BANK FOR INTERNATIONAL SETTLEMENTS

BIS Working Papers
No 239

Spanned stochastic volatility in bond markets: a reexamination of the relative pricing between bonds and bond options

by Don H Kim

Monetary and Economic Department

December 2007

Abstract

This paper reexamines the issue of unspanned stochastic volatility (USV) in bond markets and the puzzle of poor relative pricing between bonds and bond options. I make a distinction between the "weak USV" and the "strong USV" scenarios, and analyze the evidence for each of them. I argue that the poor bonds/options relative pricing in the extant literature is not necessarily evidence for the strong USV scenario, and show that a maximally flexible 2-factor quadratic-Gaussian model (a non-USV model) estimated without bond options data can capture much of the movement in bond option prices. Dropping the positive-definiteness requirement for nominal interest rates and adopting "regularized" estimations turn out to be important for obtaining sensible results.

JEL Classification Numbers: G12, G13, E43

Keywords: term structure of interest rates, unspanned stochastic volatility, relative pricing, interest rate derivatives

Copies of publications are available from:

Bank for International Settlements
Press & Communications
CH-4002 Basel, Switzerland

E-mail: publications@bis.org

Fax: +41 61 280 9100 and +41 61 280 8100

This publication is available on the BIS website (www.bis.org).

ISSN 1020-0959 (print)

ISSN 1682-7678 (online)

SPANNED STOCHASTIC VOLATILITY IN BOND MARKETS:
A REEXAMINATION OF THE RELATIVE PRICING BETWEEN BONDS AND BOND OPTIONS

DON H. KIM*

ABSTRACT. This paper reexamines the issue of unspanned stochastic volatility (USV) in bond markets and the puzzle of poor relative pricing between bonds and bond options. I make a distinction between the "weak USV" and the "strong USV" scenarios, and analyze the evidence for each of them. I argue that the poor bonds/options relative pricing in the extant literature is not necessarily evidence for the strong USV scenario, and show that a maximally flexible 2-factor quadratic-Gaussian model (a non-USV model) estimated without bond options data can capture much of the movement in bond option prices. Dropping the positive-definiteness requirement for nominal interest rates and adopting "regularized" estimations turn out to be important for obtaining sensible results.

1. INTRODUCTION

An outstanding problem in modeling the term structure of interest rates is to characterize the variation in the volatility of interest rates and its possible relation to the factors that affect yield curve movements. Some early models, including the well-known Cox-Ingersoll-Ross (CIR) model, assumed that the volatility of interest rates is positively related to the level of interest rates. While the episode of 1979-83 (during which both the level and volatility of interest rates were high) seemed to provide some support for this, more recent studies, such as Duffee (2002), indicated that the relation between the interest rate uncertainty and the factors underlying the term structure movements is more complicated. Perhaps the most striking result in this regard is that of Collin-Dufresne and Goldstein (2002, henceforth CDG), who found that the changes in bond option prices[1] are poorly explained by the changes in bond yields and argued that this implies the presence of a stochastic volatility in interest rates that does not affect the

Date: November 28, 2007.

* Federal Reserve Board (phone: +1 202 452 5223, email: don.h.kim@frb.gov). This paper was written while I was on leave at the Bank for International Settlements. I thank the seminar participants at the BIS and the Federal Reserve Board, Scott Joslin, Frank Packer, Haibin Zhu, and especially Peter Hoerdahl, for helpful comments, and Benson Durham, Richhild Moessner and Bill Nelson for providing data used in some of the plots in this paper. The opinions in this paper do not necessarily reflect those of the BIS or the Federal Reserve Board.

[1]In this paper, I shall use the term "bond options" to refer to a broad range of interest rate derivatives that have option characteristics, including Treasury futures options, eurodollar futures options, interest rate swaptions, and interest rate caps.

cross-section of the yield curve. CDG dubbed the effect "unspanned stochastic volatility" (USV) and constructed models that have such a feature.

