathcal{M}_3$. Beyond the first three models, the posterior odds ratios of all other models are above 10. Based on the guidelines for interpreting Bayes factors in Kass and Raftery (1995), there is substantial evidence against all models but the first three.

As a reality check for the model-selection results, the last two columns of table 5 report information criteria—the Akaike Information Criterion (AIC) and the Schwartz-Bayes Information Criterion (SBIC)—based on maximum likelihood estimates of the restricted models.[27] While AIC gives a somewhat different ranking, the ranking of the models according to the (theoretically superior) SBIC is consistent with the ranking based on Bayesian model selection. This shows that the results in Tables 4 and 5 are actually driven by information in the data, and not by the choice of priors, the Lindley-Bartlett paradox, or some feature of the sampling algorithms. On the other hand, the disadvantages of information criteria are also evident: In particular, they do not tell us how much more plausible one model is over another and have no intuitive interpretation, besides the fact that they cannot be used to address the issue of model uncertainty.

We can learn a lot about the pricing of risks in Treasury markets from the model-selection results, since we can systematically investigate any hypothesis about risk-pricing. A key question is whether only level risk of Treasury bonds is priced, as argued for example by Cochrane and Piazzesi (2008). In the model, level risk is priced if the first element of λ_0 or any element in the first row of λ_1 is non-zero. According to the GVS sample, the posterior probability for this hypothesis is 99.99%. In contrast, the posterior probabilities that slope or curvature risk is priced is only 22% and 18%, respectively.[28] Hence, the evidence here supports the view that level risk is priced, while slope risk and curvature risk is not. In economic terms, this means that investors only worry about (and require risk compensation for) shocks to the level of the yield curve. The risk of shocks that affect only the slope or curvature of the yield curve is not priced, i.e., these shocks are not correlated with expected excess bond returns.

Are risk prices and term premia time-varying, and if yes, what drives this time variation? The posterior probability that the price of level risk is time-varying is 99.8%, based on the

[27] Instead of full MLE of all restricted models, which is computationally costly, these are obtained using a shortcut: Since the $\mathbb{Q}$-parameters and Σ are largely unaffected by the risk-price restrictions, I take these as given and use the same values for all models—they are obtained from the estimation of the unrestricted model. For each restricted model the optimal values for λ can then be obtained using least squares.

[28] These probabilities are estimated using the relative frequency of how often the corresponding row of $[\lambda_0, \lambda_1]$ is non-zero.

frequency of draws with at least one element in the first row of λ_1 being non-zero. This is very strong evidence against the expectations hypothesis.[29] Furthermore, the evidence is strong that changes in the slope of the yield curve drive variation in the price of level risk, and there is some modest evidence that changes in the level of yields contribute to this variation as well: according to Table 4, the posterior probabilities are 97% for the former hypothesis, and 21% for the latter. Overall, the data overwhelmingly suggests that the price of level risk varies with changes in the slope factor. The finding that changes in the slope affect term premia goes back to Fama and Bliss (1987) and Campbell and Shiller (1991). It is of course comforting that the results here are consistent with a long tradition of regression results on Treasury yields. But more importantly, while previous work in the DTSM literature has essentially ignored these restrictions, we will see in Section 5 how the estimates and economic implications of a DTSM are changed, once the risk-price restrictions that are so strongly supported by the data are incorporated into an otherwise standard DTSM for Treasury yields.

4.3 Sensitivity to prior dispersion

Model selection is sensitive to the specification of the prior. In particular, disperse parameter priors lead to results that favor highly restricted models. One may be concerned that the evidence above in favor of tightly restricted model specifications could be due to high prior dispersion for λ.

The prior distribution for λ is parameterized in terms of the hyperparameter g, which is equal to 100 in the baseline setting. To understand how the selection of risk-price restrictions is affected by the choice of the prior, I carry out the model selection exercise, using the GVS sampler, for different values of g. Table 6 displays key statistics describing the results in each case. In particular, it shows the estimated posterior probability of model $\mathcal{M}_1$, the number of models visited by the sampler with frequency of at least 1% and in total, the posterior mean number of unrestricted risk-price parameters, and the posterior probability that not more than two risk-price parameters are non-zero.

Increasing g leads to a more peaked posterior distribution over models: more probability mass is put on the modal model; there are fewer high-probability models; and less models are visited by the sampler overall. Lower values of g flatten out the posterior model distribution.

The results of the model selection exercise are robust to a wide range of values for g. Model $\mathcal{M}_1$ remains the modal specification in all cases. The evidence in favor of tight restrictions is strengthened with a more disperse parameter prior, as one would expect. More importantly,

[29] The posterior probability that *any* risk price is time-varying is 99.9%.

this evidence also remains robust to making the prior much less disperse. Even in the case that $g = 10$, which is a value much lower than commonly used in practice, the evidence strongly favors parsimonious models. In this case, the prior probability of at most two unrestricted risk-price parameters is updated substantially by the data, from 0.3% to 13.7%.

It is important to vary prior dispersion in practical applications of model selection methods. For the case at hand, the key finding—the evidence in favor of tight restrictions on risk-price parameters—remains robust.

5 Economic implications

This section discusses the economic implications of restrictions on risk prices. It compares the results for the unrestricted model $\mathcal{M}_0$, the restricted models $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$, as well as the results for BMA.

The models have almost identical $\mathbb{Q}$-dynamics, i.e., estimates for $k_\infty^{\mathbb{Q}}$ and $\phi^{\mathbb{Q}}$ are very similar (results omitted). All models have essentially the same cross-sectional fit, with a the root-mean-squared fitting errors across all observations and maturities is about three basis points. Restrictions on risk pricing essentially have no effect on the cross-sectional fit to the yield curve (see also Joslin et al., 2014). However, they substantially affect the $\mathbb{P}$-dynamics, that is, the time-series properties of the model, and consequently the properties of expectations and term premia, as will be discussed in the following.

5.1 Persistence and volatilities

The estimated persistence of risk factors and interest rates is of crucial importance for the economic implications of a DTSM, because it determines the properties of short-rate expectations and term premia. Table 7 compares the estimated persistence across models, under both probability measures, $\mathbb{Q}$ and $\mathbb{P}$, measured by the largest eigenvalue of the mean-reversion matrix in each case ($\Phi^{\mathbb{Q}}$ and Φ). It also reports model-implied volatilities of monthly changes in five-to-ten-year forward rates (short-rate expectations under $\mathbb{Q}$), in risk-neutral forward rates (short-rate expectations under $\mathbb{P}$), and in the corresponding forward term premium. For each statistic, I report posterior means and 95%-CIs.

The $\mathbb{Q}$-persistence is very similar across models, since the $\mathbb{Q}$-dynamics are largely unaffected by risk-price restrictions. Consequently, the volatility of fitted forward rates does not vary across models. The $\mathbb{Q}$-persistence is generally high, to match the empirical fact that long-term forward rates are quite variables (see, for example, Gürkaynak et al., 2005).

Under the real-world probability measure $\mathbb{P}$, interest rates are less persistent than under $\mathbb{Q}$, and this is true for all models. Consequently, short-rate expectations are less variable than forward rates. But there are important differences across models. Restricted models generally exhibit higher persistence under $\mathbb{P}$ than the maximally-flexible model $\mathcal{M}_0$. Model $\mathcal{M}_2$, which implies lower $\mathbb{P}$-persistence than $\mathcal{M}_0$, is an exception.[30] All restricted models imply more volatile short-rate expectations than the maximally-flexible model. Across all models, as reflected in the BMA results, the $\mathbb{P}$-persistence and the volatility of short-rate expectations are higher than in the maximally-flexible model. For BMA, this volatility is almost twice as large as for $\mathcal{M}_0$. The low volatility of long-horizon expectations of short rates is a counter-factual implications of conventional DTSM estimates (e.g., Kim and Orphanides, 2012). The evidence here shows that plausible risk-price restrictions raise the volatility of short-rate expectations, so that they play a more important role for movements in long-term forward rates.

Since conventional DTSMs typically imply very stable long-horizon short-rate expectations, they attribute a large role to the term premium for explaining movements in long rates. The large volatilities of term premia has been criticized as a short-coming of these models. Table 7 shows that risk-price restrictions often, though not for all models, lower the volatility of term premia. For BMA the term premium volatility is 10% lower than for $\mathcal{M}_0$, and the term premium accounts for an about equal amount of volatility in long rates as do short-rate expectations. That is, the puzzle of an implausibly large role for term premia in explaining variation in long rates is to some extent alleviated when plausible restrictions are imposed on an otherwise standard DTSM.

Beyond changing point estimates of $\mathbb{P}$-persistence and term premium volatility, the restrictions also improve precision. Using only time-series information, it is difficult to estimate the speed of mean reversion of interest rates. This is reflected in large CIs in Table 7 for the unrestricted model. However, the precision can be improved by imposing risk-price restrictions. The CIs are generally narrower for the restricted models. The intuition is that information in the cross section is used for estimating the dynamic properties of interest rates. For the BMA results the CIs are wider than for any individual restricted model. This demonstrates that it may be problematic to focus on one restricted model, as do Cochrane and Piazzesi (2008) and Joslin et al. (2014), because this can significantly understate the statistical uncertainty. Model uncertainty should be taken into account whenever possible.

