

FEDERAL RESERVE BANK OF SAN FRANCISCO

WORKING PAPER SERIES

## **Term Premia and the News**

Michael D. Bauer  
Federal Reserve Bank of San Francisco

January 2011

Working Paper 2011-03

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

The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Federal Reserve Bank of San Francisco or the Board of Governors of the Federal Reserve System.

# Term Premia and the News\*

Michael D. Bauer<sup>†</sup>

First draft: November 2, 2009

This version: January 18, 2011

## Abstract

How do monetary policy expectations and term premia respond to news? This paper provides new answers to this question by means of a dynamic term structure model (DTSM) in which risk prices are restricted. This leads to more precise and more reliable estimates of expectations and term premium components. I provide a new econometric framework for DTSM estimation that allows the researcher to select plausible constraints from a large set of restrictions, to correctly quantify statistical uncertainty, and to incorporate model uncertainty in the inference about risk pricing. 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 and for essentially all of their procyclical response to macroeconomic news. At both high and low frequencies, term premia are more stable than implied by a DTSM with unconstrained risk prices. The apparent disconnect between long-term rates and policy rates that has puzzled macroeconomists for some time is resolved by appropriately restricting the risk adjustment in models for bond pricing.

*Keywords:* term structure of interest rates, macroeconomic news, term premium, no-arbitrage, market prices of risk, Bayesian inference

*JEL Classifications:* E43, E44, E52, G12

---

\*For their insights and feedback I would like to thank Jens Christensen, John Cochrane, James Hamilton, Bruce Lehmann, Sydney Ludvigson, Ioannis Ntzoufras, Dimitris Politis, Glenn Rudebusch, Ken Singleton, John Taylor, Irina Telyukova, Allan Timmermann, Jonathan Wright, seminar participants at the Banque de France, the International Monetary Fund, the Federal Reserve Bank of New York, the Federal Reserve Bank of San Francisco, the Massachusetts Institute of Technology, the Stanford Institute of Economic Policy Research, the University of California, San Diego, the University of Chicago Graduate School of Business, the University of South Carolina, and Queen Mary College at the University of London, as well as conference participants at the Eastern Economic Association 2010 Annual Conference, the Econometric Society World Congress 2010, and the Spanish Economic Association 2009 Annual Conference. All errors are mine.

<sup>†</sup>Federal Reserve Bank of San Francisco, michael.bauer@sf.frb.org

# 1 Introduction

Policymakers and academic researchers are keenly interested in estimating term premia in government bond yields. To properly interpret the information in the yield curve, identifying the expectations component and the risk premium component 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. In practice the issue is often to decompose changes over short intervals: How did the expected path of monetary policy change since the last FOMC meeting? How did the recent economic data release affect policy expectations and term premia? How can we interpret yesterday’s movements in forward rates? These are pressing questions in policy discussions, and this paper provides an econometric framework to answer them.

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.<sup>1</sup> Unfortunately this presents a tough statistical problem: the near-unit-root behavior of interest rates makes it very difficult to determine their long-term properties, which crucially affect 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. Furthermore, the speed of mean reversion is overestimated because of the well-known bias in estimates of the largest root of a time series that is nearly integrated.

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 (see e.g. Joslin et al., 2010a, JSZ). 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 pre-

---

<sup>1</sup>An interesting alternative to imposing rational expectations is to construct subjective expectations from surveys (Piazzesi and Schneider, 2008).

mia (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 makes sense if one believes in no-arbitrage: cross-sectional and dynamic behavior of interest rates *should* be close to each other. However, most studies using DTSMs impose few or no zero restrictions.<sup>2</sup> 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 key methodological 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 the alternative (non-nested) restricted model specifications. The relevant selection criterion, the posterior model probability, has a rigorous statistical justification and an intuitive interpretation, in contrast to the commonly used information criteria. An important challenge is the large number of possible specifications. To deal with this problem I develop a new algorithm that is based on Gibbs variable selection (Dellaportas et al., 2002), which enables me to narrow down the model space to a handful of preferred candidate specifications without having to estimate every 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., 2006). I document that the economic implications of a DTSM depend in important ways on the risk price restriction. A Bayesian framework can deal with specification uncertainty easily: one can average across models using posterior model probabilities as weights, a procedure which is called Bayesian model averaging (BMA). This averages out the specification and the inference, simply speaking, is only conditional on the data and not on both the data and the

---

<sup>2</sup>Examples are Dai and Singleton (2002), Ang and Piazzesi (2003), and Kim and Wright (2005).

model. I use BMA to average over different restricted DTSMs, in this way obtaining term premium estimates which 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., 2010b, CDR), which amounts to imposing three overidentifying restrictions, including a unit root, on the risk-neutral dynamics. While these do not restrict the forecasts of the short rate if prices of risk are unrestricted (see JSZ), they become powerful when I restrict the risk adjustment. Now the restrictions determine whether the model also has a unit root under the objective 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. Eurodollar futures rates are often quoted in the press when it comes to revisions of policy expectations. The instruments I include are quarterly contracts out to a four-year maturity. I show how they are priced and how changes in their rates decompose into changes in policy expectations and term premia.

The key empirical contribution is to show that the data call for tight restrictions on the market prices of risk and that under these restrictions term premia are much more stable than in unrestricted term structure models. Changes in short rate expectations account for most of the high frequency changes/volatility in forward rates across all maturities considered in this paper. The procyclical response of long forward rates to macro news 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, 2007) 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.

In a sense the paper revives the expectations hypothesis (EH). Under the (weak) EH, term premia are constant. While this is not the literal truth, the other extreme—a completely unrestricted risk adjustment—has the unattractive implication that monetary policy and long rates are essentially disconnected and term premia overly variable. One would expect that

long rates reveal *something* about policy expectations, but, as will be shown, in the absence of risk price restrictions they do not. The findings of this paper imply that on the spectrum between the two extremes of EH and unrestricted risk-adjustment we want to be a lot closer to the EH than most existing term structure models are.

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 (Cochrane and Piazzesi, 2005). 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., 2010b, JPS) 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 (JPS estimate 2<sup>19</sup> different specifications) 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 and the role of risk prices, how to decompose rate changes and volatilities, and the pricing of Eurodollar futures. In Section 3 the econometric framework is presented. Section 4 shows the empirical results. Section 5 concludes.

## 2 Dynamic Term Structure Model

### 2.1 Prices of risk in affine Gaussian DTSMs

Denote by  $X_t$  the  $(k \times 1)$  vector of term structure factors which represents the new information that market participants obtain at time  $t$ . In this paper these factors are taken to be latent and will be filtered from interest rates. 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  $\varepsilon_t \sim N(0, I_k)$  and  $E(\varepsilon_r \varepsilon_s') = 0$ ,  $r \neq s$ . The frequency of the model is daily. The short rate  $r_t$ , the overnight rate, is specified to be an affine function of the factors,

$$r_t = \delta_0 + \delta_1' X_t.$$

This short rate is the policy instrument of the Federal Reserve, thus expectations under the physical measure of its future values correspond to expectations about future monetary policy.<sup>3</sup>

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,

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

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

$$\lambda_t = \lambda_0 + \lambda_1 X_t.$$

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  $\varepsilon_t$ .<sup>4</sup> The risk sensitivity parameters  $\lambda_0$  ( $k \times 1$ ) and  $\lambda_1$  ( $k \times k$ ) determine the behavior of risk prices and risk premia and will be crucial in this paper.

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

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

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

$$\mu^{\mathbb{Q}} = \mu - \Sigma \lambda_0, \quad \Phi^{\mathbb{Q}} = \Phi - \Sigma \lambda_1. \quad (2)$$

This “essentially affine” specification (Duffee, 2002), popularized by the seminal work of Ang

---

<sup>3</sup>I abstract from the facts that the overnight rate in the U.S., the effective fed funds rate, deviates from the target set by the monetary authority, and that the target has a step-function character. Both simplifications are inconsequential since I do not include observations of the short rate – inference is based on Eurodollar futures rates, which correspond to average forward rates over an entire quarter.

<sup>4</sup>For the shocks to  $X_t$  investors require risk compensation of  $\Sigma \lambda_t$ .

and Piazzesi (2003), has become the de facto standard for affine DTSMs.<sup>5</sup> Bond pricing is straightforward in this framework, where yields are affine functions of the state variables.<sup>6</sup>

This paper focuses on money market futures, which have a payoff that is proportional to the difference between the average short rate over a future time horizon and the contractual rate. Thus futures rates are equal to average forward rates, where a forward rate is defined as  $f_t^n = E_t^{\mathbb{Q}}(r_{t+n})$ .<sup>7</sup> Solving for the forward rates I obtain<sup>8</sup>

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

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

Note that (in addition to  $\delta_0$  and  $\delta_1$ ) 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  $\mu$  and  $\Phi$  do not appear. It is in this sense that the “risk-neutral dynamics [...] fit the cross-section of yields or forward rates” (Cochrane and Piazzesi, 2008, p. 2). Because typically many cross-sectional observations are available the  $\mathbb{Q}$ -dynamics can be precisely estimated.

If the marginal investor were risk-neutral, the forward rate  $f_t^n$  would be equal to the expected future short rate under the physical measure. This is usually called the “risk-neutral forward rate,” here denoted by  $\tilde{f}_t^n$ . 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  $\Pi_t^n$ . Since one can observe the forward rates (up to a measurement error), the problem of estimating the term premium is equivalent to the problem of inferring the risk-neutral forward rates. For risk-neutral rates it holds that

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

---

<sup>5</sup>Studies that assess the behavior of the term premium using this framework include Dai and Singleton (2003); Kim and Orphanides (2005); Rudebusch and Wu (2008); Joslin et al. (2010b).

<sup>6</sup>Note that this model belongs to the class of discrete-time models that have affine dynamics under  $\mathbb{Q}$  and time-constant volatility,  $DA_0^{\mathbb{Q}}(N)$  as defined by Dai et al. (2006). Because of the particular choice of  $\lambda_t$  the model is also affine under  $\mathbb{P}$ , an assumption which, as these authors stress, can be relaxed without giving up on tractable bond pricing formulas.

<sup>7</sup>I refer to this object as a forward rate, although this term usually denotes the rate that can be contracted at  $t$  for a loan from  $t+n$  to  $t+n+1$  by entering the appropriate bond positions. This rate is equal to  $\log \frac{P_t^n}{P_t^{n+1}} = \log \frac{E_t^{\mathbb{Q}} \exp(-r_t - r_{t+1} - \dots - r_{t+n-1})}{E_t^{\mathbb{Q}} \exp(-r_t - r_{t+1} - \dots - r_{t+n})}$  ( $P_t^n$  is the time- $t$  price of a discount bond with  $n$  days until maturity) which differs from  $f_t^n$  by a Jensen inequality term.

<sup>8</sup>The loadings obtained here closely correspond to those for one-period forward rates common in the bond pricing literature, as derived for example in Cochrane and Piazzesi (2008), with the difference that Jensen inequality terms resulting from convexity effects are absent in our case.

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

In order to construct estimates of the risk-neutral rates, which embody the expectations about future monetary policy, knowledge of  $\mu$  and  $\Phi$  is required. These are difficult to estimate because of the high persistence of interest rates (Section 3.1).

The no-arbitrage condition requires the consistency of dynamic properties (determined by  $\mu$  and  $\Phi$ ) and cross-sectional properties (determined by  $\mu^Q$  and  $\Phi^Q$ ) of interest rates, allowing for a risk adjustment, and equation (2) makes this risk adjustment explicit. Thus no-arbitrage can in principle be very useful for estimation: the (very precise) estimates of the risk-neutral dynamics can potentially help to pin down the real-world dynamics and thus term premia. For intuition consider the extreme case of the strong expectations hypothesis, 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  $\lambda_0$  and  $\lambda_1$  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).

However if  $\lambda_0$  and  $\lambda_1$  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 no cross-sectional information is used to pin down the P-dynamics. In fact most studies, with the above-mentioned exception, impose no or only minimal restrictions on risk prices. Consequently the potential of no-arbitrage is hardly ever realized. This paper will show that strong zero-restrictions on  $\lambda_1$  are supported by the data. These restrictions not only improve precision but also have important economic implications: they lead to very different conclusions about term premia than an unrestricted specification.

