Term Premia and the News[*]

Michael D. Bauer[†]

Job Market Paper
November 2, 2009

Abstract

How do short rate expectations and term premia respond to news? Dynamic term structure models typically imply that the term premium accounts for most of the pro-cyclical response of long-term interest rates, which is at odds with the conventional wisdom about bond risk premia. Bias and lack of precision in the estimated short rate dynamics make it difficult to interpret this evidence. This paper solves these problems by imposing plausible zero restrictions on the market prices of risk. The no-arbitrage assumption becomes useful for estimation, because information in the cross section helps to pin down the dynamics of the short rate. Inference about term premia is performed in a Bayesian framework based on Markov Chain Monte Carlo methods. This allows the researcher to select plausible restrictions and to correctly quantify statistical uncertainty. The main empirical result is that under the restrictions favored by the data the expectations component, and not the term premium, accounts for the majority of high-frequency movements of long rates and for essentially all of their procyclical response to macroeconomic news.

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

JEL Classifications: E43, E44, E52, G12

1   Introduction

Policy makers and academic researchers have keen interest in the estimation of term premia in long-term interest rates. Different approaches have been used for this purpose, such as return regressions,[1] no-arbitrage dynamic term structure models[2] (DTSMs) and more recently macro-finance term structure models.[3] Previous studies established that there is a sizeable term premium which varies over time. Thus the expectations hypothesis, which posits a constant term premium, is at odds with the data. Moreover the term premium seems to vary at business-cycle frequencies and this variation is countercyclical (Cochrane and Piazzesi, 2005; Piazzesi and Swanson, 2008).

An important question in this context is how the response of the term structure of interest rates to news events, such as macroeconomic announcements and policy actions, decomposes into revisions of short rate expectations and changes in term premia. More generally, how do term premia change at high frequencies? This question is of relevance both to policy makers, who need up-to-date information about market participants' expectations of the future policy path, and to investors, since optimal asset allocation requires knowing how new information changes expected returns. A recent attempt to answer this question was made by Meredith Beechey (2007), who uses the DTSM of Kim and Wright (2005) to decompose rate changes into expectations and term premium components. Her study finds that the procyclical response of long-term interest rates is mainly due to changes in term premia – short rate expectations seem to hardly react to the news. However, this strong procyclical response of term premia at high frequencies is a puzzle in light of the conventional term premium wisdom cited above.

This puzzling evidence results from two general problems with the estimation of DTSMs and term premia that are caused by the high persistence[4] of the short rate: *lack of precision* and *bias* in estimates of the short rate's dynamic properties. At high frequencies the persistence is particularly strong, so the resulting problems are even more severe. The literature has documented these issues, for example in Duffee and Stanton (2004) and Kim and Orphanides (2005), but so far has offered no solution. The present paper, which focuses on high-frequency changes in interest rates and risk premia, solves both of these problems.

Persistence leads to imprecise estimates of the unconditional mean and the speed of mean

---

[1] See for example Fama and Bliss (1987), Cochrane and Piazzesi (2005) and Piazzesi and Swanson (2008).

[2] The standard reference for affine no-arbitrage term structure models is Duffie and Kan (1996), applications to term premium estimation include Dai and Singleton (2002) and Kim and Wright (2005).

[3] Studies that incorporate macro-factors into no-arbitrage models and study term-premium properties include Ang and Piazzesi (2003); Rudebusch et al. (2006); Joslin et al. (2008).

[4] The null of a unit root can usually not be rejected for the short rate, see for example Rose (1988) and Jardet et al. (2009). Whereas the short rate is not literally I(1), since it is bounded from below and usually remains in a certain range, its largest root is certainly very close to one.

reversion of the short rate intuitively because the short rate does not revert to its mean very often. Term premium estimates rely on forecasts of the short rate, thus the high estimation uncertainty about short rate dynamics translates into equally large uncertainty about the term premium. Empirical results that proceed by using term premium estimates without accounting for their uncertainty (e.g. Beechey, 2007) should therefore be taken with a grain of salt. Unfortunately, as John Cochrane (2007, p. 278) puts it,

> we are usually treated only to one estimate based on one a priori specification, usually in levels, and usually with no measure of the huge sampling uncertainty.

An indication of the magnitude of estimation and specification uncertainty are the strikingly different term premium estimates in the literature (Swanson, 2007). I will quantify this statistical uncertainty and show its relation to the pricing of risk.

A closely related problem is an upward bias in the estimated speed of mean reversion of the short rate: The largest root of a persistent variable is generally under-estimated. This results in short rate forecasts that revert to the unconditional mean too quickly. As a consequence, changes in term premia are usually found to be the dominant source of the variation in long-term interest rates (e.g. Kim & Wright, 2005). This is implausible since we think that the term premium moves slowly.

Estimation of a DTSM requires inference about the short rate dynamics under both the risk-neutral ($\mathbb{Q}$) measure and the physical ($\mathbb{P}$) measure. The $\mathbb{Q}$-dynamics determine the loadings of the cross section of interest rates on the term structure factors and can be pinned down very precisely. The $\mathbb{P}$-dynamics describe the evolution of the factors over time and estimation is difficult for the above-mentioned reasons. Can we use the information in the cross section to improve our estimates of the $\mathbb{P}$-dynamics? No-arbitrage requires consistency between cross-sectional and dynamic properties of the term structure, allowing for a risk adjustment. If market prices of risk are unrestricted, then the parameters under the $\mathbb{Q}$ measure and under the $\mathbb{P}$ measures are estimated independently of each other – no-arbitrage has no bite in this case. The solution is to impose restrictions on the prices of risk. Whereas this fact has been recognized in the literature (Kim and Orphanides, 2007; Cochrane and Piazzesi, 2008), the important question about which restrictions we should impose has so far not been satisfactorily answered. A commonly employed approach is to restrict some risk sensitivity parameters with large standard errors to zero and then to re-estimate the model. This ad-hoc procedure, unsatisfactory for several reasons, usually leads to very few restrictions and does not solve the bias and uncertainty problems.[5]

---

[5]Prominent studies that employ this two-step approach are Dai and Singleton (2002); Ang and Piazzesi (2003); Kim and Wright (2005). None of these restrict more than three risk sensitivity parameters to zero.

This paper develops a new statistical framework for the estimation of DTSMs which allows the researcher to impose sensible restrictions on the prices of risk. The question posed by Cochrane (2007, p. 276), "Can statistics help us?", is answered with a strong affirmative. The physical dynamics of the short rate are estimated more accurately and more precisely. The resulting decompositions of rate changes are more reliable, and they also turn out to be more plausible than those of conventional DTSMs. My approach opens up a new road for estimation of term structure models, where parsimony and no-arbitrage play a more prominent role than they currently do in the term structure literature.

Estimation and inference is performed using Markov Chain Monte Carlo (MCMC) methods. A Bayesian framework is necessary not only for correctly quantifying statistical uncertainty but more importantly for appropriately restricting the market prices of risk: The tools of Bayesian model selection allow me to select those zero restrictions on the risk sensitivity parameters that are most strongly supported by the data. I develop a new algorithm related to Gibbs Variable Selection, using latent indicator variables that represent the restrictions, in order to identify plausible models. For this smaller subset of candidate models I can then precisely estimate posterior model probabilities and assess their economic implications.

The DTSM used in this paper belongs to the affine Gaussian class. Its key distinguishing characteristic is that the $\mathbb{Q}$-dynamics are specified in a way that identifies the latent factors a priori as level, slope and curvature. This is crucial for two reasons: First, the presence of a level factor, an empirical necessity, requires a unit root for the short rate under $\mathbb{Q}$. Because the model is parameterized in terms of $\mathbb{Q}$-dynamics and prices of risk, it then depends on the restrictions on the prices of risk whether there is a unit root under $\mathbb{P}$. Hence I let the data choose between a stationary or integrated specification for the short rate, instead of imposing this a priori. Second, having labeled factors allows for an economic interpretation of risk premia, based on the inference about, for example, the importance of level vs. slope risk. The DTSM in this paper parallels the Dynamic Nelson-Siegel model in Christensen et al. (2007), but gives rise to new Nelson-Siegel-type factor loadings since it is set in discrete time.

I also differ from most previous studies in basing the analysis on Eurodollar futures, which have several practical advantages over Treasury bonds in this context: 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 (smoothed vs. unsmoothed, etc.). Second, the liquidity is very high, in fact Eurodollar futures are the most liquid futures contracts worldwide (in terms of open interest). Third, the most liquidly traded government bonds, on-the-run Treasury securities, do not cover the maturity spectrum at a similar detail. Last, the futures contracts are not affected by flight-to-quality effects or other extraordinary

forces affecting supply and demand of Treasury securities. A conceptual 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.[6]

I decompose rate changes into three components: Revisions of short rate expectations, surprise changes in the term premium and expected returns. The usual decomposition into predictable and unpredictable components overlooks that unpredictable changes have two different sources. Since the predictable component is negligible at high frequencies, the question is how much changes in short rate expectations vs. surprise term premium changes contribute to observed rate changes. Because of the bias problem unrestricted models wrongly attribute the majority of long-maturity rate changes to the term premium.

The paper also introduces some new ways to represent the implications of different models visually. I derive a "risk-neutral volatility curve", which captures the volatility of short rate expectations across maturities. Based on the analytical decompositions I graphically summarize the contribution of short rate expectations to rate changes on individual days, to the level of volatility, and to the systematic response of forward rates to macroeconomic news.

Turning to the empirical findings of the paper, a key result is that *the data support strongly restricted prices of risk.* This is plausible if we believe in no-arbitrage: physical and risk-neutral dynamics *should* be close to each other. As a consequence of the restrictions the physical dynamics are estimated with higher precision, and our inference about the term premium becomes more reliable. This is the key contribution of the paper: by imposing sensible zero restrictions on risk sensitivity parameters we can overcome the bias and uncertainty problems that most DTSMs suffer from. Those restrictions that imply a unit root for the short rate receive particularly strong empirical backing – a stochastically trending short rate evidently is a good approximation to the true data-generating process, at least at the daily frequency.

Changes in short rate expectations account for most of the daily volatility in forward rates across all maturities. Thus *the procyclical response of long forward rates to macro news is mainly due to revisions of short rate expectations* – I overturn the result that this response primarily reflects changes in the term premium. The contribution of risk premia to variation in long rates is found to be much smaller than implied by most DTSMs in the literature. My results accord with our intuition in two ways: First, we think that term premia move in a countercyclical fashion, thus their contribution to procyclical interest rate changes caused by news should be small. Second, since most macroeconomists would agree that term premia probably move at business cycle frequencies, we would not expect them to account for much

---

[6]Hence Nelson-Siegel loadings without any convexity adjustments are consistent with no-arbitrage.

variability at the daily frequency. The decompositions of daily forward rate changes into risk premium and expectations components proposed in this paper are thus more plausible than those implied by the largely unrestricted DTSMs that are common in the literature.

These results remain robust when I account for specification uncertainty. The decompositions obtained using Bayesian model averaging (BMA) confirm an important role of short rate expectations for variation in long rates. Importantly the use of BMA solves the "discontinuity problem", the stark difference between forecasts for the short rate depending on whether its largest root is less than or equal to one (Cochrane and Piazzesi, 2008; Jardet et al., 2009), since estimates effectively are averages of stationary and non-stationary specifications.

The paper is structured as follows: Section 2 describes the DTSM and shows how it can be used to decompose observed rate changes and the term structure of volatility. In Section 3 the model is estimated without restrictions on the prices of risk, revealing dramatic uncertainty and bias in decompositions of rate changes. Section 4 develops and applies a new framework for estimation of DTSMs, based on restrictions on the prices of risk. Section 5 concludes.

2 Dynamic Term Structure Model

In this section I introduce the affine Gaussian DTSM with its particular specification of the risk-neutral dynamics, present a new decomposition of changes in forward rates, introduce the "risk-neutral vol curve" and discuss the pricing of Eurodollar futures.

2.1 Affine Gaussian DTSMs

Denote by X_t the $(k \times 1)$ vector of term structure factors which represent the new information that market participants obtain at time t. Generally it can contain both latent and observable factors, but this paper uses only latent factors. 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:

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

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.[7]

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

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

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

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

The risk sensitivity parameters λ_0 $(k \times 1)$ and λ_1 $(k \times k)$ determine the behavior of risk premia. Under these assumptions the risk-neutral dynamics (see Appendix A) are given by

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

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

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

This discrete-time affine Gaussian DTSM was introduced by Ang and Piazzesi (2003) and is now widely used. While it does not allow for stochastic volatility it is rather flexible in matching risk premia (Dai and Singleton, 2003).[8]

No-arbitrage requires the consistency of dynamic properties (determined by μ and Φ) and cross-sectional properties (determined by $\mu^{\mathbb{Q}}$ and $\Phi^{\mathbb{Q}}$) of interest rates, allowing for a risk-adjustment, and (6) makes this risk-adjustment explicit. Note however that if λ_0 and λ_1 are left unrestricted, then (6) is not restrictive at all: Any estimates for the physical dynamics are consistent, for some prices of risk, with a given choice of risk-neutral dynamics. Most studies

---

[7]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.

[8]Examples of studies that assess the behavior of the term premium using this affine Gaussian framework are Duffee (2002); Ang and Piazzesi (2003); Kim and Orphanides (2005); Kim and Wright (2005); Rudebusch and Wu (2008); Joslin et al. (2008).

impose no or only minimal restrictions on λ_0 and λ_1. This paper will show that strong zero-restrictions on λ_1 are supported by the data, and that imposing these restrictions provides us with more precise and more plausible term premium estimates.

2.2 Forward rates

With this framework we can price any asset with payoff depending on the future path of the short rate, the price being given by the discounted expected future payoff under $\mathbb{Q}$. Instead of considering bonds, this paper focuses on money market futures, which pay off according to the average short rate over a future time horizon. Thus the main object of interest is the expected future short rate under the risk-neutral measure, $f_t^n = E_t^{\mathbb{Q}}(r_{t+n})$. I will refer to this object as a *forward rate*, although by this term one would usually mean the rate that can be contracted today for a loan from $t+n$ to $t+n+1$ by entering the appropriate bond positions and which includes Jensen inequality terms.[9]

Solving for these forward rates is straightforward because of the linearity of the short rate and the availability of analytical expressions for the conditional expectation of X_t. We obtain

$$
\begin{aligned}
f_t^n &= E_t^{\mathbb{Q}}(\delta_0 + \delta_1' X_{t+n}) \\
&= \delta_0 + \delta_1' \left[\sum_{i=0}^{n-1} (\Phi^{\mathbb{Q}})^i \mu^{\mathbb{Q}} + (\Phi^{\mathbb{Q}})^n X_t \right] \\
&= A_n + B_n' X_t
\end{aligned}
\tag{7}
$$

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

These loadings 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.