To understand the dynamic properties of a DTSM, it is instructive to consider volatilities

[30] The usual intuition is that risk-price restrictions generally tighten the connection between cross section and time series, and "pull up" the $\mathbb{P}$-persistence toward to $\mathbb{Q}$-persistence (see also Joslin et al., 2014). The results for $\mathcal{M}_2$ show that this is not always the case.

across maturities. Focusing on volatilities of forward rates helps to isolate the behavior of expectations at specific horizons. Figure 1 displays "term structures of volatility," showing model-implied volatilities of monthly changes in forward rates and risk-neutral forward rates for maturities from one month to ten years, the top panel for $\mathcal{M}_0$, and the bottom panel for BMA. The figure displays the posterior means of the volatilities of forward rates, as well as the posterior means and 95%-CIs of the volatilities of risk-neutral forward rates. All volatility curves show the typical hump-shaped pattern, reaching a peak at one to two years, and declining with maturity. The forward rate volatilities are similar for the two models, declining only slowly and almost leveling out for horizons longer than five years. But the risk-neutral volatility curves differ substantially. For $\mathcal{M}_0$, they show only a very slight hump and decrease quickly. Except for the very shortest maturities, risk-neutral volatilities are much lower than forward rate volatilities, implying only a limited role for changes in expectations to account for movements in interest rates. For BMA, risk-neutral volatilities rise with forward rate volatilities, and display a very similar pattern for horizons up to five years. Only for longer maturities do risk-neutral volatilities drop below forward rate volatilities. Overall, BMA attributes a much larger role to short-rate expectations for explaining interest rate volatility. Risk-price restrictions typically lead to a slower decay of risk-neutral volatility with horizon, so that far-ahead expectations are more volatile and term premia more stable. The figures also show that it is hard to estimate risk-neutral volatilities—the CIs are quite wide in both cases. While for any individual restricted model, these are much narrower (not shown) than for the maximally flexible model, taking into account the model uncertainty naturally widens the range of plausible estimates, reflecting the significant model uncertainty for DTSMs.

5.2 Historical evolution of short-rate expectations and term premia

One of the main uses of a DTSM is to decompose long-term interest rates into short-rate expectations and term premia. Figure 2 shows this decomposition for the ten-year yield, comparing $\mathcal{M}_0$ and BMA. The top panel shows the estimates of the risk-neutral yield, and the bottom panel shows the corresponding term premium, calculated as the difference between fitted and risk-neutral yield. Both panels also show the actual and model-fitted ten-year yield as a point of reference.[31] Recessions are indicated as shaded areas.

The top panel shows that the risk-neutral yields from model $\mathcal{M}_0$ are more stable than the estimates from BMA. The risk-neutral yield for BMA falls very significantly around the 2001 recession, and in the time before the onset of the Great Recession (2007–2009), whereas

[31] The fitted yields are obtained from model $\mathcal{M}_0$, which are practically identical to those obtained from the other models.

the decreases in the risk-neutral yield for $\mathcal{M}_0$ are more muted. The bottom panel shows that the yield term premium is noticeably more stable for BMA, relative to the maximally flexible model. Of note, the BMA term premium appears to exhibit a more pronounced counter-cyclical behavior, rising before and during recessions, and falling during expansions. This is appealing in light of much theoretical and empirical work suggesting that term premia behave in a counter-cyclical fashion.

Long-term interest rates have declined by a significant amount over the sample period. To which extent was this due to changes in monetary policy expectations and movements in term premia?[32] Figure 2 suggests that the models differ in their explanation of this downward trend in long-term rates over the course of the sample. Table 8 summarizes the models' implications for decomposing the decline. For both actual and risk-neutral yields, it reports the levels in 1990 and 2007, calculated as averages over each year, and the changes over this period. Also shown are 95%-CIs for levels and changes in risk-neutral forward rates. The ten-year yield declined by 3.8 percentage points (pps) over the sample period. The unrestricted model $\mathcal{M}_0$ implies that only a small share of this decline, less than one fourth, is due to declining short-rate expectations, and the CI for the decline in expectations straddles zero. The restricted models, with the exception of $\mathcal{M}_2$, imply a decline of short-rate expectations that is much more pronounced and significantly different from zero. BMA attributes more than one half of the yield decline to falling short-rate expectations. The decline in expectations is more pronounced, and it is estimated more precisely, as is evident from the narrower CIs, even in the case of BMA. The interpretation of the secular decline in long-term interest rates differs in important ways when risk-price restrictions are imposed on a DTSM. We can confidently conclude that it was not only caused by a lower term premium, but also to a significant extent by a downward shift in expectations of future monetary policy. Importantly, this findings is in line with the sizable decreases in survey-based expectations of inflation and policy rates documented in Kozicki and Tinsley (2001), Kim and Orphanides (2012), and others.[33]

Term premium estimates have also been used to analyze the puzzling behavior of interest rates during 2004 and 2005, which former Fed Chairman Alan Greenspan referred to as a "conundrum" (Greenspan, 2005). During this period, the Fed tightened monetary policy by substantially raising the federal funds rate, but long-term interest rates actually declined.[34] Over the period from June 2004 to June 2005 the three-month rate increased by 1.75pps, whereas the ten-year yield declined by 0.8pps. The right panel of Table 8 reports how the

[32] On this issue, see also Bauer and Rudebusch (2013b).

[33] A related critique of unrestricted Gaussian DTSMs is given in Bauer et al. (2014).

[34] From June 2004 until December 2005 the FOMC increased the target for the federal funds rate 13 times by 0.25 pps each. It then tightened four more times until June 2006.

different models interpret this decline. Conventional DTSMs typically explain the decline in long yields by a substantial decline in the term premium, a prominent example being Backus and Wright (2007). This interpretation is confirmed by the estimates here, which imply that short-rate expectations increased over this period, in line with the observed short rate, and term premia consequently fell substantially. In this case, risk-price restrictions do not change the interpretation of historical changes in interest rates.

5.3 Predictability of bond returns

It is a long-known fact that returns on U.S. Treasury bonds are predictable based on current interest rates (Fama and Bliss, 1987; Campbell and Shiller, 1991). Cochrane and Piazzesi (2005) have documented the substantial predictable component of annual excess holding-period returns. This evidence supports the view that term premia in long-term yields are time-varying, and conflicts with the expectations hypothesis. Maximally-flexible affine Gaussian DTSMs have been shown to successfully capture this feature of interest rate data (Dai and Singleton, 2002; Singleton, 2006). The flexibility of such models enables them to match not only the cross section of yields, but also the dynamic properties of yields and returns. In the restricted models proposed in the present paper, term premia and expected returns are more stable than in unrestricted models, so return predictability is more limited. Therefore, the question arises whether DTSMs with tight restrictions on risk pricing are also able to match the predictability of bond returns that we see in the data.

To answer this question, I run predictability regressions for returns on long-term bonds and check whether the estimated R^2 are matched by those implied by the models, both in population and in small samples (as in Singleton, 2006, Sec. 13.3.1.). The return regressions are similar to those of CP in that I regress annual excess bond returns on current interest rates. Specifically, the regression equation is

$$rx_{t,t+12}^{(n)} = \alpha^{(n)} + \beta^{(n)} X_t + \nu_t^{(n)}, \tag{12}$$

where $rx_{t,t+12}^{(n)}$ are annual holding-period returns, in excess of the one-year yield, on a bond with maturity n, and $\nu_t^{(n)}$ is the prediction error. The predictors here are the first three PCs of the yield curve, and hence correspond to the risk factors in the models.[35] Equation (12) is estimated for bonds with maturities of two, five, seven, and ten years, on $T - 12 = 204$ monthly observations. The R^2 are reported in the first column of Table 9. In the data, annual

[35] While CP use five instead of three yield curve variables (forward rates in their case), this is neither necessary nor possible when evaluating three-factor DTSMs (see also Singleton, 2006, p. 352).

excess returns are strongly predictable; about 35 percent of their variation is explained by level, slope, and curvature of the yield curve. The amount of predictability is substantial, and is consistent with the evidence in Cochrane and Piazzesi (2005) for a strong deviation from the expectations hypothesis. Any plausible yield curve model needs to be able to capture this important pattern in the data.

The model-implied population R^2 for these regressions can easily be obtained analytically. Since the regressions use X_t as the predictors, the model-R^2 are simply the ratio of the variance of expected excess returns to the variance of realized excess returns. Appendix E derives the model-implied population variances. For most models and maturities, the model-implied population R^2 are below the value for the actual data. The discrepancy is generally more pronounced for those restricted models with less variable term premia (in particular, $\mathcal{M}_1$), because excess returns in these models are also less variable. The BMA estimates imply population R^2 that are quite substantially below those in the data.

Predictability in population may differ from estimated predictability in small samples. To obtain small-sample R^2, I simulate 1000 yield data sets of the same length as the original data ($T = 216$), using the point estimates (posterior means) of each model's parameters. The risk factors are initialized using a draw from their unconditional distribution. For each simulated data set, I construct annual excess returns and run the same regressions as in the data. This results in an empirical distribution for the small-sample R^2. and Table 9 reports means and standard deviations for these distributions. The small-sample R^2 are notably higher than the population values, and generally close to their data counterparts. The difference between the data and the small-sample values never exceeds one standard deviation. Evidently, small-sample bias, which is due to the persistent nature of the predictors (Stambaugh, 1999), plays an important role in this context. In small samples, all models under consideration, including those with tightly restricted risk prices, are consistent with the empirical evidence on bond return predictability.[36]

6 Conclusion

This paper has introduced an novel Bayesian econometric framework to estimate DTSMs under restrictions on risk pricing. It allows for a systematic choice among a large number of restrictions and for parsimony in otherwise overparameterized models. For the specific data

[36]The models with restricted risk pricing are also consistent with the evidence in Campbell and Shiller (1991). Along the lines of the analysis in Dai and Singleton (2002), the models can both replicate the Campbell-Shiller patterns as well as fix them based on model-implied expected returns. The results are available from the author upon request.

set and model under consideration, the data support tight restrictions on risk prices. This stands in contrast to the common practice of leaving most or all of the risk-price parameters unrestricted. The restrictions change the economic implications in important ways, because they increase the estimated persistence of interest rates and therefore make short-rate forecasts (and risk-neutral rates) significantly more variable. This resolves the puzzle of implausibly stable short-rate expectations shared by most conventional DTSM models.