## 2.2 Decomposing rate changes and volatilities

By construction rate changes decompose into changes in risk-neutral rates and changes in forward risk premia,  $f_{t+1}^n - f_t^{n+1} = \tilde{f}_{t+1}^n - \tilde{f}_t^{n+1} + \Pi_{t+1}^n - \Pi_t^{n+1}$ . Using a DTSM one can provide a more detailed decomposition:

$$\begin{aligned} f_{t+1}^n - f_t^{n+1} &= A_n + B'_n X_{t+1} - A_{n+1} - B'_{n+1} X_t \\ &= B'_n \Sigma \varepsilon_{t+1}^Q = B'_n \Sigma (\varepsilon_{t+1} + \lambda_t) \\ &= \tilde{B}'_n \Sigma \varepsilon_{t+1} + (B_n - \tilde{B}_n)' \Sigma \varepsilon_{t+1} + B'_n \Sigma \lambda_t. \end{aligned} \tag{3}$$

Note that this rate change corresponds to the one-period (absolute) return of a hypothetical futures contract which pays the difference between the realized future short rate and the contractual rate.<sup>9</sup> Notably it is an *excess return*, because the risk-neutral expected return,  $E_t^Q(f_{t+1}^n - f_t^{n+1})$  is zero. Expression (3) decomposes this return into three components:

1. *Revisions to short rate expectations*: The first component corresponds to the change in the expectation of the future short rate,  $(E_{t+1} - E_t)r_{t+n+1} = \tilde{B}'_n \Sigma \varepsilon_{t+1}$ . 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,  $\Pi_{t+1}^n - E_t \Pi_{t+1}^n = (B_n - \tilde{B}_n)' \Sigma \varepsilon_{t+1}$ .
3. *Expected returns*: The third component is equal to the expected change in the forward risk premium,  $E_t \Pi_{t+1}^n - \Pi_t^{n+1} = B'_n \Sigma \lambda_t$ . This term captures the predictable part of the return and corresponds to the *return risk premium*.<sup>10</sup>

To perform this decomposition one needs estimates of both the parameters of the models and of the latent factor shocks. Of course the decomposition will inherit the statistical uncertainty from the inference about unknown parameters and factors.

To foreshadow a key empirical result, models with few or no restrictions on risk prices attribute most of daily rate changes at medium and long maturities to surprise changes in the term premium (the second component), whereas a model with constrained risk pricing typically implies an important role for policy expectations (the first component).

Equation (3) also provides the basis for understanding sources of volatility. The term structure of volatility, the “vol curve,” describes the volatility of changes in yields or forward rates across maturities. Based on equation (3) the variance of forward rate changes is given by

$$\text{Var}(f_{t+1}^n - f_t^{n+1}) = B'_n \Sigma (I_k + \text{Var}(\lambda_t)) \Sigma' B_n.$$

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. Since

---

<sup>9</sup>Specifically it is the absolute return if one enters at  $t$  into a long position in a contract that pays  $r_{t+n+1} - f_t^{n+1}$  at maturity, and liquidates the position at  $t + 1$ . Note that money market futures usually pay the difference between the contractual rate and the short rate, in which case the above rate change corresponds to the return on a short position.

<sup>10</sup>In the language of Cochrane and Piazzesi (2005),  $B'_n \Sigma \lambda_1 X_t$  is the *return-forecasting factor*, which generally differs across maturities. It differs across maturities only by a factor of proportionality if only one element in the vector  $\lambda_t$  is non-zero; see for example the model of Cochrane and Piazzesi (2008).

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

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

and I will call the square root of this expression for varying  $n$  the “risk-neutral vol curve.”

## 2.3 Eurodollar futures

For the purpose of inferring changes in monetary policy expectations, Eurodollar futures<sup>11</sup> have several practical advantages over Treasury bonds. First the futures rates directly reflect the forward rate curve, whereas forward rates derived from bond prices depend on the algorithm used to infer zero rates from observed bond prices. 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. Last, the futures contracts arguably are not as affected by flight-to-quality effects or other extraordinary forces affecting supply and demand of Treasury securities.<sup>12</sup>

These futures contracts have payoffs that are based on the three-month LIBOR rate on the settlement day. This settlement rate can be modeled by the average expected short rate (under  $\mathbb{Q}$ ) for the three-month period starting on the settlement day:  $S_t = N^{-1} \sum_{h=0}^{N-1} E_t^{\mathbb{Q}}(r_{t+h})$ , where  $N$  is the number of days in the quarter, taken to be 91.<sup>13</sup> Eurodollar futures contracts involve no cost today and have a payoff proportional to the difference between the contractual rate and the settlement rate.<sup>14</sup> For the Eurodollar futures contract that settles at the end of quarter  $i$ ,

---

<sup>11</sup>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 08/23/2010).

<sup>12</sup>A technical advantage is that the payoffs of money market futures depend linearly on future short rates, thus convexity terms, which necessarily arise for yields and forward rates implied by bond prices, are absent in the pricing formulas for these securities.

<sup>13</sup>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.

<sup>14</sup>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.

where  $i = 1$  corresponds to the current quarter, the pricing equation is:

$$0 = E_t^{\mathbb{Q}}(ED_t^{(i)} - S_{T(i,t)}),$$

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 thus given by

$$\begin{aligned} ED_t^{(i)} &= E_t^{\mathbb{Q}}(S_{T(i,t)}) = N^{-1} \sum_{n=iN-d(t)}^{(i+1)N-d(t)-1} E_t^{\mathbb{Q}}r_{t+n} = N^{-1} \sum_{n=iN-d(t)}^{(i+1)N-d(t)-1} (A_n + B'_n X_t) \\ &= a_i + h'_i X_t. \end{aligned} \tag{4}$$

Note that the futures rate is simply an average of the relevant forward rates.<sup>15</sup> The scalar  $a_i$  and the vector  $h_i$  are the averages of  $A_n$  and  $B_n$ , respectively, over the relevant period.<sup>16</sup>

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

$$\tilde{ED}_t^{(i)} = \tilde{a}_i + \tilde{h}_i X_t \tag{5}$$

where  $\tilde{a}_i$  and  $\tilde{h}_i$  are the averages of  $\tilde{A}_n$  and  $\tilde{B}_n$ . The corresponding forward risk premium for contract  $i$  is given by  $\Pi_t^{(i)} = ED_t^{(i)} - \tilde{ED}_t^{(i)}$ . The decomposition for changes in forward rates in equation (3) analogously holds for changes in futures rates:

$$ED_{t+1}^{(i)} - ED_t^{(i)} = \tilde{h}'_i \Sigma \varepsilon_{t+1} + (h_i - \tilde{h}_i)' \Sigma \varepsilon_{t+1} + h'_i \Sigma \lambda_t.$$

The term structure of volatility and the risk-neutral vol curve for Eurodollar futures are analogous to those for forward rates, with  $B_n$  and  $\tilde{B}_n$  replaced by  $h_i$  and  $\tilde{h}_i$ .

---

<sup>15</sup>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.

<sup>16</sup>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.

## 2.4 The arbitrage-free Nelson-Siegel model

Since not all parameters of the DTSM are identified, normalization restrictions need to be imposed (Dai and Singleton, 2000). A DTSM is called “canonical” or “maximally flexible” if there are no overidentifying restrictions, which for the case  $k = 3$  amounts to 22 free parameters in  $(\delta_0, \delta_1, \mu^{\mathbb{Q}}, \Phi^{\mathbb{Q}}, \mu, \Phi, \Sigma)$ . 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:

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

where the parameter  $\rho$  is restricted to be less than one in absolute value. This implies a Nelson-Siegel-type forward rate curve<sup>17</sup> given by

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

The AFNS specification amounts to imposing three overidentifying restrictions on a canonical affine Gaussian DTSM, including a unit eigenvalue of  $\Phi^{\mathbb{Q}}$ .<sup>18</sup> 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

<sup>17</sup>There is no convexity/yield-adjustment term 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).

<sup>18</sup>Strictly speaking there is an arbitrage opportunity in the presence of a unit root under  $\mathbb{Q}$ : The long forward rate  $\lim_{n \rightarrow \infty} f_t^n = X_t^{(1)}$  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  $\epsilon$  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.

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  $\lambda_1$ , 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.<sup>19</sup>

### 3 Econometric Methodology

The data set consists of daily observations of Eurodollar futures rates. I include contracts maturing at the end of the current and the following 15 quarters, denoted by ED1 to ED16, essentially covering the forward rate curve up to a maturity of four years. The sample starts on 1 January 1990 and ends on 29 June 2007, before the start of the financial crisis. The number of days in the sample is  $T = 4401$ .

#### 3.1 Persistence of interest rates and the role of no-arbitrage

A state-space representation forms the basis for estimation. Equation (1) is the transition equation for the  $3 \times 1$  state vector, which I reproduce here:

$$X_t = \mu + \Phi X_{t-1} + v_t, \quad v_t \sim N(0, Q), \quad E(v_t v_s') = 0, \quad t \neq s, \quad (6)$$

introducing the notation  $v_t = \Sigma \varepsilon_t$  and  $Q = \Sigma \Sigma'$ . For future reference let  $X = (X_1, \dots, X_T)$ . The observation equation is

$$Y_t = a + H' X_t + w_t, \quad w_t \sim N(0, R), \quad E(w_t w_s') = 0, \quad t \neq s, \quad (7)$$

where  $Y_t$  contains the futures rates:  $Y_t = (ED_t^{(1)}, \dots, ED_t^{(16)})'$ . The intercept vector  $a = 0$  because of the normalization  $\mu^{\mathbb{Q}} = 0$ . The  $3 \times 16$  coefficient matrix  $H = (h_1, \dots, h_{16})$  is determined by  $\Phi^{\mathbb{Q}}$ . The vector  $w_t$  contains measurement errors, included to avoid stochastic

---

<sup>19</sup>An attractive alternative to the AFNS specification is the canonical form of JSZ, who take observable yield portfolios (e.g., principal components) as the factors, and parameterize the risk-neutral dynamics in terms of the long-run  $\mathbb{Q}$ -mean of the short rate and the eigenvalues of  $\mathbb{Q}$ . This also allows for a level/slope/curvature interpretation of risk premia. The AFNS restrictions on the eigenvalues under  $\mathbb{Q}$  could be included in the set of restrictions that are evaluated using my model selection presented below.

singularity as is common in the DTSM literature. They are assumed to be serially uncorrelated, as usual, and for the sake of parsimony I impose  $R = \sigma_w^2 I_{16}$ .

The high persistence of interest rates makes estimation of the parameters governing the dynamic system,  $\mu$  and  $\Phi$ , very problematic, as has been documented by Duffee and Stanton (2004) and Kim and Orphanides (2005), among others. The first problem is a *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 an upward *bias* in the estimated speed of mean reversion of the short rate: The closer the largest autoregressive root of a time series is to one, the more pronounced is the downward bias in its estimate.<sup>20</sup> The largest root of  $\Phi$  is very close to one, thus we are likely to significantly underestimate it. These two problems have serious consequences in that they cause our short rate forecasts and term premium estimates to be imprecise and biased. Specifically, short rate forecasts are likely to be too close to the unconditional mean and thus too stable, which results in term premium estimates that are too variable.

The big advantage of using a no-arbitrage model to obtain short rate forecasts and term premium estimates, as opposed for example to an unrestricted VAR, is that the cross section of interest rates can provide information about their dynamics. Intuitively, estimation of equation (7) pins down the Q-dynamics very precisely, and no-arbitrage restrictions tie these to the parameters in equation (6). However, absent restrictions on the prices of risk, these two equations are estimated separately from each other. Estimates of  $\mu$  and  $\Phi$  are in this case no different from OLS.<sup>21</sup> Only if risk prices are in some way restricted does no-arbitrage become in any way useful to tie down the parameters of the dynamic system. This paper asks the question whether some specific restrictions on risk prices, namely zero restrictions on elements of  $\lambda_1$ , are supported by the data. Fortunately it turns out that they are: the data favor tight restrictions on risk prices, so that the cross section becomes very helpful in the estimation of the physical dynamics. In this way, restrictions on the prices of risk help overcome the problems caused by the persistence of interest rates.

I parameterize the model in terms of the Q-dynamics and the risk sensitivity parameters.<sup>22</sup> This is because (i) the Q-dynamics are chosen according to the AFNS specification, and (ii) the estimation will focus on inference and restrictions on the prices of risk. The parameters to be estimated are  $\theta = (\rho, \lambda_0, \lambda_1, \Sigma, \sigma_w^2)$ . Since this paper is concerned with the variability of risk premia and not their average level, the focus is on restrictions on  $\lambda_1$ . Depending on the

---

<sup>20</sup>See Kendall (1954) and Shaman and Stine (1988). For the multivariate case see Nicholls and Pope (1988).