The USV effect however remains controversial. For example, Fan, Gupta, and Ritchken (2003) have argued against USV, based on their finding that Heath-Jarrow-Morton(HJM)-type models of the yield curve seem to hedge risks in bond options well. Bikbov and Chernov (2005) have also reported that their test of CDG (2002)'s specific restriction that guarantees USV is rejected in a 3-factor affine model. Recently Joslin (2007) has argued against USV by showing that the general 4-factor (non-USV) affine models fit bond option prices better than the USV counterparts. However, some other recent papers conclude in support of USV. In particular, Li and Zhao (2006, henceforth LZ) find that a fairly rich non-USV model (3-factor quadratic-Gaussian model) still cannot match the observed market cap prices well, while Collin-Dufresne, Goldstein, and Jones (2004, henceforth CDGJ) report that their 4-factor USV model fits data better than 3-factor non-USV models. Furthermore, Andersen and Benzoni (2006, henceforth AB) show that their measure of intraday volatility of yields is poorly explained by term structure factors, contrary to the implication of the non-USV models.

The evidence in support of USV notwithstanding, the USV models and the USV phenomenon itself present a puzzle. Consider, for specificity, CDG (2002)'s original affine model displaying USV. This model has restrictions imposed on the general affine model such that the volatility of interest rates does not enter the expression for bond yields. But we *should* expect bond yields to depend on volatility on theoretical and intuitive grounds. One concrete channel this would happen is the so-called Jenssen's inequality effect (also called "convexity bias"), arising from the fact that bond price is an expectation of a convex function of the short rate; see, e.g., Burghardt and Hoskins (1995). The convexity bias increases with the amount of interest rate uncertainty, and is quantitatively small for yield maturities up to ten years (typically about 10 basis points for the 10-year yield), though it is important at very long maturities. Even if one were to restrict attention to the below-10-year maturities, the convexity bias is a reminder that volatility can affect the term structure, as a matter of principle.[2] Another channel that is perhaps more important quantitatively is the "term premium effect": intuitively we should expect risk premium (term premium) on bonds to depend not only on the *price* of uncertainty (market price of risk) but also on the *amount* of uncertainty. Because this term premium effect would have certain maturity dependence, we would expect the yield curve itself to have a dependence on the amount of uncertainty.[3] Thus the independence between volatility and the yield curve suggested by the USV models seems strange.

The purpose of this paper is to reexamine the existing evidence on the USV effects and to present some new evidence that may help clarify the USV puzzle as well as the related puzzle of poor relative pricing between bonds and bond options. To this end I

[2]This point is emphasized by Joslin (2007).

[3]In order to have a situation like the $A_1(3)$ USV model of CDG (2002), the volatility dependence of the term premium component of the yield curve (including convexity bias) has to cancel mysteriously with that of the expectations component.

make a distinction between the "weak USV" and the "strong USV" scenarios. I refer to CDG (2002)'s original USV definition (namely that there is some component of bond option prices or instantaneous yield volatilities that is not spanned by the yield curve factors) as the weak USV condition, since this condition by itself might not have strong implications for modeling fundamental risks in the economy. For example, the weak USV scenario could arise from relatively high-frequency effects associated with certain institutional features of bond and option markets. Thus it is also useful to consider a stronger condition, "strong USV": there exists a lower-frequency (macroeconomic) variation in interest rate uncertainty that is unrelated to the yield curve, as in the $A_1(3)$ and $A_1(4)$ USV models of CDG (2002) and CDGJ (2004). The evidence for USV in the literature, such as CDG (2002)'s and AB (2006)'s regressions, supports the weak USV scenario but not necessarily the strong USV scenario.

One could argue that there *is* evidence for the strong USV scenario, namely the poor relative pricing between bonds and bond options in the existing literature: studies including Jagannathan, Kaplin, and Sun (2003, henceforth JKS) and LZ (2006) have found that term structure models estimated only with yields data have difficulty in capturing not just the high-frequency but also the lower-frequency variations in observed bond option prices. However, I shall argue that these studies may have had problems with specification and estimation (e.g., normalization issues, potential problems with conventional estimation techniques, daily sampling of term structure data with short time span). Addressing these issues (using a normalization that forgoes the positive-definiteness of interest rates and employing "regularized" estimations), I obtain quite good relative pricing results using a quadratic-Gaussian model (non-USV model) with just two factors. This is a welcome finding by itself, but it also casts doubt on the strong USV scenario.

The plan of the remainder of the paper is as follows. After setting up some notations and briefly reviewing term structure models, Section 2 defines the "weak USV" and the "strong USV" scenarios and discusses the evidence for them. Section 3 discusses potential problems with several commonly made assumptions in the specification and estimation of term structure models, taking a close look at some relevant aspects of the term structure data. With these caveats in mind, Section 4 reexamines the relative pricing between bonds and bond options in the context of the 2-factor quadratic-Gaussian model, and Section 5 concludes.