Estimation and specification uncertainty are often ignored but are significant in term structure models. In the words of Cochrane (2007, p. 278), "when a policymaker says something that sounds definite, such as '[...] risk premia have declined,' he is really guessing." The present paper quantifies the uncertainty around short rate forecasts and term premia. I document that model uncertainty is substantial and should not be ignored in practical applications of DTSMs.

The framework is applicable beyond the specific types of restrictions and class of models that the paper focuses on. One could study a broader class of restrictions, including rank restrictions on the risk sensitivity matrix of the type suggested by Cochrane and Piazzesi (2005, 2008) and Joslin et al. (2011). Enlarging the class of restrictions can open up additional potential for the no-arbitrage assumption to pin down term premium estimates. The approach can be extended to affine DTSMs that include (spanned or unspanned) stochastic volatility, such as the models described in Creal and Wu (2015). This is particularly promising because in such models, variation in term premia is not exclusively attributed to changes in prices of risk, but also to variation in the quantity of risk (i.e., changes in second moments). We may therefore gain more insight from such models about risk pricing and the sources of variation in term premia. Another important extension would be to non-affine models, such as the shadow-rate DTSMs in Kim and Singleton (2012) and Bauer and Rudebusch (2013a) which incorporate the zero lower bound on nominal interest rates. All of these extensions are in principle straightforward, because in a Bayesian framework with block-wise sampling, the estimation of risk prices parameters—conditional on the risk factors and the other parameters of these models—remains a simple RVAR problem. I leave these extensions for future work.

An area of particular promise are macro-finance DTSMs, which include macro variables as risk factors. These models face the challenge of putting more structure on risk prices, since the number of parameters is large and the joint dynamics of term structure and macro variables are typically overfitted (Kim, 2007). In addition to the much-needed parsimony and more reliable term premia, my framework would deliver posterior inference on risk-price restrictions in macro-finance models that could help answer the questions about which macroeconomic variables drive variation in term premia and which macroeconomic shocks carry risk. These

are among the most pressing questions in macro-finance.

References

Ang, Andrew, Jean Boivin, Sen Dong, and Rudy Loo-Kung (2011) "Monetary Policy Shifts and the Term Structure," *Review of Economic Studies*, Vol. 78, pp. 429–457.

Ang, Andrew, Sen Dong, and Monika Piazzesi (2007) "No-Arbitrage Taylor Rules," NBER Working Paper 13448, National Bureau of Economic Research.

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

Backus, D.K. and J.H. Wright (2007) "Cracking the conundrum," *Brookings Papers on Economic Activity*, Vol. 2007, pp. 293–329.

Bartlett, M. S. (1957) "A comment on D. V. Lindley's statistical paradox," *Biometrika*, Vol. 44, pp. 533–534.

Bauer, Michael D. and Christopher J. Neely (2014) "International channels of the Fed's unconventional monetary policy," *Journal of International Money and Finance*, Vol. 44, pp. 24–46.

Bauer, Michael D. and Glenn D. Rudebusch (2013a) "Monetary Policy Expectations at the Zero Lower Bound," Working Paper 2013-18, Federal Reserve Bank of San Francisco.

——— (2013b) "What Caused the Decline in Long-term Yields?" *FRBSF Economic Letter*, Vol. 19.

Bauer, Michael D., Glenn D. Rudebusch, and Jing Cynthia Wu (2012) "Correcting Estimation Bias in Dynamic Term Structure Models," *Journal of Business and Economic Statistics*, Vol. 30, pp. 454–467.

——— (2014) "Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset: Comment," *American Economic Review*, Vol. 104, pp. 1–16.

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

Carlin, Bradley P. and Siddartha Chib (1995) "Bayesian Model Choice via Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 57, pp. 473–484.

Chib, S. and B. Ergashev (2009) "Analysis of Multifactor Affine Yield Curve Models," *Journal of the American Statistical Association*, Vol. 104, pp. 1324–1337.

Chib, Siddhartha and Edward Greenberg (1994) "Bayes Inference in Regression Models With ARMA(p,q) Errors," *Journal of Econometrics*, Vol. 64, pp. 183–206.

Chib, Siddhartha and Kyu Ho Kang (2014) "Change-Points in Affine Arbitrage-Free Term Structure Models," *Journal of Financial Econometrics*, Vol. 12, pp. 237–277.

Chipman, Hugh, Edward I. George, and Robert E. McCulloch (2001) "The Practical Implementation of Bayesian Model Selection," in P. Lahiri ed. *IMS Lecture Notes – Monograph Series*, Vol. 38: Institute of Mathematical Statistics, pp. 65–116.

Clyde, Merlise and Edward I. George (2004) "Model Uncertainty," *Statistical Science*, Vol. 19, pp. 81–94.

Cochrane, John (2007) "Commentary on "Macroeconomic implications of changes in the term premium"," *Federal Reserve Bank of St. Louis Review*, pp. 271–282.

Cochrane, John H. and Monika Piazzesi (2005) "Bond Risk Premia," *American Economic Review*, Vol. 95, pp. 138–160.

——— (2008) "Decomposing the Yield Curve," unpublished manuscript, Chicago Booth School of Business.

Creal, Drew D. and Jing Cynthia Wu (2015) "Estimation of affine term structure models with spanned or unspanned stochastic volatility," *Journal of Econometrics*, Vol. 185, pp. 60–81.

Dai, Qiang and Kenneth J. Singleton (2000) "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, Vol. 55, pp. 1943–1978.

——— (2002) "Expectation puzzles, time-varying risk premia, and affine models of the term structure," *Journal of Financial Economics*, Vol. 63, pp. 415–441.

Dellaportas, Petros and Jonathan J. Forster (1999) "Markov chain Monte Carlo model determination for hierarchical and graphical log-linear models," *Biometrika*, Vol. 86, pp. 615–633.

Dellaportas, Petros, Jonathan J. Forster, and Ioannis Ntzoufras (2002) "On Bayesian model and variable selection using MCMC," *Statistics and Computing*, Vol. 12, pp. 27–36.

Duffee, Gregory R. (2002) "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, Vol. 57, pp. 405–443.

——— (2011a) "Forecasting with the Term Structure: the Role of No-Arbitrage," Working Paper January, Johns Hopkins University.

——— (2011b) "Information In (and Not In) the Term Structure," *Review of Financial Studies*, Vol. 24, pp. 2895–2934.

Duffee, Gregory R. and Richard H. Stanton (2012) "Estimation of Dynamic Term Structure Models," *Quarterly Journal of Finance*, Vol. 2.

Fama, Eugene F. and Robert R. Bliss (1987) "The Information in Long-Maturity Forward Rates," *The American Economic Review*, Vol. 77, pp. 680–692.

Gamerman, Dani and Hedibert F. Lopes (2006) *Markov Chain Monte Carlo*: Chapman & Hall/CRC, 2nd edition.

George, Edward I. and Robert E. McCulloch (1993) "Variable Selection Via Gibbs Sampling," *Journal of the American Statistical Association*, Vol. 88, pp. 881–889.

George, Edward I., Dongchu Sun, and Shawn Ni (2008) "Bayesian stochastic search for VAR model restrictions," *Journal of Econometrics*, Vol. 142, pp. 553–580.

Geweke, John (1996) "Variable selection and model comparison in regression," in J. M. Bernando, J. O. Berger, A. P. Daviw, and A. F. M. Smith eds. *Bayesian Statistics 5*: Oxford University Press, pp. 609–620.

Godsill, Simon J. (2001) "On the Relationship Between Markov Chain Monte Carlo Methods for Model Uncertainty," *Journal of COmputational and Graphical Statistics*, Vol. 10, pp. 230–248.

Goodsill, Simon J. (2001) "On the Relationship Between Markov Chain Monte Carlo Methods for Model Uncertainty," *Journal of Computational and Graphical Statistics*, Vol. 10, pp. 230–248.

Green, Peter J. (1995) "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, Vol. 82, pp. 711–732.

Greenspan, Alan (2005) "Semiannual Monetary Policy Report to Congress," January 16.

Gürkaynak, Refet S., Brian P. Sack, and Eric T. Swanson (2005) "The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models," *American Economic Review*, Vol. 95, pp. 425–436.

Hamilton, James D. and Jing Cynthia Wu (2012) "Identification and estimation of Gaussian affine term structure models," *Journal of Econometrics*, Vol. 168, pp. 315–331.

Joslin, Scott, Anh Le, and Kenneth J. Singleton (2013) "Why Gaussian Macro-Finance Term Structure Models Are (Nearly) Unconstrained Factor-VARs," *Journal of Financial Economics*, Vol. 109, pp. 604–622.

Joslin, Scott, Marcel Priebsch, and Kenneth J. Singleton (2014) "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," *Journal of Finance*, Vol. 69, p. 11971233.

Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu (2011) "A New Perspective on Gaussian Dynamic Term Structure Models," *Review of Financial Studies*, Vol. 24, pp. 926–970.

Kass, Robert E. and Adrian E. Raftery (1995) "Bayes Factors," *Journal of the American Statistical Association*, Vol. 90, pp. 81–94.

Kim, Don H. (2007) "Challenges in macro-finance modeling," BIS Working Papers 240, Bank for International Settlements.

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

Kim, Don H. and Kenneth J. Singleton (2012) "Term Structure Models and the Zero Bound: An Empirical Investifation of Japanese Yields," *Journal of Econometrics*, Vol. 170, pp. 32–49.

Kim, Don H. and Jonathan H. Wright (2005) "An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates," Finance and Economics Discussion Series 2005-33, Federal Reserve Board of Governors.

Kozicki, S. and P.A. Tinsley (2001) "Shifting endpoints in the term structure of interest rates," *Journal of Monetary Economics*, Vol. 47, pp. 613–652.

Kuo, Lynn and Bani Mallick (1998) "Variable selection for regression models," *Sankhya Ser. B*, Vol. 60, pp. 65–81.