2.3 The arbitrage-free Dynamic Nelson-Siegel model

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

---

[9]The actual forward rate based on bond positions is equal to $\log \frac{P_t^n}{P_t^{n+1}} = \log \frac{E_t^{\mathbb{Q}} \exp(-r_t - r_{t+1} - \ldots - r_{t+n-1})}{E_t^{\mathbb{Q}} \exp(-r_t - r_{t+1} - \ldots - r_{t+n})}$ where P_t^n is the time-t price of a discount bond with n days until maturity.

canonical form of Joslin et al. (2009) (JSZ), who impose the restrictions $\mu^{\mathbb{Q}} = 0, \delta_1 = \iota_k$ and parameterize the $\mathbb{Q}$-dynamics in terms of δ_0 and the eigenvalues of $K^{\mathbb{Q}} = \Phi^{\mathbb{Q}} - I_k$, which is taken to be in real ordered Jordan form.[10]

In canonical DTSMs with only latent factors the role of each factor is a priori left unidentified. On the other hand the widely used yield-curve parametrization of Nelson and Siegel (1987) posits a simple factor structure for forward rates involving three factors that correspond to level, slope and curvature. The dynamic version of the original 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, as shown by Christensen et al. (2007). My discrete-time analogue to their model, using a discretization scheme with one-period increments, is given by the following specification:

$$\delta_0 = 0, \ \delta_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \ \mu^{\mathbb{Q}} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \ \Phi^{\mathbb{Q}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \rho & 1-\rho \\ 0 & 0 & \rho \end{pmatrix} \tag{8}$$

where the parameter ρ is restricted to be less than one in absolute value. This is the Dynamic Nelson-Siegel (DNS) specification in discrete time. Straightforward algebra based on equation (7) leads to a Nelson-Siegel-type forward rate curve given by

$$f_t^n = X_t^{(1)} + \rho^n X_t^{(2)} + n(1-\rho)\rho^{n-1} X_t^{(3)}. \tag{9}$$

Notably there is no convexity term since I consider $f_t^n = E_t^{\mathbb{Q}}(r_{t+n})$.

The DNS specification in (8) is equivalent to the JSZ normalization with three over-identifying restrictions (JSZ, Section 4.2). The first restriction is the unit eigenvalue of $\Phi^{\mathbb{Q}}$, i.e. a zero eigenvalue of $K^{\mathbb{Q}}$. The empirical evidence (e.g. Gürkaynak et al., 2005) overwhelmingly calls for a unit root under $\mathbb{Q}$ since otherwise long-horizon forward rates would be constant. The yield curve cannot have a level factor unless $\Phi^{\mathbb{Q}}$ has a unit eigenvalue. The second restriction is the zero long-run mean of the short rate under $\mathbb{Q}$ ($\delta_0 = 0$). Given that $X_t^{(1)}$ serves as a level factor there is no need for a non-zero unconditional mean. The unit root and zero mean under $\mathbb{Q}$ are highly plausible and empirically necessary restrictions on the JSZ

---

[10]All normalizations are imposed on the $\mathbb{Q}$-dynamics, hence for given (observable or filtered) factors and absent restrictions on the prices of risk, consistent estimates of (μ, Φ) can be obtained using ordinary least squares, as shown by JSZ.

normalization.

The third restriction of the DNS model is that the two other eigenvalues of $\Phi^{\mathbb{Q}}$ are equal, which identifies the third factor as curvature. This somewhat restrictive assumption is useful because now the factors are a priori identified as level, slope and curvature. One advantage of this is that it enables us to provide an economic interpretation of risk premia.[11] Labeling the factors by means of a particular choice of the risk-neutral dynamics, as well as introducing a unit root under $\mathbb{Q}$, are the two major advantages of the DNS specification.

An important remark about the consequence of a unit root under $\mathbb{Q}$: Because I will estimate the model based on a parametrization in terms of (λ_0, λ_1) and not in terms of (μ, Φ), and since $\Phi = \Phi^{\mathbb{Q}} + \Sigma \lambda_1$, it will depend entirely on the restrictions imposed on λ_1 whether Φ has a unit eigenvalue. Absent any restrictions, estimates will generally imply a stationary short rate. By having the data choose plausible zero restrictions on λ_1 the possibility of a unit root under $\mathbb{P}$ will be explicitly allowed for.

2.4   Risk-neutral rates and forward risk premia

The term premium can be defined in different ways, and I focus on the forward risk premium, which I relate later to the return risk premium.[12] If the marginal investor was 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 $\tilde{f}_t^n$. We have

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

with loadings given by

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

Risk-neutral rates embody the expectations about future short rates and thus about future monetary policy. They are not observable and have to be inferred by constructing forecasts for the short rate. The forward risk premium, denoted by Π_t^n, is defined as the difference

---

[11]Specifically we would like to know whether level risk or slope risk causes time-variation in the term premium, since this provides some hints about the role of different macroeconomic shocks, as pointed out by Cochrane and Piazzesi (2008).

[12]Specifically, the term premium can be defined equivalently as a yield risk premium (the difference between a bond yield and the average expected future short rate), a forward risk premium (the difference between a forward rate and the expected future short rate), or a return risk premium (the expected excess return on a bond or futures contract). For a detailed discussion of how these are related see Cochrane and Piazzesi (2008).

between the forward rate and the risk-neutral forward rate:

$$\Pi_t^n = f_t^n - \tilde{f}_t^n = A_n - \tilde{A}_n + (B_n - \tilde{B}_n)'X_t.$$

Importantly, the statistical uncertainty about μ and Φ translates into uncertainty about risk-neutral rates and forward risk premia.

2.5 Decomposing rate changes

The main question asked in this paper is to what extent daily changes in the forward rate curve are driven by changing short rate expectations. By definition, rate changes decompose into changes in risk-neutral rates and changes in forward risk premia, $f_{t+1}^n - f_t^{n+1} = \tilde{f}_{t+1}^n - \tilde{f}_t^{n+1} + \Pi_{t+1}^n - \Pi_t^{n+1}$. The DTSM can be used to provide a more detailed decomposition:

$$
\begin{aligned}
f_{t+1}^n - f_t^{n+1} &= A_n + B_n'X_{t+1} - A_{n+1} - B_{n+1}'X_t \\
&= B_n'\Sigma\varepsilon_{t+1}^{\mathbb{Q}} \\
&= \tilde{B}_n'\Sigma\varepsilon_{t+1} + (B_n - \tilde{B}_n)'\Sigma\varepsilon_{t+1} + B_n'\Sigma\lambda_t.
\end{aligned}
\tag{10}
$$

This rate change corresponds to the one-period return of a hypothetical futures contract which pays the difference between the realized future short rate and the contracted rate.[13] Notably it is an *excess return*, because the risk-neutral expected return, $\mathrm{E}_t^{\mathbb{Q}}(f_{t+1}^n - f_t^{n+1})$ is zero. Expression (10) decomposes this return into three components:

Revisions to short rate expectations – The first component corresponds to the change in the expectation of the short rate for $t + n + 1$:

$$
\begin{aligned}
(\mathrm{E}_{t+1} - \mathrm{E}_t)r_{t+n+1} &= \delta_1'(\mathrm{E}_{t+1}X_{t+n+1} - \mathrm{E}_tX_{t+n+1}) \\
&= \delta_1'\Phi^n\Sigma\varepsilon_{t+1} = \tilde{B}_n'\Sigma\varepsilon_{t+1}.
\end{aligned}
$$

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.

Surprise changes in the forward risk premium – The second component is equal to the

---

[13]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 contracted rate and short rate, in which case the above rate change corresponds to the return on a short position.

unexpected change in the forward risk premium:

$$(\Pi^n_{t+1} - \mathrm{E}_t\Pi^n_{t+1}) = (B_n - \tilde{B}_n)'(X_{t+1} - \mathrm{E}_tX_{t+1})$$
$$= (B_n - \tilde{B}_n)'\Sigma\varepsilon_{t+1}.$$

Expected returns – The third component is equal to the expected change in the forward risk premium:

$$\mathrm{E}_t\Pi^n_{t+1} - \Pi^{n+1}_t = \mathrm{E}_tf^n_{t+1} - f^{n+1}_t = A_n + B'_n\mathrm{E}_tX_{t+1} - A_{n+1} - B'_{n+1}X_t$$
$$= B'_n\mathrm{E}_t\varepsilon^Q_{t+1} = B'_n\Sigma\lambda_t = B'_n\Sigma(\lambda_0 + \lambda_1 X_t)$$

This term captures the predictable part of the daily return and corresponds to the *return risk premium*.[14] For daily rate changes this component is negligibly small. How large it is for longer holding periods is an important question which I consider in Section 4.6.

Using equation (10) changes in forward rates can be decomposed into expectations and risk premium components. Importantly this decomposition will inherit the statistical uncertainty from the inference about the unknown parameters and factors. I will show below the dramatic uncertainty in DTSMs that lack the appropriate restrictions on the prices of risk.

To foreshadow a key empirical result: Typically DTSMs imply that for long maturities the first component is small and the second component accounts for most of daily rate changes. This is puzzling given the conventional wisdom about bond risk premia. If restrictions are imposed on the prices of risk a much larger share of rate changes is attributed to the first component, changes in expectations, implying little daily variability of the term premium.

2.6 Term Structure of Volatility

The term structure of volatility, the "vol curve", describes the volatility of changes in yields or forward rates across maturities, either in the sample or in population. Given the decomposition in (10) the variance of forward rate changes is given by

$$Var(f^n_{t+1} - f^{n+1}_t) = B'_n\Sigma(I_k + \mathrm{Var}(\lambda_t))\Sigma'B_n$$

The term structure of volatility 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,

---

[14]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 λ_t is non-zero, see for example the model of Cochrane and Piazzesi (2008).

and by a predictable component, the variation in the prices of risk. It will turn out that the predictability of daily changes is very small, thus $Var(f_{t+1}^n - f_t^{n+1}) \approx B_n' \Sigma \Sigma' B_n$.

Our framework allows us to assess the importance of changes in short rate expectations and forward risk premia for variability in forward rates. Specifically, we can calculate the term structure of volatility that would prevail if forward rates were only driven by changes in short rate expectations, i.e. if term premia were constant. The variance of changes in risk-neutral forward rates is

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

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

2.7   Eurodollar futures

In order to estimate the term structure model and for all subsequent empirical analysis, this paper uses Eurodollar futures contracts.[15] These instruments settle based on the 3-month LIBOR rate at some future date (the settlement day). This settlement rate can safely be taken to be the average expected short rate (under $\mathbb{Q}$) for the 3-month period following the settlement day: $S_t = N^{-1} \sum_{h=0}^{N-1} \mathrm{E}_t^{\mathbb{Q}}(r_{t+h})$, where N is the number of days in the quarter, taken to be 91 throughout this paper.[16] Eurodollar futures contracts involve no cost today and have payoff proportional to the difference between contracted rate and settlement rate.[17] For the Eurodollar futures contract that settles at the end of quarter i, where $i = 1$ corresponds to the current quarter, we have the following fundamental pricing equation:

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

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

---

[15]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 08/29/2009).

[16]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 counter-party when lending at LIBOR. This spread was very small (around 8 basis points) and showed little variability throughout the period of our data set, which ends before the start of the recent financial turmoil.

[17]We 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 our context.

thus given by

$$
\begin{aligned}
ED_t^{(i)} &= \mathrm{E}_t^{\mathbb{Q}}(S_{T(i,t)}) = N^{-1} \sum_{n=iN-d(t)}^{(i+1)N-d(t)-1} \mathrm{E}_t^{\mathbb{Q}} r_{t+n} = N^{-1} \sum_{n=iN-d(t)}^{(i+1)N-d(t)-1} (A_n + B_n' X_t) \quad (11) \\
&= a_i + h_i' X_t.
\end{aligned}
$$

Note that the futures rate is simply the average of the N relevant forward rates for the three-month period starting on the settlement day. The scalar a_i and the vector h_i are the averages of A_n and B_n, respectively, over the relevant period.[18]

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

$$
\tilde{ED}_t^{(i)} = \tilde{a}_i + \tilde{h}_i X_t \tag{12}
$$

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 (10) analogously holds for changes in futures rates:

$$
ED_{t+1}^{(i)} - ED_t^{(i)} = \tilde{h}_i' \Sigma \varepsilon_{t+1} + (h_i - \tilde{h}_i)' \Sigma \varepsilon_{t+1} + h_i' \Sigma \lambda_t. \tag{13}
$$

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 $\tilde{h}_i$. We now have the necessary theoretical foundations and can proceed by estimating the model.

3 Estimation of the unrestricted model

Turning to the estimation of the model, I first focus on a specification of the DTSM without any restrictions on the prices of risk. This serves as a benchmark and will reveal the bias and the large uncertainty underlying conventional term premium estimates.

The data set consists of daily observations on the rates for Eurodollar futures contracts maturing at the end of the current and the following 15 quarters, denoted by ED1 to ED16, thus covering the forward rate curve up to a maturity of about four years. The sample starts on 1 January 1990 and ends on 29 June 2007, before the start of the financial crisis. The

---

[18]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.

number of days in the sample is $T = 4401$.

3.1 Econometric methodology

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

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

introducing the notation $v_t = \Sigma \varepsilon_t$ and $Q = \Sigma \Sigma'$. The observation equation is

$$Y_t = a + H' X_t + w_t, \quad w_t \sim N(0, R), \; E(w_t w_s') = 0, \; t \neq s, \tag{14}$$

where Y_t is an $m \times 1$ vector of observations at time t. We have $m = 16$, the observations being the rates of the 16 Eurodollar futures contracts, $Y_t = (ED_t^{(1)}, \ldots, ED_t^{(16)})'$. The vector a stacks the intercepts a_1 to a_{16}, and given the normalization $\mu^{\mathbb{Q}} = 0$ we have $a = 0$. The $k \times m$ coefficient matrix $H = (h_1, \ldots, h_{16})$ is determined by $(\mu^{\mathbb{Q}}, \Phi^{\mathbb{Q}})$. The vector w_t contains measurement errors, included to avoid stochastic singularity as is common in the DTSM literature, which are serially uncorrelated and orthogonal to X_t. The variance-covariance matrix of w_t, denoted by R, is diagonal, and for the sake of parsimony I impose $R = \sigma_w^2 I_m$.

I parameterize the model in terms of the $\mathbb{Q}$-dynamics and the risk sensitivity parameters, which Bertholon et al. (2008) call the *back-modeling strategy*, for two reasons: First, the $\mathbb{Q}$-dynamics are chosen according to the DNS specification in order to a priori identify the factors as level, slope and curvature. Second, the focus of this paper is on inference and restrictions on the prices of risk, thus the parameters of the model need to explicitly include λ_0 and λ_1. The parameters of our model, under the aforementioned normalization restrictions, are $\theta = (\rho, \lambda_0, \lambda_1, \Sigma, \sigma_w^2)$. The parameter ρ determines $\Phi^{\mathbb{Q}}$, $\mu^{\mathbb{Q}} = 0$, and together with the prices of risk and Σ the parameters of the physical dynamics are determined.