<sup>21</sup>This intuitive result has been formalized by JSZ, who use this fact to simplify estimation of unrestricted affine models.

<sup>22</sup>Bertholon et al. (2008) call this the *back-modeling strategy*.

restrictions that are imposed there are between 11 and 20 free parameters. Since the short rate has a unit root under  $\mathbb{Q}$  and  $\Phi = \Phi^{\mathbb{Q}} + \Sigma\lambda_1$ , the restrictions on  $\lambda_1$  will determine whether  $\Phi$  has a unit eigenvalue. Absent any restrictions, estimates will generally imply a stationary short rate. By having the data choose the zero restrictions on  $\lambda_1$  the possibility of a unit root under  $\mathbb{P}$  will be explicitly allowed for.

### 3.2 Bayesian vs. frequentist estimation

For estimation the researcher can choose between a frequentist approach, typically maximum likelihood (ML) estimation in the context of affine models, and a Bayesian approach. Using Markov chain Monte Carlo (MCMC) methods to learn about the posterior distribution of the model parameters and performing Bayesian inference is becoming more popular for estimation of DTSMs. There are three reasons for choosing the Bayesian approach in this context, of which I believe only the first one has been recognized in the term structure literature.

The first reason is that estimation by ML is computationally difficult: The likelihood function is high-dimensional, strongly nonlinear in the model parameters, and has multiple local maxima. Hence finding the global maximum is challenging.<sup>23</sup> Specifically Kim and Orphanides (2005) note that the likelihood function often has “multiple inequivalent local maxima which have similar likelihood values but substantially different implications for economic quantities of interest” (p. 10). My attempts to estimate the model with ML confirmed this: the likelihood function has several local maxima with different prices of risk and different economic implications.<sup>24</sup> Of course without informative priors the posterior density is just equal to the likelihood function, but even in this case MCMC is helpful because with clever sampling one can reasonably hope to explore all relevant regions of the support (Ang et al., 2007, 2009). Furthermore if one is willing to impose informative priors, this can lead to a posterior density that is better behaved than the likelihood function (Chib and Ergashev, 2009). In any case MCMC is less challenging than ML since one just successively draws parameters from their conditional posterior distributions, and then can diagnose whether the algorithm has converged. However for ML estimation we are faced with troublesome numerical optimization and hardly ever know whether our estimates really come from a global maximum.

The second reason for putting the Bayesian hat on is that frequentist approaches cannot always easily quantify estimation uncertainty. First, the (high-dimensional) covariance matrix

---

<sup>23</sup>Among the many studies that have documented these issues are Ang and Piazzesi (2003), Duffee and Stanton (2004) and Hamilton and Wu (2010).

<sup>24</sup>While CDR claim that the a priori identification of the factors in the AFNS model helps in the estimation, my own experience is that troublesome local maxima are still present.

of the estimator can only be approximated using asymptotic theory. Second, the objects of interest are usually highly nonlinear functions of the parameters, and approaches such as the delta-method are again only valid asymptotically. Third, quantifying the uncertainty that results from inference about both parameters and latent factors is practically infeasible in a frequentist setting, since latent factors are inferred conditional on given point estimates of the model parameters (Kim and Nelson, 1999, chap. 8). On the other hand accounting for estimation uncertainty is very simple using MCMC: the algorithm provides a sample from the joint posterior distribution of parameters and latent factors, and I simply calculate the object of interest for each draw and in this way obtain a sample from its posterior distribution. Summary statistics for this sample then provide the correct inference.

The third reason is that for consistently dealing with the issues of model selection and model uncertainty, a Bayesian framework is required. While classical hypothesis tests can evaluate the credibility of a restriction compared to the unrestricted model, the evaluation of non-nested alternatives is difficult. One usually resorts to information criteria like AIC and BIC for this purpose, as in Christensen et al. (2010a) and JSZ, but these are ad hoc measures of model-fit and generally have no statistical interpretation. In contrast, the use of posterior model probabilities is theoretically sound, allows for an intuitive interpretation, and provides a solution to deal with model uncertainty. Furthermore, a frequentist approach to model selection always requires estimation of all alternative specifications. If the number of models is large, as in the context of this paper, this is practically infeasible. The Bayesian solution to this problem, described in Section 3.4, does not require this. Finally, a Bayesian framework allows the researcher to conveniently deal with model uncertainty by means of Bayesian model averaging (Section 3.5).

This paper uses a Bayesian framework for DSTM estimation (i) to learn about the uncertainty inherent in term premium estimates, and (ii) to assess the role of restrictions on the prices of risk. Since the goal is to understand what the data can tell us about term premia, I impose rather diffuse priors. The algorithm used to estimate the model *conditional on a particular set of restrictions on risk prices* is a simple blockwise Metropolis-Hastings (M-H) sampler. Details on the priors and on the estimation are given in Appendix B. The algorithm provides us with a sample from the joint posterior distribution of model parameters and latent factors. Notably  $\lambda_1$  has zero restrictions imposed according to the specification I choose to estimate. The next section explains how this output is used to perform inference about policy expectations and term premia before turning to the key issue of how to select the restrictions in Section 3.4.

### 3.3 Inference about the impact of news on the term structure

The estimation results allow me to decompose futures rates into risk-neutral rates and forward risk premia. First consider decomposing the changes in futures rates *on a given day*. What we observe are the changes in the rates for each of the 16 contracts on this day. The model implies fitted rates for given values of factors and parameters. For each draw from the posterior I can calculate the *changes in the fitted rates*, using equation (4). This results in a (posterior) distribution of fitted rate changes for each contract. The sample mean of this distribution is a Bayesian point estimate: the posterior mean. The 2.5%- and 97.5%-quantiles then provide an interval estimate, the 95% credibility interval. Since these intervals are rather tight (remember that 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*, using equation (5). I can thus infer what happened to monetary policy expectations and to term premia in response to the news on this particular day. Here the interval estimates will be crucial because they reveal the uncertainty about the role of short rate expectations.

The term structure of volatility (vol curve) and the risk-neutral vol curve will reveal the role of short rate expectations for interest rate volatility. For both of these objects I can obtain the posterior distributions by calculating model-implied volatilities of futures rates and risk-neutral futures rates for each draw from the MCMC sample.

Of particular interest is 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,<sup>25</sup> 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 for long forward rates, movements in the risk premium account for the majority of the response. 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 uncertainty (which I will show to be dramatic in the absence of risk price restrictions). Second, the Kim-Wright model essentially imposes no restrictions on risk prices, which gives reason to believe that the term premium estimates could be problematic. For these reasons I take a fresh look at the question of how policy expectations and term premia react to macro news.

The typical approach is to regress changes in interest rates on a measure of the surprise in the announcement, usually taken to be the difference between released and forecast values. If the left-hand-side variable is itself an estimate, such as changes in risk-neutral rates, then

---

<sup>25</sup>For example Fleming and Remolona (1999) and Gürkaynak et al. (2005).

the standard errors on the regression coefficients understate the overall uncertainty. However, the MCMC sample provides the whole posterior distribution of risk-neutral rates on any day. 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  $s_t^{(1)}$  to  $s_t^{(r)}$ , scalars that contain the surprise component on day  $t$  for each of  $r$  different macroeconomic data releases.<sup>26</sup> As in a standard Bayesian regression, calculate mean and standard errors of the posterior distribution for the regression parameters, which is the conjugate normal posterior.<sup>27</sup>
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 actual 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.

### 3.4 Restricting risk prices using Bayesian model selection

To unleash the power of no-arbitrage I want to impose restrictions on risk prices. The question of how to do this has not been satisfactorily answered. A commonly employed approach is to first estimate a model without restrictions on risk prices and then, in a second step, re-estimate the model by setting some parameters with large standard errors to zero. This ad hoc

---

<sup>26</sup>Note that it is necessary to include all data release series in the regression to partial out the impact of releases that occur on the same day. Further note that equation-by-equation least squares estimation is efficient because the regressors are the same in each equation.

<sup>27</sup>I specify the prior for the regression parameters to be independently normal with mean zero and large variance. The prior for the error variance is taken to be inverse gamma.

procedure usually leads to very few restrictions and does not solve the bias and uncertainty problems.<sup>28</sup> Furthermore it is unappealing from a statistical perspective for several reasons: Choosing restrictions based on individual standard errors ignores the off-diagonal elements in the covariance matrix of the estimates—a joint restriction is chosen without considering joint significance. The choice of a significance level required for inclusion of the parameter is necessarily arbitrary. 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.<sup>29</sup>

Bayesian model selection provides the right toolbox to assess the plausibility of alternative restrictions.<sup>30</sup> A challenge is that there are many possible model specifications: considering all possible zero restrictions on the  $3 \times 3$  matrix  $\lambda_1$  leads to  $2^9 = 512$  possible models. While this might seem manageable, including restrictions on  $\lambda_0$  or considering more factors easily makes this number explode. Thus a method is needed that avoids having to estimate each model separately. The framework I propose consists of three steps. First, plausible candidate specifications are identified using a new MCMC algorithm that involves latent indicator variables. The second step is to estimate each specification separately, providing estimates of parameters, model implications, and posterior model probabilities. The third step, detailed in Section 3.5, is to deal with the specification uncertainty by means of Bayesian model averaging.

The problem of selecting a particular restriction is related to the variable selection problem in multivariate regression analysis: In both cases I can introduce a vector of indicator variables that summarizes which parameters are allowed to be nonzero. For the regression context Dellaportas et al. (2002) developed the method of “Gibbs variable selection” (GVS) which delivers a sample from the joint posterior distribution of the regression coefficients and indicators; I adapt this method to the context of DTSM estimation. Let  $\gamma$  be a  $k^2 \times 1$  vector of indicator variables, each of which is equal to one if the corresponding element of  $\text{vec}(\lambda_1)$  is allowed to be nonzero. The goal is to obtain the joint distribution of  $(\gamma, \theta, X)$ . Since the conditional posterior of  $\gamma$  given  $\theta$  and  $X$  can be derived (up to a constant of proportionality), blockwise M-H can be used to obtain draws from this distribution. To assess the plausibility of a joint restriction on  $\lambda_1$ , represented by a specific value of  $\gamma$ , say  $\bar{\gamma}$ , we can consider the posterior probability  $P(\gamma = \bar{\gamma})$ , which is easily estimated by counting the number of draws for which  $\gamma = \bar{\gamma}$ . The algorithm is developed in Appendix C, and I will refer to it as GVS,

---

<sup>28</sup>Prominent studies that employ this two-step approach are Dai and Singleton (2002), Ang and Piazzesi (2003), and Kim and Wright (2005). None of these restrict more than three risk sensitivity parameters to zero.

<sup>29</sup>Kim and Orphanides (2005) report that in the context of their model some of the “different choices of parameters to be set to zero [...] exhibited economically significant quantitative differences” (p.11).

<sup>30</sup>For review articles on Bayesian model selection see Kass and Raftery (1995) and Clyde and George (2004).

although it shares only the idea of latent indicator variables with the original GVS algorithm. The key feature of this algorithm 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 appear quickly in the GVS algorithm.<sup>31</sup>

With the sample from the joint posterior for  $(\gamma, \theta, X)$  at hand, I select those models with a Bayes factor in comparison to the favored model of at most 20. The Bayes factor is equal to the ratio of posterior model probabilities, and a value larger than 20 can be considered strong evidence against the model (Kass and Raftery, 1995). This will lead to the inclusion of only a handful of models. In addition I include the specification without restrictions on  $\lambda_1$ , which I call the “unrestricted model.” From now on the model space will consist only of these models, the number of which I denote by  $J$ .

Having identified  $J$  candidate models, the second step is to estimate each of them individually, so-called within-model estimation, using the algorithm in Appendix B. With the MCMC samples at hand, I can perform inference and consider the economic implications of each model. To see how much support each model specification receives, I estimate posterior model probabilities based on marginal likelihood approximations. 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).<sup>32</sup>

### 3.5 Specification uncertainty and Bayesian model averaging

If one specification received overwhelming support in the data (for example 99% posterior model probability) or if all specifications had very similar economic implications, then we could stop here. However neither will turn out to be the case: some significant specification uncertainty will remain. BMA is the appropriate theoretical framework to incorporate specification uncertainty into our inference. 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 data and not conditional on the data and a particular model specification.<sup>33</sup>

A conceptually and computationally straightforward way to perform BMA in this context

---

<sup>31</sup>This point was made by George and McCulloch (1993).