2. USV: MODELS AND EVIDENCE

2.1. **Review of Model Specifications.** Many of the "time-consistent" models discussed in connection with the USV debate belong to the affine class of models.[4] Because

[4]The "time-consistent" designation here is meant to emphasize the distinction from "time-inconsistent" models (e.g., the HJM-type models used by Fan et al (2003)) which are re-calibrated every time they are used, taking the yield curve as an input. The time-inconsistent models are not considered in this paper, because they do not have much to say about term premia (being restricted to

affine models have been discussed extensively elsewhere in the literature, I discuss here only some examples that will be useful later in the paper. The feature that gives the affine models its name is that yields depend linearly on a vector of state variables (risk factors) $x_t = [x_{1t}, ..., x_{nt}]'$. That is, $y_{t,\tau}$, the zero coupon yield at time t with time-to-maturity τ, takes the form

$$y_{t,\tau} = a_\tau + b'_\tau x_t, \tag{1}$$

where b_τ is an n-dimensional vector of factor loadings. Dai and Singleton (2000) classify affine models according to the number of "volatility factors", the $A_m(n)$ model denoting an n-factor model with m stochastic volatility factors.[5]

Two cases of affine models, the $A_0(n)$ model and the $A_n(n)$ model, are particularly familiar: the $A_0(n)$ model, often called the "affine-Gaussian model" or "multi-factor Vasicek model," has Gaussian risk factors, i.e., the state vector x_t follows the multivariate Ornstein-Uhlenbeck process,

$$dx_t = \mathcal{K}(\mu - x_t)dt + \Sigma dB_t, \tag{2}$$

where $\mathcal{K}$ and Σ are $n \times n$ constant matrices, μ is a constant n-vector, and B_t is an n-vector of standard Brownian motion. A version of the $A_n(n)$ model, called the multi-factor CIR model, has been also studied for long in the literature. In this model, all of the risk factors follow the (independent) square-root process, i.e.,

$$dx_{it} = \kappa_i(\mu_i - x_{it})dt + \sigma_i\sqrt{x_{it}}dB_{it} \tag{3}$$

for $i = 1, ..., n$.

The instantaneous volatility of yields $v_{t,\tau}$ $(= \sqrt{(dy_{t,\tau})^2/dt})$ is straightforward to calculate. We have (from eqs. (1) and (2))

$$v^2_{t,\tau} = b'_\tau \Sigma\Sigma' b_\tau \tag{4}$$

for the affine-Gaussian model, and

$$v^2_{t,\tau} = \sum_{i=1}^{n} b^2_{\tau,i}\sigma^2_i x_{it} \tag{5}$$

for the multi-factor CIR model. The volatilities in the affine-Gaussian model are time-invariant, while in the CIR model the instantaneous yield variance $v^2_{t,\tau}$ depends linearly on the state variables. It is also straightforward to show that conditional variance of yield, $var_t(y_{t+u,\tau})$, takes the same form (i.e., time-invariant for the affine-Gaussian model, and linear in x_t for the multi-factor CIR model).

---

risk-neutral modeling) and because it is not clear how to make a connection between these models and the macroeconomy.

[5]Duffee (2002) enriched the market price of risk specification of Dai and Singleton (2000), calling the resulting models "essentially affine" models, and Cheridito et al (2007) have made further enrichments. To simplify discussion, in this paper I shall not make a distinction between the affine and the essentially affine models, denoting both cases $A_m(n)$.

Neither the affine-Gaussian model nor the multi-factor CIR model has the USV feature: the affine-Gaussian model does not have any stochastic volatility (either spanned or unspanned) at all, while in the multi-factor CIR model the factors that affect the variation of $v_{t,\tau}$ also appear in the expression for yields. CDG (2002) show that in the case of certain $A_{i(>0)}(n)$ models, we could have a situation in which $v_{t,\tau}^2$ depends on a factor that does not affect the cross section of yields, for example,

$$v_{t,\tau}^2 = \alpha_\tau + \beta_\tau x_{1t}, \qquad y_{t,\tau} = a_\tau + b_{\tau,2}x_{2t} + b_{\tau,3}x_{3t}. \tag{6}$$

This particular example can be viewed as a special case of the general $A_1(3)$ model; CDG (2002) provide a set of restrictions on the parameters of the general model that lead to such a feature (USV).