The model could in principle be estimated using classical statistical methods such as maximum likelihood (ML) estimation, but there are several shortcomings of this approach for estimation of DTSMs. The likelihood function generally has many dimensions (there are 20 parameters in our case) and is highly non-linear, which makes numerical optimization expensive and finding the global maximum difficult (Duffee and Stanton, 2004; Duffee, 2009). More specifically, the likelihood function of a DTSM oftentimes has "multiple inequivalent local maxima which have similar likelihood values but substantially different implications for economic quantities of interest" (Kim and Orphanides, 2005, p.10). Attempts to estimate the

model with ML confirmed this: the likelihood function has several local maxima with different prices of risk, which confirms that the physical dynamics are hard to pin down.[19] Joslin et al. (2009) have shown that for their canonical model the estimates of the $\mathbb{P}$-dynamics are given by ordinary least squares, which simplifies estimation and can solve some of the above problems. However, their approach is based on the separation of risk-neutral and physical dynamics and thus is only applicable if the prices of risk are unrestricted.

Another important shortcoming of classical methods in the context of DTSM estimation is that one cannot correctly quantify the estimation uncertainty inherent in the calculations that use the model output. If for example we obtain ML estimates and calculate the risk-neutral volatility curve, which is a highly nonlinear function of the model parameters, quantifying the estimation uncertainty requires approximation based on the delta-method, which would likely be unreliable. More importantly, quantifying the uncertainty that results from inference about both parameters and latent factors is not possible with classical methods, since latent factors are inferred conditional on given point estimates of the model parameters (Kim and Nelson, 1999, chap. 8).[20]

A Bayesian approach employing Markov Chain Monte Carlo (MCMC) methods overcomes these challenges.[21] It has numerous advantages: Computationally it is less challenging since it amounts to successively drawing parameters (from their conditional posterior distributions) instead of numerically maximizing a high-dimensional and strongly non-linear likelihood function. We can diagnose whether the MCMC algorithm is likely to have converged, whereas for MLE it is rather difficult to assess whether a solution is a global maximum or not. Instead of leading to multiple local maxima, the fact that the risk sensitivity parameters are hard to estimate is appropriately reflected in a rather flat posterior. The two biggest advantages of the MCMC approach in the present context are the possibility to correctly account for estimation uncertainty, since the algorithm provides a sample from the joint posterior of the model parameters and the latent factors, and the fact that it will allow us to flexibly handle the issues of model choice and model uncertainty (Section 4).

The algorithm used to estimate the model is a block-wise Metropolis-Hastings (M-H) sampler, the details of which are described in Appendix B. It provides us with a sample from

---

[19]Christensen et al. (2007) claim that the a priori identification of the factors as level, slope and curvature solves this problem, however my own experience with the DNS model of this paper indicates that this is not the case – troublesome local maxima are still present.

[20]As an example consider the decomposition of rate changes according to equation (10): The uncertainty underlying the decomposition is due to uncertainty both in our estimates of latent factors (which determine ε_{t+1} and $\varepsilon_{t+1}^{\mathrm{Q}}$) and of the parameters (which determine B_n, $\tilde{B}_n$ and Σ).

[21]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). Other authors that have used MCMC methods in the estimation of DTSMs are Ang et al. (2007) and Boivin et al. (2009).

the joint posterior distribution of model parameters and latent factors.

3.2   Parameter estimates

Table 1 presents the parameter estimates of the DTSM without any restrictions on the prices of risk. For each parameter I report the posterior mean and a 95% credibility interval (CI), obtained by taking the sample mean and appropriate quantiles from the MCMC sample.

The risk-neutral dynamics are estimated very precisely, as was to be expected: The loadings in H are determined by ρ, and the information in the cross-section of futures rates pins down this parameter very precisely.

The risk sensitivity parameters on the other hands are estimated very imprecisely. The CIs are large relative to the magnitudes of the estimates and for most parameters the CIs contain zero. The reason is the high persistence of the short rate: Intuitively, because the short rate does not revert to its mean very often it is hard to estimate its unconditional mean and its speed of mean reversion. Hence μ and Φ are estimated very imprecisely (see for example Duffee & Stanton, 2004). Since for given $(\mu^{\mathbb{Q}}, \Phi^{\mathbb{Q}})$ the $\mathbb{P}$-dynamics are determined by λ_0 and λ_1, there is large estimation uncertainty about the risk sensitivity parameters.

Since the factors are a priori identified as level, slope and curvature, these estimates can help to understand the sources of time-variation in risk premia. The driving force seems to be changes in the prices of slope risk and curvature risk – only elements in the second and third row of λ_1 are significantly different from zero. The price of level risk on the other hand does not seem to change over time. This contrasts with the results of Cochrane and Piazzesi (2008) who find that in their data set and model *only the level factor* seems to carry risk. This issue will be discussed further below (Section 4.1).

With regards to the shock covariance matrix $Q = \Sigma\Sigma'$ standard deviations and correlations of the factor shocks are shown.[22] The shock covariance matrix is estimated rather precisely. There is strong correlation between the factor shocks. The measurement error variance is relatively small – it implies that the error is less than 12 basis points 95% of the time.

3.3   Term premium estimates

Our estimation results now allow us to decompose futures rates into risk-neutral rates and forward risk premia. First we consider average levels: Figure 1 shows for each of the 16

---

[22]An advantage of MCMC is that inference on non-linear functions of model parameters is just as precise as the inference on the original parameters. One simply calculates the parameters of interest (here standard deviations and correlations) from the fundamental parameters (Σ in this case) for each draw in the MCMC sample, which delivers a sample from the marginal posterior for these derived parameters.

contracts the average level of the empirical and model-implied futures rates together with posterior means and 95%-CIs for the average level of the risk-neutral rates.[23] The posterior mean of the average risk-neutral rate curve is flat at about 4.5%, which indicates that the short rate was, on average in our sample, expected to remain about constant. The horizontal distance between risk-neutral rate curve and futures rate curve corresponds to the forward risk premium, and the point estimate for the average forward risk premium in the ED16 rate is about 200 basis points. This seems reasonable for a four-year forward risk premium in light of the existing evidence, see for example Kim and Orphanides (2007). However, the graph reveals that estimates of average risk-neutral rates, and hence of average forward risk premia, are extremely imprecise: The CI for the risk-neutral rate at the long end extends from about 2.5% to almost 7%. The large amount of uncertainty implies that the true average four-year forward risk premium could be anywhere between about -20 and +400 basis points.

How does the risk-neutral rate behave over time? Figure 2 plots for the ED16 contract time series of the fitted rates (these are indistinguishable from the actual rates) and the risk-neutral rates, for which point-wise posterior means and 95%-CIs are shown. The risk-neutral rate is much less variable than the actual rate and hovers around 3.5% to 5.5%. The resulting forward risk premium thus accounts for most of the variability of the actual futures rate. The actual rates have declined over the period in the sample, and based on the point estimate for the risk-neutral rate series it seems that this trend was entirely accounted for by a decline in the forward risk premium. But the CIs show a large amount of uncertainty underlying this point estimate. Looking at recent values for example, we cannot say whether the forward risk premium was +150 or -200 in the first half of 2007. 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 approach presented in Section 4 will be able to reduce this uncertainty.

Turning to high-frequency changes in risk premia, the focus of this paper, it is instructive to first consider some specific days with news events. We will consider four days, two with large positive payroll surprises (03/08/1996, +408,500 and 04/02/2004, +208,000) and two with monetary policy surprises (04/18/1994 and 03/22/2005). The effects of these news events on the term structure are best visualized by showing changes of actual futures rates across maturities, as well as changes in risk-neutral rates, which represent the revisions to short rate expectations in response to the news events. Figure 3 shows the actual and model-implied rate changes together with posterior means and 95%-CIs for the changes in risk-neutral rates

---

[23]For given values for the parameters and latent factors, I calculate the risk-neutral rates using equation (12) and average them across time. This is repeated for each draw in the MCMC sample, so that for each contract we have a sample from the posterior distribution of the average risk-neutral rate.

across contracts. This is, to my knowledge, a new way to graphically analyze the effects of news events on the term structure of interest rates.

According to the point estimates changes in the short-maturity contracts are mostly due to changing short rate expectations, whereas for long maturities the *changes are attributed entirely to changing term premia*. Since revisions to short rate expectations are estimated to be close to zero at the long end of the term structure, changing forward risk premia alone account for the procyclical changes in long rates. Also evident from the figure is the *dramatic estimation uncertainty for changes in risk-neutral rates* – the CIs are very large. We cannot say with any confidence what happened to term premia in response to these news events when we appropriately account for the estimation uncertainty.

Figure 4 shows the empirical vol curve, i.e. the sample standard deviations of daily futures rate changes, and the model-implied vol curve. Furthermore it shows posterior means and 95%-CIs for the risk-neutral vol curve. The well-known hump shape (Dai and Singleton, 2003) is clearly present in the vol curve of Eurodollar futures. The vol curve of only the unpredictable component of rate changes (not shown) is virtually indistinguishable from the actual vol curve, indicating that the predictable component of daily rate changes in very small.

What is the relative importance of variation in short rate expectations and forward risk premia for the volatility of futures rates? According to the posterior mean of the risk-neutral vol curve, while changes in risk-neutral rates alone drive the volatility at the short end, they account for only less than half of the volatility at the long end. Notably the estimation uncertainty for the risk-neutral vol curve is extraordinarily large.

The above results show two important problems with the term premium estimates of conventional, unrestricted DTSMs. The first one is obvious: The estimation uncertainty underlying estimates of levels and changes in risk-neutral rates is tremendous. This is due to the lack of precision in our estimates of the physical dynamics. I term this the "uncertainty problem". My approach allows to quantify the uncertainty, which constitutes a challenge for most DTSMs but is hardly ever explicitly recognized.

The second issue is that the decompositions are implausible: The unrestricted model implies that forward risk premia account for the majority of rate changes at the long end of the term structure. The case studies showed that the entire procyclical response of long forward rates to the news events is attributed to forward risk premia. The risk neutral vol curve implies that for daily changes in long forward rates forward risk premia are a more important source of volatility than short rate expectations. However, we think that the term premium moves slowly, at business cycle frequencies, thus it should not account for a lot of daily movements of interest rates. Also we expect term premium movements to be countercyclical,

hence its strong procyclical response to the news events comes as a surprise. The reason for the implausible decompositions is what I term the "bias problem": Unrestricted DTSMs with stationary ℙ-dynamics imply a high estimated speed of mean reversion for the short rate. Hence shocks die out quickly, far-ahead expectations of the short rate hardly move at all, and most variation in long rates is attributed to risk premia. But since the short rate is very persistent, the speed of mean-reversion is likely to be significantly over-estimated. The closer the largest autoregressive root of a time series is to one, the more pronounced is the downward bias in its estimate – see Kendall (1954) or, more recently, Jardet et al. (2009) and the references therein. Hence the close-to-zero long-run revisions cannot be taken at face value, since they are due to biased estimates of the physical dynamics.

The uncertainty problem and the bias problem have been recognized by other researchers, particularly by Duffee and Stanton (2004), Kim and Orphanides (2005) and Kim (2007). Before presenting a new statistical framework to overcome these problems, I now turn to the systematic response of short rate expectations and term premia to macroeconomic news.

3.4   The impact of macroeconomic news

Studies of the response of the term structure of interest rates to macroeconomic announcements (Fleming and Remolona, 1999; Gürkaynak et al., 2005) have found strong procyclical responses, with a distinct hump shape and a significant sensitivity of far-ahead forward rates. The common approach is to regress changes in yields or forward rates on a measure of the surprise in the announcement, usually taken to be the difference between released and forecast values. Estimates of a DTSM can be used to assess how much of these responses are due to changing short rate expectations and changes in risk premia, respectively, by using model-implied changes in risk-neutral rates or risk premia as the dependent variables in these regressions. This approach is employed by Beechey (2007), who uses estimates of risk-neutral rates from Kim and Wright (2005) and finds that the forward risk premium in long forward rates responds strongly procyclical to macro news and seems to account for the majority of the total response of forward rates.

There are two important problems with this approach: First, it does not account for the uncertainty underlying estimates of risk-neutral rates and forward risk premia. In this section I perform inference that appropriately incorporates this estimation uncertainty and show that Beechey's point estimates cannot be taken at face value. Second, the results are driven by the fact that Kim & Wright's decomposition leads to implausibly high variability of term premia (see Beechey's figure 1), typical for DTSMs without or with only minimal restrictions on the prices of risk. Section 4 will present an approach that overcomes both of these problems and

leads to different conclusions.

The term structure innovations under the physical measure on a given day can be calculated as $v_t = X_t - \mu - \Phi X_{t-1}$, based on a set of parameter estimates and values for the latent factors. In order to assess the impact of macro news on the term structure we simply project the factor innovations on measures of the macro surprises using the following system:

$$v_t = \alpha + \beta^{(1)} s_t^{(1)} + \ldots + \beta^{(r)} s_t^{(r)} + \eta_t, \tag{15}$$

where $s_t^{(1)}$ to $s_t^{(r)}$ are scalars that contain the surprise component on day t for each of r different macroeconomic data releases, α and $\beta^{(1)}$ to $\beta^{(r)}$ are $k \times 1$ vectors of unknown parameters, and η_t is a vector of innovations.[24] Equation-by-equation least squares is efficient, despite the innovations η_t being correlated across equations, because the regressors are the same in each equation (Zellner, 1962). The resulting estimates $\hat{\beta}^{(j)}$ show the response for each of the k innovations to a one unit surprise in release j. If a specific piece of news tends to always have a similar impact on the term structure, then this will be reflected by the value of the corresponding $\hat{\beta}^{(j)}$ vector. The change in the risk-neutral rate for contract i caused by a unit surprise in a specific news release is $\tilde{h}_i' \hat{\beta}^{(j)}$, corresponding to the first term in equation (13).

The approach of Beechey amounts to calculating the innovations based on the DTSM's point estimates for the parameters and smoothed estimates for the factors, and then performing the regression in equation (15) taken as given these innovations.[25] However, this ignores the fact that the physical innovations v_t are not known but instead estimated in a first step. We can appropriately account for this uncertainty using the posterior sample for parameters and factors, as detailed in the following algorithm:

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

2. Calculate the loadings h_i and $\tilde{h}_i$ from the parameters, and the innovations from the factors and parameters using the fact that $v_t = X_t - \mu - \Phi X_{t-1}$.

3. For each of the k factor innovations, sample the regression coefficients, corresponding to the relevant equation of the system in (15), from the conjugate normal posterior.[26]

---

[24]It is necessary to include all data release series in the regression in order to partial out the impact of releases that occur on the same day.

[25]Beechey regresses changes in implied risk-neutral rates and forward risk premia on the surprise measure, which is equivalent to the approach outlined above since model-implied rate changes are driven by just k underlying term structure factors.

[26]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.

4. Calculate the predicted response of model-implied futures rates ($h_i'\hat{\beta}^{(j)}$) and of risk-neutral rates ($\tilde{h}_i'\hat{\beta}^{(j)}$) for each futures contract to each of the r different news releases.

