FEDERAL RESERVE BANK OF SAN FRANCISCO

WORKING PAPER SERIES

Bayesian Estimation of Dynamic Term Structure Models under Restrictions on Risk Pricing

Michael D. Bauer
Federal Reserve Bank of San Francisco

November 2011

Working Paper 2011-03
http://www.frbsf.org/publications/economics/papers/2011/wp11-03bk.pdf

Bayesian Estimation of Dynamic Term Structure Models under Restrictions on Risk Pricing*

Michael D. Bauer†

First draft: November 2, 2009
This version: November 29, 2011

Abstract

This paper performs Bayesian estimation of affine Gaussian dynamic term structure models (DTSMs) in which the risk price parameters are restricted. A new econometric framework for DTSM estimation allows the researcher to select plausible constraints from a large set of restrictions, to correctly quantify statistical uncertainty, and to incorporate model uncertainty. The main empirical result is that under the restrictions favored by the data the expectations component, and not the term premium, accounts for the majority of high-frequency movements of long-term interest rates. At lower frequencies, term premia are counter-cyclical and more stable than implied by DTSMs without risk price restrictions.

Keywords: dynamic term structure model, no-arbitrage, term premium, market prices of risk, Bayesian inference

JEL Classifications: C52, E43, G12

---

*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.

†Federal Reserve Bank of San Francisco, 101 Market St. MS 1130, San Francisco, CA 94105, phone: (415) 974-3299, e-mail: michael.bauer@sf.frb.org

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, identifying the expectations and risk premium components in long-term interest rates is crucial. 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) are potentially powerful tools for this purpose. However, these models only add real value if the risk pricing, which connects the cross section to the time series of interest rates, is in some way restricted. This paper provides a statistical framework to choose the most plausible restrictions among a large set of candidates, and to incorporate these restrictions into the inference about expected future short rates and term premia.

A widely used approach to infer expectations of future monetary policy is to estimate a time series model for interest rates and use the implied forecasts of the short-term interest rate as a proxy for the market's policy expectations. Unfortunately this presents a tough statistical problem: the near-unit-root behavior of interest rates makes it difficult to determine their long-term properties, which determine forecasts of the short rate and hence estimates of term premia (Kozicki and Tinsley, 2001; Rudebusch, 2007; Cochrane and Piazzesi, 2008). Specifically, 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 upwardly biased estimates of the speed of mean reversion.

A dynamic term structure model (DTSM) imposes absence of arbitrage, which can help alleviate these problems and provide more reliable term premium estimates than those based on, for example, an unrestricted Vector Autoregression (VAR). Because the no-arbitrage condition requires that the cross section of interest rates reflect forecasts of future short rates, allowing for a risk adjustment, the cross-sectional observations can help pin down the conditional expectation of the short rate. However, the risk adjustment loosens this connection between cross-sectional and dynamic properties. In fact, absent any restrictions on the risk pricing, the cross section provides no information for determining short rate expectations (Joslin et al., 2011). Thus DTSMs can suffer from the same problems as simple time series models. This explains why the resulting term premium estimates are often puzzling from a macroeconomic perspective, attributing most of long-term interest rate variability to risk premia (Kim and Orphanides, 2007) and implying large procyclical responses to macroeconomic news (Beechey, 2007).

The potential of no-arbitrage in term structure models is only unleashed if the risk adjustment can in some way be restricted. This paper introduces a decision-theoretic framework that enables the researcher to choose restrictions on the market prices of risk. It turns out that the data call for tight restrictions on risk prices. This result stands in stark contrast to many existing studies that impose few or no such restrictions—prominent examples are Dai and Singleton (2002), Ang and Piazzesi (2003), and Kim and Wright (2005). The term premium estimates obtained using a restricted model are more precise in terms of reduced statistical uncertainty and more reliable in terms of reduced bias than those obtained from an unrestricted version of the model. And as it turns out they are also more plausible from a macroeconomic perspective.

The main contribution of this paper is to provide a new econometric framework for estimation of DTSMs. The framework is Bayesian, which has two major advantages in this context: First, we can correctly quantify the estimation and model uncertainty inherent in term premium estimates. Second, this provides the appropriate tools to assess alternative restricted model specificiations. The relevant selection criterion, the posterior model probability, has a rigorous statistical justification and an intuitive interpretation, in contrast to frequentist information criteria. An important challenge is the large number of possible specifications. To deal with this problem I develop a new algorithm, a version of the product-space sampling approach to model selection of Carlin and Chib (1995), which enables me to narrow down the model space to a handful of plausible candidate specifications *without having to estimate every possible specification separately.* My estimation framework has broad applicability beyond the particular set of restrictions (zero restrictions on risk pricing) and the particular model (affine Gaussian term structure model) that this paper considers. Its potential lies in the ability to rigorously assess a large set of alternative restrictions and in this way to impose parsimony on otherwise overparameterized models.

Specification uncertainty is an important issue for term premium estimation, and the strikingly different term premium estimates in the literature bear witness to this (Rudebusch et al., 2007). The paper documents how the economic implications of a DTSM depend in important ways on the risk price restrictions. My framework deals with this issue by means of Bayesian model averaging. Here, estimates of policy expectations and term premia are calculated as averages across different restricted DTSM specifications, using posterior model probabilities as weights. Hence, estimates of economic objects of interest are not conditional on one set of restrictions but instead take into account model uncertainty.

The term structure model is a discrete-time affine Gaussian DTSM with three latent factors. I employ the arbitrage-free Nelson-Siegel (AFNS) specification of (Christensen et al., 2011, CDR), which amounts to imposing three overidentifying restrictions, including a unit root, on the risk-neutral dynamics. While this does not restrict the forecasts of the short rate if prices of risk are unrestricted, it becomes powerful when I restrict the risk adjustment. Now the restrictions determine whether the model also has a unit root under the physical measure. In this way the paper deals with the near-integrated behavior of the short rate.

Instead of pricing Treasury bonds, the paper considers Eurodollar futures, because these give a direct and detailed view of the forward rate curve. They are often quoted in the press when it comes to documenting how the market supposedly revised its expectations of future monetary policy. The paper shows how DTSM estimates can be used to decompose daily changes in Eurodollar futures, daily volatilities, and responses to macroeconomic data releases into expectations and term premium components.

For the specific model and dataset used in this paper, the data call for tight restrictions on the market prices of risk. Under these restrictions term premia are much more stable than in an unrestricted DTSM. Changes in short rate expectations account for most of the daily changes and volatility in forward rates across all maturities considered in this paper. The procyclical response of long forward rates to macroeconomic data surprises is attributed mainly to revisions of policy expectations, which stands in contrast to previous findings that this response primarily reflects changes in the term premium (Beechey, 2007).

Many macroeconomists have the prior that risk premia move slowly and in a countercyclical fashion. The surprisingly high variability of term premia (Kim and Orphanides, 2005) and their seemingly procyclical response to macro news (Beechey, 2007) have thus been viewed as a puzzle. This paper shows that when those restrictions are imposed on risk prices that a rigorous decision-theoretic framework calls for, a standard DTSM actually delivers term premium estimates that are much more plausible in light of this conventional macro wisdom: They are (i) rather stable, (ii) countercyclical at low frequencies and (iii) not strongly procyclical in response to macro news.

One way to interpret the results of this paper is to think about the properties of term premia in a DTSM on a spectrum between two extremes. One extreme is the expectations hypothesis (EH). Under the EH, term premia are constant, and all changes in interest rates are driven by changes in short rate expectations. The other extreme is a completely unrestricted risk adjustment. This has the unattractive implication that monetary policy and long rates are essentially disconnected and term premia overly variable. Intuition suggests that we should be somewhere on this spectrum between the two extremes—one would expect that long rates reveal *something* about policy expectations. The findings of this paper imply that on this spectrum we want to be closer to the EH than most existing term structure models are. Term premia show some important time variation but are more stable than long-term interest rates, and the connection between physical and risk-neutral dynamics is relatively tight.

Two other studies in the DTSM literature systematically impose restrictions on risk prices. In earlier work, Cochrane and Piazzesi (2008) derive the restrictions from a careful analysis of return predictability, so that all variation in risk premia are due to changes in their tent-shaped return-forecasting factor. My approach differs in that I use statistical tools to choose restrictions, which allow me to consider a large set of possible constraints, including, in principle, the specific restrictions that Cochrane and Piazzesi impose. In parallel work, Joslin et al. (2010) estimate restricted versions of their macro-finance DTSM by means of maximum likelihood, imposing zero restrictions on risk sensitivity parameters, and then pick a specification on the basis of information criteria. My statistical approach has some distinct advantages: First, posterior model probabilities/Bayes factors are valid in small samples and have an economic interpretation, whereas information criteria do not. Second, my framework does not require estimation of every possible model but instead allows me to quickly identify promising candidates. Third, I systematically deal with the issue of model uncertainty, which is important in the context of DTSM estimation.

The paper is structured as follows: Section 2 describes the DTSM, the role of risk price restrictions, and the econometric framework. Section 3 shows the results of the model selection exercise, and compares the time series of the forward term premium across models. In Section 4 the estimation results are used to decompose daily changes in interest rates and their volatilities into expectations and term premium components. Section 5 presents a methodology and the results for an analysis of the model-implied responses of short rate expectations to macroeconomic news. In Section 6 I show that the model estimates are consistent with model-free regression results about the predictability of returns and departures from the EH. Section 7 concludes.

2 Econometric framework

2.1 Model specification

Denote by X_t the $(N \times 1)$ vector of term structure factors which represents the new information that market participants obtain at time t. These risk factors can be observable or latent, as in this paper, in which case they are filtered from observables. Assume that X_t follows a first-order Gaussian vector autoregression under the physical measure $\mathbb{P}$:

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

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

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

Assuming absence of arbitrage, there exists a risk-neutral probability measure, denoted by $\mathbb{Q}$, which prices all financial assets. A stochastic discount factor (SDF) defines the change of probability measure between the physical and the risk-neutral world. The one-period SDF, M_{t+1}, is specified to be exponentially affine,

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

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

$$\lambda_t = \lambda_0 + \lambda_1 X_t. \tag{4}$$

An intuitive interpretation of these risk prices is that they measure the additional expected return that the marginal investor requires per unit of risk in each of the shocks in ε_t—for the shocks to X_t investors require risk compensation of $\Sigma\lambda_t$. The risk sensitivity parameters λ_0 $(N \times 1)$ and λ_1 $(N \times N)$ determine the behavior of risk prices and term premia and will be crucial in this paper.

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

$$X_t = \mu^{\mathbb{Q}} + \Phi^{\mathbb{Q}} X_{t-1} + \Sigma \varepsilon_t^{\mathbb{Q}}, \tag{5}$$

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:

$$\mu^{\mathbb{Q}} = \mu - \Sigma\lambda_0, \quad \Phi^{\mathbb{Q}} = \Phi - \Sigma\lambda_1. \tag{6}$$

In such an affine DTSM, prices of zero-coupon bonds are exponentially affine and yields, as well as rates on implied forward contracts or money market futures, are affine in the state variables (see, e.g., Joslin et al., 2011). Collecting the rates on the M instruments that are used for estimation in a vector Y_t, I specify the observation equation as

$$Y_t = A + BX_t + e_t, \tag{7}$$

where e_t is a vector of measurement errors that is iid normal with diagonal covariance matrix

R, included to avoid stochastic singularity. I will impose $R = \sigma_e^2 I_M$, which implies that the model tries equally hard to fit the rates of the M instruments in a least-squares sense. The M-vector A and the $M \times N$-matrix B contain the model-implied loadings of the instruments on the risk factors. These are determined by the parameters δ_0, δ_1, $\mu^{\mathbb{Q}}$, and $\Phi^{\mathbb{Q}}$, which can therefore be called "cross-sectional parameters", as well as Σ. Equations (1) and (7) together form the state space representation of the model. For future reference, define $Y = (Y_1', \dots, Y_T')'$ and $X = (X_1', \dots, X_T')'$.

Since not all parameters of the DTSM are identified, normalization restrictions need to be imposed (Dai and Singleton, 2000). A DTSM is "maximally flexible" if it is identified and there are no overidentifying restrictions. In this paper I use the three-factor model Christensen et al. (2011), detailed in Appendix B. Their "arbitrage-free Nelson-Siegel" (AFNS) specification imposes the necessary normalizing restrictions on the cross-sectional parameters, as well as three overidentifying restrictions which imply that the risk factors take the role of level, slope and curvature. In this model, $\mu^{\mathbb{Q}}$ and $\Phi^{\mathbb{Q}}$ are uniquely determined by one scalar, ρ. The specification imposes a unit root for the short rate under the $\mathbb{Q}$ measure, which implies that distant forward rates with arbitrarily long maturity can still be time-varying. While I focus on the AFNS specification in this paper, my estimation framework is applicable to any affine model specification. The model is parameterized in terms of the $\mathbb{Q}$-dynamics and the risk sensitivity parameters. The parameters to be estimated are $\theta = (\rho, \lambda_0, \lambda_1, \Sigma, \sigma_e^2)$. Absent any further restrictions, the model has 20 free parameters.

2.2 Restrictions on risk pricing

Absence of arbitrage requires the consistency of time series (physical, $\mathbb{P}$) dynamics and cross-sectional (risk-neutral, $\mathbb{Q}$) dynamics of interest rates, allowing for a risk adjustment. Equation (6) makes this risk adjustment explicit, showing how $\mathbb{Q}$-parameters are tied to $\mathbb{P}$-parameters.

If λ_0 and λ_1 are left unrestricted then any estimates for the physical dynamics are consistent, for some prices of risk, with a given choice of risk-neutral dynamics. In this case, only the time series dimension of interest rates is used to pin down the $\mathbb{P}$-dynamics, and no use is made of the cross-sectional information. This point is particularly obvious in a maximally-flexible model with observable risk factors, since then the maximum likelihood (ML) estimates of μ and Φ are equivalent to the ordinary least squares (OLS) estimates of the VAR parameters (Joslin et al., 2011). The intuition is that ML estimation of the model without risk price restrictions essentially amounts to separately estimating equation (1) by OLS to determine μ and Φ, and then choosing the remaining model parameters by minimizing the pricing errors in equation (7). The first step uses only time series information, and the second step uses only cross-sectional information. However, high persistence of interest rates makes it problematic to estimate μ and Φ using only time series information, as has been documented by Duffee and Stanton (2004) and Kim and Orphanides (2005), among others. The first problem is lack of precision: Because interest rates are highly autocorrelated and revert to their unconditional mean slowly, this mean and the speed of mean reversion are estimated with low statistical precision. The second problem is small-sample bias present in ML estimates of autoregressive systems, which causes the speed of mean reversion of the short rate to be overestimated. The typically small samples of highly persistent observations makes this bias particularly severe in the DTSM context (Bauer et al., 2011). These arguments suggest that in DTSMs without risk price restrictions, estimates of the $\mathbb{P}$-parameters are potentially unreliable. Furthermore, these issues are clearly related to the computational difficulties with finding the ML estimates of affine models.[1]

In any empirical DTSMs, the $\mathbb{Q}$-parameters are typically estimated very precisely, because the data contains a lot of cross-sectional information. If the risk pricing in the model is somehow restricted, then this information will also be used to pin down the time series dynamics. For intuition consider the extreme case of the strong EH, where prices of risk and term premia are zero. A cross-sectional observation of rates in this case shows exactly what the real-world expectations of future short rates are. More generally, if λ_0 and λ_1 are strongly restricted, as for example in Cochrane and Piazzesi (2008), then "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 estimation problems that the DTSM literature has been struggling with.