For the later discussion, it is useful to describe also the so-called quadratic-Gaussian (QG) models. In the QG model, the short rate r_t depends quadratically on the state vector x_t that follows the multivariate O-U process, eq. (2), and the market price of risk λ_t depends linearly on x_t, i.e.,[6]

$$\begin{aligned} r_t &= \phi + \rho' x_t + x_t' \Psi x_t \\ \lambda_t &= \lambda_a + \Lambda_b x_t, \end{aligned} \tag{7}$$

where ϕ is a constant, ρ and λ_a are n-dimensional constant vectors, and Ψ and Λ_b are $n \times n$ constant matrices. (Ψ is a symmetric matrix.) With this specification, bond yields of general maturity also take the quadratic form

$$y_{t,\tau} = a_\tau + b_\tau' x_t + x_t' C_\tau x_t, \tag{8}$$

where the factor loadings a_τ, b_τ, C_τ can be expressed as a solution of a set of ordinary differential equations, called the Riccati equation; see, for example, Ahn, Dittmar and Gallant (2002, p255) and Leippold and Wu (2002, eq. (8)).

It is straightforward to show (using the Ito's lemma and eqs. (2) and (8)) that instantaneous yield volatility $v_{t,\tau}$ in the QG model is given by

$$v_{t,\tau}^2 = (b_\tau' + 2x_t' C_\tau)\Sigma\Sigma'(b_\tau + 2C_\tau x_t). \tag{9}$$

Note that this is an even richer form for describing the volatility variation than that of the multi-factor CIR model, not only because the functional form (9) is richer than (5) but also because the risk factors $(x_{1t}, x_{2t}, ...)$ in the QG model can have a general correlation. But again, it is in general inconsistent with USV, as a factor that moves $v_{t,\tau}$ would also appear in the expression for bond yields.

[6]These define the (continuous-time) pricing kernel M_t as $dM_t = -M_t r_t dt - M_t \lambda_t' dB_t$, where B_t is the shocks (Brownian motions) that drive state variables (as in eq. (2)). The price of a zero-coupon bond with time-to-maturity τ is given by $P_{t,\tau} = E_t(M_{t+\tau})/M_t$.

2.2. **Weak USV and Strong USV.** To interpret the evidence on USV, it is useful to make a distinction between the "weak USV" and the "strong USV" scenarios.

By weak USV, I refer to the case in which there is *some* component of bond option price variation or instantaneous yield volatility variation that is not spanned by the yield curve factors; this corresponds to the original USV condition of CDG (2002). I label this condition "weak", in the sense that it might not imply a strong or useful constraint on term structure models insofar as the modeling of fundamental risks are concerned. For example, certain institutional aspects of bond markets or option markets (market organization, trading rules, etc.) may give rise to a relatively short-lived ("high-frequency") USV effect, which, although possibly important to some traders, might not be central to the discussion of basic risk and return in bond markets.

Therefore, I also consider a stronger case of USV ("strong USV"), in which there is a "fundamental" variation in bond option prices or in yield volatility that is not spanned by the yield curve factors. Here I have in mind the kind of variation at a time scale of a few quarters or longer (lower-frequency variation) that would be important for discussions about the macroeconomy and asset pricing.[7] Note that the specific $A_1(n)$ USV models of CDG (2002) and CDGJ (2004) have not only the weak USV feature but also the strong USV feature. In fact, they imply a very strong form of USV, since they do not have a "spanned" component of volatility at all.

To illustrate some contexts in which the strong USV debate matters, let us now discuss the issue of interpreting the variation of the bond market term premia. As is well known, long-term yields in the US were substantially higher in the 1980s than in 2000s. This may be partly due to the long-term yields in the earlier period containing higher term premia, which, in turn, may have been due in part to the larger *amount* of risk (higher uncertainty about macroeconomy and monetary policy), which came down over time since then (a phenomenon often referred to as the "Great Moderation"). Such a trend-like variation in the interest rate uncertainty may be difficult to detect from one-month or one-week changes in option-implied volatilities.

Besides the "trend" variation, "intermediate-frequency" variation in interest rate uncertainty (related to business cycles and other macro effects) may have also contributed to term premium variations. Figure 1 shows the width of a 90% confidence interval for the distribution of the 1-year-ahead short rate based on the eurodollar futures options.[8] While it shows many short-term fluctuations, broader variations (over yearly or longer time scales) are also visible. In recent years, the "unusually low" level of long-term

[7] "Strong USV" is not a precise concept but a heuristic one, as concepts like "fundamental" and "lower-frequency" are not easy to define. For concreteness, however, think of the bond option price or the instantaneous volatility as a function $f(x^s_{1t}, x^s_{2t}, ..., x^u_{1t}, x^u_{2t}, ...)$, where x^s_{it}'s are "spanned" factors (yield curve factors) and x^u_{it}'s are "unspanned" factors. If any of x^u_{it}'s have a characteristic time scale (or half-life) that exceeds some number (say, a few quarters) and have a non-negligible weight, we could call that "strong USV."