Litterman, Robert and J. Scheinkman (1991) "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, Vol. 1, pp. 54–61.

Ludvigson, Sydney C. and Serena Ng (2009) "Macro Factors in Bond Risk Premia," *Review of Financial Studies*, Vol. 22, pp. 5027–5067.

Lütkepohl, Helmut (2006) *New introduction to multiple time series analysis*: Springer Verlag.

O'Hara, Robert B. and Mikko J. Sillanpää (2009) "A Review of Bayesian Variable Selection Methods: What, How and Which," *Bayesian Analysis*, Vol. 4, pp. 85–118.

Rudebusch, Glenn (2007) "Commentary on "Cracking the Conundrum"," *Brookings Papers on Economic Activity*, Vol. 38, pp. 317–325.

Singleton, Kenneth J. (2006) *Empirical Dynamic Asset Pricing*: Princeton University Press.

Sisson, Scott A. (2005) "Transdimensional Markov Chains: A Decade of Progress and Future Perspectives," *Journal of the American Statistical Association*, Vol. 100, pp. 1077–1090.

Stambaugh, Robert F. (1999) "Predictive regressions," *Journal of Financial Economics*, Vol. 54, pp. 375–421.

Zellner, Arnold (1962) "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, Vol. 57, pp. 348–368.

Zellner, Arnold and Tomohiro Ando (2010) "A direct Monte Carlo approach for Bayesian analysis of the seemingly unrelated regression model," *Journal of Econometrics*, Vol. 159, pp. 33–45.

A Change of measure

To show what kind of process the term structure factors follow under $\mathbb{Q}$, I derive the conditional Laplace transform of X_{t+1} under $\mathbb{Q}$. I define the one-period stochastic discount factor (pricing kernel) as

$$M_{t+1} = \exp\left(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\varepsilon_{t+1}\right).$$

For any one-period pricing kernel the change of measure is implied by

$$M_{t+1} = \exp(-r_t) f^{\mathbb{Q}}(X_{t+1}|X_t)/f^{\mathbb{P}}(X_{t+1}|X_t).$$

Note that the Radon-Nikodym derivative, which relates the densities under the physical and risk-neutral measure, is given by

$$\frac{f^{\mathbb{P}}(X_{t+1}|X_t)}{f^{\mathbb{Q}}(X_{t+1}|X_t)} = \left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)(X_{t+1};\lambda_t) = \exp\left(\frac{1}{2}\lambda_t'\lambda_t + \lambda_t'\varepsilon_{t+1}\right).$$

I obtain for the risk-neutral conditional Laplace transform

$$\begin{aligned}
E^{\mathbb{Q}}(\exp(u'X_{t+1})|X_t) &= \int \exp(u'X_{t+1}) f^{\mathbb{Q}}(X_{t+1}|X_t) dX_{t+1} \\
&= \int \exp\left(u'X_{t+1} - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\varepsilon_{t+1}\right) f^{\mathbb{P}}(X_{t+1}|X_t) dX_{t+1} \\
&= E\left[\exp\left(u'(\mu + \Phi X_t + \Sigma\varepsilon_{t+1}) - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\varepsilon_{t+1}\right)|X_t\right] \\
&= \exp\left[u'(\mu - \Sigma\lambda_t + \Phi X_t) + \frac{1}{2}u'\Sigma\Sigma'u\right]
\end{aligned}$$

which is recognized as the conditional moment-generating function of a multivariate normal distribution with mean $\mu - \Sigma\lambda_t + \Phi X_t = (\mu - \lambda_0) + (\Phi - \lambda_1)X_t$ and variance $\Sigma\Sigma'$.

The physical innovations ε_t, which are a vector martingale-difference sequence (m.d.s.) under $\mathbb{P}$, are related to the innovations under $\mathbb{Q}$ by

$$\varepsilon_t^{\mathbb{Q}} = \varepsilon_t + \lambda_{t-1}. \tag{13}$$

Note that the risk-neutral innovations, while being m.d.s. under $\mathbb{Q}$, can have non-zero mean and be predictable under $\mathbb{P}$, depending on the risk price specification.

B Affine bond pricing

Under the assumptions of Section 2.1, model-implied bond prices are exponentially affine functions of the pricing factors:

$$\hat{P}_t^m = e^{\mathcal{A}_m + \mathcal{B}_m X_t},$$

and the loadings $\mathcal{A}_m = \mathcal{A}_m(\mu^{\mathbb{Q}}, \phi^{\mathbb{Q}}, \delta_0, \delta_1, \Sigma)$ and $\mathcal{B}_m = \mathcal{B}_m(\phi^{\mathbb{Q}}, \delta_1)$ follow the recursions

$$\begin{aligned} \mathcal{A}_{m+1} &= \mathcal{A}_m + (\mu^{\mathbb{Q}})'\mathcal{B}_m + \frac{1}{2}\mathcal{B}_m'\Sigma\Sigma'\mathcal{B}_m - \delta_0 \\ \mathcal{B}_{m+1} &= (\phi^{\mathbb{Q}})'\mathcal{B}_m - \delta_1 \end{aligned}$$

with starting values $\mathcal{A}_0 = 0$ and $\mathcal{B}_0 = 0$. Model-implied yields are determined by $\hat{y}_t^m = -m^{-1}\log P_t^m = A_m + B_m X_t$, with $A_m = -m^{-1}\mathcal{A}_m$ and $B_m = -m^{-1}\mathcal{B}_m$.

Risk-neutral yields, the yields that would prevail if investors were risk-neutral, can be calculated using

$$\tilde{y}_t^m = \tilde{A}_m + \tilde{B}_m X_t, \quad \tilde{A}_m = -m^{-1}\mathcal{A}_m(\mu, \phi, \delta_0, \delta_1, \Sigma), \quad \tilde{B}_m = -m^{-1}\mathcal{B}_m(\phi, \delta_1).$$

Risk-neutral yields reflect policy expectations over the life of the bond, $m^{-1}\sum_{h=0}^{m-1} E_t r_{t+h}$, plus a convexity term. The yield term premium is defined as the difference between actual and risk-neutral yields, $ytp_t^m = y_t^m - \tilde{y}_t^m$.

C Risk prices as parameters in a restricted VAR

Estimation of λ, for given values of all other parameters, amounts to estimation of the VAR in (1) subject to the linear constraints in (7). This section lays out the specifics: how to obtain the least squares estimates of λ, which maximize the likelihood for given values of the other parameters (C.1), how to carry out Bayesian inference about λ using the exact conditional posterior (C.2), and how to specify a g-prior in the RVAR context (C.3).

An alternative to RVAR estimation for obtaining estimates of λ is often possible. Subtracting $\mathbb{Q}$-measure expectations from both sides of the VAR in equation (1), the dynamic system becomes

$$X_t - E_{t-1}^{\mathbb{Q}} X_t = \lambda_0 + \lambda_1 X_{t-1} + \Sigma\varepsilon_t,$$

which is a system of Seemingly Unrelated Regressions (SUR) (Zellner, 1962). If the DTSM is estimated using frequentist methods, this formulation can be used to concentrate out λ_0 and λ_1 from the likelihood function—this approach is used, for example, by Joslin et al.

(2014). If the estimation is Bayesian, the conditional posterior for λ can be easily derived (see Zellner and Ando, 2010, and the references therein). For those cases where all rows of (λ_0, λ_1) have at least one non-zero element, the SUR approach is identical to the restricted-VAR approach. However, when one or more rows contain only zeros, the SUR approach is not applicable, because it effectively ignores the equation where all parameters are restricted to zero, whereas the restricted-VAR approach correctly accounts for the inter-equation dependencies of the residuals.

C.1 Likelihood function and least squares estimates

Lütkepohl (2006) describes frequentist estimation of restricted VARs (Section 5.2). Here, I show how inference about λ in a DTSM maps into this context. The VAR in equation (1) can be written in full-data matrix notation as $X = BZ + U$, where $X = (X_1, \ldots, X_T)$, $Z = (Z_0, \ldots, Z_{T-1})$, $Z_t = (1, X_t')'$, $B = (\mu, \Phi)$, and $U = (u_1, \ldots, u_T)$, $u_t = \Sigma\varepsilon_t$. The linear constraints are

$$\beta := vec(B) = \lambda + r = S\lambda_\gamma + r,$$

where S is a known $N(N+1) \times a$ matrix of zeros and ones, λ_γ is an a-vector with the unrestricted elements of λ (according to γ), and $r = vec(\mu^{\mathbb{Q}}, \Phi^{\mathbb{Q}})$. To clarify: λ_γ contains the a unrestricted risk prices, λ contains these as well as $N(N+1) - a$ zeros, and S transforms one into the other ($\lambda = S\lambda_\gamma$). If there are no zero restrictions on λ—as in the case of a maximally-flexible DTSM, and for the SSVS model selection approach—we have $S = I_{N(N+1)}$, $\lambda = \lambda_\gamma$, and there are $a = N(N+1)$ risk prices to estimate.