<sup>32</sup>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). For the volume I use 5% in both cases and I estimate the mode by taking that parameter draw which maximizes the posterior.

<sup>33</sup>Of course inference is always conditional on the class of models under consideration.

is to include a model indicator which identifies the different candidate models,  $j \in \{1, \dots, J\}$ , as an additional parameter. This is a case of “joint model-parameter sampling,” something I also did when identifying candidate models. Reversible-jump MCMC (RJMCMC), initially developed by Green (1995), is the appropriate way to sample. This method is characterized by the ability to “jump” between models with parameter spaces of different dimensionality. The sampler, detailed in Appendix D, provides draws that approximately come from the joint posterior distribution  $P(j, \theta_j, X|Y)$ . BMA is now particularly easy: the model indicator is “averaged out” if I simply ignore its value for each draw. Hence I can perform inference on term premia *unconditional of model specification* by calculating the object of interest using the RJMCMC sample. This amounts to averaging over the different specifications, weighing each one appropriately by its posterior model probability.

## 4 Empirical Results

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 on term premia that is not conditional on a particular model specification.

### 4.1 Restrictions on the market prices of risk

The first step of the econometric analysis delivers six model specifications that receive particular support in the data. These are shown in Table 1 as models  $M_2$  to  $M_7$ , together with the unrestricted model,  $M_1$ . The second column shows the frequency of how often each model was visited by the GVS sampler, with a cumulative frequency of 75.8%. Columns three to five indicate which elements in the respective columns of  $\lambda_1$  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. Column six shows the largest eigenvalue of  $\Phi$ . Column seven shows the long-run revision of short rate expectations,  $\lim_{h \rightarrow \infty} (E_{t+1} - E_t)r_{t+h}$ , in response to a unit level shock—details about this calculation are given in Appendix E. The remaining columns show alternative estimates of the posterior model probabilities based on GVS 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 reversible-jump MCMC (column 11).<sup>34</sup>

---

<sup>34</sup>The various estimates of the posterior model probabilities show some numerical differences. One might still be worried about lack of convergence of the different MCMC sampler. While there is certainly room for improvement of the efficiency of my sampling algorithms, the numbers in the last four columns are usually in

The unrestricted model’s posterior probability is estimated to be zero, in fact it is not visited at all by the GVS 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  $\lambda_1$  restricted to zero. Such tight restrictions are of course very plausible if one believes in the absence of arbitrage: long rates *should* have some relation to expected future short rates, thus physical and risk-neutral dynamics *should* be close to each other. Notably even the weak expectations hypothesis (EH), model  $M_5$ , receives some support in the data. With the data strongly favoring tight zero restrictions on  $\lambda_1$ , 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.

To put these results into perspective, consider the parameter estimates of the unrestricted model, presented in Table 2. While risk-neutral dynamics ( $\rho$ ),  $\Sigma$ , and  $\sigma_w^2$  are estimated very precisely, the estimates of the risk sensitivity parameters in  $\lambda_0$  and  $\lambda_1$  lack precision, which reflects the fact that  $\mu$  and  $\Phi$  are hard to estimate. For only three of the nine parameters in  $\lambda_1$  the 95% credibility intervals do not straddle zero, i.e., are “significant” in frequentist parlance. Most readers will agree that in light of these estimates one would be inclined to restrict some of the parameters to zero. 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 model uncertainty under the carpet. A rigorous model selection procedure is needed.

Returning to Table 1, it is evident that under most of the preferred specifications, the short rate exhibits a unit root: as column (6) shows, all models except for  $M_1$  and  $M_6$  have physical dynamics with a largest eigenvalue of unity. There is strong support in the data for a stochastic trend in the short rate. While the short rate 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 in a model of daily frequency an integrated specification approximates the true data-generating process better than a stationary specification.

If the short rate is specified as I(1) then shocks to the level factor lead to revisions of far-ahead short rate expectations not only under  $\mathbb{Q}$  but also under  $\mathbb{P}$ . The models have very 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 the long-run revision is negative, i.e., the forward risk premium increases by more than one in response

---

the same ballpark, and the economic conclusions that one would draw are not materially different depending on which column one looks at.

to a unit level shock. Evidently there is specification uncertainty about this important aspect of the model. However, the models for which the number in column 7 is close to or equal to one have about 80% posterior probability: there is significant evidence for the hypothesis that the long-run revision of real-world expectations is similar to the revision of risk-neutral expectations, implying that forward risk premia at three to four years maturity move very little at a daily frequency.

A note on the role of priors: Since I used a 50%-50% prior for each element of  $\lambda_1$  to be restricted, one could say that the joint prior puts a lot of mass on restricted models. While this is true, each specification has prior probability of only  $(\frac{1}{2})^9 \approx .2\%$ , and the data update these priors significantly, putting almost all the weight on models with very few unrestricted elements in  $\lambda_1$ .

So is it level risk or slope risk that drives the variation in risk premia? While Cochrane and Piazzesi (2008) argue that only level risk is priced, the evidence here points to an important role for variation in slope risk: almost all preferred specifications have some nonzero elements in the second row of  $\lambda_1$ , which determines the time variation of slope risk. The present framework allows one to assess the hypotheses of whether level or slope risk varies over time: the posterior probability of each hypothesis can be estimated by the relative frequency of visits of the GVS sampler to models with nonzero elements on the first or second row of  $\lambda_1$ , respectively. In this way the probability that level risk varies over time is found to be 17.3%, whereas the probability for variation in slope risk is 89.3%. Since monetary policy is associated with slope shocks, it seems to be the variation in risk associated with policy shocks that drives variation in term premia. This is consistent with the results of JPS, who also find an important role for changes in slope risk. Assessing the sources of this variation using the present framework in the context of a macro-finance term structure model is a promising avenue for future research.

## 4.2 Economic implications of risk price restrictions

Turning to the economic implications of the different models, I will compare the unrestricted model ( $M_1$ ) to that specification which is favored by the data ( $M_2$ ). Figure 1 visualizes what  $M_1$  implies for term premia, and Figure 2 does so for model  $M_2$ .

The first panel of each figure decomposes the changes in interest rates on March 8, 1996, a day with a strongly positive surprise in nonfarm payroll employment (job creation in February exceeded the market's expectation by more than 400,000 jobs). It shows actual and fitted changes in futures rates, as well as point and interval estimates of how monetary policy expectations were revised, i.e., the posterior mean and 95% credibility intervals of the inferred

changes in risk-neutral rates.

The unrestricted model 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. 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. The reason for this result of course is the upward bias in the estimated speed of mean reversion. This causes the revision to die out faster than it should. 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.

In stark contrast to this, the favored specification 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).

The second panel visualizes how the models decompose interest rate volatility. It shows actual and model-implied volatilities, together with point and interval estimates for the risk-neutral volatilities. The unrestricted model implies that more than half of daily volatility at the long end is due to changing term premia, with very large estimation uncertainty. The favored model on the other hand attributes the vast majority of volatility at all maturities to changing short rate expectations, with a rather precisely estimated risk-neutral vol curve.

The third panel shows how interest rates and policy expectations respond to macro news. In addition to the responses of actual and fitted futures rates to a one standard deviation surprise in nonfarm payroll employment, it includes point and interval estimates of the responses of risk-neutral rates. 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. The conclusion would be that forward premia account for most of the procyclical reponse of long-term interest rates to economic news, in line with the findings of Beechey (2007). The favored model however implies that this procyclical response is entirely due to revisions of short rate expectations. Again the statistical uncertainty is large for the inference about risk-neutral rates using model  $M_1$ , whereas the restricted model  $M_2$  allows for very precise inference.

The difference between the two models' economic implications is striking. Since the restricted model  $M_2$  is favored by the data, whereas the unrestricted model has zero posterior

model probability, the evidence points to a dominant role for the expectations component to explain daily changes in interest rates and their response to macroeconomic news.

Based on a prior that risk premia move slowly and in a countercyclical fashion, it has been viewed as a puzzle that the term premium implied by unrestricted DTSMs show strongly procyclical responses to macroeconomic news (Beechey, 2007). My results indicate that such puzzling term premium movements at high frequencies are due to a lack of restrictions on the underlying term structure model. If we restrict the risk pricing and in this way make use of no-arbitrage in pinning down short rate forecasts and term premia, the puzzling procyclical responses of premia to macro news essentially disappear.

### 4.3 Specification uncertainty and Bayesian model averaging

Models other than  $M_2$  also receive some empirical support, and Table 1 indicates that these have materially different economic implications (column 7). Figure 3, which shows the decompositions of changes, volatilities, and response to macro news *across models*, confirms this. While they have 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  $M_5$  volatilities and macro responses for all maturities are mainly due to changes in expectations, models  $M_1$ ,  $M_3$ ,  $M_6$ , and  $M_7$  imply that at the long end forward risk premia are the driving force.

Because (a) several models receive non-negligible empirical support in terms of posterior model probabilities and (b) these models have materially different economic implications, there is economically important specification uncertainty. Thus we cannot simply pick one model and analyze term premium behavior on the basis of this model. Instead inference about premia needs to incorporate model uncertainty. The statistically sound way to deal with such specification uncertainty is to *average out the specification* using BMA), as described in Section 3.5. The results are visualized in Figure 4.

The conclusion is that short rate expectations are an important driving force for the daily volatility of the entire term structure and its procyclical responses to macro news. Forward risk premia move rather little at this frequency. Thus the implications of *BMA* are similar to those of the favored model  $M_2$ , but this was not at all obvious a priori. To make sound economic conclusions it is important to correctly account for model uncertainty.

Averaging across models instead of picking one favorite model increases the uncertainty about expectations and risk premia. This is to be expected. However, when comparing Figures 1 and 4 it becomes evident that averaging across the restricted specifications leads to slightly *lower overall uncertainty* about changes in risk-neutral rates than for the unrestricted model,

in particular when considering the risk-neutral vol curve. Restricting the prices of risk, even after accounting for model uncertainty, improves the precision of estimates of the short rate dynamics.

#### 4.4 Variability and cyclicity at the business cycle frequency

Evidently the implications of risk price restrictions on the variability of term premia at high frequencies are quite significant. What about the implications at lower frequencies and over the course of the business cycle? It is common to graph the time series of the risk premium inherent in a long-term interest rate, which makes it easy to visually assess secular trends, relative variability, and behavior over the business cycle. 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 Bayesian model averaging.<sup>35</sup>

The two-year forward premium implied by the unrestricted model is highly variable at the frequencies visible in the graph, and it mostly moves in tandem with the long-term interest rate. The risk premium implied by this specification accounts for most of the variability of the forward rate. By contrast, the restricted models  $M_2$  and *BMA* imply a much more stable two-year forward premium. Correspondingly, the variation in the forward rate is attributed to a large extent to expectations about monetary policy. Thus, not only at the daily frequency but also at business cycle frequencies do term premia become much more stable when risk price restrictions are imposed.

Interest rates have declined significantly over the course of the sample, and this also holds for the two-year forward rate. What is the source of this decline? While all three measures of the forward premium show a downward trend over the sample period, the decomposition of the secular decline in the forward rate is very different 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  $M_2$  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.<sup>36</sup>

A related issue is that the unrestricted estimates repeatedly turn negative toward the end of the sample. This is not unreasonable per se: if bonds pay off better in bad times,

---

<sup>35</sup>The graph is comparable to Figure 5.b in Kim and Orphanides (2005) and Figure 5.a in JPS.

<sup>36</sup>Specifically, the *BMA*-implied two-year-ahead short rate expectations have declined from 6.4% to 4.2% over the course of the sample, whereas under  $M_1$  the expectations have actually slightly increased from 4.8% to 5.3%.

they serve as an insurance, and in that case investors will be willing to forgo some yield to take advantage of this insurance. However it seems unlikely that the payoff profile of bonds changed so dramatically over the course of the sample, with the covariance of returns and marginal utility, which determines the sign of term premia, switching signs several times. The consistently positive term premium implied by the preferred model specifications seem more plausible in this light.

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.<sup>37</sup> Thus, while the variability is very different across models, the cyclicalities at low frequencies is similar. However, as shown in Sections 4.2 and 4.3, at higher frequencies the term premium estimates implied by  $M_1$  behave strongly procyclically, whereas the premium estimates implied by  $M_2$  and  $BMA$  do not.