In this paper I will focus on zero restrictions on λ_1. This parameter determines the variability of term premia. The weak EH, where term premia are constant, corresponds to $\lambda_1 = 0_{N\times N}$—actual and risk-neutral rates will in this case move one for one. The other extreme is to have all N^2 elements of λ_1 unrestricted, with the consequences discussed above. The estimation approach proposed here is to let the data decide which zeros should be imposed, using model selection methods to find plausible restrictions.

Let γ be a vector of indicator variables, each of which corresponds to a specific model parameter and is equal to one if that parameter is allowed to be nonzero. For the case that I focus on, the value of γ describes the zero restrictions on λ_1, and can take on $2^{N^2} = 512$ different values—if additional parameter restrictions are considered or a high-dimensional model is used, this number will explode quickly. A crucial feature of the estimation framework

[1] Among the many studies that report such difficulties are Ang and Piazzesi (2003), Duffee and Stanton (2004), Kim and Orphanides (2005), and Hamilton and Wu (2010).

developed below is that we do not need to estimate every one of these specifications separately.

2.3 Likelihood

The full likelihood of the model is $P(Y|\theta,\gamma) = P(Y|\theta,\gamma,X) \times P(X|\theta,\gamma)$. For the likelihood of the factors, which I call the $\mathbb{P}$-likelihood, we have

$$P(X|\theta,\gamma) = P(X|\rho,\lambda_0,\lambda_1,\gamma,\Sigma). \tag{8}$$

This is simply the likelihood of a Gaussian VAR. It depends on $(\rho,\lambda_0,\lambda_1)$ since these parameters, together with Σ, determine the physical dynamics μ and Φ—see equation (6). For the likelihood of the observations Y conditional on the factors X (the cross-sectional/$\mathbb{Q}$-likelihood), we have

$$P(Y|\theta,\gamma,X) = P(Y|\rho,\sigma_e^2,X). \tag{9}$$

For the AFNS specification ρ by itself determines all cross-sectional loadings in A and B, which together with the measurement error variance determine the $\mathbb{Q}$-likelihood. Naturally, the risk price restrictions do not affect the cross-sectional fit.

2.4 Priors

Estimation will proceed in a Bayesian setting, and we need to specify prior distributions. Since the goal is to understand what the data can tell us about term premia, priors are chosen as sufficiently diffuse. This stands in contrast to the approach of Chib and Ergashev (2009), who use prior information to overcome the typical problems in DTSM estimation.

Priors need to be specified for the parameters $\theta = (\rho,\lambda_0,\lambda_1,\Sigma,\sigma_e^2)$ and for γ. I specify ρ to be uniformly distributed over the unit interval. The prior for $\Sigma\Sigma'$ is inverse Wishart (IW) and the prior for σ_e^2 is inverse gamma (IG), both very dispersed.

The elements of λ_0 and λ_1 are taken to be normally distributed and a priori independent, with a prior mean of zero. For the choice of the prior variance there is an important trade-off: On the one hand, we want to impose as little prior information as possible and let the data speak about the risk price parameters. On the other hand, the focus is on model selection, and very diffuse priors will lead to results that necessarily favor the restricted model—the Lindley-Bartlett paradox (Bartlett, 1957). Hence, the results might be very sensitive to the prior variance of λ_1. I deal with this issue by specifying a prior variance that is trying to strike a middle ground, and then assessing the robustness of my results to changing this variance by orders of magnitude. For now I choose a prior variance of unity, which is not very informative given the magnitude of the ML estimates (not shown) and also not excessively diffuse.

For the model identifier γ I specify the elements γ_i to be a priori independent. They are naturally Bernoulli distributed, and I set the prior probability to 50%. Thus, every model specification has prior probability of $.5^9 \approx .195\%$. While this gives considerable prior weight to models with zero restrictions on λ_1, a comparison of prior and posterior model probabilities will show the amount of support in the data for restricted model specifications.

The $\mathbb{P}$-dynamics should not be allowed to be explosive. Therefore the joint prior $P(\theta, \gamma)$ also imposes the restriction that the eigenvalues of Φ are at most one in absolute value, preventing explosive dynamics.

A important aspect of this prior specification is that there is a point mass on specifications with a unit root under $\mathbb{P}$. To see that the prior explicitly allows for a unit root under $\mathbb{P}$, note that the short rate has a unit root under $\mathbb{Q}$, and the value of γ determines whether Φ has a unit eigenvalue. Without inclusion of γ and explicitly considering zero restrictions on λ_1, any conventional prior specification would imply a continuous distribution for the largest eigenvalue of Φ, which would prevent the posterior estimates to ever exhibit a unit root for the short rate under the $\mathbb{P}$ measure. On the other hand by having point mass on a unit root under $\mathbb{P}$ a priori, the data will reveal whether an integrated or stationary specification provides a better fit.

2.5 Model selection and estimation

To estimate the model under plausible risk price restrictions, I develop a Bayesian model selection framework.[2] This enables me to find those restricted specifications that have the strongest support in the data, to quantify estimation and specification uncertainty, and to perform inference about short-rate expectations and term premia without conditioning on a specific model (using Bayesian model averaging).

The conventional approach to Bayesian model selection is to estimate a set of models, to calculate marginal likelihoods for each one, and then to select models based on Bayes factors (Kass and Raftery, 1995). Since I assign equal prior probability to each model specification, the posterior odds ratio for comparing two models is equal to the Bayes factor, which in turn is equal to the ratio of posterior model probabilities. Calculation of the marginal likelihood is required for all models under consideration in order to calculate the posterior model probability of a given model. If the number of candidate models is too large, this is not feasible. The purpose of the estimation framework here is to avoid having to estimate each model separately,

[2]For review articles on Bayesian model selection see Kass and Raftery (1995) and Clyde and George (2004).

and instead to identify a small set of plausible candidate specifications using an Markov chain Monte Carlo (MCMC) algorithm that samples across models and parameters. Using this smaller set of models, the estimation and model selection problem becomes manageable.

The estimation approach suggested here consists of four steps:

1. Estimate the unrestricted model
2. Identify the most plausible models
3. Estimate each model in the constrained model set
4. Obtain a reversible-jump MCMC sample

The second step is crucial in that it reduces the set of models under consideration to a small manageable number.

2.5.1 Estimating the unrestricted model

The first step is to estimate the DTSM without restrictions on the risk pricing. The main reason to do so is to have a benchmark against which to compare the restricted model estimates, so that we can assess if and how risk price restrictions change the economic implications. Another reason is that the posterior distributions for the parameters of the unrestricted model are useful for designing proposal distributions in the subsequent sampling algorithms. Furthermore, the credibility intervals for the elements of λ_1 are an important "reality check" for the results of the model selection exercise.

Estimation of the unrestricted model is conceptually straight-forward. The joint posterior distribution of the model parameters and the latent factors is proportional to the product of the likelihood function for the data, the likelihood function for the factors, and the joint prior:

$$P(\theta, X|Y, \gamma = 1) \propto P(Y|\theta, \gamma = 1, X)P(X|\theta, \gamma = 1)P(\theta), \tag{10}$$

where $\gamma = 1$ indicates that all elements of λ_1 are unrestricted. In order to obtain a sample which is approximately distributed according to the posterior $P(\theta, X|Y, \gamma = 1)$, a block-wise Metropolis-Hastings MCMC algorithm is used. The parameters are grouped into five blocks, namely X, ρ, (λ_0, λ_1), Σ and σ_e^2. In each iteration of the algorithm only one block is updated, chosen at random according to specified probabilities. Such a "random scan" algorithm has the advantage that it increases the efficiency of the algorithm since we can fine-tune how often each block is sampled: Blocks for which the sampler does not mix as well as for others are updated more frequently. The five blocks are sampled with probability 10%, 20%, 50%, 10% and 10%, respectively. Details on how each block is drawn are provided in Appendix C.

Iterating on this blockwise algorithm, the first B observations are discarded (the burn-in sample) so that the effect of the starting values becomes negligible. Of the following iterations, only every sth draw is retained, so that the number of iterations necessary for a sample of size G is $B + s \cdot G$. I choose $B = 20000$, $G = 5000$ and $s = 40$. These values result from a careful inspection of convergence plots under different configurations, given the restrictions of computational costs and memory constraints. Notably not more than several thousand draws can be saved since every draw contains not only the parameters but also $T \cdot N$ values for the sampled paths of the latent factors. After having obtained the MCMC sample I check several convergence diagnostics, described in Appendix D, which indicate that the sample is likely to be from the stationary distribution of the Markov chain.

2.5.2 Identifying the most plausible specifications

Joint model-parameter sampling is a convenient tool to identify a smaller set of plausible model specifications without having to estimate all candidate models. A model identifier, γ in this context, is included as an additional parameter to be sampled from. An appropriate MCMC algorithm will allow us to draw from the joint posterior distribution of (γ, θ, X). The great advantage of joint model-parameter sampling is that not all models need to be visited by the sampler: we are only interested in those models with high posterior probability, and exactly these are most likely to be visited most often by the sampling algorithm.

The idea of using indicator variables to represent zero restrictions goes back to the "Gibbs variable selection" (GVS) method of Dellaportas et al. (2002), which is a special case of the product-space sampling of Carlin and Chib (1995). Product-space sampling means that in each iteration one keeps track of the parameters of all models, not only of those that are included in the current model. This implies that the dimensionality of the space being sampled from remains the same across models, which allows standard MCMC sampling. What is particular to GVS is that the models are nested. Therefore, the complete set of parameters is simply the entire λ_1 matrix, in addition to the other model parameters, which are assumed to be shared among models. Computationally, keeping track of all parameters just means that λ_1 always contains N^2 non-zero elements, but when calculating the likelihood for a restricted model only those elements of λ_1 that are "switched on" according to γ are taken to be non-zero. The algorithm here is closely related to the original GVS algorithm. Dellaportas et al. (2002) considered a model with a set of parameters and corresponding indicators, and any parameter could be restricted to zero. Here the GVS idea is applied to one parameter block, (λ_1, γ), as part of a larger block-wise M-H algorithm. I believe that this approach, which is novel, has great potential beyond the DTSM context.

The sampling algorithm is again random-scan block-wise M-H. It has six blocks: (λ_1, γ), X, ρ, λ_0, Σ, σ_e^2, which are updated at random with probabilities 50%,10%,10%,10%,10%,10%, respectively. Only the drawing of the block (λ_1, γ) requires explanation, all other blocks are drawn exactly as in the MCMC sampler for estimation of the unrestricted model. For this block, I iterate through the N^2 elements of λ_1 and γ in random order. In each iteration I first update the element of λ_1 and then the element of γ.

First, let's consider drawing an element of λ_1, denoted by λ_i, conditional on γ, the remaining parameters of the model θ_-, the latent factors X and the data Y. One needs to distinguish whether a particular parameter is currently included in the model, and thus the draw is informed by the data, or whether it is currently excluded. In the latter case the data is not informative and one samples from a "pseudo-prior" or "linking density," (Carlin and Chib, 1995). The conditional posterior of λ_i is given by

$$P(\lambda_i|\lambda_{-i}, \gamma_i = 1, \gamma_{-i}, \theta_-, X, Y) \propto P(X|\theta, \gamma)P(\lambda_i|\gamma_i = 1) \quad (11)$$

$$P(\lambda_i|\lambda_{-i}, \gamma_i = 0, \gamma_{-i}, \theta_-, X, Y) \propto P(\lambda_i|\gamma_i = 0), \quad (12)$$

where θ_- denotes all parameters in θ other than λ_1, and λ_{-i} (γ_{-i}) contains all elements of λ_1 (γ) other than λ_i (γ_i).[3] If element i is currently included in the model, one samples from the conditional posterior in (11), using Metropolis-Hastings since the normalizing constant is unknown—I use a fat-tailed RW proposal, with scaling chosen to tune the acceptance probability. The $\mathbb{Q}$-likelihood does not appear on the right-hand-side because it does not depend on λ_1. My prior specification implies prior conditional independence in the sense that $P(\lambda_i|\lambda_{-i}, \theta_i, \gamma_i = 1) = P(\lambda_i|\gamma_i = 1)$. Therefore, only the prior $P(\lambda_i|\gamma_i = 1)$ needs to be calculated here. If λ_i is not currently included, i.e., if $\gamma_i = 0$, it is drawn from the pseudo-prior $P(\lambda_i|\gamma_i = 0)$. I take this distribution to be normal with mean and variance given by the sample moments of the marginal posterior draws of λ_i for the unrestricted model.

The conditional posterior distribution of an element of the vector of indicators is Bernoulli and the success probability is easily calculated based on:

$$\frac{P(\gamma_i = 1|\gamma_{-i}, \theta, X, Y)}{P(\gamma_i = 0|\gamma_{-i}, \theta, X, Y)} = \frac{P(X|\gamma_i = 1, \gamma_{-i}, \theta)}{P(X|\gamma_i = 0, \gamma_{-i}, \theta)} \frac{P(\lambda_i|\gamma_i = 1)}{P(\lambda_i|\gamma_i = 0)} \frac{P(\gamma_i = 1, \gamma_{-i})}{P(\gamma_i = 0, \gamma_{-i})}. \quad (13)$$

Because of the fair Bernoulli prior for γ_i the last term cancels out. Denoting the above

[3] These conditional distributions parallel the ones in equations (9) and (10) of Dellaportas et al. (2002).

ratio by q, the conditional posterior probability of $\gamma_i = 1$ is given by $q/(q+1)$. A subtlety, which is ignored in the above notation, is that the joint prior $P(\gamma, \theta)$ imposes non-explosive $\mathbb{P}$-dynamics, for any choice of γ and λ_1. This is easily implemented in the algorithm: If including a previously excluded element would lead to explosive dynamics then I simply do not include it, i.e., I set $\gamma_i = 0$, and vice versa.

The algorithm is run with a burn-in sample of size $B = 50,000$ and a sample size of $G = 100,000$, using every fifth draw from a chain of length 500,000. In order to get an idea of the convergence properties of this algorithm, I run several chains from different starting values and make sure that the results are similar. I verify that the results are reasonable given the sample from the posterior for λ_1 for the unrestricted model, based on individual credibility intervals and based on highest-posterior-density regions resulting from a normal approximation to the joint posterior. Another important issue is the prior for λ_1, since a too diffuse prior might favor restricted models. To assess the sensitivity of the results I rerun the analysis by changing the prior variance for λ_1 by orders of magnitude. I find that the results of the model selection exercise are robust to alternative choices of the prior variance.