Plugging in the restrictions and vectorizing the VAR equation we have

$$\begin{aligned} x := vec(X) &= vec(BZ) + vec(U) \\ &= (Z' \otimes I_N)\beta + u \\ &= (Z' \otimes I_N)(S\lambda_\gamma + r) + u \\ z := x - (Z' \otimes I_N)r &= (Z' \otimes I_N)S\lambda_\gamma + u. \end{aligned}$$

The likelihood for z is $N((Z' \otimes I_N)S\lambda_\gamma, I_T \otimes \Omega)$, with $\Omega = \Sigma\Sigma'$. This likelihood is maximized at the generalized least-squares estimate (see Lütkepohl, 2006, eq. 5.2.6) which is given by

$$\hat{\lambda}_\gamma = \left[S'(ZZ' \otimes \Omega^{-1})S\right]^{-1} S'(Z \otimes \Omega^{-1})z. \tag{14}$$

Its estimated covariance matrix is

$$\hat{V}_\gamma = \left[S'(ZZ' \otimes \Omega^{-1})S\right]^{-1}. \tag{15}$$

C.2 Bayesian inference in a restricted VAR

If we assume a natural conjugate prior for λ, which is a normal prior denoted by $N(\underline{\lambda}_\gamma, \underline{V}_\gamma)$, the conditional posterior can be derived in closed form:

$$\begin{aligned}
P(\lambda_\gamma|z,\ldots) &\propto \exp\left\{-.5\left[(\lambda_\gamma - \underline{\lambda}_\gamma)'\underline{V}_\gamma^{-1}(\lambda_\gamma - \underline{\lambda}_\gamma)\right.\right. \\
&\quad \left.\left. +(z-(Z'\otimes I_N)S\lambda_\gamma)'(I_T\otimes\Omega)^{-1}(z-(Z'\otimes I_N)S\lambda_\gamma)\right]\right\} \\
&\propto \exp\{-.5[\lambda_\gamma'\underline{V}^{-1}\lambda_\gamma - 2\underline{\lambda}_\gamma\underline{V}^{-1}\lambda_\gamma + \lambda_\gamma' S'(ZZ'\otimes\Omega^{-1})S\lambda_\gamma - 2z'(Z'\otimes\Omega^{-1})S\lambda_\gamma]\} \\
&\propto \exp\left\{-.5\left[\lambda_\gamma'(\underline{V}^{-1} + S'(ZZ'\otimes\Omega^{-1})S)\lambda - 2(\underline{\lambda}_\gamma'\underline{V}^{-1} + z'(Z'\otimes\Omega^{-1})S)\lambda_\gamma\right]\right\}.
\end{aligned}$$

The last expression is the Kernel of a Normal distribution with covariance matrix

$$\overline{V}_\gamma = (\underline{V}_\gamma^{-1} + S'(ZZ'\otimes\Omega^{-1})S)^{-1}$$

and mean

$$\overline{\lambda}_\gamma = \overline{V}_\gamma(\underline{V}_\gamma^{-1}\underline{\lambda}_\gamma + S'(Z\otimes\Omega^{-1})z).$$

The (conditional) posterior mean can also be written as

$$\overline{\lambda}_\gamma = \overline{V}_\gamma\left\{\underline{V}_\gamma^{-1}\underline{\lambda}_\gamma + \hat{V}_\gamma^{-1}\hat{\lambda}_\gamma\right\}.$$

These results will be used for drawing λ from its conditional posterior distribution, i.e., using a Gibbs step, in the various MCMC samplers used in this paper.

C.3 g-prior in for a restricted VAR

A g-prior for the parameters of a restricted VAR has covariance matrix

$$\underline{V} = g\left[S'(ZZ'\otimes\Omega^{-1})S\right]^{-1} = g\hat{V}$$

so that the posterior mean becomes

$$\overline{\lambda} = \frac{1}{1+g}\underline{\lambda} + \frac{g}{1+g}\hat{\lambda}$$

and the posterior covariance is

$$\overline{V} = \frac{g}{1+g}\left[S'(ZZ' \otimes \Omega^{-1})S\right]^{-1}.$$

The prior used for λ in this paper is an "orthogonalized g-prior" with zero mean. In particular, I calculate $[S'(ZZ' \otimes \Omega^{-1})S]^{-1}$ for the model without restrictions on λ—in which case $S = I_{N(N+1)}$—at the MLE estimates of the remaining parameters (the $\mathbb{Q}$-parameters are needed to calculate Z, and Σ is needed to calculate Ω), and use the diagonal values of the resulting matrix, multiplied by g, as the prior variances.

D Alternative model-selection samplers

This appendix describes two model-selection samplers that can be used as alternatives to the GVS sampler of Section 2.6. In the paper—see Sections 3 and 4—they are used to ensure the robustness of the model-selection results to different sampling algorithms.

D.1 Stochastic Search Variable Selection

Stochastic Search Variable Selection (George and McCulloch, 1993, SSVS) was the first MCMC sampling approach to variable selection. It was developed by George and McCulloch (1993) and has since been applied extensively, including for Bayesian VAR estimation (George et al., 2008). The idea here is that those parameters that are excluded from the model are taken to be from a prior that is a tight distribution around zero.

Formally, the parameter prior is specified as the following normal mixture: $P(\lambda_i|\gamma_i) = (1-\gamma_i)N(0, \tau_{0i}^2) + \gamma_i N(0, \tau_{1i}^2)$. This can be obtained from the multivariate normal prior $P(\lambda|\gamma) = N(0, D_\gamma R D_\gamma)$, where D_γ is a diagonal matrix with elements $(1-\gamma_i)\tau_{0i} + \gamma_i\tau_{1i}$, and R is a correlation matrix, here equal to the identity matrix due to prior conditional independence. Intuitively, for included elements the prior variance, τ_{1i}^2, is chosen large so that the posterior estimate is informed by the data, while for excluded elements the prior variance, τ_{0i}^2, is set to a small value, so that posterior estimates are close to zero. Following common practice, I choose these variances to be multiples of the variance of the least-squares estimates for the unrestricted model: $\tau_{0i} = c_0\hat{\sigma}_{\lambda_i}$ and $\tau_{1i} = c_1\hat{\sigma}_{\lambda_i}$, where c_0 and c_1 are tuning parameters. Specifically, I will set $c_1 = \sqrt{g}$ (in line with the prior specification for the other samplers), and $c_0 = 1/c_1$. These choices leads to good convergence speed).

In iteration j, first λ is drawn conditional on the values for all other parameters from iteration $j-1$. All elements of λ are drawn simultaneously in a Gibbs step from the normal

conditional posterior distribution, whose moments are given in Appendix C.2. Note that the prior covariance matrix is $D_\gamma R D_\gamma$ and no zeros are imposed on λ.

The elements of γ are drawn successively, in random order, from their conditional posterior distributions. Denote by $\gamma_i^{(j)}$ the element that is sampled, and by $\gamma_{-i}^{(j)}$ all remaining elements. The success probability of the Bernoulli conditional posterior is determined by:

$$\begin{aligned}\frac{P(\gamma_i^{(j)}=1|\lambda^{(j)},\theta_-^{(j-1)},\gamma_{-i}^{(j)})}{P(\gamma_i^{(j)}=0|\lambda^{(j)},\theta_-^{(j-1)},\gamma_{-i}^{(j)})} &= \frac{P(\lambda^{(j)}|\gamma_i^{(j)}=1,\gamma_{-i}^{(j)},\theta_-^{(j-1)})}{P(\lambda^{(j)}|\gamma_i^{(j)}=0,\gamma_{-i}^{(j)},\theta_-^{(j-1)})}\cdot\frac{P(\gamma_i^{(j)}=1)}{P(\gamma_i^{(j)}=0)}\\ &= \frac{(\tau_{1i})^{-1}\exp(-.5(\lambda_i^{(j)}/\tau_{1i})^2)}{(\tau_{0i})^{-1}\exp(-.5(\lambda_i^{(j)}/\tau_{0i})^2)},\end{aligned}$$

where the second line follows from the prior conditional independence of the elements of λ and the equal prior model probabilities. This ratio does not depend on the data because in the SSVS approach, γ affects the likelihood only through λ.

D.2 Reversible-jump Markov chain Monte Carlo

The Reversible-jump Markov chain Monte Carlo (RJMCMC) sampler developed by Green (1995) allows moves between parameter spaces of different dimensionality, and therefore can deal with the setting in which the prior of the excluded elements of λ is simply a point mass at zero. This approach is extremely flexible, allowing for different types of moves between models. I adapt the local reversible-jump sampler of Dellaportas and Forster (1999), where only local moves are considered. Denote the current state of the chain by $(\lambda^{(j)},\gamma^{(j)},\theta_-^{(j)})$. First the decision is made between a "null move" (a within-model move) and a "jump move" (a between-model move). With 25% probability a null move is undertaken, in which γ remains unchanged and λ is updated using a Gibbs step. If a jump move is attempted, a proposal for (λ',γ') is constructed by *changing only one parameter*. The index of the element to be changed, i, is randomly chosen. If $\gamma_i^{(j)}=0$, then the element is added to the model, otherwise it is deleted. When the move involves adding a parameter to the model, the proposal is $\lambda'=g(\lambda^{(j)},u)$, where u is a scalar drawn from the proposal density $q_i(u)$, taken to be $N(\mu_i,\sigma_i^2)$, and $g(\cdot)$ is the identity transformation (such that $\lambda_i'=u$). The acceptance probability is

$$\begin{aligned}\alpha(\lambda^{(j)},\gamma^{(j)},\theta_-^{(j)},\lambda',\gamma') &= \frac{P(Y|\lambda',\gamma',\theta_-^{(j)})}{P(Y|\lambda^{(j)},\gamma^{(j)},\theta_-^{(j)})}\frac{P(\lambda'|\gamma')}{P(\lambda^{(j)}|\gamma^{(j)})}\frac{P(\gamma')}{P(\gamma^{(j)})}\frac{1}{q_i(u)}\\ &= \frac{P(Y|\lambda',\gamma',\theta_-^{(j)})}{P(Y|\lambda^{(j)},\gamma^{(j)},\theta_-^{(j)})}\frac{v_i^{1/2}\exp(-u^2/v_i)}{\sigma_i\exp(-(u-\mu_i)^2/\sigma_i^2)}\end{aligned}$$

where the second line follows from prior conditional independence and equal prior model probability.[37] When an element is deleted from the model, we have $(\lambda', u') = g^{(-1)}(\lambda^{(j)})$, meaning that λ' now has a zero at the i-th position, and $\lambda_i^{(j)} = u'$. In this case,

$$\alpha(\lambda^{(j)}, \gamma^{(j)}, \theta_-^{(j)}, \lambda', \gamma') = \frac{P(Y|\lambda', \gamma', \theta_-^{(j)})}{P(Y|\lambda^{(j)}, \gamma^{(j)}, \theta_-^{(j)})} \frac{\sigma_i \exp(-(u' - \mu_i)^2/\sigma_i^2)}{v_i^{1/2} \exp(-u'^2/v_i)}.$$

The choice of the proposal distributions plays an important role for the efficiency of the sampler. Ideally, they should be close to the conditional posteriors (Goodsill, 2001, Sec. 2.3.2). Here, the parameter conditionals for each element of λ are in fact available in closed form, based on standard results.[38] I use the moments of these conditional posterior distributions, calculated anew in each iteration for the chosen element i, for μ_i and σ_i^2. The resulting sampler achieves good efficiency in that it jumps between models frequently.