“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.<sup>38</sup> 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 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 as well. 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.

To put these findings into perspective, a key problem with unrestricted DTSMs is that the estimated term premia are too variable and short rate expectations at long horizons are too stable. Such estimates are implausible for three reasons: First, macroeconomists have the prior that risk premia are slow-moving, based on, among others, models with habit-driven variation in risk compensation (Campbell and Cochrane, 1999; Wachter, 2006). Second, far-ahead expectations of nominal short rates should reflect the secular trends that are present in

---

<sup>37</sup>The cyclical behavior of the term premium would be much more pronounced had macro factors been included in the model. To analyze the behavior of term premia over the business cycle, one would generally prefer to use a macro-finance term structure model such as JPS.

<sup>38</sup>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.

inflation expectations (Kozicki and Tinsley, 2001). And third, evidence from surveys shows quite significant variability in expectations of future monetary policy (Kim and Orphanides, 2005).

Imposing restrictions on risk pricing in a DTSM leads to term premium estimates that are much more stable than in the absence of such restrictions. Furthermore short rate expectations implied by the preferred models have decreased over the sample period, consistent with decreasing long-term inflation expectations (Kozicki and Tinsley, 2001). Hence the framework for term structure estimation I suggest here is attractive not only because it increases parsimony and statistical precision in term premium estimation. It also leads to economic implications that, instead of being puzzling and implausible, are very much in agreement with some salient features of our models and evidence about risk premia, monetary policy, and inflation expectations.

## 5 Conclusion

This paper has demonstrated that the data support tight restrictions on risk prices in affine term structure models, and that 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.

The econometric framework for estimation of DTSMs that I propose enables the researcher to systematically choose among a large set of restrictions and to impose parsimony on an otherwise overparameterized model. Estimation and specification uncertainty are often ignored but are significant in term structure models. In the words of Cochrane (2007, p. 278), “when a policymaker says something that sounds definite, such as ‘[...] risk premia have declined,’ he is really guessing.” The present paper (a) quantifies the dramatic uncertainty around short rate forecasts and term premia in the absence of risk price restrictions, (b) shows how this uncertainty can be greatly reduced by constraining the risk adjustment, and (c) systematically deals with model uncertainty by incorporating this uncertainty into the statistical inference about risk premia. The framework is more generally applicable beyond the class of models and specific restrictions that the paper focuses on.

Given the empirical conclusion of low variability in forward premia, an important question is how this relates to the vast body of evidence against the EH, which finds substantial variation in expected returns (Campbell and Shiller, 1991; Cochrane and Piazzesi, 2005). Are the expected returns generated by a model with highly constrained risk adjustment too stable to generate the high  $R^2$  found in the data? Preliminary results show that they are not, and that Stambaugh’s (1999) bias, which leads to higher  $R^2$  in small samples than what the predictability is in the population, can be significant. A thorough analysis of these issues in the more common context of monthly Treasury yields is a work in progress. A related issue is to extend the framework to consider restrictions other than zeros on  $\lambda_1$ , 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 (JPS). 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, Andrew and Monika Piazzesi, “A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables,” *Journal of Monetary Economics*, May 2003, 50 (4), 745–787.
- , Jean Boivin, and Sen Dong, “Monetary Policy Shifts and the Term Structure,” NBER Working Paper 15270, National Bureau of Economic Research, Inc August 2009.

- , **Sen Dong**, and **Monika Piazzesi**, “No-Arbitrage Taylor Rules,” NBER Working Paper 13448, National Bureau of Economic Research, Inc September 2007.
- Backus, D.K. and J.H. Wright**, “Cracking the conundrum,” *Brookings Papers on Economic Activity*, 2007, *2007* (1), 293–329.
- Bartlett, M. S.**, “A comment on D. V. Lindley’s statistical paradox,” *Biometrika*, 1957, *44* (3-4), 533–534.
- Beechey, Meredith**, “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.) 2007.
- Bertholon, H., A. Monfort, and F. Pegoraro**, “Econometric Asset Price Modelling,” *Journal of Financial Econometrics*, 2008, *6* (4), 407–458.
- Brooks, Stephen P. and Gareth O. Roberts**, “Convergence assessment techniques for Markov chain Monte Carlo,” *Statistics and Computing*, Dec 1998, *8* (4), 319–335.
- Campbell, John Y and Robert J Shiller**, “Yield Spreads and Interest Rate Movements: A Bird’s Eye View,” *Review of Economic Studies*, May 1991, *58* (3), 495–514.
- Campbell, J.Y. and J.H. Cochrane**, “By force of habit: A consumption-based explanation of aggregate stock market behavior,” *Journal of political Economy*, 1999, *107* (2), 205–251.
- Carlin, Bradley P. and Siddhartha Chib**, “Bayesian Model Choice via Markov Chain Monte Carlo Methods,” *Journal of the Royal Statistical Society. Series B (Methodological)*, September 1995, *57* (3), 473–484.
- Carter, C. K. and R. Kohn**, “On Gibbs sampling for state space models,” *Biometrika*, 1994, *81* (3), 541–553.
- Chib, S. and B. Ergashev**, “Analysis of Multifactor Affine Yield Curve Models,” *Journal of the American Statistical Association*, 2009, *104* (488), 1324–1337.
- Chib, Siddhartha**, “Markov chain Monte Carlo methods: computation and inference,” in J.J. Heckman and E.E. Leamer, eds., *Handbook of Econometrics*, Vol. 5, Elsevier, 2001, chapter 57, pp. 3569–3649.
- and **Edward Greenberg**, “Understanding the Metropolis-Hastings Algorithm,” *The American Statistician*, Nov 1995, *49* (4), 327–335.

- **and** –, “Markov Chain Monte Carlo Simulation Methods in Econometrics,” *Econometric Theory*, Aug 1996, *12* (3), 409–431.
- Christensen, Jens H. E., Jose A. Lopez, and Glenn D. Rudebusch**, “Inflation Expectations and Risk Premiums in an Arbitrage-Free Model of Nominal and Real Bond Yields,” *Journal of Money, Credit and Banking*, 2010, *42*, 143–178.
- Christensen, Jens H.E., Francis X. Diebold, and Glenn D. Rudebusch**, “The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models,” *forthcoming in Journal of Econometrics*, 2010.
- Clyde, Merlise and Edward I. George**, “Model Uncertainty,” *Statistical Science*, 2004, *19* (1), 81–94.
- Cochrane, John**, “Commentary on “Macroeconomic implications of changes in the term premium”,” *Federal Reserve Bank of St. Louis Review*, Jul 2007, pp. 271–282.
- Cochrane, John H. and Monika Piazzesi**, “Bond Risk Premia,” *American Economic Review*, March 2005, *95* (1), 138–160.
- **and** –, “Decomposing the Yield Curve,” manuscript 2008.
- Cowles, Mary Kathryn and Bradley P. Carlin**, “Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review,” *Journal of the American Statistical Association*, 1996, *91*, 883–904.
- Dai, Qiang and Kenneth J. Singleton**, “Specification Analysis of Affine Term Structure Models,” *The Journal of Finance*, Oct 2000, *55* (5), 1943–1978.
- **and** –, “Expectation puzzles, time-varying risk premia, and affine models of the term structure,” *The Journal of Financial Economics*, Mar 2002, *63* (3), 415–441.
- **and** –, “Term Structure Dynamics in Theory and Reality,” *Review of Financial Studies*, 2003, *16* (3), 631–678.
- **, Anh Le, and Kenneth J. Singleton**, “Discrete-time Dynamic Term Structure Models with Generalized Market Prices of Risk,” working paper March 2006.
- Dellaportas, Petros, Jonathan J. Forster, and Ioannis Ntzoufras**, “On Bayesian model and variable selection using MCMC,” *Statistics and Computing*, January 2002, *12* (1), 27–36.

- DiCiccio, Thomas J., Robert E. Kass, Adrian Raftery, and Larry Wasserman**, “Computing Bayes Factors by Combining Simulation and Asymptotic Approximations,” *Journal of the American Statistical Association*, Sep 1997, *92* (439), 903–915.
- Duffee, Gregory R.**, “Term Premia and Interest Rate Forecasts in Affine Models,” *Journal of Finance*, 02 2002, *57* (1), 405–443.
- **and Richard H. Stanton**, “Estimation of Dynamic Term Structure Models,” working paper, Haas School of Business Mar 2004.
- Dybvig, P.H., J.E. Ingersoll Jr, and S.A. Ross**, “Long forward and zero-coupon rates can never fall,” *Journal of Business*, 1996, *69* (1), 1–25.
- Fleming, Michael J. and Eli M Remolona**, “The term structure of announcement effects,” BIS Working Papers 71, Bank for International Settlements June 1999.
- Gamerman, Dani and Hedibert F. Lopes**, *Markov Chain Monte Carlo*, second ed., Chapman & Hall/CRC, 2006.
- Gelman, Andrew and Donald B. Rubin**, “Inference from Iterative Simulation Using Multiple Sequences,” *Statistical Science*, 1992, *7* (4), 457–472.
- George, Edward I. and Robert E. McCulloch**, “Variable Selection Via Gibbs Sampling,” *Journal of the American Statistical Association*, September 1993, *88* (423), 881–889.
- Geweke, John**, “Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments,” in “In Bayesian Statistics 4” Oxford University Press 1992, pp. 169–193.
- Green, Peter J.**, “Reversible jump Markov chain Monte Carlo computation and Bayesian model determination,” *Biometrika*, 1995, *82* (4), 711–732.
- Gürkaynak, Refet S., Brian P. Sack, and Eric T. Swanson**, “The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models,” *American Economic Review*, March 2005, *95* (1), 425–436.
- Hamilton, James D. and Jing (Cynthia) Wu**, “Identification and Estimation of Affine Term Structure Models,” working paper, University of California, San Diego July 2010.
- Jegadeesh, Narasimhan and George G Pennacchi**, “The Behavior of Interest Rates Implied by the Term Structure of Eurodollar Futures,” *Journal of Money, Credit and Banking*, August 1996, *28* (3), 426–46.

- Johannes, Michael and Nicholas Polson**, “MCMC Methods for Continuous-Time Financial Econometric,” in Yacine Ait-Sahalia and Lars Hansen, eds., *Handbook of Financial Econometrics*, Vol. 2, Elsevier, 2009, chapter 1.
- Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu**, “A New Perspective on Gaussian DTSMs,” *forthcoming in Review of Financial Studies*, 2010.
- , **Marcel Pribsch, and Kenneth J. Singleton**, “Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks,” working paper May 2010.
- Kass, Robert E. and Adrian E. Raftery**, “Bayes Factors,” *Journal of the American Statistical Association*, jun 1995, *90* (430), 81–94.
- Kendall, M. G.**, “A note on bias in the estimation of autocorrelation,” *Biometrika*, 1954, *41* (3-4), 403–404.
- Kim, Chang-Jin and Charles R. Nelson**, *State-Space Models with Regime-Switching*, The MIT Press, 1999.
- Kim, Don H.**, “Challenges in macro-finance modeling,” BIS Working Papers 240, Bank for International Settlements December 2007.
- **and Athanasios Orphanides**, “Term Structure Estimation with Survey Data on Interest Rate Forecasts,” *Computing in Economics and Finance 2005* 474, Society for Computational Economics November 2005.
- **and** – , “The bond market term premium: what is it, and how can we measure it?,” *BIS Quarterly Review*, June 2007.
- **and Jonathan H. Wright**, “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.) 2005.
- Kozicki, S. and P.A. Tinsley**, “Shifting endpoints in the term structure of interest rates,” *Journal of Monetary Economics*, 2001, *47* (3), 613–652.
- Nelson, Charles R. and Andrew F. Siegel**, “Parsimonious Modeling of Yield Curves,” *The Journal of Business*, 1987, *60* (4), 473–489.