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 to any of the r news releases. This distribution importantly takes into account the different sources of uncertainty: the first-step uncertainty underlying estimates of v_t as well as the second-step uncertainty from the regression analysis.

Figure 5 shows the results obtained for six different macroeconomic releases: Non-farm payroll employment, the unemployment rate, hourly earnings, Core CPI and Core PPI (Bureau of Labor Statistics, BLS) as well as retail sales (Department of Commerce). The surprise component in the data release is calculated as the difference between the actually released number and the value expected by the market, which is then standardized to have unit variance in order to make the different news releases comparable. To measure the market expectation I take the median market forecast, which is compiled by Money Market Services the Friday before the announcement. The figure shows the empirical responses of futures rates to macro news with 95% confidence intervals, the responses of model-implied rates, and the posterior means and 95%-CIs for the responses of risk-neutral rates. Note that the model satisfactorily captures the response of futures rates to the news – the model-implied responses are within the confidence intervals for the empirical responses in all cases.

The responses of risk-neutral rates to macro announcements, while significant at the short end of the term structure, are estimated to decrease to zero quickly with maturity for all six news releases. Thus at the long end of the term structure short rate expectations seem to not respond at all. This replicates the Beechey-result that the procyclical responses of long rates are mainly attributed to changes in forward risk premia. However, this neglects the two problems of uncertainty and bias in the estimates of the risk-neutral rates.

With regard to the uncertainty problem, the graphs reveal that because of the estimation uncertainty we cannot say with much confidence how strongly short rate expectations at long horizons respond to macro news: The CIs for the revisions caused by news are very large. The conclusion that "movements in term premia, not expected future short rates, account for most of the reaction" (Beechey, 2007, p. 2) is not warranted when we appropriately account for the uncertainty in the estimates of risk-neutral rates. The bias problem is reflected in the implausibly small responses of short rate expectations at the long end of the term structure.

4 Restrictions on prices of risk

An unrestricted DTSM has unsatisfactory implications for estimates of the term premium: The estimation uncertainty is dramatic, due to a lack of precision in estimates of the physical dynamics. Furthermore term premia have implausibly high volatility, since the short rate's speed of mean reversion is over-estimated. Most DTSMs, such as the ones in Dai and Singleton (2002), Ang and Piazzesi (2003) and Kim and Wright (2005), suffer from these issues.

The remedy against both of these problems is to incorporate additional information to pin down the term premium, taking the form of either additional data or constraints on the model.[27] This paper suggests to impose constraints on the market prices of risk. In their absence the physical dynamics and the cross-sectional dynamics are estimated independently of each other, hence the no-arbitrage assumption does not restrict the estimates at all. However if we restrict prices of risk then the information in the cross section of interest rates, which pins down the risk-neutral dynamics very precisely, is brought to bear on our estimates of the physical dynamics.[28] The big question of course is: Which restrictions are reasonable?

Common practice in the term structure literature is to first estimate a DTSM without restrictions, and then in a second step to re-estimate the model by imposing zero restrictions on those risk sensitivity parameters which are insignificantly different from zero or have the largest relative standard errors.[29] There are several problems with this approach. 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. A related problem (and probably the reason why it is uncommon to test joint restrictions on the parameters of risk) is that the MLE standard errors for the parameters of a DTSM are rather unreliable.[30] Moreover, the choice of a significance level required for inclusion of the parameter or of a cutoff for the relative standard error is necessarily arbitrary. Most importantly, alternative sets of restrictions lead to economically significant differences in results, as I will show below, and this approach offers no guidance on which set of restrictions is more credible.[31]

---

[27]Examples of studies that use additional data are Kim and Orphanides (2005), who include interest rate forecasts from surveys, and Campbell et al. (2009), who proxy for the price of risk using a dividend/price ratio.

[28]This fact has been noted for example by Kim and Orphanides (2005) and Cochrane and Piazzesi (2008).

[29]Among the numerous studies employing this approach are the influential papers of Duffee (2002), Dai and Singleton (2002), Ang and Piazzesi (2003) and Kim and Wright (2005).

[30]There are several reasons to doubt these standard errors: There are multiple local maxima, the asymptotic approximation might not be valid, and the numerical approximations to gradient and Hessian of the likelihood function are imprecise.

[31]Kim and Orphanides (2005) report that in the context of their DTSM some of the "different choices of parameters to be set to zero [...] exhibited economically significant quantitative differences" (p.11).

So far the literature has not developed an econometric approach to select restrictions on the prices of risk.[32] This paper provides a new framework for choosing plausible restrictions, which is based on Bayesian model selection.[33] One challenge is that there are many possible restrictions, which is overcome by first identifying plausible candidates using a new MCMC algorithm that involves latent indicator variables (Section 4.1). For the smaller set of candidate specifications we can then estimate the parameters and posterior model probabilities more precisely (Section 4.2). This shows the economic implications of each model (Section 4.3), and how much support each specification receives from the data. Since no single model clearly dominates all other candidates, I employ, in a third step, Bayesian model averaging (BMA) in order to perform inference that incorporates model uncertainty (Section 4.4).

The resulting estimates turn out to be both more precise and more plausible than those of the unrestricted model. In particular far-ahead short rate expectations are found to be significantly more variable, implying a slow-moving term premium. Furthermore the pro-cyclical response to news is found to be mainly due to changing short rate expectations, and there is no puzzling pro-cyclical term premium response. These results accord well with the conventional wisdom about bond risk premia.

The model-implied decompositions into expectations and risk premium components are independently verified by assessing forecast accuracy of the models (Section 4.5), as well as model-implied return-predictability (Section 4.6), with encouraging results.

4.1 Identifying candidate specifications

We would like to know which zero restrictions on λ_1 are supported most strongly by the data – since this paper is concerned with the time variation in risk premia, we will leave the vector λ_0 unrestricted. The problem of selecting a particular restriction is related to the variable selection problem in multivariate regression analysis: In both cases we 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 γ be a $k^2 \times 1$ vector of indicator variables, each of which is equal to one if the corresponding element of vec(λ_1) is allowed to be nonzero. The goal is to obtain the joint distribution

---

[32]Cochrane and Piazzesi (2008) have taken the question of risk-price restriction more seriously, however they choose their restrictions based on very specific evidence on expected excess returns for a particular frequency and data set, thus their approach is not generally applicable.

[33]For review articles on the topic of Bayesian model selection see for example Kass and Raftery (1995) and Clyde and George (2004).

of (γ, θ, X). Since the conditional posterior of γ given θ and X can be derived, block-wise M-H can be used to obtain draws from this distribution. To assess the plausibility of a joint restriction on λ_1, represented by a specific value of γ, 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, although it shares only the idea of latent indicator variables with the original GVS algorithm.

The matrix λ_1 has k^2 elements thus there are 2^{k^2} possible specifications, a large number even for the case of only a few factors. Thus a very large MCMC sample would be necessary to precisely estimate the posterior model probabilities of all models. However, those specifications with high posterior probability are likely to appear quickly in the GVS algorithm. Our goal is to identify the most promising specifications, and running the sampler for a limited number of iterations will achieve this goal.[34]

With the sample from the joint posterior for (γ, θ, X) at hand I select those models with a Bayes factor in comparison to the most plausible 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 leads to inclusion of six models, which together account for a total posterior model probability of 75.8%. To be clear: of the $2^9 = 512$ possible specifications in the model space of the GVS sampler, the six most frequent specifications plus the unrestricted model are the only ones that the subsequent analysis will consider – from now on the model space will consist only of these seven models.

Table 2 shows the candidate models. The unrestricted model is denoted by M_1 and the restricted models M_2 to M_7. The second column shows for the restricted models the frequency of each model in the GVS sampler – the unrestricted model is not visited by the sampler. Since these estimated model probabilities add up to 75.8% and not to 100%, I report the normalized numbers in column eight. Columns three to five indicate which elements in the respective columns of λ_1 are restricted (0) and unrestricted (1), e.g. for model M_2 only the element in the second row of the first column is unrestricted.

Importantly, the sample from the posterior distribution of γ allows us to perform inference on the determinants of time-variation in risk premia. Cochrane and Piazzesi (2008) find, based on an analysis of excess returns, that only the price of level risk seems to vary. Since monetary policy mainly has slope effects on the term structure they conclude that it cannot be the risk associated with policy shocks that varies over time. We can assess whether this finding is supported by our sample. Time-variation in the price of level risk corresponds to the presence of non-zero elements in the first row of λ_1. The posterior probability of this hypothesis is

---

[34]This point was made by George and McCulloch (1993).

estimated to be 17.3%, indicating that it is rather unlikely that level risk varies.[35] The price of slope risk varies if the second row of λ_1 is non-zero, and the posterior probability of this hypothesis is estimated to be 89.3%. Thus, in contrast to Cochrane and Piazzesi (2008), I find that the price of slope risk and not the price of level risk seems to vary over time. Further analysis is necessary to reconcile this difference, which could be due to the sample choice, the frequency of the model, or the method of inference. My results suggest that the compensation for slope risk, and thus possibly for the risk associated with policy shocks, seems to play an important role for variation in excess returns and term premia.

4.2   Within-model simulation and posterior model probabilities

Having identified the candidate models I proceed by estimating each model individually. The purpose of this second step of my estimation approach is to estimate the parameters of each model more precisely. Their posterior distributions are needed for estimating posterior model probabilities by marginal likelihood methods, for a precise assessment of the different specifications' economic implications, and for constructing efficient proposals for the joint model-parameter sampling by means of RJMCMC (Section 4.4). The MCMC algorithm for estimating the restricted models closely corresponds to the one for the unrestricted model, with the only difference that the each element of λ_0 and λ_1 is sampled separately.

After performing within-model estimation, posterior model probabilities can be estimated 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).[36] The resulting estimates are given in column nine and ten of Table 2. Comparing these probabilities to the ones obtained from the GVS algorithm in column eight we see that they provide a similar ranking and weighting of the models: The unrestricted model is extremely unlikely, model M_2 is strongly preferred, M_3 to M_5 receive about 10-20% probability each, and M_6 and M_7 are least likely. The correspondence between the results from GVS and from marginal likelihood approximations is reassuring – numerical differences are due to the above-mentioned imprecision of GVS and to the different approximations employed to estimate marginal likelihoods.

The unrestricted model's posterior probability is estimated to be zero. Instead, strongly

---

[35]This is the relative frequency of draws with at least one element of the first row of λ_1 being non-zero.

[36]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.

restricted specifications are supported by the data – none of the preferred models have more than three unrestricted elements in λ_1. The fact that *the data clearly supports tight restrictions on the prices of risk* is very plausible if we believe in the absence of arbitrage: Long rates should have some relation to expected future short rates, meaning that the physical and risk-neutral dynamics *should* be close to each other, but unrestricted prices of risk completely disconnect the two. Only when prices of risk are restricted does the no-arbitrage assumption have any bite. My results clearly speak in favor of such restrictions.

A crucial characteristic of the candidate models is whether they imply a stationary or an integrated short rate. The largest eigenvalue of Φ for each model is shown in column six. If for a particular model this is unity, then the physical dynamics are non-stationary, the short rate contains a unit root, and far-ahead expectations of the short rate are affected by current factor shocks. On the other hand for stationary $\mathbb{P}$-dynamics the short rate is mean-reverting and far-ahead short rate expectations are constant (in the limit). Remember that $\Phi^{\mathbb{Q}}$ has a unit eigenvalue, that $\Phi = \Phi^{\mathbb{Q}} + \Sigma\lambda_1$, and that we impose non-explosive dynamics (no eigenvalue of Φ can be larger than one). Hence whether the system is stationary or not depends entirely on the restrictions on λ_1: For unrestricted λ_1 and for most restrictions the system will be stationary, however some strongly restricted specifications lead to non-stationary dynamics. Importantly, *most preferred models exhibit a unit root for the short rate.* In particular model M_2, which is strongly favored by the data, implies non-stationary physical dynamics. The support in the data for a stochastic trend in the short rate stands in stark contrast to the implication of the unrestricted model that the short rate relatively quickly reverts to its unconditional mean. While the true short rate process 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. My evidence suggests that in a DTSM of daily frequency an integrated specification approximates the true data generating process better than a stationary specification.

In those restricted models that contain a unit root shocks to the level factor lead to revisions of far-ahead short rate expectations. Column seven shows these long-run revisions, $\lim_{h\to\infty}(\mathrm{E}_{t+1} - \mathrm{E}_t)r_{t+h}$, in response to a unit level shock – details about this calculation are given in Appendix D. 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 to a unit level shock. Evidently there is specification uncertainty about this important aspect of the model.

4.3 Economic implications of alternative models

The previous analysis has revealed that model M_2 receives strong support from the data. What exactly does this model imply for the properties of short rate expectations and risk premia? Figures 6 and 7 graphically summarize the implications of the unrestricted model (M_1) and the favorite model (M_2).[37] The first panel of each figure shows the model-implied changes in actual and risk-neutral rates on 8 March 1996, where markets saw a strong positive payroll surprise of +408,500. The second panel shows the actual and risk-neutral term structure of volatility. The third panel shows the responses of actual and risk-neutral futures rates to a one-standard-deviation positive payroll surprise. Credibility intervals for the changes, volatilities and responses of risk-neutral rates indicate the estimation uncertainty.

Model M_2 implies that *changes in short rate expectations play a dominant role for changes in actual futures rates* across the entire maturity spectrum. According to the point estimates for risk-neutral rate changes, the employment surprise in March 1996 caused futures rates to increase mainly because the expected future short rate path was revised upwards, and forward risk premia hardly changed on this day. The risk-neutral vol curve implied by M_2 shows that most of the daily variability in futures rates results from changing short rate expectations. Furthermore model M_2 attributes the pro-cyclical response of futures rates to payroll surprises mainly to revisions of the expected short rate path. The differences between models M_1 and M_2 are dramatic, with the implications of the latter model being more plausible in light of the conventional term premium wisdom.

The figure also reveals that the *decomposition obtained using the restricted model exhibits low estimation uncertainty*, indicated by the narrow credibility intervals. A policy-maker that believes in this model would not be guessing but instead could make confident statements about changes in policy expectations and risk premia. The strongly restricted prices of risk lead to high precision in the inference about short rate expectations.

If we were convinced that model M_2 is the right specification, then we could stop here. However the other candidate models also receive some support from the data, hence we should consider their implications with regard to short rate expectations and risk premia. Figure 8 compares the models by showing in the first panel what happened on 8 March 1996, in the second panel actual and risk-neutral vol curves, and in the third panel the responses to payroll surprises. The models have identical implications for the changes in actual futures rates, since the estimated risk-neutral dynamics are essentially the same for all models. However they differ significantly in their implications for risk-neutral rates.[38]

---

[37]Note that Figure 6 simply pulls together what was shown in Figures 3, 4 and 5.