[8] Note that this uncertainty information is about the risk-neutral measure (since it's from options), but we can expect qualitatively similar behaviors in the physical measure.

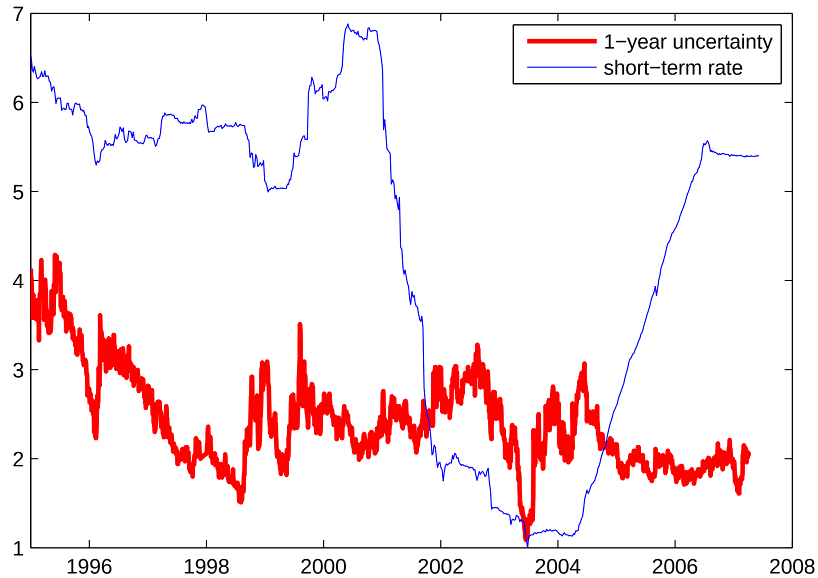

Figure 1: Width of a 90% confidence interval for the distribution of the short rate 1-year ahead, based on the eurodollar futures option prices (Source: internal data, Division of Monetary Affairs, Federal Reserve Board). The 3-month LIBOR rate (a short-term interest rate) is also shown for comparison.

yields since Federal Reserve's policy tightening of June 2004 received much attention. One proposed partial explanation is that a reduction in the uncertainty about interest rates lowered the term premium in yields. (See, for example, Backus and Wright (2007).) Indeed, the interest rate uncertainty has been generally low in the past few years, as can be seen in Figure 1.

Kim and Orphanides (2007) present some evidence for a positive relation between term premium and the uncertainty about monetary policy at intermediate- and low- frequencies. However, their measure of monetary policy uncertainty is based on the *dispersion* of survey forecasts, which, some might argue, is not a water-tight proxy for uncertainty. In addition, their term premium modeling is based on the affine-Gaussian model, which has a limitation that interest rate uncertainty does not vary over time. To examine the potential relation between term premium and interest rate uncertainty in an internally consistent manner, we need a no-arbitrage model that can jointly describe the relevant volatility variations and term premium variations. In particular, the model should be able to capture the kind of lower- and intermediate-frequency volatility variations we have just discussed, even if it misses the details of higher-frequency variations.

In writing down such a "candidate" model, we are faced with the following questions: Are the models that do not have the USV feature (e.g., the QG models) suitable for that purpose? Should we require that the model have the strong USV feature (as in the affine USV models of CDG (2002) and CDGJ (2004))? A potential concern with models that have the strong USV feature is that they might rule out a meaningful relation between term premia and volatility by design, as these models "decouple" the volatility dynamics and term structure dynamics to some extent. Therefore, answers to some very basic questions about risk and return in bond markets hinge on whether the strong USV condition holds or not.