- Nicholls, DF and AL Pope**, “Bias in the estimation of multivariate autoregressions,” *Australian & New Zealand Journal of Statistics*, 1988, 30 (1), 296–309.
- Piazzesi, Monika and Eric T. Swanson**, “Futures prices as risk-adjusted forecasts of monetary policy,” *Journal of Monetary Economics*, May 2008, 55 (4), 677–691.
- **and Martin Schneider**, “Bond positions, expectations, and the yield curve,” Working Paper 2008-02, Federal Reserve Bank of Atlanta January 2008.
- Raftery, Adrian E. and Steven Lewis**, “How Many Iterations in the Gibbs Sampler?,” in “In Bayesian Statistics 4” Oxford University Press 1992, pp. 763–773.
- Rudebusch, Glenn**, “Commentary on “Cracking the Conundrum”,” *Brookings Papers on Economic Activity*, 2007, 38 (1), 317–325.
- Rudebusch, Glenn D. and Tao Wu**, “A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy,” *Economic Journal*, 07 2008, 118 (530), 906–926.
- , **Brian P. Sack, and Eric T. Swanson**, “Macroeconomic implications of changes in the term premium,” Working Paper Series 2006-46, Federal Reserve Bank of San Francisco 2006.
- Shaman, Paul and Robert A. Stine**, “The Bias of Autoregressive Coefficient Estimators,” *Journal of the American Statistical Association*, 1988, 83 (403), 842–848.
- Stambaugh, Robert F.**, “Predictive regressions,” *Journal of Financial Economics*, December 1999, 54 (3), 375–421.
- Svensson, L.E.O.**, “Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994,” NBER Working Papers 4871, National Bureau of Economic Research, Inc September 1994.
- Wachter, J.A.**, “A consumption-based model of the term structure of interest rates,” *Journal of Financial economics*, 2006, 79 (2), 365–399.

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

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

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 Basic MCMC algorithm and convergence diagnostics

For surveys on the use of MCMC methods in econometrics see Chib and Greenberg (1996) and Chib (2001). On estimation of asset pricing models using MCMC see Johannes and Polson (2009).

## B.1 Likelihood functions

Denote by  $X$  the latent factors for all time periods, and by  $Y$  the full sample of observed futures rates. The likelihood of the factors is

$$\begin{aligned} P(X|\theta) &= P(X|\rho, \lambda_0, \lambda_1, \Sigma) \\ &= \prod_{t=2}^T (2\pi)^{-\frac{k}{2}} |Q|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} v_t' Q^{-1} v_t\right) \end{aligned}$$

where  $v_t = X_t - \mu - \Phi X_{t-1}$ . Note that  $\Sigma$  determines not only the factor covariance matrix  $Q = \Sigma\Sigma'$  but also affects the physical dynamics  $\mu$  and  $\Phi$  (see equation (2)). For the distribution of the observations  $Y$  conditional on the factors  $X$  the likelihood is

$$\begin{aligned} P(Y|\theta, X) &= P(Y|\rho, \sigma_w^2, X) \\ &= \prod_{t=1}^T \prod_{i=1}^m (2\pi\sigma_w^2)^{-\frac{1}{2}} \exp\left(-\frac{(ED_t^{(i)} - a_i - h_i' X_t)^2}{2\sigma_w^2}\right). \end{aligned}$$

## B.2 Blockwise Metropolis-Hastings

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:

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

A blockwise Metropolis-Hastings (M-H) algorithm is used to obtain draws from this posterior distribution: At each iteration one draws from the full conditional posterior distribution for each block of parameters conditional on the other parameter values. If this distribution is not known in closed form, an M-H step is used to obtain the desired draw, otherwise one can directly draw from the conditional posterior (this is called a Gibbs step). The latent factors are drawn using the filter-forward-sample-backward (FFSB) algorithm developed by Carter and Kohn (1994). Iteratively drawing the blocks in this way leads to a sample which is approximately distributed according to the posterior  $P(\theta, X|Y)$ , which is the stationary distribution of the Markov chain (Chib and Greenberg, 1995).

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  $s$ th draw is retained, so that the number of iterations necessary for a sample of size  $G$  is  $B + s \cdot G$ . For the basic MCMC algorithm used to estimate a single DTSM specification the configuration is  $B = 20000$ ,  $G = 5000$  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  $T \cdot k$  values for the sampled paths of the latent factors.

Priors need to be specified for the parameters  $(\rho, \lambda_0, \lambda_1, \Sigma, \sigma_w^2)$ . While for the purpose of

estimation they could be taken to be diffuse or improper this would lead to problems when I turn to model selection—posterior distributions tend to be much less sensitive to the choice of priors than Bayes factors (Kass and Raftery, 1995).<sup>39</sup> For example, improper priors lead to undefined Bayes factors. Furthermore, very diffuse but proper priors will lead to results that necessarily favor the restricted model (the Lindley-Bartlett paradox; see Bartlett, 1957). Given the focus on restrictions on  $\lambda_1$  this prior should not be too diffuse.

I specify  $\rho$  to be uniformly distributed over the unit interval. The prior for  $Q = \Sigma\Sigma'$  is inverse Wishart (IW) and the prior for  $\sigma_w^2$  is inverse gamma (IG), both rather dispersed. The elements of  $\lambda_0$  and  $\lambda_1$  are normally distributed, independent, with mean zero and unit variance. The absolute magnitudes of the estimates for these parameters are small, thus despite the unit variance the prior is not very informative. Sensible alternative choices hardly affect the estimates I obtain. The joint prior  $P(\theta)$  also imposes the restriction that the eigenvalues of  $\Phi$  are at most one in absolute value, thus preventing explosive dynamics.

Instead of successively drawing every block in each iteration, one can randomize which block is sampled next (random scan M-H). I choose this method since one can fine-tune how frequently each block is sampled: The blocks that are more problematic in terms of mixing properties are sampled more frequently, and the blocks with parameters that mix well are sampled less frequently. This greatly increases the efficiency of the algorithm. Specifically I sample only one block in each iteration, and the five blocks  $X$ ,  $\rho$ ,  $(\lambda_0, \lambda_1)$ ,  $\Sigma$  and  $\sigma_w^2$  are sampled with probability 10%, 20%, 50%, 10% and 10%, respectively. In the following I describe how each block is sampled.

### Drawing the latent factors ( $X$ )

Given  $\theta$ , 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, 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 ( $\rho$ )

The risk-neutral dynamics are given by  $\Phi^Q$ , since I impose  $\mu^Q = 0$ . The prices of risk are taken as given when drawing this block, so drawing  $\Phi^Q$  affects not only  $a$  and  $H$  but also the transition matrix of the physical dynamics,  $\Phi$ . The matrix  $\Phi^Q$  is completely determined by the root  $\rho$ , for which the conditional posterior is

$$P(\rho|\theta_-, X, Y) \propto P(Y|\theta, X)P(X|\theta)P(\theta)$$

where  $\theta_-$  denotes all other parameters except  $\rho$ . Since one cannot sample directly from this distribution— $\rho$  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

---

<sup>39</sup>A Bayes factor is the ratio of the posterior model probabilities for two competing models/hypotheses.

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  $\zeta_\rho$  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 an RW step, the acceptance probability is given by

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

For the case that the prior restrictions ( $0 < \rho < 1$  and non-explosive  $\Phi$ ) 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 ( $\lambda_0$ and $\lambda_1$ )

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  $\theta_-$  denotes all parameters except for  $\lambda_0$  and  $\lambda_1$ , since the likelihood of the data for given risk-neutral dynamics does not depend on the prices of risk. The parameters enter the likelihood for the latent factors in a highly nonlinear fashion, thus I cannot directly sample from the conditional posterior distribution. I tried both RW and independent Metropolis proposals, and found the former to work better in this context. If there are no restrictions imposed on  $\lambda_1$  then I draw  $\lambda_0$  and each column of  $\lambda_1$  separately. The innovation for the RW is then a  $k \times 1$  vector of independent  $t_4$ -distributed innovations (one could of course use a multivariate t-distribution). For the case that some elements of  $\lambda_1$  are restricted to zero, I draw each non-zero element of  $\lambda_0$  and  $\lambda_1$  separately, using a univariate RW with  $t_4$ -distributed innovations. 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  $\Sigma$  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  $\theta_-$  denotes all parameters except  $\Sigma$ , 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  $\Sigma$ . Since I need successive draws of  $\Sigma$  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  $\Sigma$  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

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

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 ( $\sigma_w^2$ )

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  $\sigma_w^2$  is the natural conjugate IG distribution.

## B.3 Convergence diagnostics

After having obtained a sample using the described 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.<sup>40</sup> 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  $\rho$  and the elements of  $\lambda_0$  and  $\lambda_1$  traverse

---

<sup>40</sup>For surveys on convergence diagnostics see Cowles and Carlin (1996) and Brooks and Roberts (1998).

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.<sup>41</sup>

## C MCMC algorithm: Latent indicator variables

The algorithm developed here is based upon 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). What is particular to GVS is that the models are nested. The idea of product-space sampling is rather simple: 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 blockwise M-H sampling, in contrast to RJMCMC where the dimensionality differs between models. Since the models are nested, the complete set of parameters is simply the entire  $\lambda_1$  matrix, in addition to the other model parameters,  $(\rho, \lambda_0, \Sigma, \sigma_w^2)$ , which are also assumed to be shared among models. Keeping track of all parameters then just means that  $\lambda_1$  always contains  $k^2$  non-zero elements, but when calculating the likelihoods conditional on a specific set of restrictions, only those elements of  $\lambda_1$  that are ‘‘switched on’’ according to  $\gamma$  are taken to be non-zero.

When sampling the elements of  $\lambda_1$ , conditional on  $\gamma$ , one needs to distinguish whether a particular element 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,’’ a concept introduced to the theory of Bayesian model selection by Carlin and Chib (1995). More precisely, the conditional posterior of an arbitrary element of  $\lambda_1$ , which I denote by  $\lambda_i$ , is given by

$$P(\lambda_i | \lambda_{-i}, \gamma_i = 1, \gamma_{-i}, \theta_-, X, Y) \propto P(X | \theta, \gamma) P(\lambda_i | \gamma_i = 1) \quad (8)$$

$$P(\lambda_i | \lambda_{-i}, \gamma_i = 0, \gamma_{-i}, \theta_-, X, Y) \propto P(\lambda_i | \gamma_i = 0), \quad (9)$$

where  $\theta = (\rho, \lambda_0, \lambda_1, \Sigma, \sigma_w^2)$  as before,  $\theta_-$  denotes all parameters in  $\theta$  other than  $\lambda_1$ , and  $\lambda_{-i}$  ( $\gamma_{-i}$ ) contains all elements of  $\lambda_1$  ( $\gamma$ ) other than  $\lambda_i$  ( $\gamma_i$ ). These conditional distributions parallel the ones in equations (9) and (10) of Dellaportas et al. (2002). I assume prior conditional independence of the elements of  $\lambda_1$  given  $\gamma$ .

For the case that  $\lambda_i$  is currently included, I sample from equation (8). Note that the conditional posterior only depends on the latent factors  $X$  and not on the data  $Y$ , since all parameters that determine the likelihood of the data (e.g., cross-sectional dynamics and measurement error variance) are in the conditioning set. The prior for each price of risk parameter,  $P(\lambda_i | \gamma_i = 1)$ , is taken to be standard normal. A difference between the DTSM context and the original GVS context of Dellaportas et al. (2002) is that the conditional

---

<sup>41</sup>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).

posterior in (8) is not known analytically. Hence I employ Metropolis-Hastings to obtain the draws. I use a fat-tailed RW proposal, with scaling chosen to tune the acceptance probability.

If  $\lambda_i$  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  $\lambda_i$  for the full, unrestricted model. Carlin and Chib (1995) recommend choosing a distribution for the pseudo-prior close to the actual posterior, which for the elements of  $\lambda_1$  is likely to be similar between full model and restricted models.

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

$$\frac{P(\gamma_i = 1|\gamma_{-i}, \theta, X, Y)}{P(\gamma_i = 0|\gamma_{-i}, \theta, X, Y)} = \frac{P(X|\gamma_i = 1, \gamma_{-i}, \theta)}{P(X|\gamma_i = 0, \gamma_{-i}, \theta)} \frac{P(\lambda_i|\gamma_i = 1)}{P(\lambda_i|\gamma_i = 0)} \frac{P(\gamma_i = 1, \gamma_{-i})}{P(\gamma_i = 0, \gamma_{-i})}. \quad (10)$$

Since I use an uninformative prior, putting equal weight on  $\gamma_i = 1$  and  $\gamma_i = 0$ , the last term cancels out. Denoting the above ratio by  $q$ , the probability with which I draw  $\gamma_i = 1$  is given by  $q/(q + 1)$ .<sup>42</sup>

The MCMC algorithm used to produce a sample from the joint posterior for  $(\gamma, \theta, X)$  is again random-scan blockwise Metropolis-Hastings: In each iteration the block to be updated, either  $X$ ,  $\rho$ ,  $\Sigma$ ,  $\sigma_w^2$ ,  $\lambda_0$ , or  $(\lambda_1, \gamma)$ , is selected at random, then the parameters in the block are drawn from their full conditional posterior distribution. Only the last block needs further explanation, the others are updated exactly as in the full model. Conditional on  $\theta_-, X$ , and the data,  $(\lambda_1, \gamma)$  is drawn as follows: First the elements of  $\lambda_1$  are updated conditional on the value of  $\gamma$  from the previous iteration. Second, the elements of  $\gamma_i$  are drawn conditional on  $\gamma_{-i}$ ,  $\theta$ ,  $X$ , and the data. I implement two different versions of the algorithm: In the first version I update all elements of  $\lambda_1$  and  $\gamma$  in each step. In the second version I randomly choose to update only one pair  $(\lambda_i, \gamma_i)$ .