[38]The graph shows the posterior means for changes (first panel), volatilities (second panel) and responses

According to models M_2, M_4 and M_5, the volatility of futures rates across all maturities (second panel) as well as their pro-cyclical response to macro news (first and third panel) are mainly due to changes in short rate expectations. Since these models account for about 70-80% of the posterior model probability mass this constitutes evidence that far-ahead short rate expectations are quite variable, respond significantly to the news, and hence play an important role for determining rate changes.

However, models M_1, M_3, M_6 and M_7 imply that changes in forward risk premia account for most or all of the volatility and pro-cyclical responses of long rates – for the stationary models M_1 and M_6 this is due to the fact that shocks die out in the limit and do not affect far-ahead short rate expectations, whereas for the non-stationary models M_3 and M_7 the explanation is that a positive level shock is associated with a negative long-run revision. Since the candidate models have economically different implications, there remains some specification uncertainty.

4.4   Accounting for model uncertainty using reversible-jump MCMC

When a model indicator is included as a parameter to be sampled using MCMC we speak of "joint model-parameter sampling." Another approach than the product-space sampling introduced by Carlin and Chib (1995), of which GVS is a special case, is Reversible-jump MCMC, initially developed by Green (1995). This method is characterized by the ability to "jump" between models with parameter spaces of different dimensionality. Sampling simultaneously across model- and parameter-space using RJMCMC constitutes the third step of my estimation approach, the goal being to deal with specification uncertainty.

A model indicator $j \in \{1, \ldots, J\}$ is included as an additional parameter, and the sampler, which is detailed in Appendix E, approximates the joint posterior $P(j, \theta_j, X|Y)$. Because we have previously estimated all candidate models separately, we can use our previous results to choose very efficient proposal distributions.

Table 2 shows in the last column the estimated posterior model probabilities obtained using RJMCMC. The finding of model M_2 being strongly favored by the data is confirmed, with its posterior probability estimated to be around 50%. The unrestricted model is never visited by the sampler; probabilities for the other models are estimated to be between 4% and 18%. These results support our previous conclusions.[39]

(third panel) for the risk-neutral rates, as in previous figures. Calculating these objects of interest at the posterior means of the parameters leads to slightly different results (not shown) because of the non-linearity in the parameters – e.g. the posterior mean volatility is different from the volatility at the posterior mean. This aspect of estimation uncertainty is ignored in classical estimates of term premium characteristics.

[39]Note that the GVS and RJMCMC algorithms use the same priors and thus, despite the different model spaces, should deliver the same weightings of the candidate models. The differences indicate lack of conver-

The sample obtained using the RJMCMC algorithm allows us to perform inference that accounts for specification uncertainty. This is crucial given the economically important differences between the models, and since we cannot be 100% sure that model M_2 is the correct one, rather only about 50-60%, simply put. Bayesian model averaging (BMA) is the appropriate theoretical framework to incorporate specification uncertainty into our inference about risk premia: The posterior distribution of some object of interest, say the risk-neutral vol curve, conditional on only the data and not on a specific model, is obtained by averaging out the model indicator using posterior probabilities.[40] The RJMCMC sample conveniently delivers draws from the relevant posterior distribution of the parameters – the model indicator is "averaged out" if we simply ignore its value for each draw. I denote by BMA the model estimates obtained in this way.

Figure 9 shows the properties of the risk-neutral rates inferred using BMA. It corresponds to figures 6 and 7 except that now we do not condition on a specific model but instead average across models. The point estimates for changes and responses of risk-neutral rates and for the risk-neutral vol curve obtained using BMA confirm the earlier conclusion: *Short rate expectations are the more important driving force for the daily volatility of the entire term structure and its pro-cyclical responses to macro news – forward risk premia move very little at the daily frequency.* This result stands in stark contrast to the implications of an unrestricted DTSM, where risk-neutral rates hardly move at all at the long end of the term structure and all daily variation is due to changing risk premia, as was shown in Section 3. The model in Kim and Wright (2005), being largely unrestricted, leads to the same implications, as shown by Beechey (2007), despite being augmented by survey forecasts. The decomposition of rate changes I obtain under sensible restrictions on the market prices of risk are more plausible: Since the term premium seems to move mainly at business cycle frequencies we would not expect it to account for much variability at the daily frequency. Furthermore the conventional wisdom, empirically and theoretically founded, has it that the term premium moves in a countercyclical way. Thus its contribution to the procyclical interest rate changes caused by news should be small. This is exactly what we find if we impose the restrictions on the prices of risk that are suggested by the data.

The previous conclusion was based on the point estimates obtained using BMA. Figure 9 also shows that averaging across models, instead of choosing a favored restricted model a priori, increases the uncertainty about short rate expectations and risk premia. This was of course to be expected. However, when comparing figures 6 and 9 it becomes evident that averaging

---

gence, the reason being, as mentioned above, that the large number of possible models in the GVS algorithm would require a huge number of iterations to achieve complete convergence.

[40]For an introduction to BMA see Hoeting et al. (1999).

across our restricted DTSM 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.

Another advantage of the approach developed in this section is that it solves the "discontinuity problem" documented by Cochrane and Piazzesi (2008) and Jardet et al. (2009): The implications of a DTSM for risk premia dramatically differ depending on whether the largest root of the short rate is equal to or less than unity. In reality this root is very close to but slightly less than one. A stationary model will underestimate this root as argued above, but a root of unity implied by an integrated specification is not the literal truth either. The discontinuity problem is thus tantamount to the bias problem. Jardet et al. (2009) solve it using a "near-cointegrated VAR" based on the averaging estimators proposed by Hansen (2009). Effectively they average a stationary and a non-stationary specification. While this overcomes the discontinuity problem it does not solve the uncertainty problem – no restrictions are imposed on the prices of risk, so no-arbitrage is not brought to bear on estimation of the $\mathbb{P}$-dynamics. The estimation uncertainty, which is not reported by the authors, is likely to be very large, as usual for unrestricted DTSMs. Using BMA on restricted specifications of a DTSM also amounts to averaging between stationary and non-stationary specifications, thus solving the discontinuity problem, but at the same time it solves the uncertainty problem. The result is that we obtain more precise and more plausible decompositions of interest rate levels and changes into expectations and risk premium components.

4.5 Forecast accuracy

Claiming that specific risk premium estimates are more plausible than others is equivalent to saying that the model more accurately captures the market's short rate expectations. One way to evaluate this claim, based on the notion that market participants construct the best possible forecasts, is to assess the forecast accuracy of the model's predictions for the short rate (e.g. Duffee, 2002; Dai et al., 2006).

As a simple reality check for the model's forecasts, I construct in-sample forecast errors for the model-implied short rate $r_t = X_t^{(1)} + X_t^{(2)}$. The choice of in-sample forecasts is made for simplicity and data-availability – because of the parsimony of the restricted DTSMs we would expect these models to perform even better out-of-sample. I compare root mean squared forecast errors (RMSEs) across model specifications, including the forecasts based on BMA. To construct forecasts of the term structure factors, filtered values for X_t are used. As a point of reference, I include the RMSEs based on forecasts using a random walk for the short

rate. Since we are mainly interested in long-horizon forecasts of several years, the horizons considered are 900, 1200 and 1500 days.

Table 3 shows the results. All models perform better than a random walk for the short rate. The restricted models produce superior forecasts of the short rate compared to the unrestricted model, particularly at longer horizons.[41] While it is not the case that our favorite model (M_2) produces the best forecasts, this was to be expected, since the estimation and model selection procedures consider only one-step-ahead forecast errors. Almost without exception all restricted models beat the unrestricted model at all forecast horizons. Notably the models with stationary $\mathbb{P}$-dynamics, M_1, M_3 and M_7, perform worse the longer the forecast horizon – the reason is that their forecasts are close to the unconditional mean of the short rate. The forecasts based on BMA perform very well, being a close second or third for each horizon. Since these are averages of the individual forecasts, their good performance reflects the generally good accuracy of forecast combinations (Timmermann, 2006).

A more detailed assessment of the forecasting performance of DTSMs with restricted prices of risk is warranted: The use of out-of-sample forecasts, rigorous inference about forecast accuracy, and inclusion of other forecasting methods will provide more detailed evidence as to whether this modeling approach can help improve interest rate forecasts. Because of the improved precision in estimates of the $\mathbb{P}$-dynamics and the parsimony of the restricted models this seems to be a promising direction for future research.

The results above support the claim that the term premium estimates of restricted specifications are more accurate than those based on unrestricted prices of risk. Notably this analysis is about the *level* of short rate expectations and risk premia. The following analysis will assess the plausibility in terms of *changes* in risk premia, which have been the main focus of this paper.

4.6   Return predictability

In the presence of time-varying risk premia, changes in futures rates, like bond returns, are partly predictable. Using predictive regressions, Cochrane and Piazzesi (2005)(CP) find that one-year bond returns are explained by current forward rates with R^2 of up to 44%, and similar results obtain for Eurodollar futures rates, as I will show below. On the other hand, the specification and parameter estimates of a DTSM have very concrete implications for the predictability of returns. This suggests that we can check the plausibility of model-based risk premium estimates by comparing the model's implications to the regression-based findings

---

[41]The puzzling fact that forecast errors are smaller at longer horizon is due to the fact that the sample periods differ in each case.

about return-predictability. In particular, the question is whether the restricted DTSMs imply similar predictability as we find in the data.

Which holding period should be considered? The predictable component of *daily* changes in futures rates is negligibly small, as mentioned above – daily changes are mainly driven by surprise changes in short rate expectations and risk premia. Inference about the predictable component of returns thus needs to be based on longer holding periods. For this reason we will consider one-year changes in Eurodollar futures rates, which correspond to absolute returns on positions in futures contract that are rolled over for four quarters and then liquidated. If I denote by N the length of this holding period[42] then the relevant return is $ED_{t+N}^{(i)} - ED_t^{(i+4)}$.

To assess predictability using return regressions, the one-year changes are projected onto current rates, using daily observations. Since the inclusion of all 16 futures contracts as explanatory variables leads to perfect multicollinearity, I restrict attention to only three contracts, namely ED4, ED9 and ED13 – these capture essentially all of the predictability.[43] Hence the regression specification is

$$ED_{t+N}^{(i)} - ED_t^{(i+4)} = \beta_0 + \beta_1 ED_t^{(4)} + \beta_2 ED_t^{(9)} + \beta_3 ED_t^{(13)} + u_t^{(i)}.$$

Table 4 shows in the first column the R^2 for $i = 4, 8, 12$. A large share of the variance in returns, namely 45-52%, is predictable based on current futures rates. These numbers are in the ballpark of the results of CP.

In order to compare these regression-based results to the models' implications, I estimate the R^2 based on simulated data for futures rates, assuming that the specific model is the true data-generating process. Specifically, for each set of parameters in the MCMC sample, I simulate time series for futures rates of length $T = 4000$ (similar to the actual data), and run the same regressions as for actual futures rates. This leads to a sample from the posterior distribution for the R^2 estimated using the typical regression approach.

The remaining columns table 4 show the simulation-based R^2 with 95%-CIs for models M_1, M_2 and BMA. The predictability in the simulated data is similar to what we found in the data. BMA implies simulation-based R^2 of 25-27%, with CIs comfortably straddling the values found in the data. Hence risk premium estimates based on a DTSM with tight restrictions on the prices of risk estimates are plausible from the perspective of return regressions as well.

---

[42]In the theoretical calculations this is taken to be equal to 260, the approximate number of business days in one year. In the data, where I take the holding period to be exactly one year, N varies between observations, a fact that the notation ignores for simplicity.

[43]These contracts are selected by choosing those three contracts with the highest predictive power for one-year changes, considering the average across contracts, i.e. $\frac{1}{12} \sum_{i=1}^{12} \left(ED_{t+N}^{(i)} - ED_t^{(i+4)} \right)$. Including more contracts as explanatory variables in addition to ED4, ED9 and ED13 hardly changes the R^2.

5   Conclusion

This paper shows that conventional term structure models, which do not restrict market prices of risk, lead to unsatisfactory implications for short rate expectations and forward risk premia: The estimation uncertainty is too large for us to make any useful statements about bond risk premia. Furthermore term premia have implausibly high variability. These issues, which are particularly serious when we want to decompose changes at the daily frequency, are due to a disconnect between risk-neutral and physical dynamics. With unrestricted prices of risk the no-arbitrage assumption, which requires consistency between cross-sectional and dynamic properties of the term structure, does not restrict our estimates. I develop an approach that allows us to rigorously assess which restrictions on the market prices of risk are plausible. Estimation of restricted models brings in the information in the cross section to improve the precision of our estimates of the physical dynamics of the short rate. The inference about short rate expectations is thus much more precise than in a model without restrictions. The two main empirical results are: (1) The data supports tight restrictions on the prices of risk. (2) Under these restrictions short rate expectations, and not changing risk premia, account for the majority of daily volatility at the long end of the term structure and of the response to macroeconomic news. This contrasts with existing results and is more plausible in light of the conventional wisdom about the term premium.

   A promising application of my framework is the context of macro-finance term structure models, which, as noted by Kim (2007), face the important challenge of putting more structure on the prices of risk: A key problem of these models is that the number of parameters is very large and the joint dynamics of term structure and macro variables are over-fitted. My approach can help overcome this problem since it greatly reduces the number of free parameters. In addition to leading to more parsimony and to more precise estimates of the factor dynamics, my framework allows researchers to perform rigorous inference on risk premia. 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), can be answered by testing restrictions on the prices of risk. The statistical framework I present in this paper allows to assess such restrictions, wherefore it is the right setting to tackle these questions.

References

Ang, Andrew and Monika Piazzesi, "A no-arbitrage vector autoregression of term struc-

ture dynamics with macroeconomic and latent variables," *Journal of Monetary Economics*, May 2003, *50* (4), 745–787.

_ , **Sen Dong, and Monika Piazzesi**, "No-Arbitrage Taylor Rules," NBER Working Papers 13448, National Bureau of Economic Research, Inc September 2007.

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.

Boivin, Jean, Sen Dong, and Andrew Ang, "Monetary Policy Shifts and the Term Structure," *AFA 2009 San Francisco Meetings Paper*, March 2009.

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., Adi Sunderam, and Luis M. Viceira, "Inflation Bets or Deflation Hedges? The Changing Risks of Nominal Bonds," NBER Working Papers 14701, National Bureau of Economic Research, Inc Feb 2009.

Carlin, Bradley P. and Siddartha 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, 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., Francis X. Diebold, and Glenn D. Rudebusch, "The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models," Working Paper Series 2007-20, Federal Reserve Bank of San Francisco 2007.

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 Kenneth Singleton**, "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 Mar 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.

_ , "Forecasting with the Term Structure: the Role of No-Arbitrage," Working Paper 2009.

_ **and Richard H. Stanton**, "Estimation of Dynamic Term Structure Models," working paper, Haas School of Business Mar 2004.

Duffie, Darrell and Rui Kan, "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, 1996, *6* (4), 379–406.

Fama, Eugene F. and Robert R. Bliss, "The Information in Long-Maturity Forward Rates," *The American Economic Review*, 1987, *77* (4), 680–692.

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.