With the MCMC sample at hand, one can assess which model specifications are the most plausible ones. The relative frequency with which a given model is visited by the sampler provides an estimate of its posterior model probability. Based on this, we can constrain the model set and focus our attention on those models for which this posterior model probability is sufficiently high. An important question of course is where to draw the line. Kass and Raftery (1995) suggest to consider a Bayes factor comparing models i and j above 20 as strong evidence against model j. I compare each model to the favored model (the one with the highest posterior model probability) and include it in the constrained model set only if the Bayes factor does not exceed 20. If the data is at all informative for the purpose of model choice, this will generally lead to the inclusion of only a handful of models. The models included in the restricted model space are denoted by M_1 to M_J.

2.5.3 Estimating each model in the constrained model set

After having reduced the number of models to J, I estimate each model individually. While theoretically the MCMC sample from the previous step could be used to learn about the posterior distributions of the parameters, in practice it is preferable to re-estimate each model. This will provide more reliable approximations to the posteriors for each model, since the efficiency of a sampler is higher when we focus on one model specification at the time.

The algorithm used to estimate an individual restricted model is very similar to the one used for estimation of the unrestricted specification. The only difference is that now the elements of λ_1 are drawn separately, i.e., a separate RW M-H step is used for each risk price parameter that is not restricted to zero. For all other parameters the algorithm is the same and details are in Appendix C.

Posterior model probabilities, denoted by $P(M_j|Y)$, can now be calculated by approximating the marginal likelihoods, denoted by $P(Y|M_j)$. I use two different approximations, a Bartlett-adjusted Laplace estimator and a version of Candidate's estimator, both of which are described in detail in DiCiccio et al. (1997). The first approximation, $\hat{C}_B^*$ in those authors' notation, is a localized (i.e. volume-corrected) version of the Bartlett-adjusted Laplace estimator (see DiCiccio et al., 1997, section 2.2). The second approximation, $\hat{C}^C$, a Candidate's estimator, is based on a simple kernel density estimate of the posterior distribution (see DiCiccio et al., 1997, section 2.5)—this latter estimate in fact corresponds to the method of Chib (1995) with the posterior ordinate replaced by a normal approximant. For the volume I use 5% in both cases and I estimate the mode by taking that parameter draw which maximizes the posterior. For each of these two ways to approximate marginal likelihoods, posterior model probabilities are calculated using Bayes law as

$$P(M_j|Y) = \frac{P(Y|M_j)P(M_j)}{\sum_{i=1}^{J} P(Y|M_i)P(M_i)} = \frac{P(Y|M_j)}{\sum_{i=1}^{J} P(Y|M_i)}, \tag{14}$$

where the second equality is due to equal prior model probabilities.

2.5.4 Obtaining a reversible-jump MCMC sample

The first three steps of the estimation provide everything necessary for inference about each individual model and for model choice. If one model received overwhelming support in the data, i.e., $P(M_j|Y) \approx 1$, or if all J models had very similar economic implications, then we could stop here. This will generally not be the case, hence we are confronted with specification uncertainty. Bayesian model averaging (BMA) is the appropriate way to incorporate specification uncertainty. The idea is to "average out" the model specification by calculating objects of interest as averages across models, using posterior model probabilities as weights. The inference is then conditional only on the class of models under consideration, and not on a particular model specification.

For this purpose, the last step of the estimation is joint sampling across the J models and their parameters using reversible-jump MCMC (RJMCMC). While conceptually this step is not necessary to perform BMA—we have posterior model probabilities and posterior distributions for the parameters of each model—there are two good reasons to include it: First, it provides us with an additional alternative set of estimates of posterior model probabilities as a verification of our previous results. Second, BMA is particularly straightforward if a joint-model-parameter-sample is available, since simply ignoring the value of the model indicator amounts to averaging out the model specification.

As always for joint-model-parameter sampling, a model indicator is included as an additional parameter to be drawn in the MCMC algorithm. Since only J models are considered now, this indicator will be $j \in \{1, \ldots, J\}$. The set of models is small, hence sampling can be done using RJMCMC, a method initially developed by Green (1995), which is characterized by the ability of the sampler to jump between models with parameter spaces of different dimensionality. The sampler will provide draws that are approximately from the joint posterior distribution $P(j, \theta_j, X|Y)$, where the parameter vector for each model, θ_j, differs in length because of restrictions.[4]

The algorithm in each iteration randomly chooses with equal probability between a "within-model" step and a "model-jump" step. For a within-model step, the parameters of the current model are updated as in the algorithm to estimate each model separately. A model-jump step is a M-H step where a new model and corresponding parameters are proposed and then either accepted or rejected. The proposed model, denoted by j', differs from the current model j, and is chosen randomly with equal probability $1/(J-1)$.

An important question is how to propose values for the parameters of model j', denoted by $\theta_{j'}$. I decide for models to share the parameters ρ, Σ, and σ_e^2, denoted here by θ_-, but not to share any elements of λ_0 and λ_1. This is different from the algorithm used to identify plausible models, where models shared λ_0 and those elements of λ_1 that are unrestricted in both models. Sharing these parameters was necessary there, because product-space sampling would have otherwise required to keep track of too large a number of parameters. The advantage of RJMCMC is that one does not need to keep track of all parameters in the product space, which allows us the luxury to share less parameters among models. Not sharing parameters among models improves efficiency if posterior distributions for these parameters differ across models, and this is the case for some elements of (λ_0, λ_1). To construct $\theta_{j'}$ I take θ_- from the current model together with proposed values for all non-zero elements of (λ_0, λ_1) in model j', denoted by $\lambda_{j'}$. To propose values for $\lambda_{j'}$ I use the normal approximation to the posterior distribution of $\lambda_{j'}$, which is available from the previous estimation step.

The idea of reversible-jump MCMC is that reversibility of the Markov chain is ensured by matching the dimensions between candidate parameter-vector and proposed parameter-vector.

[4]To be precise the model is now parameterized as (j, θ_j, X_j), but since the latent variables carry over between models, I write (j, θ_j, X).

The acceptance probability for the proposed jump is given by the minimum of one and

$$\frac{P(Y|j',\theta_{j'},X)P(X|j',\theta_{j'})P(\theta_{j'}|j')P(j')}{P(Y|j,\theta_j,X)P(X|j,\theta_j)P(\theta_j|j)P(j)} \times \frac{q(u'|\theta_{j'},j',j)q(j'\to j)}{q(u|\theta_j,j,j')q(j\to j')}\left|\frac{\partial g_{j,j'}(\theta_j,u)}{\partial(\theta_j,u)}\right|, \tag{15}$$

the product of model ratio (likelihood ratio times prior ratio) and proposal ratio. The parameter values for the candidate model are determined using $(\theta_{j'},u') = g_{j,j'}(\theta_j,u)$, a bijection that ensures the dimension-matching. In this context $(\theta_{j'},u') = (\theta_-,\lambda_{j'},u') = (\theta_-,u,\lambda_j) = g_{j,j'}(\theta_j,u)$ – the g-function is an identity function that simply matches the correct elements. Intuitively, u provides proposal values for all parameters in model j' that are not shared with model j, i.e. $u = \lambda_{j'}$, and u' takes on the values of the parameters in model j that are not used in model j', i.e. $u' = \lambda_j$. Thus, recognizing the uniform prior over models, the equal jump probabilities $q(j \to j') = 1/(J-1)$ for all $j \neq j'$, and the fact that the $\mathbb{Q}$-likelihood only depends on θ_- and thus does not change between jumps, the above ratio simplifies to

$$\frac{P(X|j',\theta_{j'})P(\theta_{j'}|j')}{P(X|j,\theta_j)P(\theta_j|j)} \times \frac{q(u'|\theta_{j'},j',j)}{q(u|\theta_j,j,j')}. \tag{16}$$

Note that since $u' = \lambda_j$, the distribution $q(u'|\theta_{j'},j',j) = q(\lambda_j|j)$ is the normal distribution with moments obtained from the sample from the posterior for model j, and correspondingly for $q(u|\theta_j,j,j') = q(\lambda_{j'}|j')$. Again, the minimum of the above expression and one is the probability with which I accept the proposed jump $(j,\theta_j) \to (j',\theta_{j'})$.

I run the sampler for $B = 100,000$ burn-in iterations and then create a sample of length $G = 5,000$ by using one out of every $s = 200$ iterations. This is motivated by the fact that memory constraints make it impossible to save more draws of (j,θ_j,X), yet the sampler needs to be running for a considerable amount of iterations. Separate runs based on different starting values indicate that the chain has satisfactory convergence properties.

2.6 Comparison to alternative approaches

Before turning to the data and estimation results, a discussion of alternative statistical approaches for estimating DTSMs with risk price restrictions is warranted. The most common approach is to first estimate a DTSM without such restrictions and then, in a second step, to re-estimate the model by setting some parameters with large standard errors to zero. Prominent studies that employ this two-step approach are Dai and Singleton (2002), Ang and Piazzesi (2003), and Kim and Wright (2005). It is unappealing for several reasons. First, choosing restrictions based on individual standard errors amounts to imposing a joint restriction without considering joint significance. Second, the choice of a significance level required for inclusion of the parameter is necessarily arbitrary. Third, and most importantly, alternative sets of restrictions often lead to economically significant differences in results, and this approach offers no guidance on which set of restrictions is more credible—Kim and Orphanides (2005) report that in their estimation “different choices of parameters to be set to zero [...] exhibited economically significant quantitative differences” (p.11). All classical approaches to selecting risk price restrictions are faced with this issue, such as the use of F-statistics to test joint restrictions on risk prices, or the use of information criteria, such as the Akaike information criterion (AIC) and the Schwarz-Bayes information criterion (BIC), as in Joslin et al. (2010).

Only a Bayesian framework enables the researcher to satisfactorily deal with model uncertainty, i.e., with the problem that alternative model specifications have different economic implications. Classical hypothesis tests or information criteria can identify one or several preferred models, but the researcher cannot easily incorporate the statistical uncertainty about the model. On the other hand the use of Bayesian methods allows the researcher to take into account this uncertainty using BMA. A second advantage is that at the model selection stage, Bayes factors and posterior model probabilities, in contrast to information criteria, have a sound theoretical foundation and allow for an intuitive interpretation.

If estimation uncertainty is to be taken seriously, Bayesian estimation is the only way to go. Given the use of Bayesian methods, what other possibilities are there relative to the approach I propose here? It is crucial to reduce the model space, since otherwise a large number of models would have to be estimated separately. The computational costs of parameter estimation and marginal likelihood approximation for each model are prohibitively large. This is the case for the 512 specifications under consideration here, and even more so for higher-dimensional problems. Any viable approach therefore needs to address the problem of a large model space.

There are some alternative ways to reduce the model space relative to what I propose. One could use an RJMCMC sampler to jump between all models. This, however, is very inefficient without good proposal distributions, which would have to come from separate estimation of each model. One could use an alternative product-space sampler where models do not share parameters. This leads to an excessively large sampling space that would be hard to deal with. Another alternative is to use the posterior distribution of λ_1 from the unrestricted estimation to come up with a set of plausible restrictions. This would require a normal approximation would have to be used to construct highest-posterior-density regions. In contrast, the approach I have suggested for reducing the model space is easy to implement, theoretically appealing, and works well in the present context. The other steps of my estimation framework go naturally hand-in-hand with this step.

3 Model estimates

Turning to the empirical results of the paper, I will first focus on the support that different risk price restrictions receive in the data, describe the economic implications of alternative specifications, and then proceed to perform inference about policy expectations and term premia.

3.1 Data

The data set consists of daily observations of Eurodollar futures rates. I use these futures contracts instead of Treasury yields for three reasons: First, they directly reflect the forward rate curve, whereas Treasury yields depend on the algorithm used to infer zero-coupon rates from observed bond prices (unsmoothed vs. smoothed, etc.). Second, the liquidity is very high: Eurodollar futures are the most liquid futures contracts worldwide. Third, the most liquidly traded government bonds, on-the-run Treasury securities, do not cover the maturity spectrum at similar detail. Pricing Eurodollar futures is straight-forward using an affine DTSM. For details and the exact affine loadings refer to Appendix E.

I include contracts maturing at the end of the current and the following 15 quarters—I denote the contracts by ED1 to ED16, and the futures rates by $ED_t^{(m)}$, $m = 1, \ldots, 16$. These contracts cover the forward rate curve up to a maturity of about four years. The sample consists of daily observations from 1 January 1990 and to 29 June 2007, before the start of the financial crisis. The number of days in the sample is $T = 4401$.

3.2 Unrestricted model

First the DTSM is estimated without restrictions on risk prices. The posterior means and 95% credibility intervals for each of the model's parameters are presented in Table 1. The point and interval estimates are very close to the results of ML estimation (not shown), because the priors are not very informative.

The risk-neutral dynamics, i.e., ρ, as well as Σ and σ_e^2 are estimated very precisely. On the other hand, the estimates of the risk sensitivity parameters in λ_0 and λ_1 evidently lack precision, since the intervals around the point estimates are large. This reflects the fact that μ and Φ are hard to estimate in light of the high persistence of the risk factors (on this issue see also Kim and Orphanides, 2005).

Notably, for only three of the nine parameters in λ_1 do the credibility intervals not straddle zero, i.e., are "significant" in frequentist parlance. This indicates that a more parsimonious model, with some of the risk price parameters constrained to zero, might be a better choice. However, settling on one specification based on these first-step results would be misleading, since this ignores much of the available information on which restrictions are plausible, and sweeps under the rug model uncertainty.

3.3 Model selection

The second step of my estimation framework enables me to identify a small set of relevant models. I include the unrestricted model, denoted by M_1, and the six models that are visited most frequently by the product-space sampling algorithm, denoted by M_2 to M_7. Information for these $J = 7$ models is shown in Table 2. The second column shows the frequency of how often each model was visited by the sampler. The cutoff is a Bayes factor of at most 20 in comparison to M_2, the favored model.[5] These seven models, out of 512 possible specifications, have a cumulative frequency of 75.8%, meaning that the data strongly support these models. Posterior model probabilities that are reported subsequently are to be interpreted relative to this restricted model space.

Columns three to five indicate which elements in the respective columns of λ_1 are restricted (0) and unrestricted (1), e.g., for model M_2 only the element in the second row of the first column is unrestricted. The unrestricted model's posterior probability is estimated to be zero, in fact it is not visited at all by the sampler. In contrast, tight restrictions on risk prices are supported by the data: the six preferred models all have between six and nine elements of λ_1 restricted to zero. This says that the physical and risk-neutral dynamics are rather close to each other, and the risk adjustment is tightly restricted. Notably even the weak EH, model M_5, receives some support in the data. With the data strongly favoring tight zero restrictions on λ_1, we evidently want to be a lot closer to the EH than to the other end of the spectrum, where the risk adjustment is entirely unrestricted.