2.3. **Evidence for Weak USV.** CDG (2002)'s principal evidence for USV comes from the regression of the change in the price of straddle portfolios onto the change in swap rates. Their "straddle" is a combination of (at-the-money) cap and floor, and has the feature that its price is not very sensitive to changes in the level of interest rates (i.e., delta-neutral) but sensitive to changes in volatility. Denoting the monthly straddle returns $\Delta s_{t,\mathcal{T}}$, their regression can be written

$$\Delta s_{t,\mathcal{T}} = c_0 + c_1 \Delta y_{t,\tau_1} + c_2 \Delta y_{t,\tau_2} + ... + c_l \Delta y_{t,\tau_l} + \epsilon_{t,\mathcal{T}}. \tag{10}$$

In the regression of staddle returns on the change in yields of various maturities (τ_i=0.5, 1,2,3,4,5,7,10 years), they found that the regression R^2's tend to be not very high, e.g., ranging between 20% and 50% (for different straddle maturity $\mathcal{T}$) in the case of the US. This kind of R^2 is substantially lower than when the change in a bond yield is regressed onto the change in other yields, in which case R^2's exceeding 99% are common. CDG (2002) further note that the regression residuals for various $\mathcal{T}$'s ($\epsilon_{t,\mathcal{T}}$) have a common dominant component. They thus conclude that the options markets (cap and floor markets) have a systematic risk factor that is not spanned by the bond markets.

Andersen and Benzoni (2006) provide further evidence that is similar in spirit. AB's basic idea is as follows. Note that the instantaneous yield variance $v^2_{t,\tau}$ in the affine model depends linearly on the state variables, e.g., eq. (3) for the multi-factor CIR model. Because yields also depend linearly on state variables, one can "invert" the state vector from a set of yields, and express $v^2_{t,\tau}$ as

$$v^2_{t,\tau} = c_{\tau,0} + c_{\tau,1} y_{t,\tau_1} + ... + c_{\tau,n} y_{t,\tau_n}. \tag{11}$$

This is a linear regression without the residual term, in other words, a regression with very high R^2. If the yield dynamics is a diffusion and if one could sample yields at infinitesimally small time intervals, one could actually measure $v^2_{t,\tau}$. This is not feasible in practice, but one could integrate eq. (11) and obtain a testable relation,

$$\overline{v^2}_{t,\tau,h} = c_{\tau,0} + c_{\tau,1} \overline{y}_{t,\tau_1,h} + ... + c_{\tau,n} \overline{y}_{t,\tau_n,h}, \tag{12}$$

where $\overline{v^2}_{t,\tau,h} \equiv \frac{1}{h}\int_0^h v^2_{t+s,\tau} ds$ and $\overline{y}_{t,\tau,h} \equiv \frac{1}{h}\int_0^h y_{t+s,\tau} ds$. A suitable proxy for $\overline{v^2}_{t,\tau,h}$ (and $\overline{y}_{t,\tau,h}$) can be obtained from discretely measured intraday data on $y_{t,\tau}$'s. AB find that running the regression (12) generates a very low R^2s, an indication of the failure of the spanning condition.

The results in CDG (2002) and AB (2006) suggest that there is *some* component in bond option prices or in the instantaneous yield volatility that is unrelated to the yield curve. These regressions thus support *weak USV* scenarios. However, by their nature they do not have much to say about *strong USV*, as they address mainly the relatively *high-frequency* aspects of the volatility dynamics: CDG (2002)'s is based on the monthly *change* in asset prices,[9] and AB (2006)'s results are also driven by high-frequency features of the volatility dynamics (scheduled macroeconomic data releases) as I shall discuss in detail in Sec. 3.1.

2.4. **Evidence for Strong USV.** A long-standing difficulty in term structure modeling is the failure of relative pricing between bonds and bond options: time-consistent no-arbitrage term structure models estimated only with yields data tend to imply bond option prices that do not agree well with market prices. For example, JKS (2003) find that the multi-factor CIR models (up to 3 factors) generate model-implied option prices that have large pricing errors and do not show much similarity to market option prices in their time-variation even if one looks beyond the high-frequency aspects (see, e.g., their Figure 4). LZ(2006) also report similar difficulties for a richer model (3-factor QG model).[10]

The USV proposal of CDG (2002) leads to a surprising answer to the puzzle of the poor relative pricing: relative pricing fails because it is *meant* to fail; the bond market is incomplete, and hence term structure information is not sufficient to price bond options. Though striking, by itself it does not completely solve the puzzle. Indeed, we are immediately led to the following questions. How incomplete is the bond market? Even if we grant the existence of a high-frequency unspanned component in volatility, shouldn't there still be a *spanned* component of volatility that is linked to slower, macroeconomic variations in bond option prices? How much of the failure to capture the non-high-frequency movements in bond option prices (as in JKS (2003)) is due to USV, rather than due to other problems?