Hansen, Bruce E., "Averaging Estimators for Autoregressions with a Near Unit Root," *Journal of Econometrics*, forthcoming 2009.

Hoeting, Jennifer A., David Madigan, Adrian E. Raftery, and Chris T. Volinsky, "Bayesian Model Averaging: A Tutorial," *Statistical Science*, Nov 1999, *14* (4), 382–401.

Jardet, Caroline, Alain Monfort, and Fulvio Pegoraro, "No-arbitrage near-cointegrated VAR(p) term structure models, term premia and GDP growth," working paper, Banque de France Jun 2009.

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," working paper Jul 2009.

_ , **Marcel Priebsch, and Kenneth J. Singleton**, "Risk Premium Accounting in Macro-Dynamic Term Structure Models," working paper Nov 2008.

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," Finance and Economics Discussion Series 2005-48, Board of Governors of the Federal Reserve System (U.S.) 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.

Nelson, Charles R. and Andrew F. Siegel, "Parsimonious Modeling of Yield Curves," *The Journal of Business*, 1987, *60* (4), 473–489.

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.

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.

Rose, Andrew Kenan, "Is the Real Interest Rate Stable?," *Journal of Finance*, December 1988, *43* (5), 1095–1112.

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.

_ **, Eric T. Swanson, and Tao Wu**, "The Bond Yield "Conundrum" from a Macro-Finance Perspective," *Monetary and Economic Studies*, December 2006, *24* (S1), 83–109.

Swanson, Eric T., "What we do and don't know about the term premium," *FRBSF Economic Letter*, Jul 2007, *20.*

Timmermann, Allan, "Forecast Combinations," in G. Elliott, C. Granger, and A. Timmermann, eds., *Handbook of Economic Forecasting*, Vol. 1, Elsevier, 2006, chapter 4.

Zellner, Arnold, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, June 1962, *57* (298), 348–368.

A   Change of measure

In order to show what kind of process the term structure factors follow under $\mathbb{Q}$ we need to derive the conditional Laplace transform of X_{t+1} under $\mathbb{Q}$. We defined the one-period SDF, or pricing Kernel, as

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

which implies the change of measure based on the fact that for any one-period pricing Kernel we have[44]

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

Thus we obtain for the risk-neutral conditional Laplace transform

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

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

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

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

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

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 prices-of-risk specification.

---

[44]The Radon-Nikodym derivative, which relates the densities under the physical and risk-neutral measure, is given by

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

B   Basic MCMC algorithm and convergence diagnostics

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 Σ determines not only the factor covariance matrix $Q = \Sigma\Sigma'$ but also affects the physical dynamics μ and Φ (see equation (6)). For the distribution of the observations Y conditional on the factors X we have the likelihood

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

B.2   Block-wise 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 block-wise Metropolis-Hastings (M-H) algorithm is used in order 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, a M-H step is used in order to obtain the desired draw, otherwise we 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 block-wise algorithm, the first B observations are discarded (the burn-in sample) so that the effect of the starting values becomes negligible. Of the following iterations, only every sth draw is retained, so that the number of iterations necessary for a sample of size G is $B + s \cdot G$. 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 we turn to model selection – posterior distributions tend to be much less sensitive to the choice of priors than Bayes factors (Kass and Raftery, 1995).[45] 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 of this paper on selecting restrictions on λ_1 (Section 4) this prior should not be too diffuse.

I specify ρ to be uniformly distributed over the unit interval. The prior for $Q = \Sigma\Sigma'$ is Inverse Wishart (IW) and the prior for σ_w^2 is Inverse Gamma (IG), both rather dispersed. The elements of λ_0 and λ_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 Φ 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, in which case we speak of Random Scan Metropolis-Hastings. I choose this method since one can fine tune how frequently each block is sampled: Those blocks are sampled more frequently which are more problematic in terms of mixing properties, and the blocks with parameters that mix well are sampled less frequently, which increases the efficiency of the algorithm. Specifically I sample only one block in each iteration, and the five blocks X, ρ, (λ_0, λ_1), Σ and σ_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 θ, draws for the latent factors are obtained by means of the FFSB algorithm developed by Carter and Kohn (1994): Kalman filtering delivers an initial time series of the factors, 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 & Nelson (1999, chap. 8).

Drawing the risk-neutral dynamics (ρ)

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

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

where θ_- denotes all other parameters except ρ. Since we cannot sample directly from this distribution – ρ enters the density in a complicated way – we need to employ a 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 we draw the parameter according to $\rho^{(g)} = \rho^{(g-1)} + \zeta_\rho t_4$, a fat-tailed RW with t_4 being a random variable with a t-distribution with four degrees of freedom, and ζ_ρ being a scale factor used to tune the acceptance probability to be around 20-50%, which is the recommended range in the MCMC literature (see Gamerman and Lopes, 2006, p.196). Since the proposal density is symmetric for a RW step, the acceptance probability is given by

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

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

Drawing the risk sensitivity parameters (λ_0 and λ_1)

In order to draw the risk sensitivity parameters, we recognize that for their conditional posterior distribution we have

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

where θ_- denotes all parameters except for λ_0 and λ_1, 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 non-linear fashion thus we cannot directly sample from the conditional posterior distribution. I tried both RW and Independence Metropolis proposals, and found the former to work better in this context. If there are no restrictions imposed on λ_1 then I draw λ_0 and each column of λ_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 λ_1 are restricted to zero, I draw each non-zero element of λ_0 and λ_1 separately, using a univariate RW with t_4-distributed innovations. The scale factors are adjusted in order to tune the acceptance probabilities. After obtaining the candidate draw, the restriction that the physical dynamics are non-explosive is checked, and the draw is rejected if the restriction is violated. Otherwise the acceptance probability for the draw is calculated as the minimum of one and the ratio of the likelihoods of the latent factors times the ratio of the priors for the new draw relative to the old draw.

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

For the conditional posterior of Σ we have

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

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

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

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

Drawing the measurement error variance (σ_w^2)

The variance of the measurement error can be drawn directly from its conditional posterior distribution, i.e. we have standard Gibbs-sampling for this step. The reason is that conditional on the latent factors, the other parameters and the data, the measurement errors can be viewed as regression residuals, and the IG distribution is the natural conjugate prior. Since I impose the variance to be the same across the m measurement equations, the residuals from all measurement equations are pooled. The conditional posterior for σ_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, in order to verify that the draws are from a distribution that is close to the invariant distribution of the Markov chain. Differently put, the question is whether the draws that we obtain 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.[46] 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

---

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

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 sub-samples. Gelman and Rubin (1992) have suggested to run parallel chains from different starting values and to 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 in order to find out how many iterations are needed for approximate convergence and how the algorithm can be tunde in order to improve mixing. The general conclusion is that a lot of iterations are needed because ρ and the elements of λ_0 and λ_1 traverse the parameter space only very slowly. This is a result of the small innovations in the RW proposals, which are necessary to obtain reasonable acceptance probabilities. Therefore I choose long burn-in samples ($B = 20,000$) and a large number of iterations ($G \cdot s = 200,000$). Under this configuration the graphs and diagnostic statistics indicate that the chain has converged.[47]

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 we keep 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 we sample from remains the same across models, which allows standard block-wise 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 λ_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 λ_1 always contains k^2 non-zero elements, but when calculating the likelihoods conditional on a specific set of restrictions, only those elements of λ_1 that are "switched on" according to γ are taken to be non-zero.

When we sample the elements of λ_1, conditional on γ, we need to distinguish whether a particular element is currently included in the model, and thus our draw informed by the data, or whether it is currently excluded. In the latter case the data is not informative and we sample 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 λ_1, which I denote by λ_i, is given by

$$P(\lambda_i|\lambda_{-i}, \gamma_i = 1, \gamma_{-i}, \theta_-, X, Y) \quad \propto \quad P(X|\theta, \gamma)P(\lambda_i|\gamma_i = 1) \tag{16}$$
$$P(\lambda_i|\lambda_{-i}, \gamma_i = 0, \gamma_{-i}, \theta_-, X, Y) \quad \propto \quad P(\lambda_i|\gamma_i = 0) \tag{17}$$

---

[47]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).

where $\theta = (\rho, \lambda_0, \lambda_1, \Sigma, \sigma_w^2)$ as before, θ_- denotes all parameters in θ other than λ_1, and λ_{-i} (γ_{-i}) contains all elements of λ_1 (γ) other than λ_i (γ_i). 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 λ_1 given γ.

For the case that λ_i is currently included, we sample from (16). 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 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 GVS is that the conditional posterior in (16) is not known analytically. Hence we need to employ Metropolis-Hastings to obtain the draws. I use a fat-tailed RW proposal, with scaling chosen to tune the acceptance probability.

If λ_i is not currently included, i.e. if $\gamma_i = 0$, it is drawn from the pseudo-prior $P(\lambda_i|\gamma_i = 0)$. I choose this distribution to be normal with mean and variance given by the sample moments of the marginal posterior draws of λ_i for the full, unrestricted model. Carlin and Chib (1995) recommend to choose a distribution for the pseudo-prior close to the actual posterior, which for the elements of λ_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 of course Bernoulli and the success probability is easily calculated based on:

$$\frac{P(\gamma_i = 1|\gamma_{-i}, \theta, X, Y)}{P(\gamma_i = 0|\gamma_{-i}, \theta, X, Y)} = \frac{P(X|\gamma_i = 1, \gamma_{-i}, \theta)}{P(X|\gamma_i = 0, \gamma_{-i}, \theta)} \frac{P(\lambda_i|\gamma_i = 1)}{P(\lambda_i|\gamma_i = 0)} \frac{P(\gamma_i = 1, \gamma_{-i})}{P(\gamma_i = 0, \gamma_{-i})} \tag{18}$$

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 we draw $\gamma_i = 1$ is given by $q/(q+1)$.[48]

The MCMC algorithm used to produce a sample from the joint posterior for (γ, θ, X) is again random-scan block-wise Metropolis-Hastings: In each iteration the block to be updated, either X, ρ, Σ, σ_w^2, λ_0, or (λ_1, γ), 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 like in the full model. Conditional on θ_-, X and the data, (λ_1, γ) is drawn as follows: First the elements of λ_1 are updated conditional on the value of γ from the previous iteration. Second, the elements of γ_i are drawn conditional on γ_{-i}, θ, X and the data. I implement two different versions of the algorithm: In the first version I update all elements of λ_1 and γ in each step. In the second version I randomly choose to update only one pair (λ_i, γ_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 5th 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

---

[48]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 γ and λ_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. set $\gamma_i = 0$, and vice versa.

between the two algorithms, we can be confident that we have identified 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. are based on a sample of size 400,000.

I assess whether the results on the plausibility of different restrictions on λ_1 are reasonable given the sample from the posterior for λ_1 for the unrestricted model. This turns out to be the case, based on individual credibility intervals, and in particular based on highest-posterior-density regions resulting from a normal approximation to this 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 λ_1. I performed additional sensitivity analysis, for example changing the prior variance of the elements of λ_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   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

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

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

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

For the long-run revision of short rate expectations we thus obtain

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

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

E   Reversible-jump Markov chain Monte Carlo

Including the model indicator as a parameter, the model is now parameterized as (j,θ_j,X_j). Since the latent variables carry over between models, we can simply write (j,θ_j,X). The algorithm I implement in order to obtain draws from the posterior distribution for (j,θ_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 is now how to propose values for the parameters of model j', denoted by $\theta_{j'}$. I decide for the models to share the parameters ρ, Σ, and σ_w^2, denoted here by θ_-, but not to share any elements of λ_0 and λ_1. It might seem that the models naturally share λ_0 and those elements of λ_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 (λ_0, λ_1) differs between models. To construct $\theta_{j'}$ I take θ_- from the current model together with the proposed values for $\lambda_{j'}$, by which I denote all non-zero elements of (λ_0, λ_1) in model j'. To propose values for $\lambda_{j'}$ I take the normal approximation to the posterior distribution of $\lambda_{j'}$, which we have 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

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

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 our context $(\theta_{j'}, u') = (\theta_-, \lambda_{j'}, u') = (\theta_-, u, \lambda_j) = g_{j,j'}(\theta_j, u)$ – the g-function is an identity function that simply matches the correct elements. Intuitively, u provides proposal values for all parameters in model j' that are not shared with model j, i.e. $u = \lambda_{j'}$, and u' takes on the values of the parameters in model j that are not used in model j', i.e. $u' = \lambda_j$. Thus, recognizing the uniform prior over models, the equal jump probabilities ($q(j \to j') = 1/6$ for all $j \neq j'$), and the fact that the likelihood of Y given X only depends on θ_- (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 we accept the proposed jump $(j, \theta_j) \to (j', \theta_{j'})$.

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

Table 1: Parameter estimates for unrestricted model

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

Posterior means and 95% Bayesian credibility intervals (based on 2.5 and 97.5 percentiles of draws from the posterior distribution) in squared brackets for each model parameter. Estimates of risk sensitivity parameters are boldfaced if the credibility interval does not straddle zero.

Table 2: 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		111	111	111	0.9989	0	0.0%	0.0%	0.0%	0.0%
M_2	46.3%	010	000	000	1.0000	0.9495	61.1%	58.2%	56.1%	48.8%
M_3	9.4%	011	000	000	1.0000	-0.3398	12.4%	10.7%	13.9%	15.9%
M_4	8.4%	010	000	010	1.0000	0.9637	11.1%	9.4%	12.1%	18.2%
M_5	6.5%	000	000	000	1.0000	1.0000	8.6%	18.4%	12.9%	4.1%
M_6	2.9%	110	000	000	0.9988	0	3.8%	2.4%	3.3%	6.3%
M_7	2.3%	011	000	010	1.0000	-0.3398	3.0%	0.9%	1.8%	6.6%

Alternative model specifications and estimated posterior model probabilities. Columns two: frequency of specification in GVS algorithm (cumulative: 75.8%). Columns three to five indicate which elements in the respective columns of λ_1 are restricted (0) and unrestricted (1). Column six: largest eigenvalue of Φ. Column seven: long-run revision of short-run expectations in response to a unit level shock. Columns eight to eleven: estimates of the posterior model probabilities based on rescaled GVS frequencies, Laplace approximation to marginal likelihood, Candidate's estimate of the marginal likelihood, and Reversible-Jump MCMC.

Table 3: Forecast accuracy

| Horizon	RW	M_1	M_2	M_3	M_4	M_5	M_6	M_7	BMA
900	3.00	2.25	2.19	2.09	2.19	2.28	2.08	2.09	2.11
1200	2.80	2.22	1.84	2.08	1.82	1.86	1.90	2.12	1.84
1500	2.66	2.14	1.61	2.05	1.60	1.60	1.80	2.13	1.65

Root mean squared errors for in-sample forecasts of the model-implied short rate, using alternative model specifications, compared to a random walk.