Having reduced the model set I estimate the parameters and marginal likelihoods for each of these models. Based on these estimates, the table shows some simple numerical summaries of the models' economic implications. Column six shows the largest eigenvalue of Φ that is implied by the point estimates. The short rate exhibits a unit root for most preferred specifications—all models except for M_1 and M_6 have $\mathbb{P}$-dynamics with a largest eigenvalue of unity. There is strong support in the data for a stochastic trend in the short rate. While it cannot literally have a unit root since it is bounded from below and usually remains in a certain range, its largest root is certainly very close to one. The evidence here suggests that

[5] For the least likely model, M_7, the Bayes factor vs. M_2 slightly exceeds 20 ($46.4/2.3 = 20.1$) but I include it nevertheless.

in a model of daily frequency, an integrated specification provides a better fit to the data than a stationary specification.

Column seven shows the long-run revision of short rate expectations, $\lim_{h\to\infty}(\mathrm{E}_{t+1} - \mathrm{E}_t)r_{t+h}$, in response to a unit level shock. Such a shock always moves distant short rate expectations under $\mathbb{Q}$ one-for-one, but only if the short rate has a unit root can it have an impact on far-ahead short rate expectations under $\mathbb{P}$. The unit-root models have different implications for this long-run revision: For M_2, M_4, and M_5 the long-run revision is positive and close to or equal to unity, implying that far-ahead short rate expectations move about as much as far-ahead forward rates. In models M_3 and M_7 on the other hand long-run expectations respond negatively. Evidently there is specification uncertainty about this important aspect of the model, and BMA is needed to account for it. As will be seen below, the evidence is in favor of the long-run revision of $\mathbb{P}$ and $\mathbb{Q}$ expectations being relatively close to each other and long-run forward risk premia being rather stable.

The remaining columns show alternative estimates of the posterior model probabilities based on the frequencies rescaled to add up to 100% (column 8), Laplace approximation to marginal likelihood (column 9), Candidate's estimate of the marginal likelihood (column 10), and RJMCMC (column 11). The various estimates of the posterior model probabilities show some numerical differences, but they are generally rather close. That they are not identical indicates that there is room for improving the efficiency of the sampling algorithms or the accuracy of the marginal likelihood approximations. However, the economic conclusions that one would draw are not materially different depending on which estimate of posterior model probability is used.

Does the prior put large weight on restricted models? In the sense that most prior probability mass is on models with at least one restriction, this is true. But for each individual model, the prior probability is only 0.195%. For the models in the table this is updated very significantly after seeing the data, with the results that almost all the posterior probability mass is on models with few unrestricted elements in λ_1. This comparison of prior and posterior model probabilities shows that the data favor tight restrictions on risk prices.

3.4 Time series of the term premium

With regard to the economic implications of an estimated DTSM, it is helpful to plot the time series of the term premium, which illustrates its trend and business cycle behavior. Figure 5 displays the time series of the futures rate ED8, which essentially is a two-year-ahead forward rate, together with the estimates of the corresponding forward risk premium resulting from the unrestricted model M_1, the favored model M_2, and BMA. The forward risk premium is calculated as the difference between the model-fitted rate and the corresponding risk-neutral rate. The MCMC sample provides its posterior distribution at each point in time, and I take its posterior mean.

The forward premium implied by M_1 is highly variable and moves essentially in tandem with the forward rate. By contrast, M_2 and BMA imply a much more stable two-year forward premium. Correspondingly, the variation in the forward rate is attributed to a larger extent to expectations about monetary policy.

The decomposition of the secular decline in the forward rate differs significantly across models. M_1 implies a dramatic slump of the premium from almost 5% to close to -2%, and a very stable expectations component that has remained around 5%—the estimated expectations have actually slightly increased from 4.8% to 5.3%. The M_2 and BMA estimates on the other hand imply a modest decline for the premium from around 3% to about 1%, with the expectations component contributing quite significantly to the decline in rates. Specifically, the BMA-implied two-year-ahead short rate expectations have declined from 6.4% to 4.2%.

The term premium implied by M_1 repeatedly turns negative toward the end of the sample. Negative premia are not unreasonable per se: if bonds pay off better in bad times, they serve as an insurance, and in that case investors will be willing to forgo some yield to take advantage of this insurance. However, a reader that considers it unlikely that the risk profile of bonds changed so erratically over the course of the sample will be more comfortable with the consistently positive term premium implied by the preferred model specifications.

With regard to business cycle variation, all term premium measures are rising and high before and during the two recessions in the sample and are falling during expansions. Thus, while the variability is very different across models, the business cycle behavior is similar: all three term premium estimates are countercyclical at the business cycle frequency. However, the subsequent analysis will show that the models differ in their implications about the cyclical behavior at higher frequencies, in terms term premium response to macroeconomic news.

"Greenspan's conundrum" refers to the period 2004 to 2005, during which long rates declined despite significant policy tightening. Unrestricted term structure models typically explain this by declining term premia (Backus and Wright, 2007). What do the estimates of this paper imply for the conundrum period? The two-year forward rate actually slightly increased during this period, but by a lot less than the policy rate.[6] While M_1 explains this gap through a sharp decrease in the term premium, the preferred estimates show only a slight decrease in the premium during the conundrum period and leave an important role for the

[6]From June 2004 until December 2005 the FOMC increased the target for the federal funds rate 13 times by 25 basis points each. It then tightened four more times until June 2006.

expectations component. Similar results obtain for longer maturities, at which rates actually declined (not shown). The lesson to be drawn is that the contribution of the term premium to the conundrum is probably overstated by DTSMs that do not restrict risk pricing, and that the expectations component is important. The economic story (which admittedly has more punch at longer horizons than two years) is that tighter monetary policy today leads to lower inflation tomorrow and therefore is consistent with lower nominal short rate expectations.

In sum, the data favors relatively tight restrictions on the risk pricing in the affine DTSM. Under these restrictions, short rate expectations are more variable and term premium estimates more stable than implied by a model in which risk prices are hardly or not at all restricted. This leads to a larger role for the expectations component to explain interest rate variation and the secular trend in long-term interest rates.

4 Decomposing rate changes and volatilities

In this section the model estimates are used to decompose daily changes and volatilities of futures rates across maturities. This will give a particularly clear picture of the different economic implications of the estimated models.

4.1 Three components of interest rate changes

To build intuition about the different components of rate changes, consider the risk-adjusted expectations of the future short-term interest rate,

$$f_t^n = E_t^{\mathbb{Q}}(r_{t+n}) = A_n + B_n' X_t, \tag{17}$$

which is an affine function of the risk factors with loadings given in Appendix E. This rate corresponds to a forward rate for a one-period loan at $t+n$ that is contracted at t, except for a Jensen inequality term due to convexity. Expectations under $\mathbb{P}$, risk-neutral forward rates, are $\tilde{f}_t^n = E_t(r_{t+n}) = \tilde{A}_n + \tilde{B}_n' X_t$, and the forward risk premium is $\Pi_t^n = f_t - \tilde{f}_t$. Now consider the one-period change $f_{t+1}^n - f_t^{n+1}$, which corresponds to the return of a hypothetical futures contract which pays the difference between the realized future short rate and the contractual rate.[7] Notably it is an *excess return*, because the risk-neutral expected return, $\mathrm{E}_t^{\mathbb{Q}}(f_{t+1}^n - f_t^{n+1})$

[7] Specifically, it is the absolute return if one enters at t into a position in a contract that pays $r_{t+n+1} - f_t^{n+1}$ at maturity, and liquidates the position at $t+1$.

is zero. For this rate change/return we have

$$\begin{aligned} f_{t+1}^{n} - f_t^{n+1} &= A_n + B_n' X_{t+1} - A_{n+1} - B_{n+1}' X_t \\ &= B_n' \Sigma \varepsilon_{t+1}^{\mathbb{Q}} = B_n' \Sigma (\varepsilon_{t+1} + \lambda_t) \\ &= \tilde{B}_n' \Sigma \varepsilon_{t+1} + (B_n - \tilde{B}_n)' \Sigma \varepsilon_{t+1} + B_n' \Sigma \lambda_t. \end{aligned} \tag{18}$$

Expression (18) shows that this rate change has three components:

1. *Revisions to short rate expectations*: The first component corresponds to the change in the expectation of the future short rate, $(\mathrm{E}_{t+1} - \mathrm{E}_t) r_{t+n+1} = \tilde{B}_n' \Sigma \varepsilon_{t+1}$. This component, which equals the change in the risk-neutral rate $\tilde{f}_{t+1}^{n} - \tilde{f}_t^{n+1}$, captures how market participants revise their expectations of future monetary policy.

2. *Surprise changes in the forward risk premium*: The second component is equal to the unexpected change in the forward risk premium, $\Pi_{t+1}^n - \mathrm{E}_t \Pi_{t+1}^n = (B_n - \tilde{B}_n)' \Sigma \varepsilon_{t+1}$.

3. *Expected returns*: The third component is equal to the expected change in the forward risk premium, $\mathrm{E}_t \Pi_{t+1}^n - \Pi_t^{n+1} = B_n' \Sigma \lambda_t$. This term captures the predictable part of the return and corresponds to the *return risk premium*.

This decomposition shows that both short rate expectations and term premia can contribute to unpredictable changes in interest rates. In fact, for high frequencies, such as daily changes, most of the variation is unpredictable. This however does not tell us anything about the relative contribution of expectations and term premia. We need get a grip on the first component, changes in risk-neutral rates, to learn about the importance of expectations for changes in interest rates. Importantly, our estimates of this component will only be as precise as our estimates of the $\mathbb{P}$-dynamics, which determine the loadings $\tilde{B}_n$.

4.2 Changes on a specific day

To illustrate the economic implications of the models, I first consider changes in Eurodollar futures on a specific day, decomposing these analogously to equation (18), using the loadings for futures rates given in Appendix E. As an example I choose March 8, 1996, a day with a strongly positive surprise in nonfarm payroll employment—job creation exceeded the market's expectation by more than 400,000 jobs. For this day, the model implies fitted rates for given values of factors and parameters. For each draw in the MCMC sample I calculate the changes in the fitted rates, resulting in a posterior distribution of fitted rate changes for each contract. The sample mean and the 2.5%- and 97.5%-quantiles provide a point estimate and a 95%

credibility interval. Since these intervals are rather tight, given that $\mathbb{Q}$-dynamics are estimated with high precision, and of no particular interest here, I will not report them. In a similar fashion I then calculate the point estimates and credibility intervals for the changes in the risk-neutral rates. Figures 1 to 4 show in their top panels actual and fitted changes in futures rates on this day, as well as point and interval estimates of how monetary policy expectations were revised, given by the posterior mean and 95% credibility intervals of the estimated changes in risk-neutral rates.

Figure 1 shows the decomposition for the unrestricted model, M_1. It implies that forward rate changes for maturities of three to four years are entirely attributed to increases in term premia. Revisions to short rate expectations are estimated to be essentially zero at the long end. Also evident from Figure 1 is the dramatic estimation uncertainty for changes in risk-neutral rates. While this was to be expected in light of the high persistence of interest rates, the extent of the uncertainty is remarkable.

Figure 2 shows the implications of the favored model, M_2, which stands in stark contrast to M_1. It implies that rates moved up on this day almost entirely because of upward revisions of monetary policy expectations. Forward risk premia correspondingly are found to hardly have moved at all. Since the credibility intervals are small, these statement can be made with high statistical confidence (conditional on the specification).

Figure 3 compares in its top panel what all seven models imply for changes in policy expectations on this day. While they have essentially identical implications for fitted futures rates, the models differ significantly with regard to the decompositions into expectations components and term premia. Whereas according to models M_2, M_4, and M_5 the changes in expectations accounted for most of the movements across all horizons and term premia changed little, models M_1, M_3, M_6, and M_7 imply that at the long end forward risk premia are the driving force. Figure 4 shows the resulting inference if we take into account this specification uncertainty using BMA. According to this decomposition, a majority of increase in futures rates on this day, across all maturities, was driven by higher policy expectations.

These results only show the models' implications for changes in expectations and term premia on one given day. To draw more general conclusions we need to consider volatilities of the different components as well as their systematic response to macroeconomic news.

4.3 Volatilities

The term structure of volatility, the "vol curve," describes the volatility of changes in yields or forward rates across maturities. Based on equation (18) the variance of forward rate changes is given by

$$Var(f_{t+1}^n - f_t^{n+1}) = Var(B_n'\Sigma\varepsilon_{t+1} + B_n'\Sigma\lambda_t) = B_n'\Sigma(I_N + Var(\lambda_t))\Sigma' B_n. \qquad (19)$$

The term structure of volatility (in population) is the square root of this expression for varying n. Variability of forward rates is driven both by an unpredictable component, the innovations to the factors, and by a predictable component, the variation in the prices of risk.[8] To understand the importance of changing policy expectations for interest rate volatility we can calculate the term structure of volatility that would prevail if forward rates were only driven by changes in short rate expectations, i.e., if term premia were constant. The variance of changes in risk-neutral forward rates is

$$Var(\tilde{f}_{t+1}^n - \tilde{f}_t^{n+1}) = Var(\tilde{B}_n'\Sigma\varepsilon_{t+1}) = \tilde{B}_n'\Sigma\Sigma'\tilde{B}_n, \qquad (20)$$

and I will call the square root of this expression for varying n the "risk-neutral vol curve." To gage the importance of changing expectations for variation in interest rates in this way is, to my knowledge, a novelty in the term structure literature. For both vol curves I obtain the approximate posterior distributions by calculating model-implied volatilities of futures rates and risk-neutral futures rates for each draw from the MCMC sample.

The second panel of Figures 1 to 4 shows the actual and model-implied vol curves, as well as point and interval estimates for risk-neutral vol curves. Naturally, all models imply that for short maturities variation is driven mostly by expectations—term premia in short contracts are small. Let us focus on the longest contract, which has a horizon of about four years. For the unrestricted model M_1, less than half of the volatility in this interest rate is due to expectations, meaning that daily changes in forward risk premia are a more important driving force for volatility at the daily frequency. Anyone who has a prior of slow-moving risk premia might be rather uncomfortable with this result. The precision of this estimate is very low, as indicated by the very wide credibility interval (the upper boundary is not visible in the graph).

As it turns out, under the restrictions on risk prices that are supported by the data, the risk-neutral vol curve looks very different. Model M_2 implies that most of the volatility in daily changes, up to the horizons under consideration in this paper, is due to changing short rate expectations. For this model, the credibility intervals around the risk-neutral vol curve are also rather tight, so that conditional on this specification, our inference is relatively precise. Going beyond one particular specification and comparing all seven models, in the

[8]Since predictability of daily changes is empirically very small, $Var(f_{t+1}^n - f_t^{n+1}) \approx B_n'\Sigma\Sigma' B_n$.

middle panel of Figure 3, the specification uncertainty underlying the risk-neutral vol curves becomes evident. The estimated contribution of expectations to interest rate volatility differs dramatically between models. However, taking into account this model uncertainty, as shown in the middle panel of Figure 4, the role for expectations is still large, account for the majority of interest rate volatility at all horizons.