While strong USV is a potential explanation for the poor relative pricing, it is possible that we have not explored comprehensively enough the alternative explanation: the existing literature may have had specification problems (unrelated to USV). Indeed, the multi-factor CIR model used by JKS (2003) has some well-known deficiencies: for example, the factors in the CIR models are constrained to be independent, and the relation between term premium and volatility in the model may be too tight (see, e.g., Duffee (2002)). While the QG model used by LZ (2006) does not have these problems, it may have a subtler problem (with normalization) which is also shared by JKS (2003), as I shall argue in Sec. 3.3.

[9]Note that one cannot simply "un-difference" the CDG (2002)'s regression (i.e., regression in levels instead of in differences) to investigate the strong USV question: because yields and option prices are fairly persistent variables, one could get a high R^2 simply due to the spurious regression effect.

[10]After the first draft of the present paper had been finished, Joslin pointed out to me a paper of Almeida, Graveline, and Joslin (2006), which reports an encouraging result in pricing bond options with a (non-USV) $A_1(3)$ model estimated without options data. However, CDGJ (2004) obtain results that are at odds with this: they find that their estimated non-USV $A_1(3)$ model produces a yield volatility that is *negatively* correlated with a GARCH-type volatility (which implies a poor relative pricing performance of the non-USV $A_1(3)$ model).

Furthermore, there might be econometric issues with the existing studies. As I shall argue in Sec. 3.4, the conditions for commonly used techniques like the maximum likelihood (ML) estimation to be valid are often not satisfied in a term structure model estimation setting. In addition, in the case of LZ (2006) it is difficult to see whether the model produces reasonable business-cycle and low-frequency variations in volatility because of the short time span of their data (just exceeding two years). LZ's estimation may also have a subtler problem: their estimation is based on *daily* sampled data, but in this case the complexity of high-frequency volatility dynamics (in some sense highlighted by AB (2006)'s results) may lead to substantial distortions in inference.

3. Close look at data and assumptions

Let us now discuss in detail several aspects of data that may have important ramifications for the USV debate and the failure of the bonds/options relative pricing, and take a critical look at some of the commonly made assumptions in the specification and estimation of term structure models.

3.1. Inhomogeneity of Volatility Dynamics.

An empirical feature that has not been discussed much in the no-arbitrage term structure modeling literature but is well known to market practitioners is that macroeconomic economic data releases (such as the announcement of the nonfarm payroll growth and core-CPI) play a substantially greater role in bond price movements than in stock price movements. It is well known from Roll (1988) and Cutler et al (1989) that stock price movements are hard to explain even ex post; the regression of stock price changes onto identifiable news or events typically gives a low R^2. By contrast, a much higher fraction of bond price movements are explainable with macroeconomic data releases and other identifiable events (such as the release of FOMC statements and minutes and the speeches of Fed officials); see, for example, Fleming and Remolona (1997).

The reaction of bond prices to macroeconomic data releases is (practically) immediate and often quite sizeable. Thus, often on days of important data releases, the intraday time series of bond yields displays a jump-diffusion-like behavior. The intraday time series of the 10-year on-the-run Treasury yield on Apr 2, 2004, shown in Figure 2a, is a fairly clean example: on this day at 8:30am there was a payroll announcement, and a sharp move to a new level is clearly visible. However, even the jump-diffusion characterization is only approximate: one can see a more complex reaction to the announcement on some other days. For instance, Figure 2b shows the behavior of the 10-year yield on Jun 4, 2004. On this day there was also a payroll announcement at 8:30am. The 10-year yield rose immediately upon the news but quickly came back down and then fluctuated upward afterwards.

It may thus be more accurate to characterize the interest rate behavior on announcement days as "high-volatility days" (rather than "jump days"). In fact, the volatility on these days tends to be so high compared to non-announcement days that an allowance for this fact may have to be made to describe the daily volatility dynamics accurately.