Table 4: Return predictability

| Contract	data	M_1		M_2		BMA	
ED4	.52	.46	[.11, .81]	.21	[.02, .52]	.27	[.02, .74]
ED8	.50	.44	[.12, .77]	.21	[.02, .51]	.25	[.02, .61]
ED12	.45	.43	[.14, .72]	.22	[.02, .52]	.25	[.03, .57]

R^2 in projection of annual returns on current rates/factors. For details please refer to text.

Figure 1: Average levels of actual, fitted and risk-neutral rates for unrestricted model
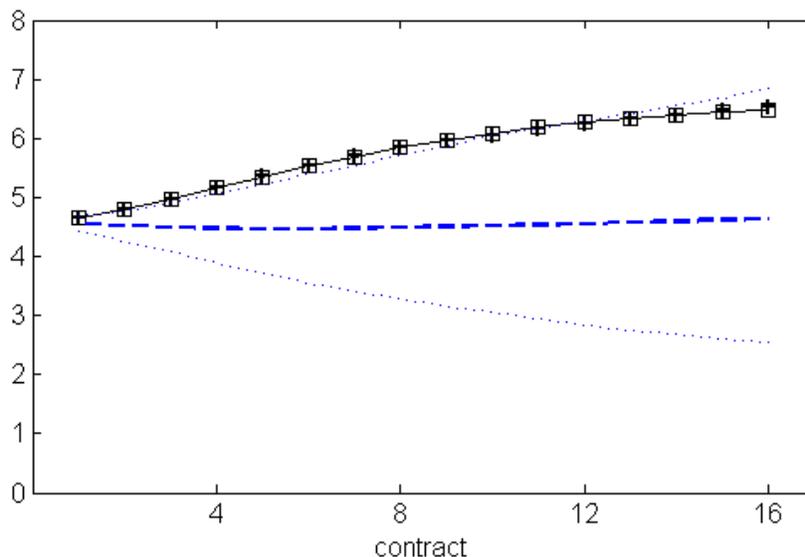
Average actual futures rates (crosses), average fitted futures rates (squares) and average risk-neutral rates (dashed line) with 95% credibility intervals (dotted lines). Units are percentage points.

Figure 2: Time series of fitted rates and risk-neutral rates for unrestricted model
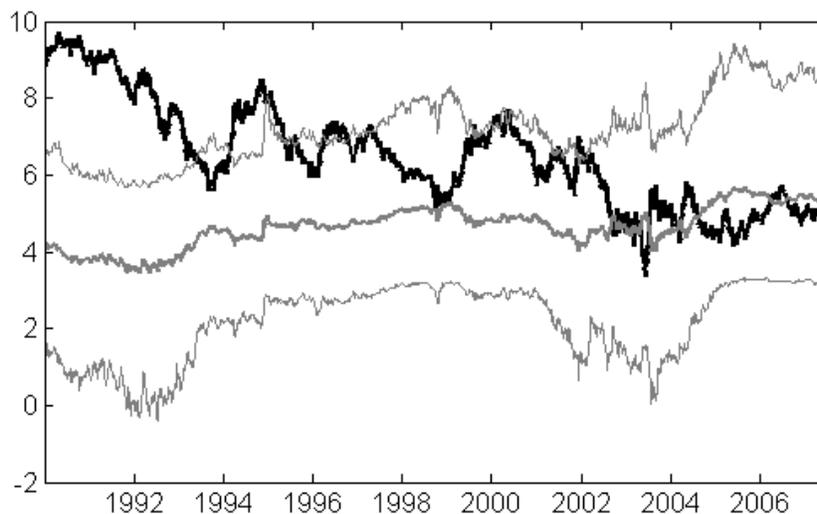
Time series of fitted rates (black line) for ED16 contract and posterior mean of corresponding risk-neutral rate (thick grey line) together with 95% point-wise credibility intervals (thin grey line). Units are percentage points.

Figure 3: Effects of specific news events (unrestricted model)
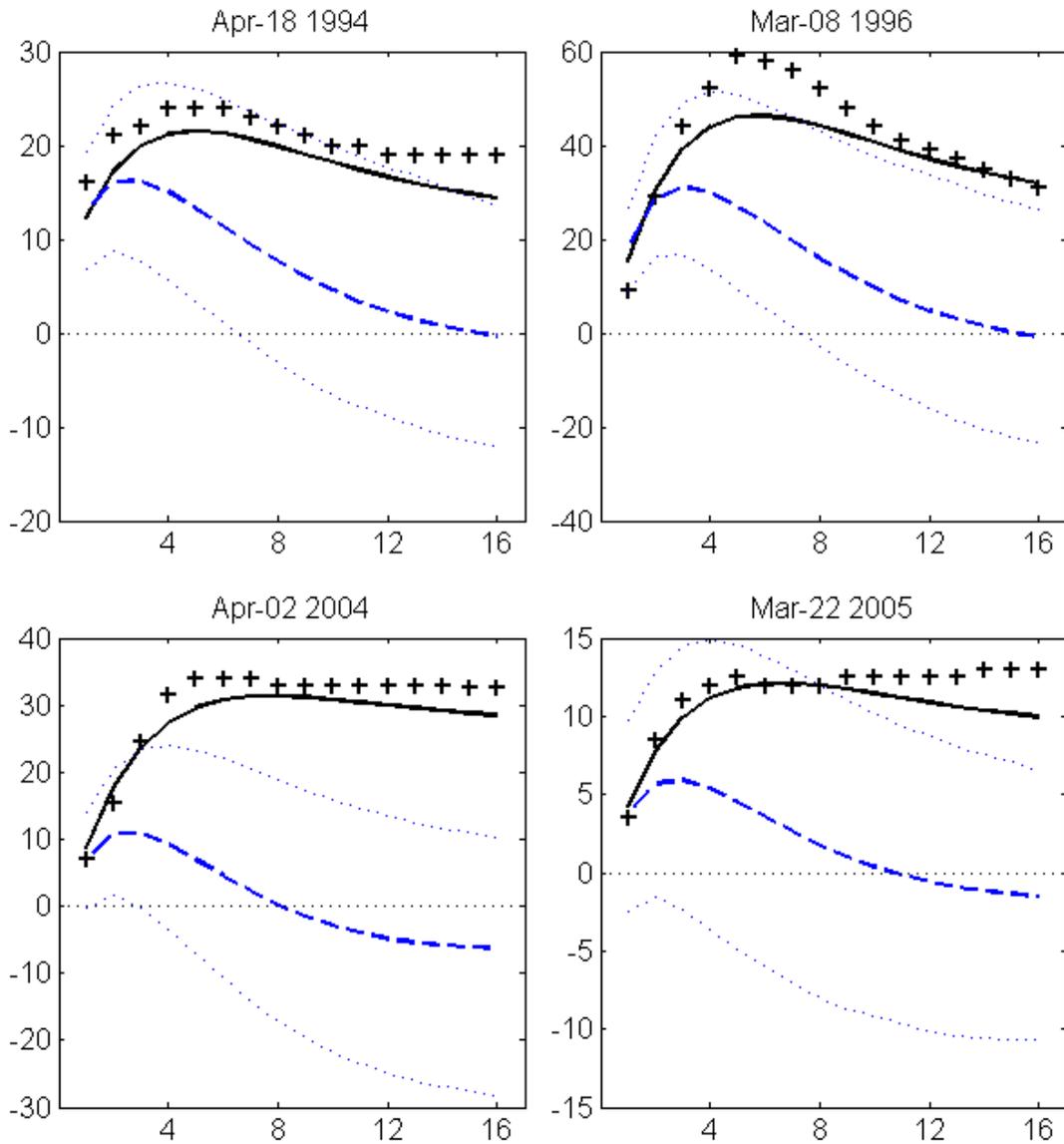
Empirical (crosses) and model-implied (solid line) rate changes in response to certain news events, together with estimates of the revisions to short rate expectations: posterior means (dashed lines) and 95% credibility intervals (dotted lines) for changes in the risk-neutral rates. Units are basis points. Description of events: Apr-18 1994 – policy action, surprise tightening 25bps; Mar-08 1996 – payroll surprise, +408,500; Apr-02 2004 – payroll surprise, +208,000; Mar-22 2005 – policy action, surprisingly hawkish FOMC statement.

Figure 4: Term structure of volatility implied by unrestricted model
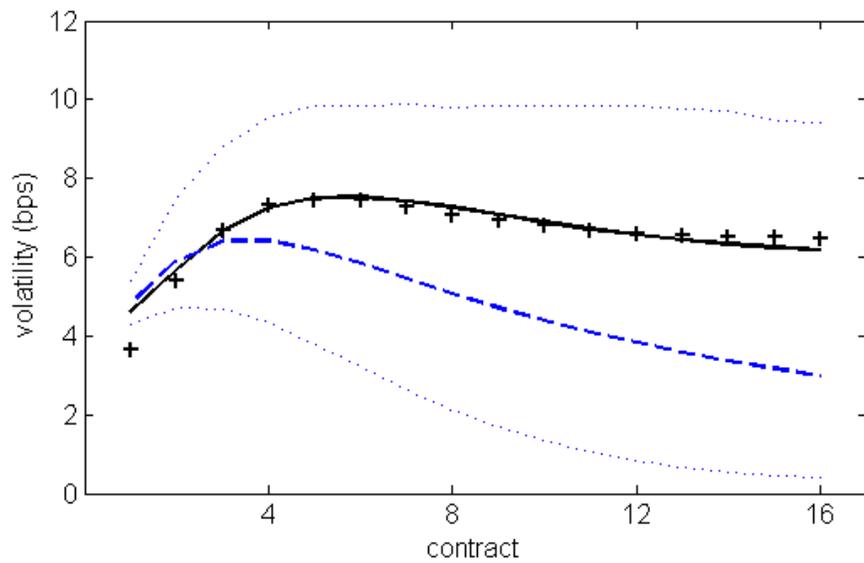
Estimates of actual and risk-neutral vol curve: Sample standard deviations (crosses) and model-implied standard deviations (solid line) of daily futures rate changes as well as posterior means (dashed line) and 95% credibility intervals (dotted lines) of standard deviations for risk-neutral rate changes. Units are basis points.

Figure 5: Responses to macro news implied by unrestricted model
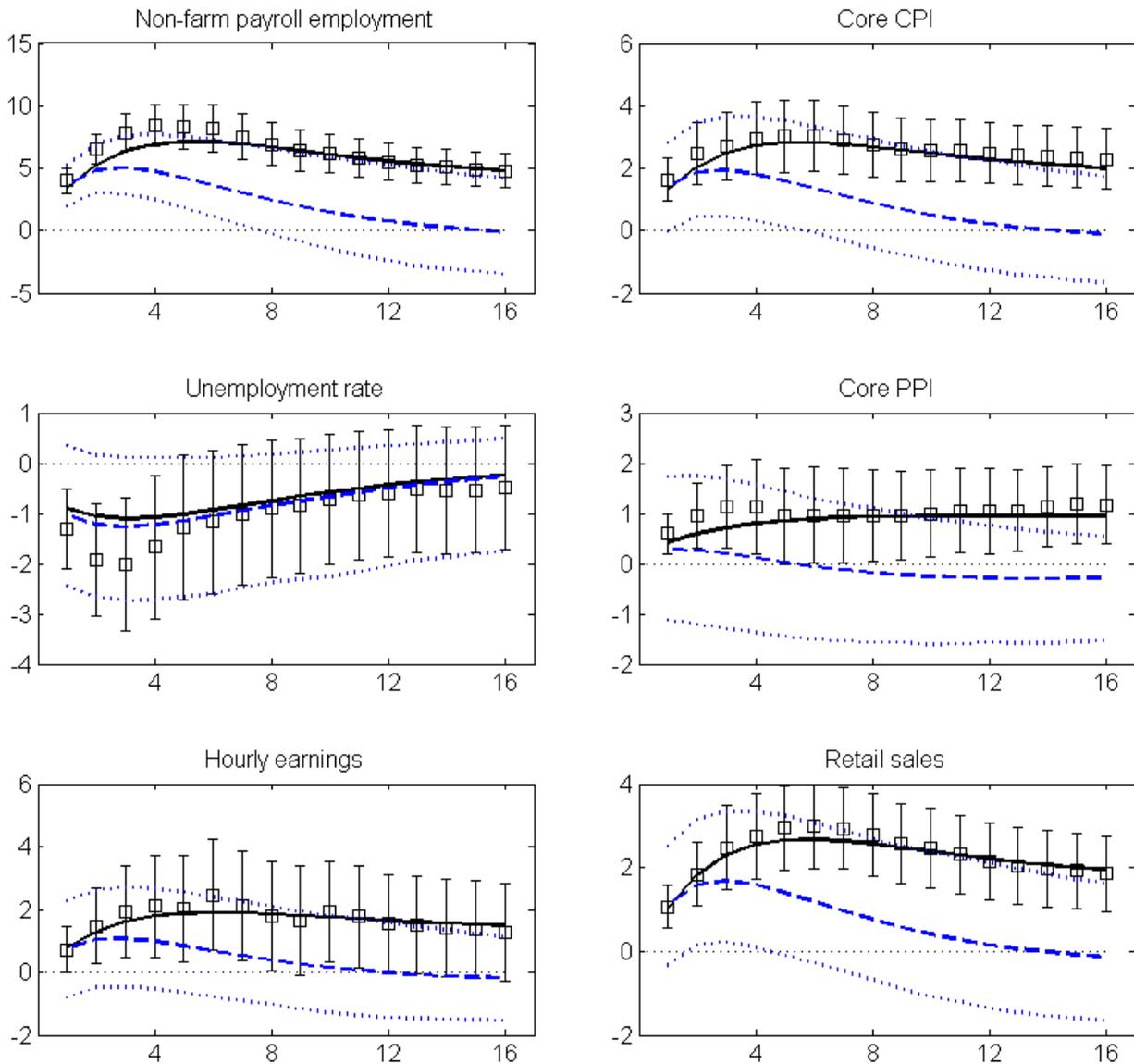
Responses to a one-standard-deviation surprise in six different macroeconomic data releases: Empirical responses of futures rates with 95% confidence intervals (error-bars), model-implied responses of futures rates (solid lines) as well as posterior means (dashed lines) and 95% credibility intervals (dotted lines) for estimated response of risk-neutral rates. Units are basis points.