E Excess bond returns

The continuously compounded return of holding an n-period bond for h periods, in excess of the risk-free rate, is

$$rx_{t,t+h}^{(n)} = -(n-h)y_{t+h}^{n-h} + ny_t^n - hy_t^h.$$

For fitted excess returns, calculated using model-implied yields, we have

$$\begin{aligned} \hat{rx}_{t,t+h}^{(n)} &= -(n-h)\hat{y}_{t+h}^{n-h} + n\hat{y}_t^n - h\hat{y}_t^h \\ &= \mathcal{A}_{n-h} - \mathcal{A}_n + \mathcal{A}_h + \mathcal{B}_{n-h}' X_{t+h} - (\mathcal{B}_n - \mathcal{B}_h)' X_t. \end{aligned}$$

Fitted excess returns are generally close to observed excess returns, because of the accurate cross-sectional yield fit of the models considered in this paper.

The time-t model-implied expected excess return is

$$\begin{aligned} E_t \hat{rx}_{t,t+h}^{(n)} &= \mathcal{A}_{n-h} - \mathcal{A}_n + \mathcal{A}_h + \mathcal{B}_{n-h}' E_t X_{t+h} - (\mathcal{B}_n - \mathcal{B}_h)' X_t \\ &= \mathcal{A}_{n-h} - \mathcal{A}_n + \mathcal{A}_h + \mathcal{B}_{n-h}' [(I_N - \Phi^h) E(X_t) + \Phi^h X_t] - (\mathcal{B}_n - \mathcal{B}_h)' X_t \\ &= \mathcal{A}_{n-h} - \mathcal{A}_n + \mathcal{A}_h + (\mathcal{B}_{n-h}' - \mathcal{B}_n' + \mathcal{B}_h') E(X_t) + (\mathcal{B}_{n-h}' \Phi^h - \mathcal{B}_n' + \mathcal{B}_h')(X_t - E(X_t)), \end{aligned}$$

[37] As for GVS, the ratio of the likelihoods only requires calculating the $\mathbb{P}$-likelihood.

[38] Consider the relevant equation in the system $X_t - E_{t-1}^{\mathbb{Q}} X_t = \lambda_0 + \lambda_1 X_{t-1} + \Sigma \varepsilon_t$. After subtracting all terms not involving the parameter of interest λ_i, the conditional posterior moments for λ_i follow from standard results for univariate Bayesian regression—see also Geweke (1996) and Kuo and Mallick (1998).

and the surprise component of the excess return is

$$\hat{rx}_{t,t+h}^{(n)} - E_t\hat{rx}_{t,t+h}^{(n)} = \mathcal{B}'_{n-h}(X_{t+h} - E_tX_{t+h}) = \mathcal{B}'_{n-h}\sum_{i=1}^{h}\Phi^{h-i}\Sigma\varepsilon_{t+i},$$

see also equation (11) of Duffee (2011b).

To gain some intuition about the model's implications for returns, and to see where equation (5) derives from, consider the one-period excess return, which can be written as follows:

$$\hat{rx}_{t,t+1}^{(n)} = -\frac{1}{2}\mathcal{B}'_{n-1}\Sigma\Sigma'\mathcal{B}_{n-1} + \mathcal{B}'_{n-1}\Sigma\lambda_t + \mathcal{B}'_{n-1}\Sigma\varepsilon_{t+1}. \tag{16}$$

The first term, which corresponds to $\frac{1}{2}Var_t(\hat{rx}_{t,t+1}^{(n)})$, is due to convexity. The second term captures the actual risk compensation for level, slope, and curvature risk. This can be seen by rewriting it as $\lambda'_tCov_t(\varepsilon_{t+1}, \hat{rx}_{t,t+1}^{(n)})$, the product of the prices of risk and the quantities of risk. In a Gaussian model, quantities of risk are constant, and time-variation in expected returns is due exclusively to movements in λ_t. The third term in equation (16) captures the surprise component of the excess return.

Consider now the predictability regression in (12), for which we want to derive the model-implied R^2. Because the regressors correspond to the risk factors in the DTSMs, the population R^2 is equal to the variability of model-implied expected excess returns relative to the variability of the model-implied realized excess returns, i.e.,

$$R^2_{pop} = \frac{Var(E_t\hat{rx}_{t,t+h}^{(n)})}{Var(\hat{rx}_{t,t+h}^{(n)})}.$$

The variance of expected excess returns is given by

$$Var(E_t\hat{rx}_{t,t+h}^{(n)}) = (\mathcal{B}'_{n-h}\Phi^h - \mathcal{B}'_n + \mathcal{B}'_h)Var(X_t)(\mathcal{B}'_{n-h}\Phi^h - \mathcal{B}'_n + \mathcal{B}'_h)'.$$

The unconditional covariance matrix is calculated using $Var(X_t) = (I_{N^2} - \Phi\otimes\Phi)^{-1}vec(\Sigma\Sigma')$. The variance of realized excess returns is

$$Var(\hat{rx}_{t,t+h}^{(n)}) = Var(E_t\hat{rx}_{t,t+h}^{h}) + \mathcal{B}'_{n-h}\left(\sum_{i=1}^{h}\Phi^{h-i}\Sigma\Sigma'(\Phi^{h-i})'\right)\mathcal{B}_{n-h}.$$

Notably, yield loadings are very similar across models, and the differences in model-implied return predictability are due to differences in Φ.

Table 1: Simulation study: risk-price restrictions and parameters

		Element of λ						Frequency
		(1)	(2)	(3)	(4)	(5)	(6)	corr. model
DGP	γ_o	0	0	0	0	1	0	
	λ_o	0.0000	0.0000	0.0000	0.0000	-0.0236	0.0000	
MCMC	$\bar{\lambda}$	-0.0467	0.0578	-0.0231	-0.0056	-0.0216	-0.0208	6%
	freq. sig.	22%	18%	36%	10%	22%	16%	
SSVS	$\bar{\gamma}$	0.4599	0.0489	0.3040	0.1095	0.5840	0.0628	54%
	$\bar{\lambda}$	-0.0603	0.0013	-0.0126	-0.0009	-0.0120	-0.0004	[32%]
	$\bar{\lambda}\|\gamma = \gamma_o$	-0.0001	0.0005	-0.0002	0.0000	-0.0249	0.0000	
GVS	$\bar{\gamma}$	0.4666	0.0556	0.2979	0.1123	0.5979	0.0724	54%
	$\bar{\lambda}$	-0.0602	0.0010	-0.0126	-0.0010	-0.0122	-0.0006	[33%]
	$\bar{\lambda}\|\gamma = \gamma_o$	0.0000	0.0000	0.0000	0.0000	-0.0246	0.0000	
RJMCMC	$\bar{\gamma}$	0.4576	0.0573	0.2426	0.0923	0.5825	0.0523	58%
	$\bar{\lambda}$	-0.0426	0.0021	-0.0067	-0.0003	-0.0148	-0.0007	[35%]
	$\bar{\lambda}\|\gamma = \gamma_o$	0.0000	0.0000	0.0000	0.0000	-0.0248	0.0000	

Risk-price specification of the data-generating process (DGP), and estimation results in simulated data. For MCMC (estimation of unrestricted model) the first row shows the average posterior mean estimates, and the second row shows fraction of samples in which the 95%-credibility interval for the element of λ did not straddle zero. For the model-selection samplers, the first row reports average estimated model indicators γ, the second row shows average parameter estimates, and the third row shows parameter estimates conditional on the true model indicators. Parameter estimates in the first two columns (λ_0) are multiplied by 1200 to convert to annualized percentage points. The last column shows the frequency—the fraction of simulated samples—that the correct model is chosen. For MCMC, the model choice is determined based on 95%-credibility intervals. For model selection samplers, the modal model is chosen. Numbers in squared brackets indicate the frequency that the model-selection samplers visited the true model, across all samples and iterations.

Table 2: Simulation study: persistence and volatilities

		Persistence		Volatilities		
		max. eigenv.	IRF(5y)	$\Delta\hat{f}_t$	$\Delta\tilde{f}_t$	Δftp_t
DGP		0.9951	0.7545	0.2574	0.2095	0.1342
MCMC	posterior mean	0.9790	0.2929	0.2581	0.0907	0.2209
	CI contains DGP	84%	62%	92%	64%	54%
SSVS	posterior mean	0.9874	0.5559	0.2574	0.1796	0.1427
	CI contains DGP	90%	92%	98%	94%	94%
GVS	posterior mean	0.9876	0.5588	0.2575	0.1796	0.1378
	CI contains DGP	90%	90%	96%	96%	94%
RJMCMC	posterior mean	0.9891	0.6022	0.2579	0.1919	0.1230
	CI contains DGP	92%	96%	96%	100%	100%

Persistence and volatilities (in population) implied by data-generating process (DGP) vs. estimates in simulated data. MCMC denotes estimation of unrestricted model. For each estimation method, the first row reports the posterior mean across all samples and iterations, and second row is the fraction of samples in which the 95%-credibility interval contains the true parameter. The persistence statistics are the maximum eigenvalue of Φ, and the impulse-response function (IRF) for the level factor in response to level shocks at the five-year horizon. The model-implied volatilities are for changes in five-to-ten year fitted forward rates, in risk-neutral forward rates (i.e., short-rate expectations), and in the forward term premium, and are reported in annualized percentage points.