5 The impact of macroeconomics news

In this section the model estimates are used to infer the systematic response of monetary policy expectations and risk premia to macroeconomic announcements. While there is ample evidence about procyclical responses of interest rates to such macro news—see, for example Fleming and Remolona (1999) and Gürkaynak et al. (2005)—it is not clear whether expectations or premia drive this phenomenon. Beechey (2007) uses estimates of risk-neutral rates and term premia from the model of Kim and Wright (2005) to shed light on this issue, finding that movements in the risk premium account for the majority of the response of long forward rates. But these results are hard to interpret for two reasons: First, the use of time series of risk-neutral rates and term premia ignores that these series are point estimates surrounded by potentially large statistical uncertainty. Second, such analysis relies on only one DTSM specification, and it is unclear how robust the results are to the use of alternative models. Here I explicitly take into account estimation and specification uncertainty.

5.1 Methodology

The typical approach in the macro announcement literature is to regress changes in interest rates on a measure of the surprise in the announcement, i.e., the difference between the released value and an estimate of the market's expectations. I will consider as dependent variables changes in fitted rates and risk-neutral rates, the latter reflecting the response of policy expectations. If the left-hand-side variable is itself an estimate of the true object of interest, the conventional regression approach will understate the statistical uncertainty around the coefficient estimates. The Bayesian framework of this paper allows me to take account for this uncertainty, because it provides the posterior distribution of the dependent variable. The following algorithm correctly accounts for the total uncertainty that results from the first-stage estimation of the DTSM and from the second-stage estimation of the responses to macro news:

1. Obtain parameters and latent factors for the current draw from the joint posterior.

2. Calculate the time series of fitted and risk-neutral futures rates for all contracts.

3. For both fitted and risk-neutral rates and for each contract, consider the regression of the daily changes on $s_t^{(1)}, s_t^{(2)}, \dots, s_t^{(r)}$, the surprise components on day t for each of r different macroeconomic data releases. Including all releases in the regression partials out the impact of releases that occur on the same day. Use standard Bayesian regression to obtain parameter estimates: Given independent normal priors for the regression parameters with mean zero and large variance, and a diffuse inverse-gamma prior for the error variance, calculate mean and standard deviation of the posterior distribution for the regression parameters from the conjugate normal posterior.

4. Obtain a draw from the posterior distribution of the regression coefficients and save it.

5. Unless the end of the MCMC sample is reached, return to step 1.

This provides a distribution of response coefficients for fitted and risk-neutral rates for each contract to any of the r news releases. I consider those macroeconomic releases that are part of the employment report of the Bureau of Labor Statistics: nonfarm payroll employment, the unemployment rate, and hourly earnings. As is common in this literature, the surprise component in the release is calculated as the difference between the actually released number and the value expected by the market, standardized to have unit variance. To measure the market expectation I take the median market forecast, which is compiled by Money Market Services the Friday before the announcement.

5.2 Results

The third panel of Figures 1 to 4 shows how interest rates and policy expectations respond to a one-standard-deviation in nonfarm payroll employment. In addition to the responses of actual and fitted futures rates, they include point and interval estimates of the responses of risk-neutral rates.

As evident from Figure 1, according to the unrestricted model policy expectations react to macro news only at the short end; around a maturity of two years the responses become insignificant and for three to four years they are essentially zero. This would indicate that forward premia account for most of the procyclical reponse of long-term interest rates to economic news. The conclusions from the point estimates are in line with the findings of Beechey (2007), who shows that the Kim-Wright model implies a large contribution of term premia to the responses of long-term rates to payroll news. Beechey calls this finding puzzling, since the strong procyclical movements of term premia stands in contrast to the common belief that risk premia are countercyclical. However, the point estimates here do not need to be regarded as puzzling: First, there is large statistical uncertainty around them. Second, they result from one particular model specification, and alternative, restricted specifications might change the economic implications.

The favored model M_2, as shown in Figure 2 implies that the procyclical response of interest rates at 3-4 years horizon is almost entirely due to revisions of short rate expectations, with much smaller statistical uncertainty around the point estimates. Forward term premia at these horizons show a small and barely significant response to payrolls according to this model. Comparing the economic implications of all seven models in Figure 3, we again see important differences across specifications, with several of the models attributing an important role to the expectations component.

Appropriate inference about the response of policy expectations to macro news needs to take into account specification uncertainty. The results for BMA are shown in the bottom panel of Figure 4. According to the point estimates, short rate expectations are the more important driving force of the procyclical response of the term structure to payroll surprises. Across all maturities, forward term premia seem to show a comparably small reaction to the news. Thus the implications of BMA are similar to those of the favored model M_2, but this was not at all obvious a priori.

5.3 Interpretation

The results in this section have made very clear the different economic implications of the estimated models. Estimates of the relative importance of expectations and premia for the procyclical response of interest rates depend crucially on risk price restrictions.

The absence of restrictions on the risk-adjustment in a standard DTSM leads to a disconnect between monetary policy and long rates: movements in forward rates at long horizons do not seem to tell us anything about policy expectations. Consequently, short rate expectations do not respond to macroeconomic data surprises at even moderate horizons, and the procyclical response of interest rates to macroeconomic news is attributed to term premium movements. In addition, the disconnect between $\mathbb{P}$ and $\mathbb{Q}$ measure leads to huge statistical uncertainty.

Under those restricted specifications that the data call for, this procyclical term premium response, which has previously puzzled researchers, disappears. Instead, revisions to expectations of future policy rates are the main driving force of the response of the term structure. Furthermore, estimation uncertainty is reduced by tying the revision of real-world expectations to the cross sectional changes.

6 Return predictability

The estimation results in this paper suggest more stable term premia than are implied by a model with few or no restrictions on risk pricing. Even the model specifications with constant premia (M_5), which corresponds to the EH, obtains some empirical support. On the other hand a large body of empirical literature has documented important time variation in expected bond returns and term premia in long term interest rates, most prominently Fama and Bliss (1987), Campbell and Shiller (1991), and Cochrane and Piazzesi (2005). Are the estimation results in this paper consistent with model-free regression results? To answer this question, I evaluate the model's implications by first estimating predictability regressions in the data, and then checking whether the results can be reproduced using data that is simulated from the model.

The regressions I estimate are most closely related to the ones in Fama and Bliss (1987), who assess predictability of annual changes in forward rates using monthly observations of the term structure. The specification used here is

$$ED_{m+12}^{(i-4)} - ED_m^{(i)} = \alpha + \beta' PC_m + u_m, \tag{21}$$

where time is now indexed by months. I use end-of-month observations, which leaves me with 198 data points, from January 1990 to July 2006. The dependent variable is the realized return on a position in Eurodollar futures contract i entered at t, held for one year and thus rolled over into a shorter contract four times. The independent variables, PC_m, are the first three principal components of the T_m monthly observations of the 16 futures rates. The regression is estimated for returns on contracts $i = 8, 12, 16$, corresponding to initial maturities of two, three, and four years, as well as for the average return across all contracts (from $i = 5$ to $i = 16$). The regression R^2 measures the predictability of the returns. Under the EH there is no predictability using time t information, so that in population $\beta = 0$.

First, the equation (21) is estimated on the actual data. Then, to assess the implications of the estimated DTSM for return predictability, the regression is estimated on data that is simulated from the model. For each estimated model specification and for the BMA estimates, I simulate 5000 time series of the same length as the data sample. For each of these simulations, I use a different draw from the joint posterior distribution of the parameters and latent factors, so that the results will reflect not only sampling uncertainty but also parameter uncertainty. The simulated factors are initialized at their unconditional mean. Based on the simulated interest rates, I construct monthly observations and annual returns in the exact same fashion as in the data, and estimate the predictability regression. This provides me with a distribution of R^2 and I report the mean and the 2.5- and 97.5-percentiles.

Table 3 shows the results. The predictability in the data is economically important, with R^2 around 40%. This indicates a strong departure from the EH, and points to time variation in term premia, in accordance with many previous studies.

The results using simulated data are consistent with the return predictability found in actual data: For models M_1, M_3 and M_7 the point estimates are around 40%, and for all models, the 95%-credibility intervals contain the empirical R^2. Evidently, the model-implied predictability in short samples, taking into account sampling and parameter uncertainty, is not at odds with the predictability in the data. This might come as a surprise, given the relatively stable term premia that the model estimates imply. My results show that we do not need a large amount of variability in term premia to reproduce regression-based evidence about return predictability.

One important question remains: How can it be that model M_5, which imposes the EH, implies significant return predictability? The answer is small-sample bias. A very long simulated interest rate sample, using the point estimates of the model parameters, leads to an estimated R^2 of essentially zero (results not shown).[9] Hence, in population the predictability is of course zero, in accordance with a time-constant λ_t. It is only for the smaller simulated samples, that return predictability seems to be present: The persistent predictors in the regression induce the well-known Stambaugh (1999) bias which typically leads to higher estimated predictability in small samples than in population. In light of this result, the empirical R^2 are almost certainly upward biased and should be taken with a grain of salt. Generally, even after Bekaert et al. (1997), the role of Stambaugh bias in generating excessive estimated return predictability remains an issue deserving further investigation. Estimated DTSMs are a useful vehicle to analyze this question.

7 Conclusion

This paper has introduced an econometric framework to estimate DTSMs by imposing restrictions on risk pricing. It enables the empirical researcher to systematically choose among a large set of restrictions and to impose parsimony on an otherwise overparameterized model. The issues of model selection and model uncertainty are dealt with in a statistically rigorous manner. 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 some-

[9]This exercise only makes sense for the model specifications without a stochastic trend, and for the model with constant risk prices—for the other models the limit of the least-squared estimator diverges.

thing 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, and shows that it can be greatly reduced by constraining the risk adjustment. The framework is more generally applicable beyond the class of models and specific restrictions that the paper focuses on.

For the specific data set and model under consideration, the data support tight restrictions on risk prices. These restrictions change the economic implications in important ways: Term premia and risk prices are more stable than in the absence of such restrictions. At high frequencies, changes in forward rates out to moderately long maturities are mainly driven by short rate expectations. The procyclical responses to macro news are due to changing policy expectations, and I do not find the puzzling and implausible procyclical term premium movements that are implied by unrestricted term structure models. At lower frequencies, term premia remain countercyclical but become less variable risk price restrictions are imposed.

One promising extension of the current framework is to consider restrictions other than zeros on λ_1, such as rank restrictions. This could provide a new perspective on the restrictions that Cochrane and Piazzesi (2005, 2008) find for return predictability. More generally, enlarging the class of restrictions can open up additional potential for no-arbitrage to pin down term premium estimates.

Macro-finance term structure models are a promising application. 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 overfitted (Kim, 2007). Imposing parsimony by constraining the risk adjustment is particularly important in this context. These models can help answer the questions about which macroeconomic variables drive variation in risk premia and which macroeconomic shocks carry risk, described as "the Holy Grail of macro-finance" by Cochrane (2007, p. 281). Recently some important steps have been made in this direction, in particular by incorporating unspanned macro risks (Joslin et al., 2010). My framework can add to this by allowing researchers to rigorously test restrictions on risk pricing, and to assess the empirical support and economic implications of alternative specifications of the joint macro-finance dynamics. Model uncertainty is a major issue in this context and needs to be accounted for. Performing inference on risk premia in this way seems to be a promising route towards new answers about the relation between the macroeconomy and risk pricing in bond markets.

References

Ang, A., Piazzesi, M., 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics 50, 745–787.

Backus, D., Wright, J., 2007. Cracking the conundrum. Brookings Papers on Economic Activity 2007, 293–329.

Bartlett, M.S., 1957. A comment on d. v. lindley's statistical paradox. Biometrika 44, 533–534.

Bauer, M.D., Rudebusch, G.D., Wu, J.C., 2011. Unbiased Estimation of Dynamic Term Structure Models. Working Paper 2011-12. Federal Reserve Bank of San Francisco.

Beechey, M., 2007. A closer look at the sensitivity puzzle: the sensitivity of expected future short rates and term premia to macroeconomic news. Finance and Economics Discussion Series 2007-06. Board of Governors of the Federal Reserve System (U.S.).

Bekaert, G., Hodrick, R., Marshall, D., 1997. On biases in tests of the expectations hypothesis of the term structure of interest rates. Journal of Financial Economics 44, 309–348.

Brooks, S.P., Roberts, G.O., 1998. Convergence assessment techniques for markov chain monte carlo. Statistics and Computing 8, 319–335.

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

Carlin, B.P., Chib, S., 1995. Bayesian model choice via markov chain monte carlo methods. Journal of the Royal Statistical Society. Series B (Methodological) 57, 473–484.

Carter, C.K., Kohn, R., 1994. On gibbs sampling for state space models. Biometrika 81, 541–553.

Chib, S., 1995. Marginal likelihood from the gibbs output. Journal of the American Statistical Association 90, 1313–1321.

Chib, S., Ergashev, B., 2009. Analysis of Multifactor Affine Yield Curve Models. Journal of the American Statistical Association 104, 1324–1337.

Christensen, J.H., Diebold, F.X., Rudebusch, G.D., 2011. The affine arbitrage-free class of nelson-siegel term structure models. Journal of Econometrics 164, 4–20.

Clyde, M., George, E.I., 2004. Model uncertainty. Statistical Science 19, 81–94.

Cochrane, J., 2007. Commentary on "macroeconomic implications of changes in the term premium". Federal Reserve Bank of St. Louis Review , 271–282.

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

Cochrane, J.H., Piazzesi, M., 2008. Decomposing the Yield Curve. unpublished manuscript.

Cowles, M.K., Carlin, B.P., 1996. Markov chain monte carlo convergence diagnostics: A comparative review. Journal of the American Statistical Association 91, 883–904.

Dai, Q., Le, A., Singleton, K.J., 2006. Discrete-time Dynamic Term Structure Models with Generalized Market Prices of Risk. working paper.

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

Dai, Q., Singleton, K.J., 2002. Expectation puzzles, time-varying risk premia, and affine models of the term structure. Journal of Financial Economics 63, 415–441.

Dellaportas, P., Forster, J.J., Ntzoufras, I., 2002. On bayesian model and variable selection using mcmc. Statistics and Computing 12, 27–36.

DiCiccio, T.J., Kass, R.E., Raftery, A., Wasserman, L., 1997. Computing bayes factors by combining simulation and asymptotic approximations. Journal of the American Statistical Association 92, 903–915.

Duffee, G.R., Stanton, R.H., 2004. Estimation of Dynamic Term Structure Models. working paper. Haas School of Business.

Dybvig, P., Ingersoll Jr, J., Ross, S., 1996. Long forward and zero-coupon rates can never fall. Journal of Business 69, 1–25.