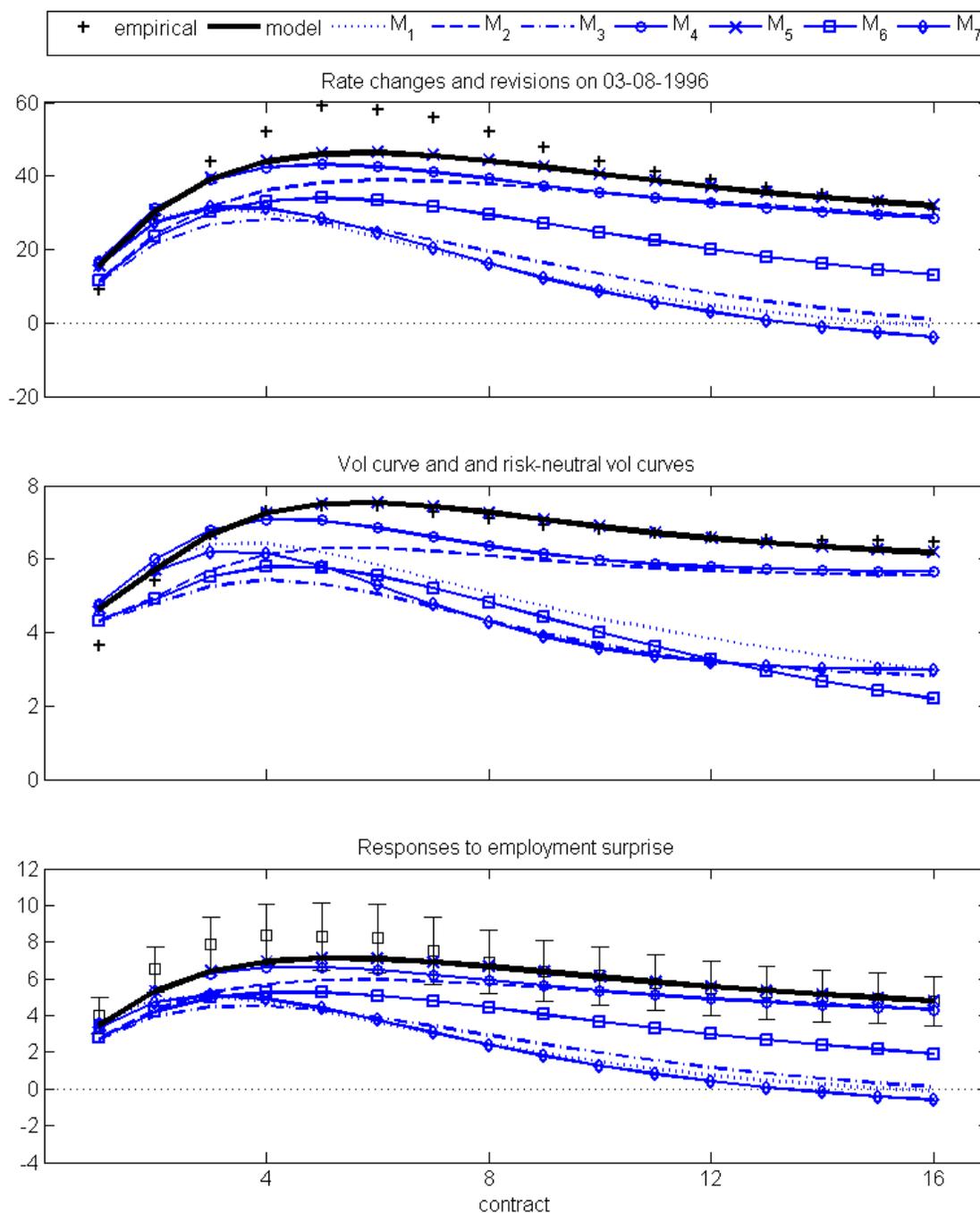Figure 6: Implications of unrestricted specification (M_1)
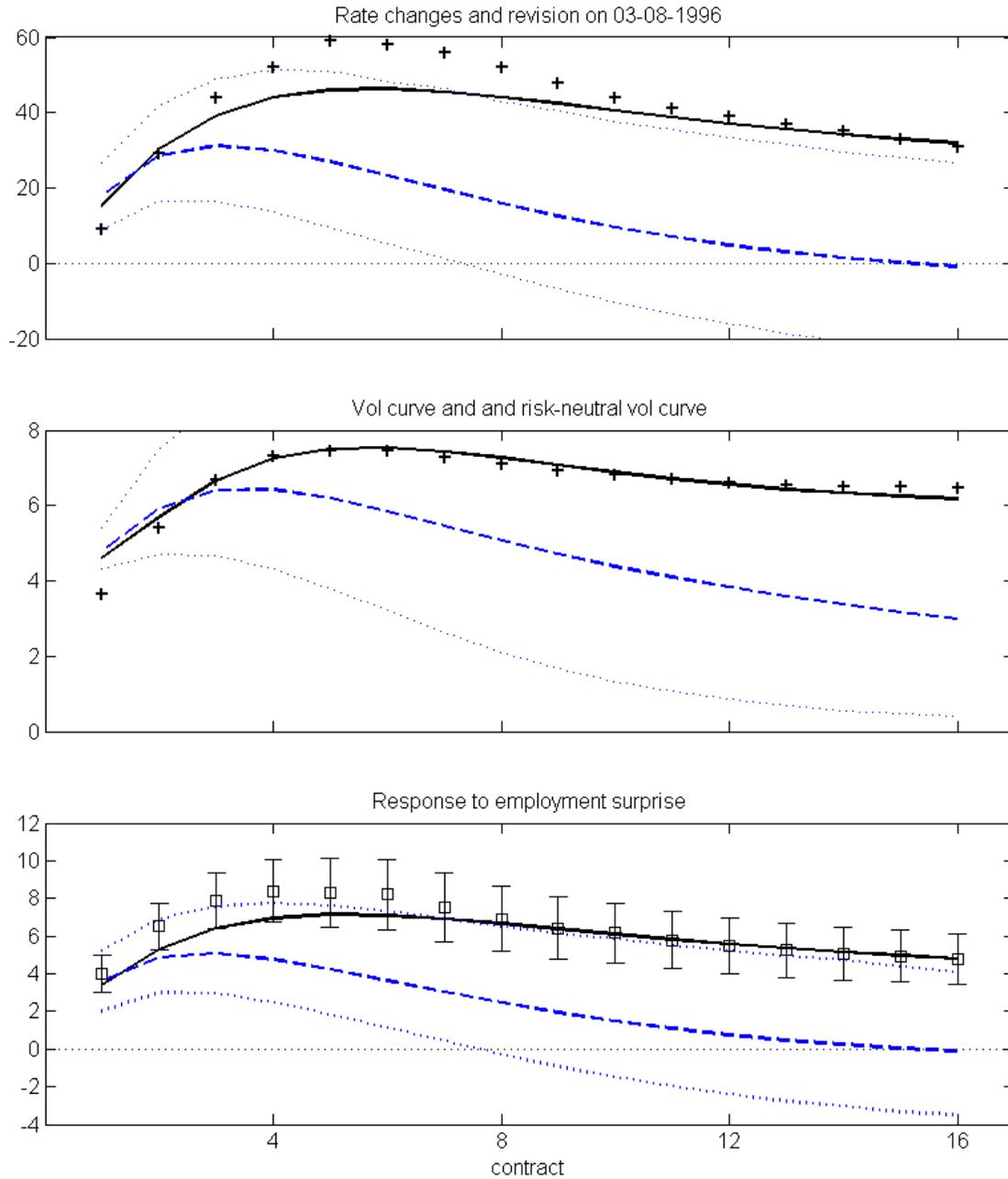
First panel: Empirical (crosses) and model-implied (solid line) rate changes on Mar-08 1996, together with estimated changes of risk-neutral rates. Second panel: Sample standard deviations (crosses) and model-implied standard deviations (solid line) for futures rate changes as well as model-implied standard 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), as well as 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.

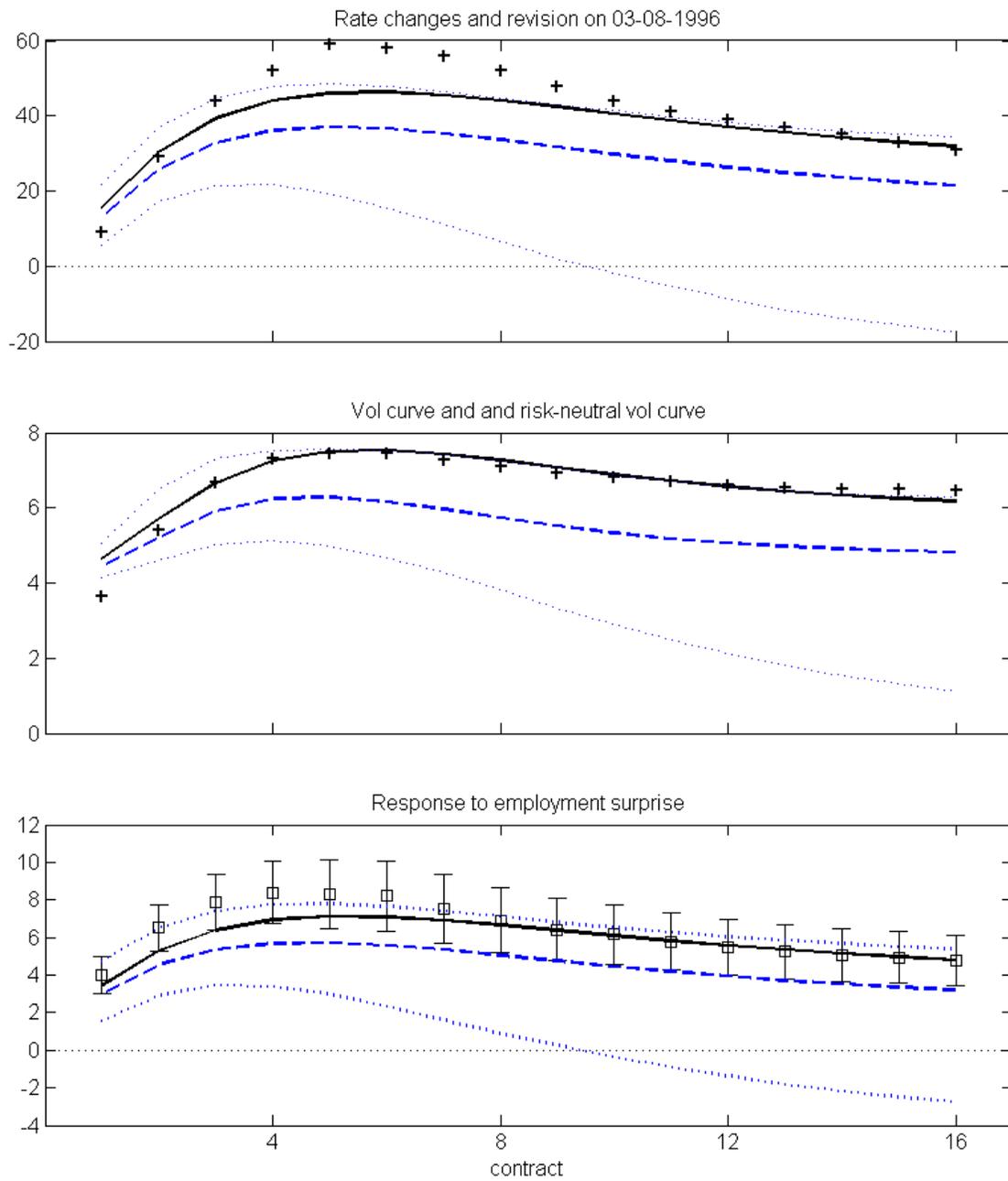Figure 7: Implications of favored specification (M_2)
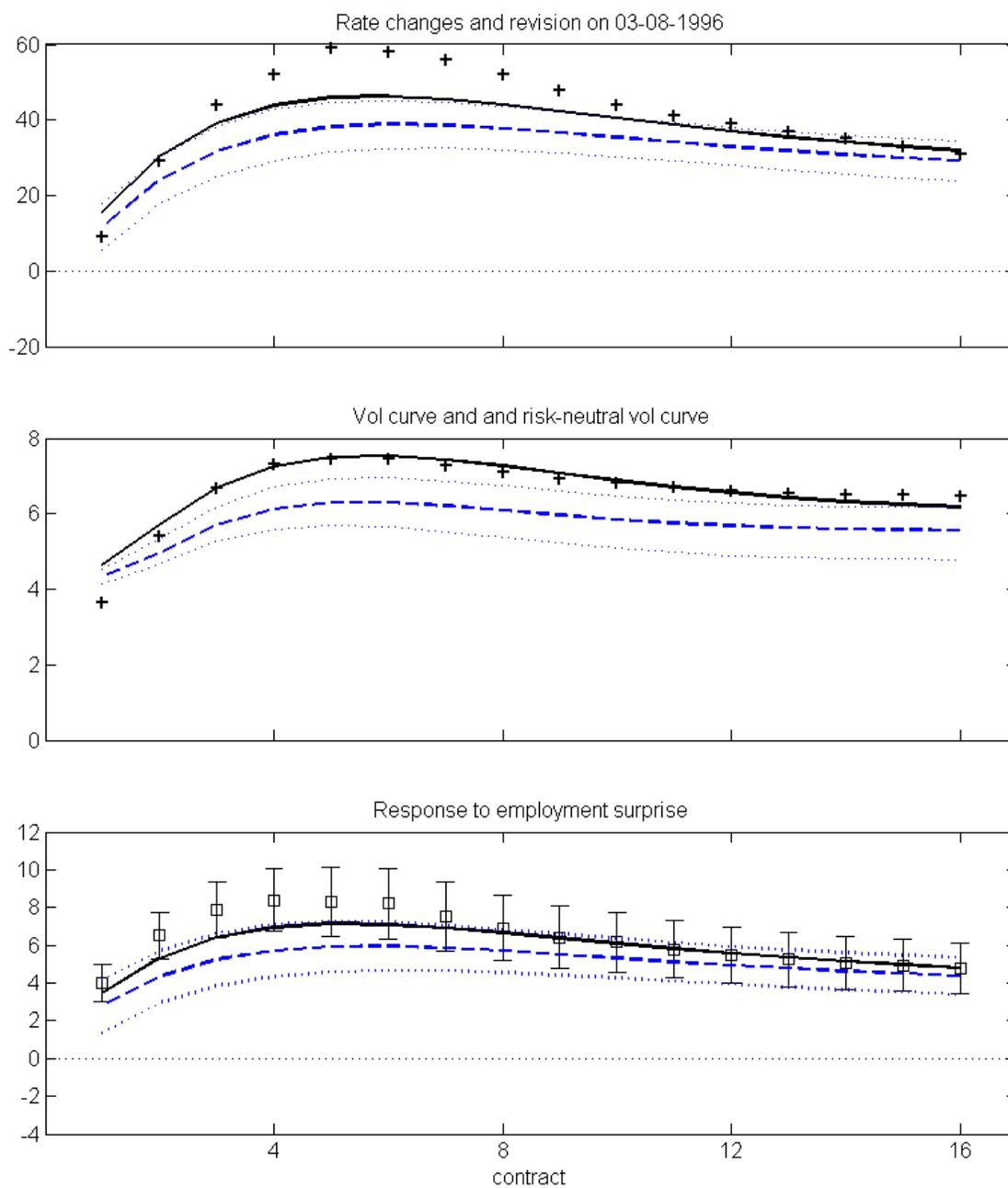
See description of Figure 6.

Figure 8: Comparison of alternative model specifications

First panel: Empirical (crosses) and model-implied (solid line) rate changes on Mar-08 1996 (payroll surprise +408,500), together 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 implied by alternative model specifications. Units are basis points.

Figure 9: Implications of Bayesian model averaging (BMA)

See description of Figure 6.