Table 3: Estimates for unrestricted model

Parameter	Prior			Posterior			Acc.	Ineff.
$k_\infty^{\mathbb{Q}}$	0			**0.0108**			89.7	255.5
	(1.0000)			(0.0030)				
$\phi^{\mathbb{Q}}$	0.5000	0.5000	0.5000	**0.9989**	**0.9625**	**0.9382**	41.2	193.1
	(0.2887)	(0.2887)	(0.2887)	(0.0005)	(0.0045)	(0.0085)		
λ_0	0	0	0	0.1165	-0.0404	0.0328		1.8
	(1.0097)	(0.6277)	(0.1736)	(0.0983)	(0.0618)	(0.0175)		
λ_1	0	0	0	-0.0212	**-0.0557**	0.1749		1.8
	(0.1225)	(0.2184)	(2.2694)	(0.0115)	(0.0216)	(0.2289)		
	0	0	0	0.0036	0.0094	0.0540		
	(0.0762)	(0.1358)	(1.4109)	(0.0076)	(0.0126)	(0.1438)		
	0	0	0	0.0000	-0.0022	**-0.0984**		
	(0.0211)	(0.0376)	(0.3902)	(0.0021)	(0.0038)	(0.0399)		
Σ				**0.2706**	0	0	31.0	25.6
				(0.0134)				
				0.0719	**0.1524**	0		
				(0.0113)	(0.0078)			
				-0.0267	-0.0024	**0.0380**		
				(0.0030)	(0.0027)	(0.0019)		
σ_e				**0.0414**				1.7
				(0.0010)				

Prior and posterior means and standard deviations, as well as average acceptance rates, where applicable, and average inefficiency factors. Parameters for which 95%-credibility intervals do not straddle zero are shown in boldface. $k_\infty^{\mathbb{Q}}$, λ_0, Σ, and σ_e are scaled by 1200 to convert to annualized percentage points.

Table 4: risk-price restrictions

	SSVS			GVS			RJMCMC		
	Mean	MCSE	Ineff.	Mean	MCSE	Ineff.	Mean	MCSE	Ineff.
1	0.292	0.007	43.1	0.301	0.007	40.1	0.342	0.017	245.2
2	0.037	0.001	3.0	0.039	0.001	2.6	0.038	0.003	34.1
3	0.117	0.006	75.8	0.089	0.005	67.2	0.085	0.010	230.8
4	0.207	0.007	49.8	0.209	0.006	42.0	0.227	0.015	239.2
5	0.067	0.001	4.9	0.067	0.001	2.1	0.067	0.003	35.0
6	0.051	0.001	6.8	0.044	0.001	5.0	0.043	0.003	36.9
7	0.967	0.003	63.3	0.971	0.003	53.0	0.973	0.005	181.2
8	0.090	0.002	6.7	0.095	0.001	1.9	0.098	0.004	31.2
9	0.060	0.001	7.1	0.052	0.001	3.7	0.055	0.004	45.0
10	0.092	0.002	9.2	0.098	0.002	5.4	0.094	0.005	64.2
11	0.049	0.001	3.9	0.049	0.001	1.6	0.052	0.003	36.3
12	0.120	0.006	74.3	0.088	0.005	70.6	0.087	0.010	219.5

Results for γ for each sampling approach: the i-th row shows posterior means, Monte Carlo standard errors, and inefficiency factors for the sampled paths of the i-th element of γ. The first three rows correspond to the indicators restricting λ_0, and rows four to twelve correspond to those for λ_1, in column-major order.

Table 5: Posterior model probabilities

	SSVS		GVS		RJMCMC		AIC	SBIC
Model	Freq.	Odds	Freq.	Odds	Freq.	Odds		
7	0.3628	1.0	0.3748	1.0	0.3536	1.0	-25317.2	-25280.1
1,4,7	0.0856	4.2	0.0866	4.3	0.0967	3.7	-25322.8	-25279.0
1,7	0.0789	4.6	0.0845	4.4	0.0912	3.9	-25319.3	-25278.8
7,8	0.0343	10.6	0.0375	10.0	0.0391	9.1	-25317.5	-25277.0
3,7,12	0.0300	12.1	0.0193	19.4	0.0192	18.4	-25320.8	-25277.0
7,10	0.0293	12.4	0.0339	11.1	0.0290	12.2	-25316.6	-25276.1
5,7	0.0233	15.6	0.0239	15.7	0.0243	14.6	-25317.1	-25276.6
4,7	0.0183	19.8	0.0181	20.7	0.0165	21.4	-25315.3	-25274.8
7,9	0.0170	21.3	0.0158	23.7	0.0150	23.6	-25315.6	-25275.1
7,11	0.0153	23.8	0.0158	23.7	0.0137	25.8	-25316.2	-25275.7
6,7	0.0130	27.9	0.0126	29.7	0.0129	27.4	-25315.8	-25275.3
3,7	0.0127	28.6	0.0122	30.7	0.0114	30.9	-25316.6	-25276.1
7,12	0.0126	28.9	0.0104	35.9	0.0114	31.1	-25315.4	-25274.9
4	0.0108	33.6	0.0118	31.7	0.0112	31.7	-25310.0	-25272.8
2,7	0.0104	34.8	0.0115	32.5	0.0109	32.5	-25316.4	-25275.9
models visited	1021 / 4096 24.9 %		1006 / 4096 24.6 %		530 / 4096 12.9 %			

Posterior model probabilities for 15 most frequently visited models, approximated by the frequency of occurrence of each model in the sampler, and posterior odds ratios relative to the best model. Models are denoted by the indices of the unrestricted elements in λ.

Table 6: Model selection and prior dispersion

	Frequency	Models visited		Posterior	
g	$\mathcal{M}_1$	freq. $\geq 1\%$	total	$E(a)$	$P(a \leq 2)$
10,000	88.9%	3	116	1.1	99.6%
1,000	73.9%	9	333	1.3	94.3%
100	37.5%	15	1006	2.1	67.5%
50	21.6%	19	1389	2.8	45.5%
25	12.1%	18	1912	3.4	30.7%
10	6.2%	13	2370	4.4	13.7%
Prior	0.02%	–	–	6	0.3%

Model selection results for different levels of prior dispersion: frequencies of model $\mathcal{M}_1$, number of models that are visited with at least 1% relative frequency, number of models visited at least once, posterior mean number of unrestricted risk-price parameters, and posterior probability of at most two unrestricted risk-price parameters. a denotes the number of unrestricted risk-price parameters. Bottom row shows prior quantities for comparison.

Table 7: Persistence and volatility

Model	Max. eigenvalue		Volatilities		
	$\mathbb{Q}$	$\mathbb{P}$	$\Delta\hat{f}_t$	$\Delta\tilde{f}_t$	Δftp_t
M0	0.9989	0.9835	0.26	0.10	0.30
	[0.9979, 0.9998]	[0.9623, 0.9988]	[0.24, 0.29]	[0.01, 0.33]	[0.17, 0.52]
M1	0.9987	0.9924	0.26	0.22	0.22
	[0.9977, 0.9997]	[0.9868, 0.9957]	[0.24, 0.28]	[0.19, 0.26]	[0.14, 0.31]
M2	1.0000	0.9790	0.26	0.12	0.32
	[1.0000, 1.0000]	[0.9706, 0.9883]	[0.24, 0.29]	[0.05, 0.27]	[0.25, 0.42]
M3	0.9987	0.9892	0.26	0.32	0.40
	[0.9978, 0.9996]	[0.9865, 0.9938]	[0.24, 0.29]	[0.20, 0.49]	[0.20, 0.61]
BMA	0.9987	0.9903	0.26	0.22	0.27
	[0.9977, 0.9997]	[0.9739, 0.9978]	[0.24, 0.29]	[0.07, 0.42]	[0.14, 0.52]

Persistence of the dynamic system under the $\mathbb{Q}$-measure and the $\mathbb{P}$-measure, and volatility of monthly changes in five-to-ten-year fitted forward rates, risk-neutral forward rates (short rate expectations), and forward term premia. Posterior mean, median (in parentheses), and 95%-credibility intervals (in squared brackets).