Fama, E.F., Bliss, R.R., 1987. The information in long-maturity forward rates. The American Economic Review 77, 680–692.

Fleming, M.J., Remolona, E.M., 1999. The term structure of announcement effects. BIS Working Papers 71. Bank for International Settlements.

Gamerman, D., Lopes, H.F., 2006. Markov Chain Monte Carlo. Chapman & Hall/CRC. second edition.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7, 457–472.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, in: In Bayesian Statistics 4, Oxford University Press. pp. 169–193.

Green, P.J., 1995. Reversible jump markov chain monte carlo computation and bayesian model determination. Biometrika 82, 711–732.

Gürkaynak, R.S., Sack, B.P., Swanson, E.T., 2005. The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models. American Economic Review 95, 425–436.

Hamilton, J.D., Wu, J.C., 2010. Identification and Estimation of Affine Term Structure Models. working paper. University of California, San Diego.

Jegadeesh, N., Pennacchi, G.G., 1996. The behavior of interest rates implied by the term structure of eurodollar futures. Journal of Money, Credit and Banking 28, 426–46.

Joslin, S., Priebsch, M., Singleton, K.J., 2010. Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks. working paper.

Joslin, S., Singleton, K.J., Zhu, H., 2011. A new perspective on gaussian dynamic term structure models. Review of Financial Studies 24, 926–970.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association 90, 81–94.

Kim, C.J., Nelson, C.R., 1999. State-Space Models with Regime-Switching. The MIT Press.

Kim, D.H., 2007. Challenges in macro-finance modeling. BIS Working Papers 240. Bank for International Settlements.

Kim, D.H., Orphanides, A., 2005. Term Structure Estimation with Survey Data on Interest Rate Forecasts. Computing in Economics and Finance 2005 474. Society for Computational Economics.

Kim, D.H., Orphanides, A., 2007. The bond market term premium: what is it, and how can we measure it? BIS Quarterly Review .

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

Kozicki, S., Tinsley, P., 2001. Shifting endpoints in the term structure of interest rates. Journal of Monetary Economics 47, 613–652.

Nelson, C.R., Siegel, A.F., 1987. Parsimonious modeling of yield curves. Journal of Business 60, 473–489.

Piazzesi, M., Swanson, E.T., 2008. Futures prices as risk-adjusted forecasts of monetary policy. Journal of Monetary Economics 55, 677–691.

Raftery, A.E., Lewis, S., 1992. How many iterations in the gibbs sampler?, in: In Bayesian Statistics 4, Oxford University Press. pp. 763–773.

Rudebusch, G., 2007. Commentary on "cracking the conundrum". Brookings Papers on Economic Activity 38, 317–325.

Rudebusch, G.D., Sack, B.P., Swanson, E.T., 2007. Macroeconomic implications of changes in the term premium. Federal Reserve Bank of St. Louis Review 89, 241–269.

Stambaugh, R.F., 1999. Predictive regressions. Journal of Financial Economics 54, 375–421.

Svensson, L., 1994. Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994. NBER Working Papers 4871. National Bureau of Economic Research, Inc.

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 $\mu - \Sigma\lambda_t + \Phi X_t = (\mu - \Sigma\lambda_0) + (\Phi - \Sigma\lambda_1)X_t$ and variance $\Sigma\Sigma'$.

Note that since X_t follows a Gaussian vector autoregression under $\mathbb{Q}$ the model is in the $DA_0^{\mathbb{Q}}(N)$ class of Dai et al. (2006).

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

$$\varepsilon_t^{\mathbb{Q}} = \varepsilon_t + \lambda_{t-1}. \tag{22}$$

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 The arbitrage-free Nelson-Siegel model

In a canonical DTSM with latent factors the role of each factor is a priori left unidentified, which makes estimation and economic interpretation difficult. In contrast, the yield-curve parameterization of Nelson and Siegel (1987), generalized by Svensson (1994), is widely used by practitioners because it posits a simple factor structure with level, slope, and curvature. The dynamic version of the original Svensson-Nelson-Siegel forward rate curve,

$$f_t^n = X_t^{(1)} + e^{-\lambda n}X_t^{(2)} + \lambda n e^{-\lambda n}X_t^{(3)},$$

is implied by a continuous-time three-factor affine DTSM with a specific choice for the risk-neutral dynamics, the Arbitrage-Free Nelson-Siegel (AFNS) model of CDR. This paper will use the discrete-time analogue of the AFNS model, which is given by the following specification:

$$\delta_0 = 0,\ \delta_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix},\ \mu^{\mathbb{Q}} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},\ \Phi^{\mathbb{Q}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \rho & 1-\rho \\ 0 & 0 & \rho \end{pmatrix}, \tag{23}$$

where the parameter ρ is restricted to be less than one in absolute value. This implies a Nelson-Siegel-type forward rate curve given by

$$f_t^n = X_t^{(1)} + \rho^n X_t^{(2)} + n(1-\rho)\rho^{n-1} X_t^{(3)}. \tag{24}$$

There is no convexity/yield-adjustment term as in CDR since I consider $f_t^n = E_t^{\mathbb{Q}}(r_{t+n})$. For these forward rates there is no difference between dynamic Nelson-Siegel and arbitrage-free Nelson-Siegel (see CDR).

The AFNS specification amounts to imposing three overidentifying restrictions on a canonical affine Gaussian DTSM, including a unit eigenvalue of $\Phi^{\mathbb{Q}}$.[10] These restrictions are attractive in the current context for two reasons. First, the factors are a priori identified as level, slope, and curvature. Thus the AFNS model allows for an economic interpretation of risk premia – one can learn about whether level or slope risk is priced, and what type of yield curve movements change risk premia. This can provide clues about the role of different macroeconomic shocks (Cochrane and Piazzesi, 2008). The second advantage is the presence of a unit root under the risk-neutral measure $\mathbb{Q}$. Because $\Phi = \Phi^{\mathbb{Q}} + \Sigma\lambda_1$, the zero restrictions on λ_1, which are chosen using a model selection framework, determine whether there is a unit root under $\mathbb{P}$. Thus, instead of a priori specifying a stationary or integrated model, which have dramatically different implications for term premia, the data will decide whether the short rate should have a unit root. In this way the current framework deals with the near-integrated behavior of interest rates.

C MCMC algorithm: drawing parameters and latent factors

This section describes how each of the five parameter blocks of the DTSM is sampled for any of the block-wise M-H algorithms used in this paper.

Drawing the latent factors (X)

Given θ, draws for the latent factors are obtained by means of the FFSB algorithm developed by Carter and Kohn (1994): Kalman filtering delivers an initial time series of the factors,

[10] Strictly speaking there is an arbitrage opportunity in the presence of a unit root under $\mathbb{Q}$: The long forward rate $\lim_{n\to\infty} f_t^n = X_t^{(1)}$ exists and can freely move around, whereas no-arbitrage requires that long rates can never fall, as shown by Dybvig et al. (1996). However, since there are no tradable bonds or futures with sufficiently long maturity to take advantage of this arbitrage opportunity, it is practically irrelevant. Also note that if the largest root is taken to be $1-\epsilon$ with very small ϵ the model is empirically indistinguishable from one with a unit root and does not violate the Dybvig-Ingersoll-Ross theorem – this point was noted by CDR.

and then one iterates backward from the last observation and successively draws values for the latent factors conditional on the following observation. A detailed explanation of the algorithm can be found in (Kim and Nelson, 1999, chap. 8).

Drawing the risk-neutral dynamics (ρ)

The risk-neutral dynamics are determined by ρ for the AFNS specification. Notably, the prices of risk are taken as given when drawing this block, so drawing ρ affects not only the cross-sectional loadings but also the transition matrix of the physical dynamics, Φ. The conditional posterior for ρ is

$$P(\rho|\theta_-, X, Y) \propto P(Y|\theta, X)P(X|\theta)P(\theta) \tag{25}$$

where θ_- denotes all other parameters except ρ. Since one cannot sample directly from this distribution—ρ enters the density in a complicated way—I employ an M-H step. Since only proposal draws that are close to the value from the previous iteration have a chance of being accepted, a random walk (RW) step is the natural choice: In iteration g I draw the parameter according to $\rho^{(g)} = \rho^{(g-1)} + \zeta_\rho t_4$, a fat-tailed RW with t_4 being a random variable with a t-distribution with four degrees of freedom, and ζ_ρ being a scale factor used to tune the acceptance probability to be around 20-50%, which is the recommended range in the MCMC literature (see Gamerman and Lopes, 2006, p. 196). Since the proposal density is symmetric for a RW step, the acceptance probability is given by

$$\alpha(\rho^{(g-1)}, \rho^{(g)}) = \min\left\{\frac{P(Y|\rho^{(g)}, \theta_-, X)P(X|\rho^{(g)}, \theta_-)P(\rho^{(g)}, \theta_-)}{P(Y|\rho^{(g-1)}, \theta_-, X)P(X|\rho^{(g-1)}, \theta_-)P(\rho^{(g-1)}, \theta_-)}, 1\right\}. \tag{26}$$

For the case that the prior restrictions ($0 < \rho < 1$ and non-explosive Φ) are satisfied—the acceptance probability is zero otherwise—this is simply equal to the ratio of the likelihoods of the new draw relative to the old draw, or one, whichever is smaller.

Drawing the risk sensitivity parameters (λ_0 and λ_1)

For the conditional posterior distribution of the risk sensitivity parameters we have

$$\begin{aligned} P(\lambda_0, \lambda_1|\theta_-, X, Y) &\propto P(Y|\theta, X)P(X|\theta)P(\theta) \\ &\propto P(X|\theta)P(\theta), \end{aligned}$$

where θ_- denotes all parameters except for λ_0 and λ_1. The second line follows from the fact that the $\mathbb{Q}$-likelihood does not depend on the prices of risk. Again a M-H step is necessary since the conditional posterior distribution is known only up to a constant, and I use RW proposals.

If there are no restrictions imposed on λ_1 I draw λ_0 and each column of λ_1 separately. The innovation for the RW proposal is then a $k \times 1$ vector of independent t_4-distributed innovations. For the case that some elements of λ_1 are restricted to zero, I draw each non-zero element of λ_0 and λ_1 separately, using a univariate RW with t_4-distributed innovations. In both cases the scale factors are adjusted to tune the acceptance probabilities.

After obtaining the candidate draw, the restriction that the physical dynamics are non-explosive is checked, and the draw is rejected if the restriction is violated. Otherwise the acceptance probability for the draw is calculated as the minimum of one and the ratio of the likelihoods of the latent factors times the ratio of the priors for the new draw relative to the old draw.

Drawing the shock covariance matrix ($\Sigma\Sigma'$)

For the conditional posterior of Σ we have

$$\begin{aligned} P(\Sigma|\theta_-, X, Y) &\propto P(Y|\theta, X)P(X|\theta)P(\theta) \\ &\propto P(X|\theta)P(\theta), \end{aligned}$$

where θ_- denotes all parameters except Σ, since by the absence of convexity effects the shock variances do not enter the arbitrage-free loadings and thus the likelihood of the data is independent of Σ. Since I need successive draws of Σ to be close to each other—otherwise the acceptance probabilities will be too small—independence Metropolis is not an option. I found element-wise RW M-H to not work particularly well. A better alternative in terms of efficiency and mixing properties is to draw the entire matrix Σ in one step. I choose a proposal density for $\Sigma\Sigma'$ that is IW with mean equal to the value of the previous draw and scale adjusted to tune the acceptance probability, which is equal to

$$\alpha(\Sigma\Sigma'^{(g-1)}, \Sigma\Sigma'^{(g)}) = \min\left\{ \frac{P(X|\Sigma^{(g)}, \theta_-)P(\Sigma\Sigma'^{(g)}, \theta_-)q(\Sigma\Sigma'^{(g)}, \Sigma\Sigma'^{(g-1)})}{P(X|\Sigma^{(g-1)}, \theta_-)P(\Sigma\Sigma'^{(g-1)}, \theta_-)q(\Sigma\Sigma'^{(g-1)}, \Sigma\Sigma'^{(g)})}, 1 \right\}. \quad (27)$$

Here $q(A, B)$ denotes the transition density, which in this case is the density of an IW distribution with mean A.

Drawing the measurement error variance (σ_e^2)

The variance of the measurement error can be drawn directly from its conditional posterior distribution, i.e., I use standard Gibbs-sampling for this step. The reason is that conditional on the latent factors, the other parameters, and the data, the measurement errors can be viewed as regression residuals, and the IG distribution is the natural conjugate prior. Since I impose the variance to be the same across the m measurement equations, the residuals from all measurement equations are pooled. The conditional posterior for σ_e^2 is the natural conjugate IG distribution.

D Convergence diagnostics

After having obtained a sample using an MCMC algorithm, convergence characteristics of the chain need to be checked to verify that the draws are from a distribution that is close to the invariant distribution of the Markov chain. Put differently, the question is whether the draws are from a chain that is mixing well.

A very simple and intuitive check of whether the chain is behaving well is to look at trace plots, i.e., plots of the successive draws for each parameter. In addition to this visual inspection, one can calculate several convergence diagnostics.[11] The autocorrelations of the draws for each parameter give a first indication of how well the chain is mixing. A commonly employed method to assess convergence, developed by Raftery and Lewis (1992), is to calculate the minimum burn-in iterations and the minimum number of runs required to estimate quantiles of the posterior distribution with a certain desired precision. Moreover one can diagnose situations where the chain has not converged, as suggested by Geweke (1992), by testing for equality of means over different subsamples. Gelman and Rubin (1992) have suggested to run parallel chains from different starting values and compare within-chain to between-chain variance, which is a simple and effective way to check for convergence. I have applied these and some other convergence checks to find out how many iterations are needed for approximate convergence and how the algorithm can be tuned in order to improve mixing. The general conclusion is that a lot of iterations are needed because ρ and the elements of λ_0 and λ_1 traverse the parameter space only very slowly. This is a result of the small innovations in the RW proposals, which are necessary to obtain reasonable acceptance probabilities. Therefore I choose long burn-in samples ($B = 20,000$) and a large number of iterations ($G \cdot s = 200,000$). Under this configuration the graphs and diagnostic statistics indicate that the chain has converged.[12]

E Money market futures

Money market futures have a payoff that is proportional to the difference between the average short rate over a future time horizon and the contractual rate. Eurodollar futures contracts have payoffs that are based on the three-month LIBOR rate on the settlement day.[13] Before turning to Eurodollar futures, it is useful to focus on a simpler, hypothetical contract, which pays off according to the future short rate. The contractual rate will be $f_t^n = E_t^{\mathbb{Q}}(r_{t+n})$. We have

$$f_t^n = \delta_0 + \delta_1' E_t^{\mathbb{Q}}(X_{t+n}) = \delta_0 + \delta_1' \left[\sum_{i=0}^{n-1} (\Phi^{\mathbb{Q}})^i \mu^{\mathbb{Q}} + (\Phi^{\mathbb{Q}})^n X_t \right] = A_n + B_n' X_t, \tag{28}$$

$$A_n = \delta_0 + \delta_1' \left[\sum_{i=0}^{n-1} (\Phi^{\mathbb{Q}})^i \mu^{\mathbb{Q}} \right], \; B_n' = \delta_1' (\Phi^{\mathbb{Q}})^n. \tag{29}$$

Note that (in addition to δ_0 and δ_1) the parameters describing the $\mathbb{Q}$-dynamics of the states, $\mu^{\mathbb{Q}}$ and $\Phi^{\mathbb{Q}}$, determine the loadings A_n and B_n, while μ and Φ do not appear.