I run the algorithm with a burn-in sample of size  $B = 50,000$  and a sample size of  $G = 100,000$ , using every fifth draw from a longer chain. In order to get an idea of the convergence properties of this algorithm, I run several chains and make sure that the results are similar across chains. Since the first several models, which account for a large share of the posterior model probability mass, are similar across different runs of the chain and between the two algorithms, the algorithm correctly identifies the specifications with high posterior model probabilities. The final results presented in the text are obtained from aggregating the samples for two runs of the first algorithm and two runs of the second algorithm, i.e., they are based on a MCMC sample of size 400,000.

I assess whether the results are reasonable given the sample from the posterior for  $\lambda_1$  for the unrestricted model. This turns out to be the case, both based on individual credibility intervals and based on highest-posterior-density regions resulting from a normal approximation to the joint posterior. This is an important reality check for the algorithm described above.

As mentioned previously, an important issue in this context is the prior for  $\lambda_1$ . I performed

---

<sup>42</sup>A subtlety, which is ignored in the above notation, is that the joint prior  $P(\gamma, \theta)$  imposes that the physical dynamics resulting from any choice of  $\gamma$  and  $\lambda_1$  can never be explosive. 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.

additional sensitivity analysis, for example changing the prior variance of the elements of  $\lambda_1$  by orders of magnitude. My findings clearly show that the results of the model selection exercise remain robust for different choices of the priors.

## D Reversible-jump Markov chain Monte Carlo

Including the model indicator as a parameter, the model is now parameterized as  $(j, \theta_j, X_j)$ . Since the latent variables carry over between models, I write  $(j, \theta_j, X)$ . The algorithm I implement to obtain draws from the posterior distribution for  $(j, \theta_j, X)$  randomly chooses, in each iteration, between a “within-model” step, where the parameters of the current model are updated just like in the algorithm to estimate each model separately, and a “model-jump” step.

If a jump is attempted, first the candidate model indicator  $j'$  is chosen randomly, with equal probability for all models other than  $j$ . The question now is how to propose values for the parameters of model  $j'$ , denoted by  $\theta_{j'}$ . I decide for the models to share the parameters  $\rho$ ,  $\Sigma$ , and  $\sigma_w^2$ , denoted here by  $\theta_-$ , but not to share any elements of  $\lambda_0$  and  $\lambda_1$ . It might seem that the models naturally share  $\lambda_0$  and those elements of  $\lambda_1$  that are unrestricted in both models. However this version of the algorithm turned out to be the more efficient than to have the models share as many as possible parameters, mainly because the posterior distribution of some elements of  $(\lambda_0, \lambda_1)$  differs between models. To construct  $\theta_{j'}$  I take  $\theta_-$  from the current model together with the proposed values for  $\lambda_{j'}$ , by which I denote all non-zero elements of  $(\lambda_0, \lambda_1)$  in model  $j'$ . To propose values for  $\lambda_{j'}$  I take the normal approximation to the posterior distribution of  $\lambda_{j'}$ , which is available from the within-model simulation.

The idea of reversible-jump MCMC is that reversibility is ensured by matching the dimensions between candidate parameter-vector and proposed parameter-vector. The acceptance probability for the proposed jump is given by the minimum of one and

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

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 ( $q(j \rightarrow j') = 1/6$  for all  $j \neq j'$ ), and the fact that the likelihood of  $Y$  given  $X$  only depends on  $\theta_-$  (which does not change between jumps) the above ratio simplifies to

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

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  $(j, \theta_j) \rightarrow (j', \theta_{j'})$ .

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  $(j, \theta_j, X)$ , 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.

## E Long-run revisions to short rate expectations

The changes in far-ahead expectations of the term structure factors, using the eigendecomposition  $\Phi = VDV^{-1}$ , are

$$\begin{aligned} \lim_{h \rightarrow \infty} (E_{t+1} - E_t)X_{t+h} &= \lim_{h \rightarrow \infty} \Phi^h \varepsilon_{t+1} \\ &= V \left( \lim_{h \rightarrow \infty} D^h \right) V^{-1} \varepsilon_{t+1}, \end{aligned}$$

which can only be non-zero if one of the eigenvalues is unity in absolute value. Since  $\Phi^Q$  has a unit eigenvalue associated with the level factor, and  $\Phi = \Phi^Q + \Sigma\lambda_1$ , if  $\Phi$  has a unit eigenvalue it will be associated with the level factor. In this case

$$\lim_{h \rightarrow \infty} (E_{t+1} - E_t)X_{t+h} = V \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} V^{-1} \varepsilon_{t+1}.$$

The long-run revision of short rate expectations is then

$$\lim_{h \rightarrow \infty} (E_{t+1} - E_t)r_{t+h} = \delta_1' V \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} V^{-1} \begin{pmatrix} \varepsilon_{t+1}^{(1)} \\ 0 \\ 0 \end{pmatrix},$$

where  $\varepsilon_{t+1}^{(1)}$  is the level shock.

Table 1: Model specifications

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

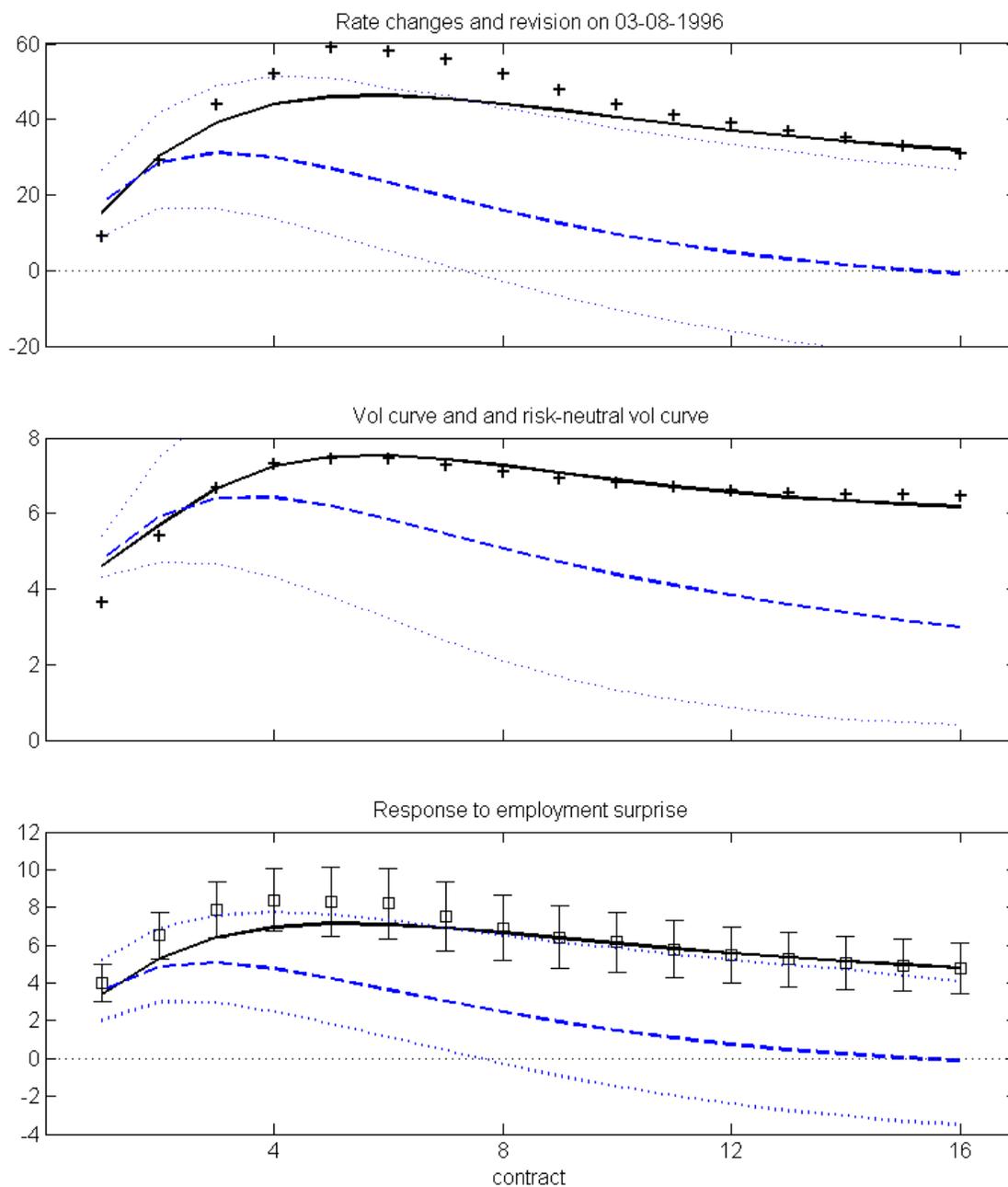
Alternative model specifications and estimated posterior model probabilities. Column two: frequency of specification in GVS algorithm. Columns three to five indicate which elements in the respective columns of  $\lambda_1$  are restricted (0) and unrestricted (1). Column six: largest eigenvalue of  $\Phi$ . 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 2: Parameter estimates for unrestricted model

<i>Risk-neutral dynamics</i>				
$\rho$	.9973	[.9973, .9973]		
<i>Risk sensitivity parameters</i>				
	$\lambda_0$	level	$\lambda_1$	curvature
	constant		slope	
level risk	.1210	-.0164	.0103	.0014
	[-.03, .28]	[-.04, .01]	[-.01, .03]	[-.02, .02]
slope risk	<b>.3042</b>	<b>-.0444</b>	-.0190	<b>.0296</b>
	[.15, .46]	[-.07, -.02]	[-.04, .00]	[.01, .05]
curvature risk	.1273	<b>-.0239</b>	.0052	-.0009
	[-.02, .28]	[-.05, -.00]	[-.02, .03]	[-.02, .02]
<i>Factor shocks</i>				
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]		
<i>Measurement errors</i>				
$\sigma_w^2$	.0039	[.0039, .0040]		