Table 8: Historical changes in long-term rates and expectations

		Sample period			Conundrum period		
		1990	2007	Change	Jun-94	Jun-95	Change
Ten-year yield		8.5	4.6	-3.8	4.7	4.0	-0.8
Exp.	M0	4.4	3.7	-0.7	2.6	3.7	1.1
		[2.7, 6.3]	[1.8, 5.6]	[-2.7, 1.2]	[0.2, 4.4]	[1.5, 5.9]	[0.1, 2.6]
	M1	4.6	2.0	-2.7	1.2	1.8	0.6
		[3.2, 6.0]	[0.9, 2.9]	[-3.1, -2.3]	[-0.1, 2.5]	[1.0, 2.6]	[0.1, 1.1]
	M2	4.7	4.4	-0.4	3.5	4.6	1.1
		[3.9, 5.7]	[3.1, 5.4]	[-1.4, 0.2]	[2.2, 4.5]	[3.2, 5.8]	[0.5, 1.9]
	M3	5.3	3.5	-1.7	2.1	3.8	1.7
		[3.7, 6.8]	[1.6, 5.4]	[-2.8, -0.7]	[0.5, 3.7]	[1.5, 6.0]	[0.4, 3.0]
	BMA	4.7	2.5	-2.2	1.6	2.4	0.8
		[2.9, 6.5]	[0.6, 5.0]	[-3.3, -0.0]	[-0.6, 4.0]	[0.5, 5.4]	[-0.2, 2.4]

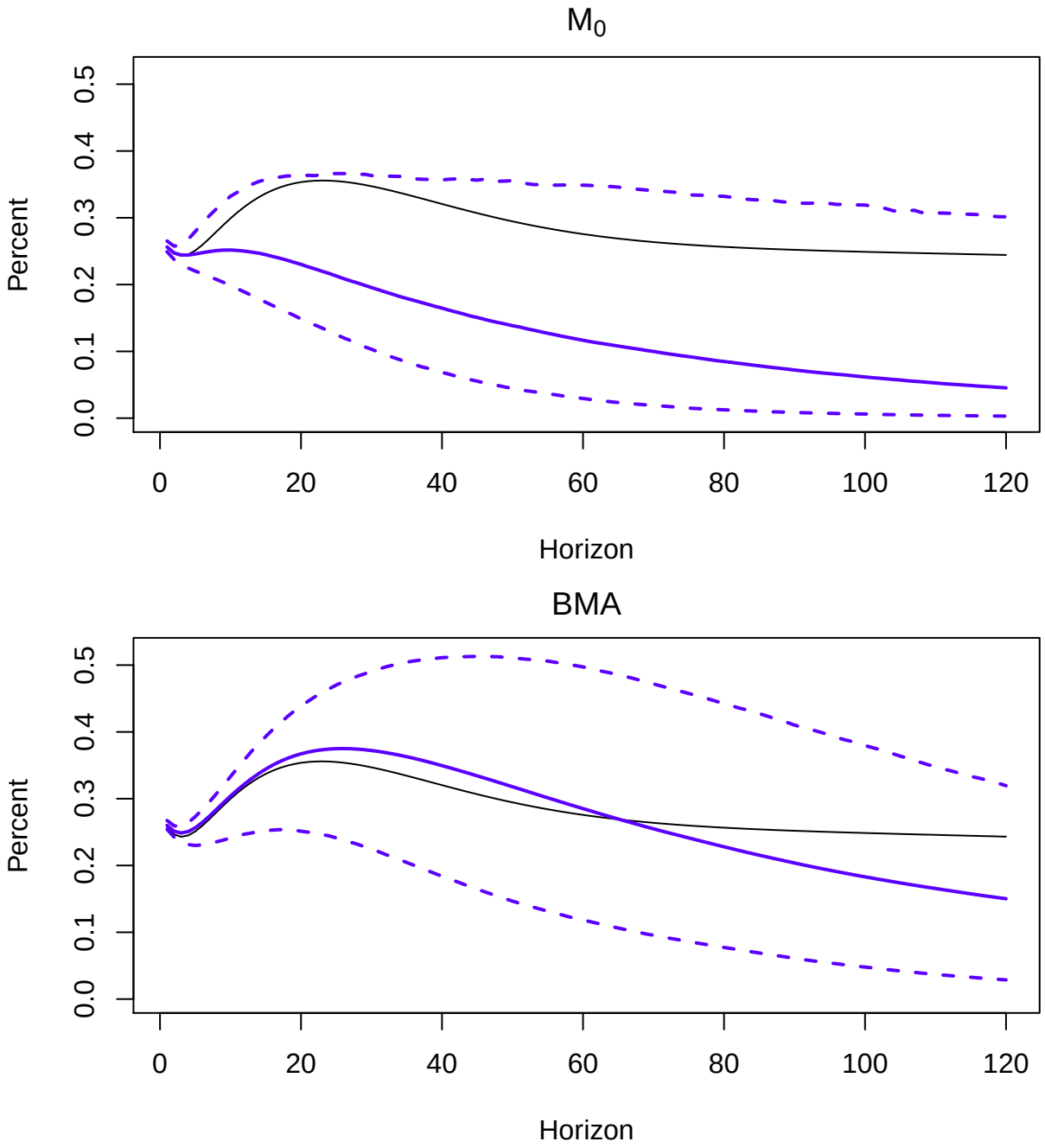
Changes in five-by-ten-year forward rates and model-implied risk-neutral rates over (i) the entire sample period (using averages over the first and last year in the sample), and (ii) the conundrum period. Numbers in squared brackets are 95%-credibility intervals. Discrepancies between levels and changes are due to rounding.

Table 9: Return predictability

		M0		M1		M2		M3		BMA	
Maturity	Data	Pop.	Smpl.	Pop.	Smpl.	Pop.	Smpl.	Pop.	Smpl.	Pop.	Smpl.
2 years	0.35	0.35	0.41	0.14	0.26	0.31	0.38	0.36	0.37	0.19	0.29
			(0.16)		(0.15)		(0.16)		(0.18)		(0.16)
5 years	0.35	0.25	0.37	0.11	0.28	0.26	0.37	0.31	0.38	0.15	0.30
			(0.13)		(0.15)		(0.15)		(0.17)		(0.15)
7 years	0.36	0.22	0.36	0.11	0.29	0.26	0.39	0.32	0.39	0.15	0.31
			(0.13)		(0.15)		(0.14)		(0.17)		(0.15)
10 years	0.35	0.20	0.35	0.11	0.30	0.26	0.40	0.32	0.40	0.15	0.32
			(0.12)		(0.14)		(0.13)		(0.16)		(0.14)

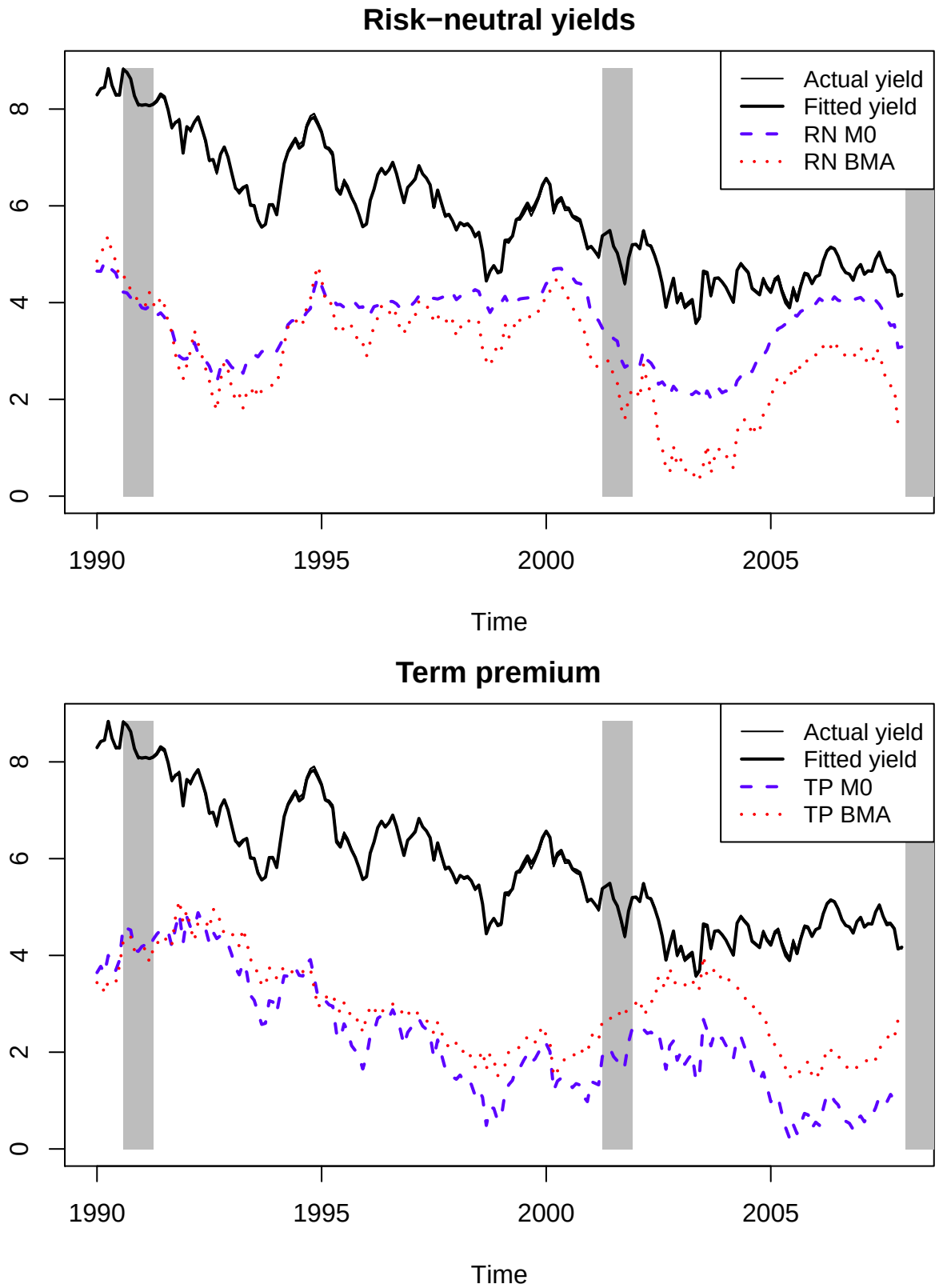
Predictability of annual excess returns, measured by R^2 for regressions of realized excess returns on the first three principal components of yields. The first column reports the R^2 for the actual yield data. Subsequent columns contain, for each model, the model-implied population R^2, and the small-sample R^2 based on repeated simulation, with standard errors in parentheses.

Figure 1: Term structures of volatility



Posterior means of volatilities of changes in fitted forward rates (thin solid line) and risk-neutral forward rates (thick solid line), as well as 95%-credibility intervals for risk-neutral volatilities (dashed lines).

Figure 2: Expectations and term premium



Actual and fitted five-to-ten-year forward rates and estimated expectations and term premium components. Top panel: forward rates and estimates of risk-neutral forward rates (RN) across models. Bottom panel: forward rates and estimates of forward term premium (TP) across models.