The rate f_t^n closely corresponds to a one-period forward rate, i.e., the rate that can be contracted at t for a loan from $t+n$ to $t+n+1$ by entering the appropriate bond positions.

[11] For surveys on convergence diagnostics see Cowles and Carlin (1996) and Brooks and Roberts (1998).

[12] There certainly remains room for improvement of the algorithm. In particular one could use methods for speeding up convergence, such as the hit-and-run algorithm, adaptive direction sampling, or simulated annealing (see Gamerman and Lopes, 2006, Section 7.4).

[13] For detailed information on Eurodollar futures contracts please refer to the Chicago Mercantile Exchange's web site at http://www.cmegroup.com/trading/interest-rates/stir/eurodollar.html (accessed 11/15/2011).

This rate is equal to

$$\log \frac{P_t^n}{P_t^{n+1}} = \log \frac{E_t^{\mathbb{Q}} \exp(-r_t - r_{t+1} - \ldots - r_{t+n-1})}{E_t^{\mathbb{Q}} \exp(-r_t - r_{t+1} - \ldots - r_{t+n})} = f_t^n + convexity, \tag{30}$$

where P_t^n is the time-t price of a discount bond with n days until maturity. The affine-model loadings for such a forward rate can be found in the literature, for example in Cochrane and Piazzesi (2008). The difference between f_t^n and a real forward rate is that Jensen inequality terms resulting from convexity effects are absent in the formula for f_t^n. For simplicity, I will in the following refer to f_t^n as a forward rate.

If the marginal investor were risk-neutral, f_t^n would be equal to the expected future short rate under $\mathbb{P}$. For this "risk-neutral forward rate" we have

$$\tilde{f}_t^n = \mathrm{E}_t(r_{t+n}) = \delta_0 + \delta_1' \mathrm{E}_t(X_{t+n}) = \tilde{A}_n + \tilde{B}_n' X_t, \tag{31}$$

$$\tilde{A}_n = \delta_0 + \delta_1' \left[\sum_{i=0}^{n-1} \Phi^i \mu \right], \; \tilde{B}_n' = \delta_1' \Phi^n. \tag{32}$$

The difference between f_t^n and $\tilde{f}_t^n$ is a measure of the term premium. This *forward risk premium* will be denoted by Π_t^n.

Eurodollar futures are simply averages of the above forward rates over the relevant horizons. The reason is that the settlement rate can be taken as the average expected short rate (under $\mathbb{Q}$) for the three-month period starting on the settlement day: $S_t = N^{-1} \sum_{h=0}^{N-1} \mathrm{E}_t^{\mathbb{Q}}(r_{t+h})$, where N is the number of days in the quarter, taken to be 91.[14] Eurodollar futures contracts involve no cost today and have a payoff proportional to the difference between the contractual rate and the settlement rate.[15] For the Eurodollar futures contract that settles at the end of quarter i, where $i = 1$ corresponds to the current quarter, the pricing equation is:

$$0 = \mathrm{E}_t^{\mathbb{Q}}(ED_t^{(i)} - S_{T(i,t)}), \tag{33}$$

where $ED_t^{(i)}$ is the futures rate and $T(i,t)$ denotes the settlement day that corresponds to contract i on day t. Settlement takes place on the last day of the quarter, therefore $T(i,t) = t + iN - d(t)$, where $d(t)$ is the day within the quarter of calendar day t. The futures rate is

[14] The LIBOR rate is usually very closely related to the average expected effective federal funds rate. The difference between the two, which is measured by the so-called LIBOR-OIS spread, stems from a small term premium and a credit risk premium due to the three-month commitment at a specific rate with a particular counterparty when lending at LIBOR. This spread was very small (around 8 basis points) and showed little variability throughout the period of this data set, which ends before the start of the recent financial turmoil.

[15] I abstract from the fact that in reality payments are made every day because of marking-to-market. Evidence in Piazzesi and Swanson (2008) indicates that this effect is likely to be negligible in this context.

thus given by

$$ED_t^{(i)} = \mathrm{E}_t^{\mathbb{Q}}(S_{T(i,t)}) = N^{-1} \sum_{n=iN-d(t)}^{(i+1)N-d(t)-1} \mathrm{E}_t^{\mathbb{Q}} r_{t+n} = N^{-1} \sum_{n=iN-d(t)}^{(i+1)N-d(t)-1} (A_n + B_n' X_t) \quad (34)$$

$$= A_i^{ED} + B_i^{ED\prime} X_t. \quad (35)$$

Note that the futures rate is simply an average of the relevant forward rates. The scalar A_i^{ED} and the vector B_i^{ED} are the averages of A_n and B_n, respectively, over the relevant horizons.[16]

The pricing formula derived here differs from Jegadeesh and Pennacchi (1996) because these authors treat LIBOR as the yield on a hypothetical three-month bond. In light of the fact that the LIBOR rate is set based on a survey of intrabank rates, which are via no-arbitrage directly determined by risk-adjusted monetary policy expectations, the approach here seems preferable.

If market participants were risk-neutral, the futures rates would be equal to expected average future short rates. This risk-neutral futures rate is given by

$$\tilde{ED}_t^{(i)} = \tilde{A}_i^{ED} + \left(\tilde{B}_i^{ED}\right)' X_t \quad (36)$$

where $\tilde{A}_i^{ED}$ and $\tilde{B}_i^{ED}$ are the averages of $\tilde{A}_n$ and $\tilde{B}_n$, respectively, over the relevant horizons. The corresponding forward risk premium for contract i is given by $\Pi_t^{(i)} = ED_t^{(i)} - \tilde{ED}_t^{(i)}$.

[16]The last equality is an approximation due to the fact that instead of having different a_i's and h_i's depending on the day of the quarter, I set $d(t)$ equal to the constant 45 (approximately the average of $d(t)$), which leads to only a very small approximation error and significantly lowers the computational burden.

Table 1: Parameter estimates for unrestricted model

Risk-neutral dynamics				
ρ	.9973	[.9973, .9973]		
Risk sensitivity	λ_0		λ_1	
parameters	constant	level	slope	curvature
level risk	.1210	-.0164	.0103	.0014
	[-.03, .28]	[-.04, .01]	[-.01, .03]	[-.02, .02]
slope risk	**.3042**	**-.0444**	-.0190	**.0296**
	[.15, .46]	[-.07, -.02]	[-.04, .00]	[.01, .05]
curvature risk	.1273	**-.0239**	.0052	-.0009
	[-.02, .28]	[-.05, -.00]	[-.02, .03]	[-.02, .02]
Factor shocks				
SD(level shock)	.0589	[.0572, .0605]		
SD(slope shock)	.0814	[.0787, .0840]		
SD(curv. shock)	.1538	[.1483, .1599]		
Corr(level, slope)	-.7450	[-.7673, -.7208]		
Corr(level, curv.)	.2611	[.2189, .3043]		
Corr(slope, curv.)	-.4217	[-.4613, -.3809]		
Measurement errors				
σ_e^2	.0039	[.0039, .0040]		

Posterior means and 95% Bayesian credibility intervals in squared brackets for model parameters. Estimates of risk sensitivity parameters are boldfaced if the credibility interval does not straddle zero.

Table 2: Model specifications

Model	Freq.	Specification			Eigenv.	LR Rev.	$P_{M_j}^{GVS}$	$P_{M_j}^{Lapl}$	$P_{M_j}^{Cand}$	$P_{M_j}^{RJ}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
		1	1	1						
M_1	0.0%	1	1	1	0.9989	0	0.0%	0.0%	0.0%	0.0%
		1	1	1						
		0	0	0						
M_2	46.3%	1	0	0	1	0.95	61.1%	58.2%	56.1%	48.8%
		0	0	0						
		0	0	0						
M_3	9.4%	1	0	0	1	-0.34	12.4%	10.7%	13.9%	15.9%
		1	0	0						
		0	0	0						
M_4	8.4%	1	0	1	1	0.96	11.1%	9.4%	12.1%	18.2%
		0	0	0						
		0	0	0						
M_5	6.5%	0	0	0	1	1	8.6%	18.4%	12.9%	4.1%
		0	0	0						
		1	0	0						
M_6	2.9%	1	0	0	0.9988	0	3.8%	2.4%	3.3%	6.3%
		0	0	0						
		0	0	0						
M_7	2.3%	1	0	1	1	-0.34	3.0%	0.9%	1.8%	6.6%
		1	0	0						

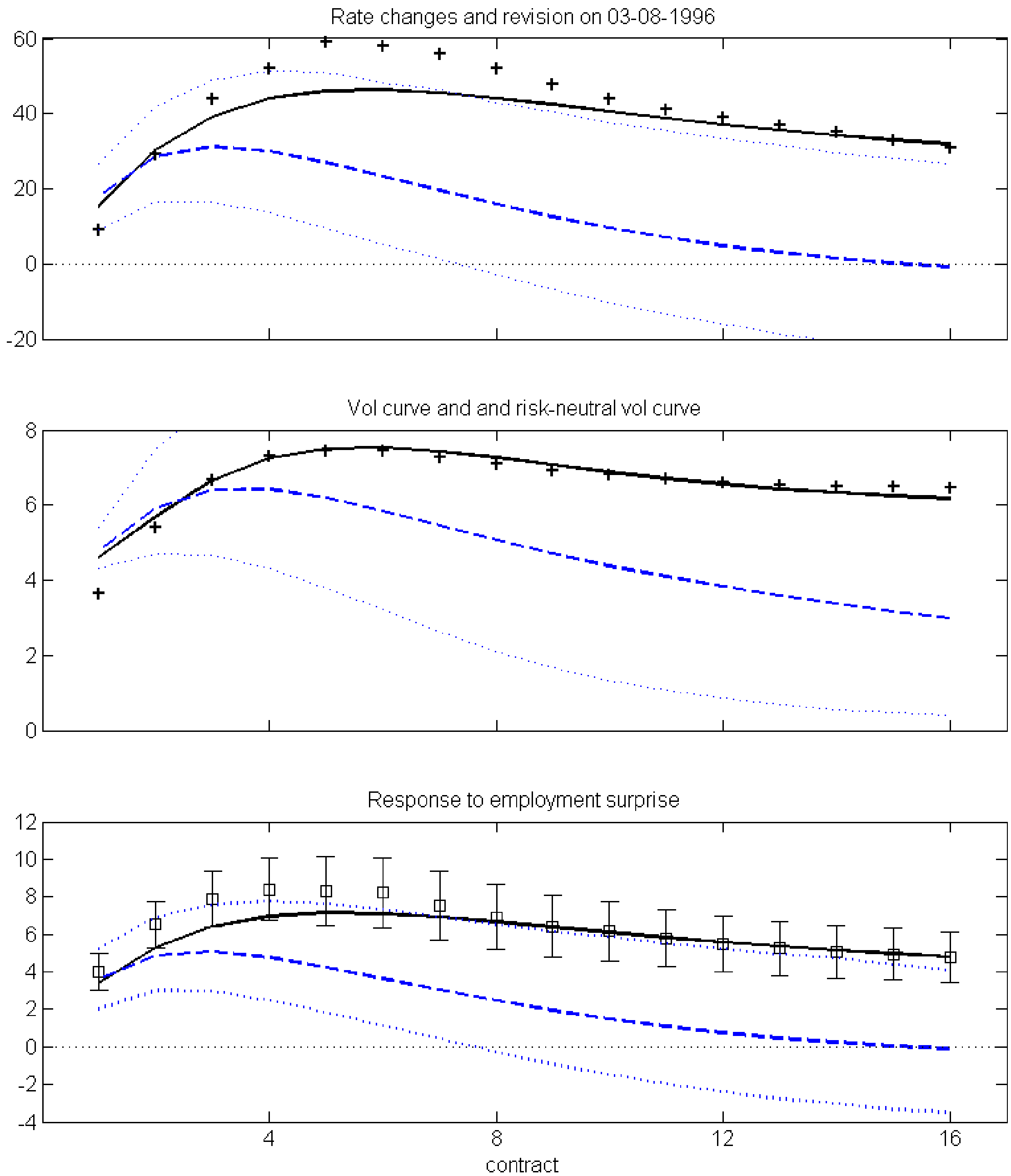
Alternative model specifications and estimated posterior model probabilities. Column two: frequency of specification in GVS algorithm. Columns three to five indicate which elements in the respective columns of λ_1 are restricted (0) and unrestricted (1). Column six: largest eigenvalue of Φ. Column seven: long-run revision of short-run expectations in response to a unit level shock. Columns eight to eleven: estimates of the posterior model probabilities based on rescaled GVS frequencies, Laplace approximation to marginal likelihood, Candidate's estimate of the marginal likelihood, and reversible-jump MCMC.