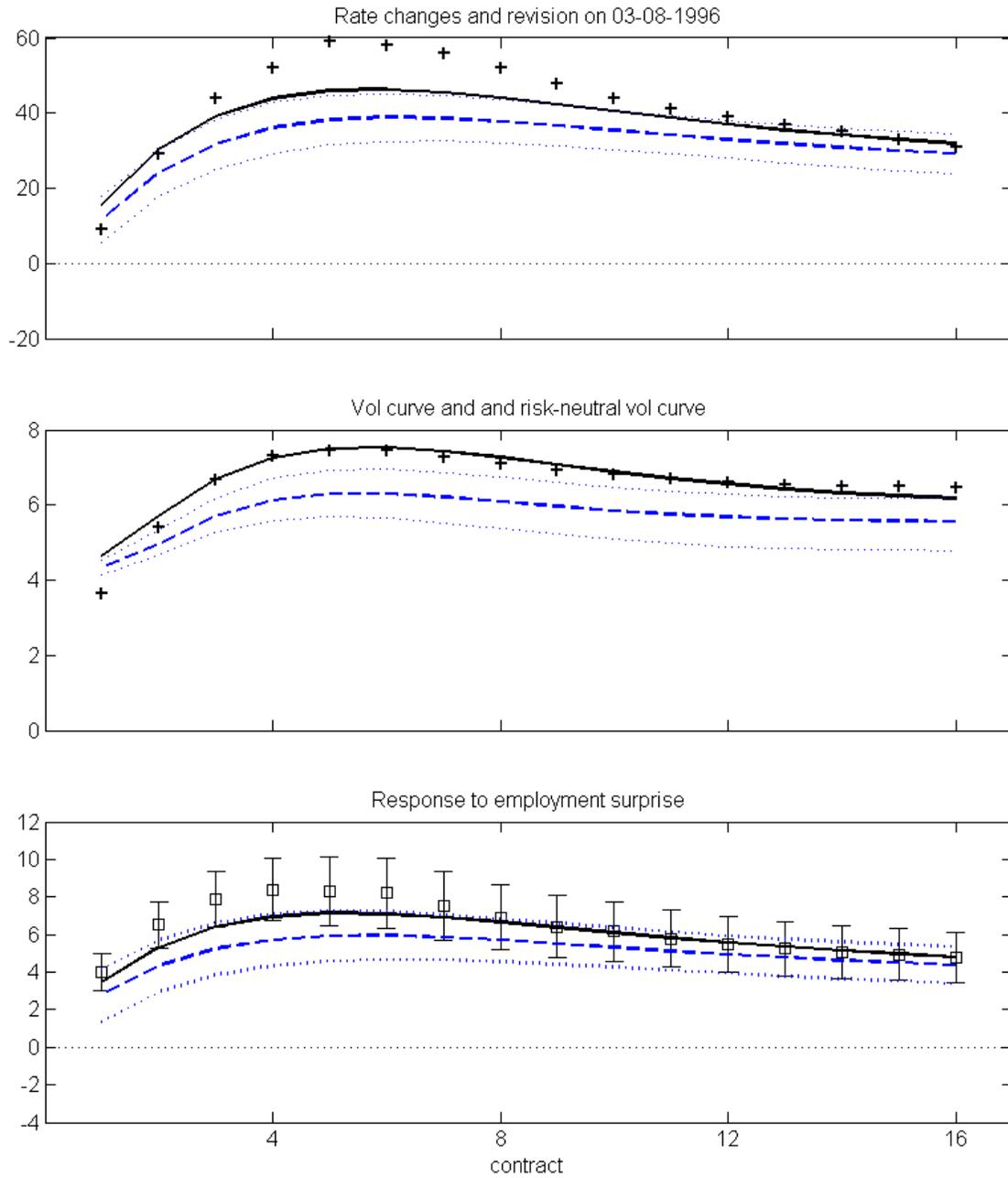
Posterior means and 95% Bayesian credibility intervals in squared brackets for model parameters. Estimates of risk sensitivity parameters are boldfaced if the credibility interval does not straddle zero.

Figure 1: Implications of unrestricted specification ( $M_1$ )



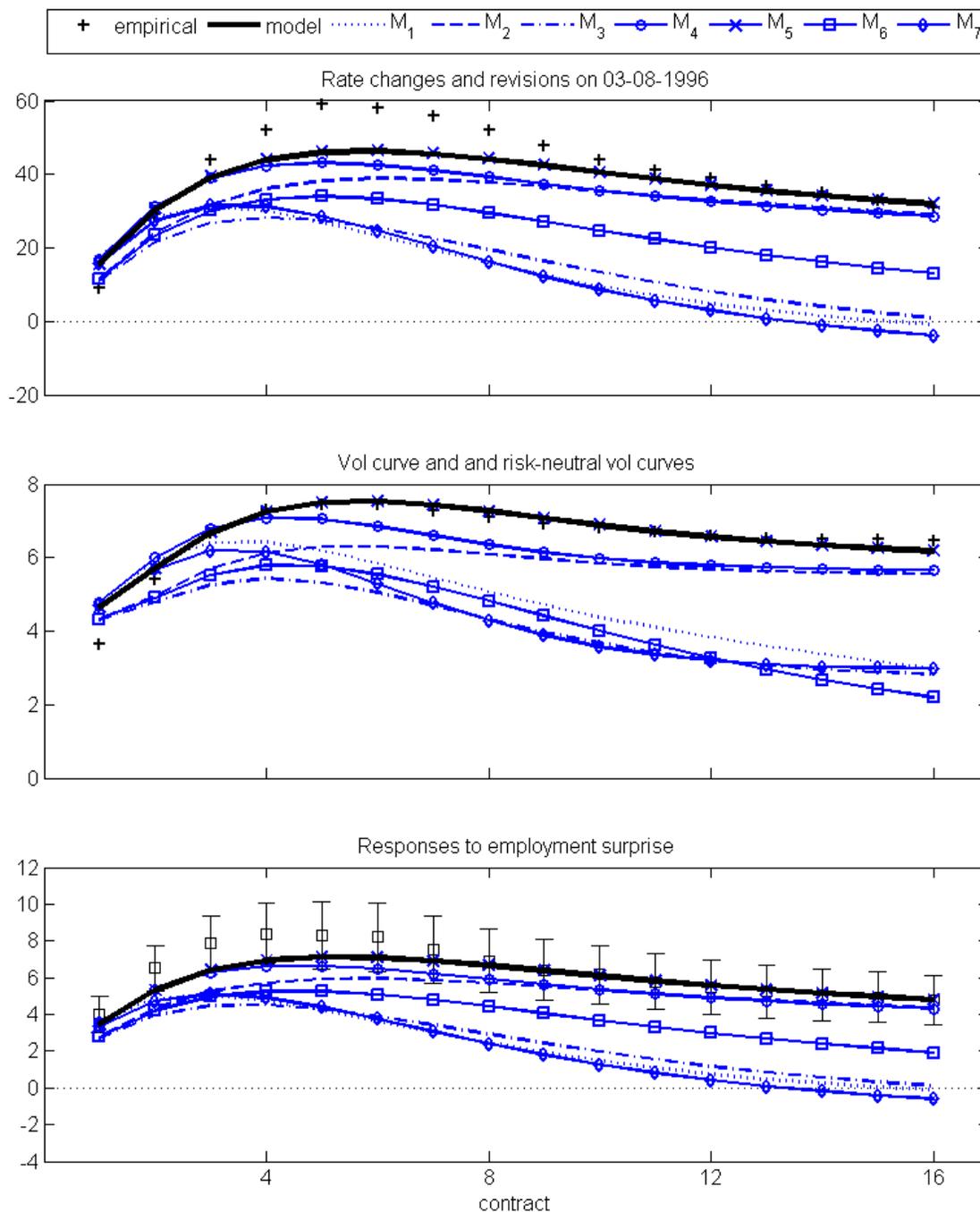
First panel: Empirical (crosses) and model-implied (solid line) rate changes on Mar-08 1996 with estimated changes of risk-neutral rates. Second panel: Sample standard deviations (crosses) and model-implied standard deviations (solid line) for futures rate changes as well as model-implied standard deviations 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 ( $M_2$ )



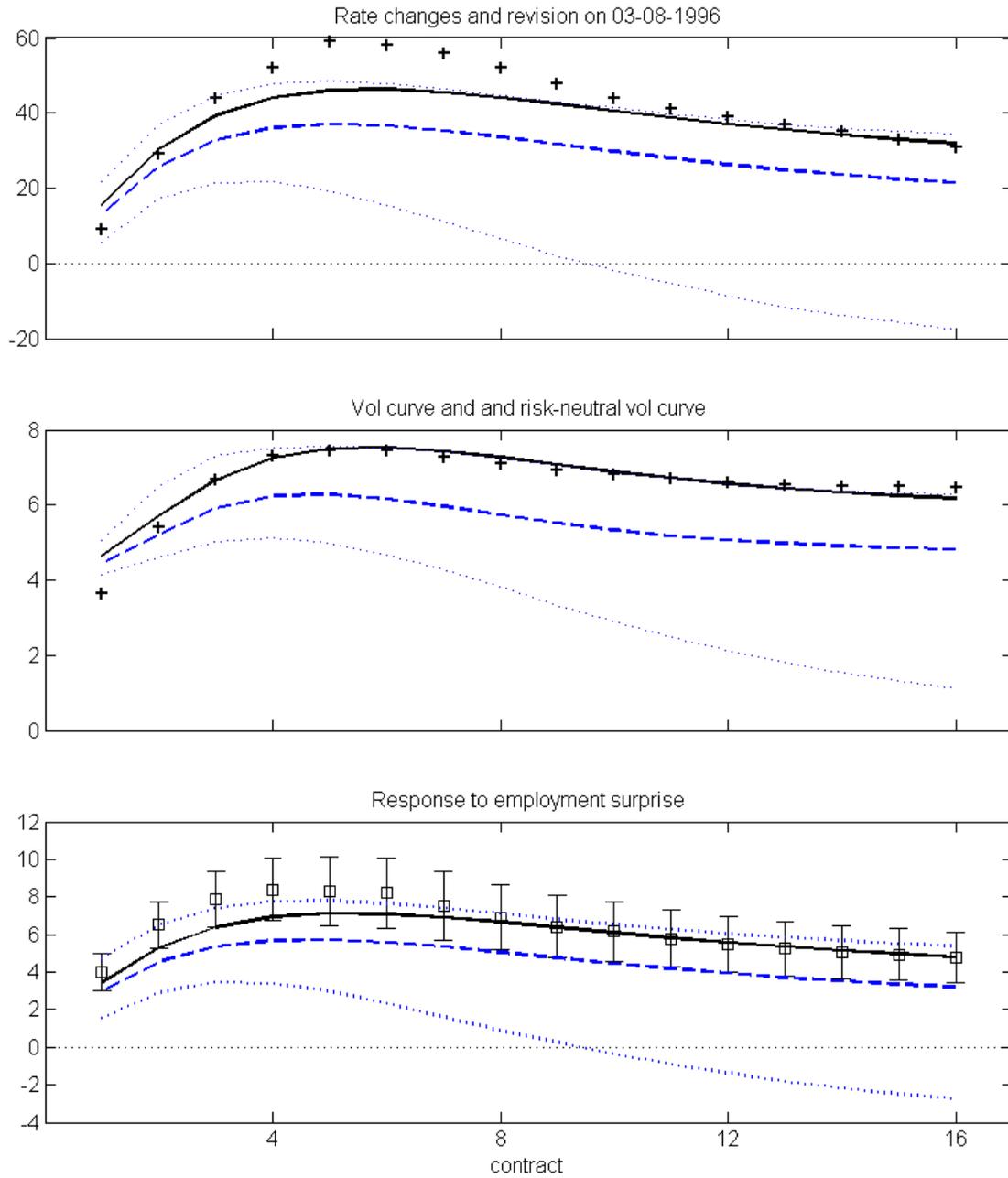
See description of Figure 1.

Figure 3: Comparison of alternative model specifications



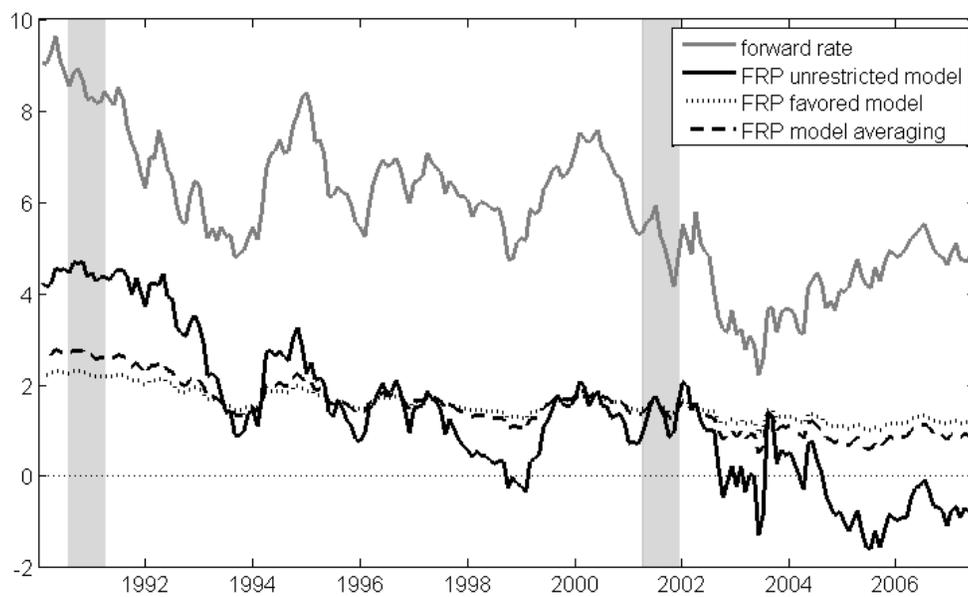
First panel: Empirical (crosses) and model-implied (solid line) rate changes on Mar-08 1996 with estimated changes in risk-neutral rates across models. Second panel: Sample standard deviations (crosses) and model-implied standard deviations (solid line) for futures rate changes as well as alternative 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 (*BMA*)



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.