Table 3: Return predictability

	$i = 8$	$i = 12$	$i = 16$	avg. return
Data	39.9	42.4	44.2	42.5
M_1	44.4 [11.0, 78.7]	42.4 [12.5, 74.7]	40.8 [13.9, 70.6]	45.4 [13.4, 78.0]
M_2	19.0 [1.6, 47.9]	19.0 [1.8, 47.5]	20.0 [1.8, 49.0]	19.9 [1.7, 49.2]
M_3	45.0 [6.7, 85.6]	35.4 [4.7, 72.4]	29.4 [3.7, 59.7]	42.9 [7.0, 81.6]
M_4	18.5 [1.7, 47.4]	18.5 [1.8, 46.1]	19.6 [1.8, 47.9]	19.4 [1.8, 48.6]
M_5	20.6 [1.8, 50.5]	21.5 [2.1, 51.2]	22.4 [2.4, 51.8]	21.5 [2.0, 51.7]
M_6	23.2 [2.5, 51.2]	24.7 [3.4, 51.8]	27.7 [5.3, 53.8]	25.8 [3.5, 53.5]
M_7	46.3 [7.4, 84.9]	35.5 [4.7, 70.0]	28.7 [3.5, 58.2]	44.0 [7.0, 81.0]
BMA	25.3 [2.1, 73.6]	23.2 [2.1, 59.8]	22.6 [2.2, 52.3]	25.6 [2.2, 69.4]

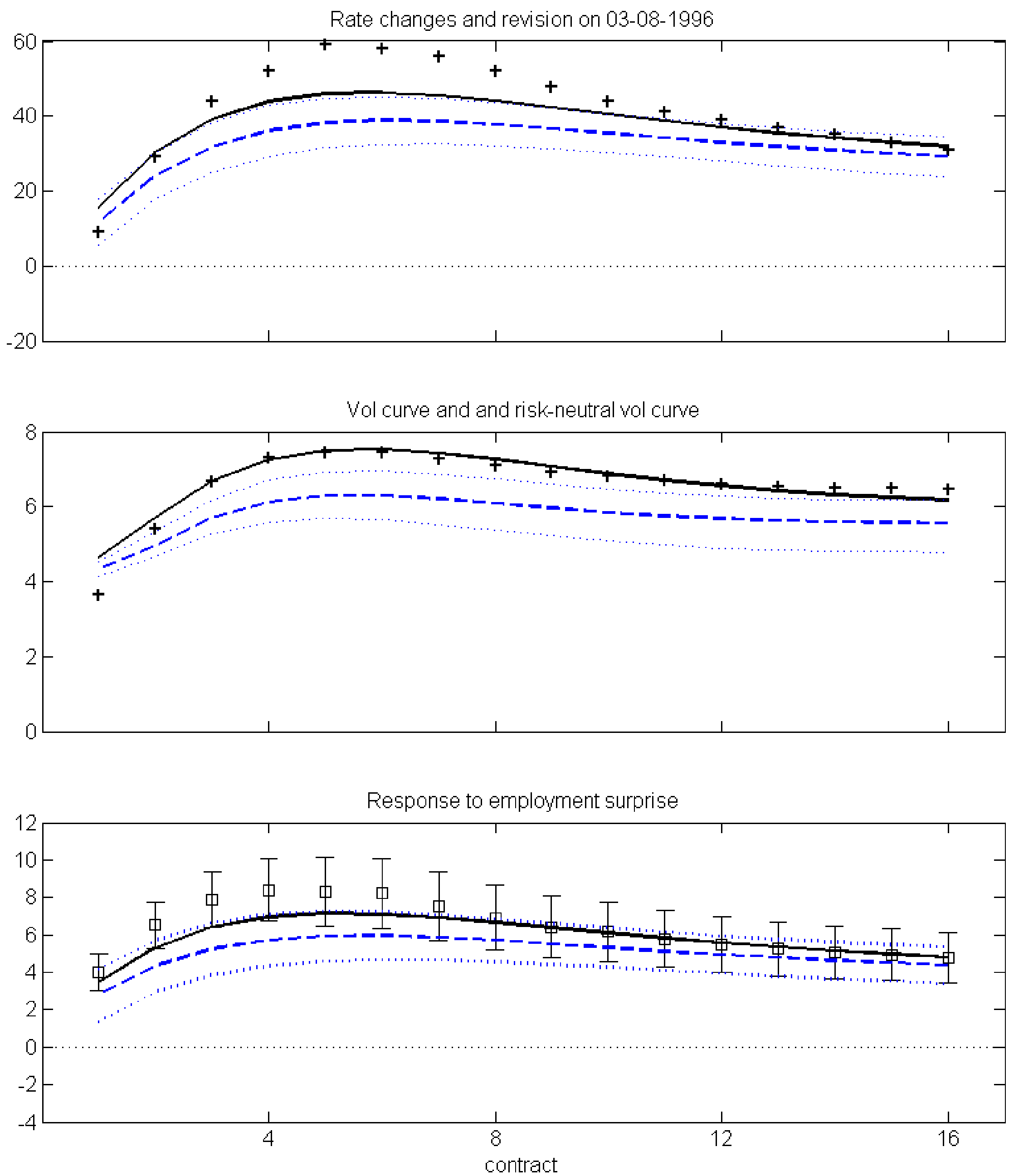
Regression R^2, in percentage points, for predictability regressions, estimated on actual and simulated data.

Figure 1: Implications of unrestricted specification (M_1)



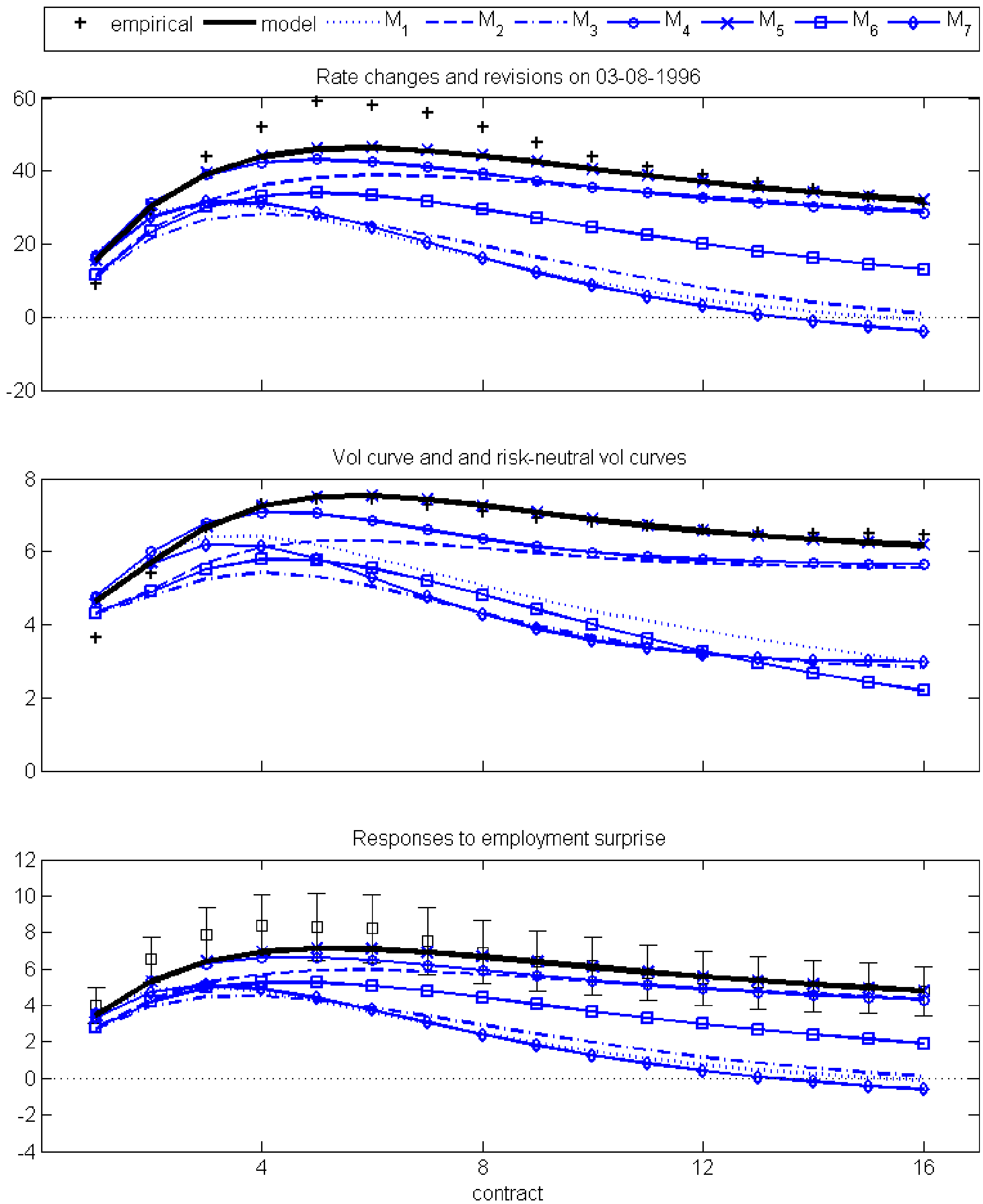
First panel: Empirical (crosses) and model-implied (solid line) rate changes on Mar-08 1996 with estimated changes of risk-neutral rates. Second panel: Sample standard deviations (crosses) and model-implied standard deviations (solid line) for futures rate changes as well as model-implied standard deviations for risk-neutral rate changes. Third panel: Empirical responses of futures rates to a one-standard-deviation payroll surprise with 95% confidence intervals (error-bars). Model-implied responses of futures rates to the news (solid lines) and estimated response of risk-neutral rates to the news. All panels show posterior means (dashed lines) and 95% credibility intervals (dotted lines) for the estimated properties of risk-neutral rates. Units are basis points.

Figure 2: Implications of favored specification (M_2)



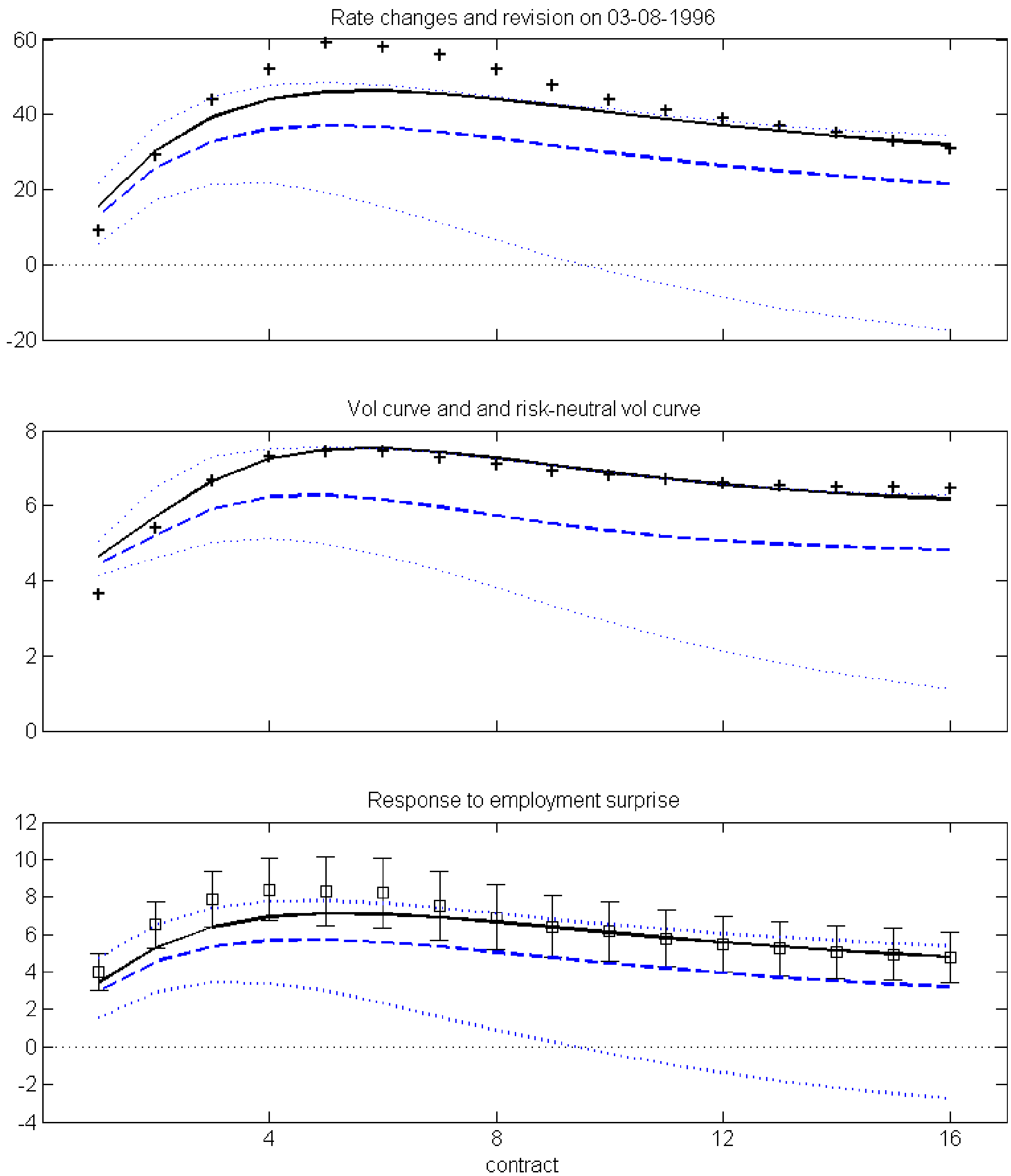
See description of Figure 1.

Figure 3: Comparison of alternative model specifications



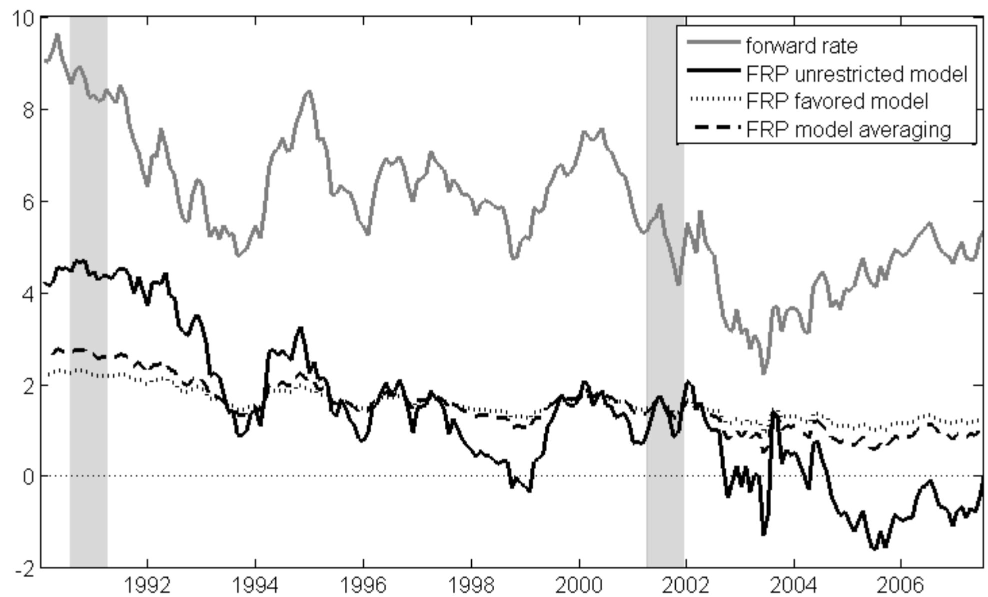
First panel: Empirical (crosses) and model-implied (solid line) rate changes on Mar-08 1996 with estimated changes in risk-neutral rates across models. Second panel: Sample standard deviations (crosses) and model-implied standard deviations (solid line) for futures rate changes as well as alternative risk-neutral vol curves. Third panel: Empirical responses of futures rates to a one-standard-deviation payroll surprise with 95% confidence intervals (error-bars), model-implied responses of futures rates to the news (solid line), and responses of risk-neutral rates across models. Units are basis points.

Figure 4: Implications of Bayesian model averaging



See description of Figure 1.

Figure 5: Time series of 2-year ahead forward risk premium



Time series of futures rate ED8 with three alternative estimates of the corresponding forward risk premium (FRP), resulting from the unrestricted model M_1, the favored model M_2, and Bayesian model averaging. Shaded areas show NBER recessions.