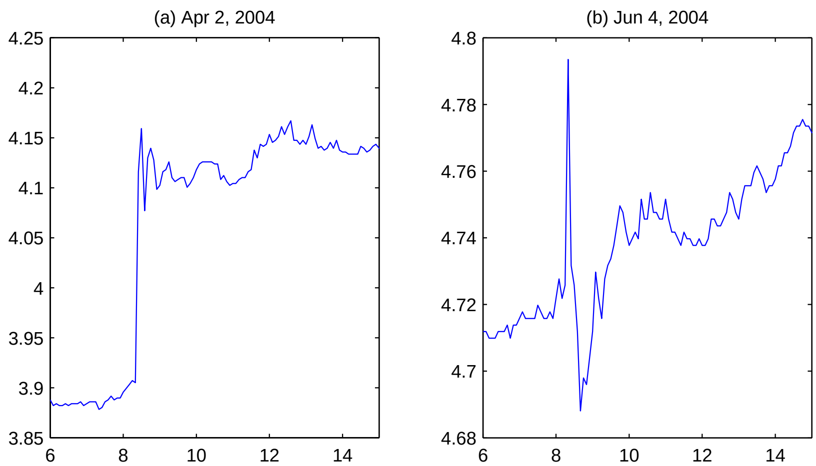

Figure 2: Intraday (5-min-interval) time series of ten-year on-the-run Treasury yield from 6 o'clock to 15 o'clock (Source: internal data, Division of Monetary Affairs, Federal Reserve Board).

For example, Jones et al (1998) model the daily Treasury bond return volatility via a GARCH model in which the volatility on macroeconomic data announcement days are larger than non-announcement days by a factor $\sqrt{1+\delta_0}$.[11] In light of the strongly inhomogeneous behavior of daily volatilities, it is not surprising that AB (2006) find very low R^2's in their regression, supporting the existence of USV.

While the effect documented in AB (2006) constitutes one channel of USV, it is not a fundamental one (from a macroeconomic point of view) but a largely "institutional" one. The dates of important macroeconomic data releases are known in advance, and traders already anticipate high volatility on these dates. Therefore, even if the volatility on an announcement day is ten times larger than the volatility on the previous day, this does not mean that the fundamental risk has increased ten times. To clarify this point further, consider a thought experiment in which the Bureau of Labor Statistics changed the date of the upcoming nonfarm payroll announcement (e.g., push back by a day). There would likely be a high volatility on the new date and not on the old date, but this change in the pattern of volatility clearly does not have any macroeconomic significance.[12] Technically speaking, the "unspanned stochastic volatility" effect in AB (2006) is in fact not fully stochastic, since much of the rise in volatility on announcement days is an *anticipated* one.

Perhaps the clearest and also the most striking indication that markets largely anticipate high volatility on important macroeconomic announcement days can be found in very-short maturity options on bonds. In the US, the monthly release of the nonfarm

[11]However, their treatment ignores the fact that some macro announcements (like the nonfarm payroll growth) are a lot more influential than some others.

[12]Note, however, that the amplitude of the bond market reaction to the news (expected volatility on the announcement day) *can* depend on the state of economy.

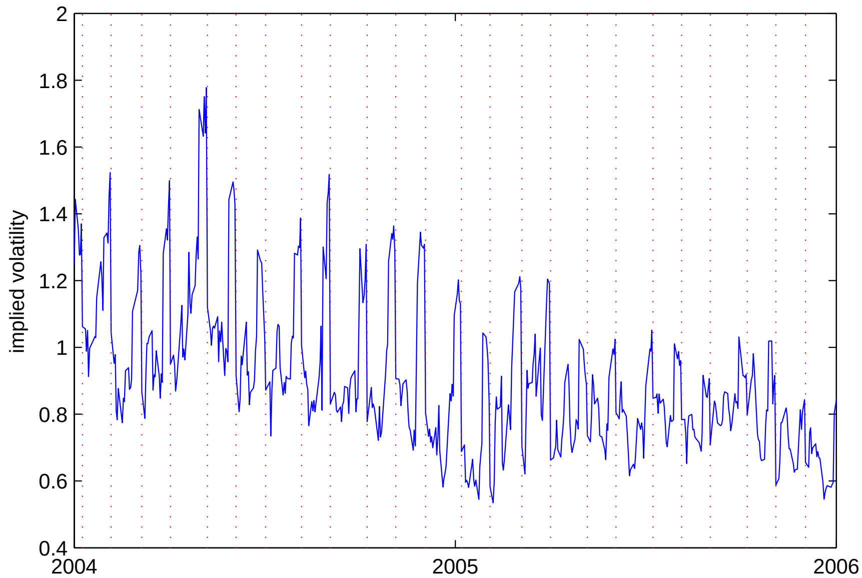

Figure 3: Implied volatility from the one-week option on the five-year Treasury note. Dotted lines show the dates of nonfarm payroll announcement. (Source: Goldman Sachs; Moessner and Nelson (2007))

payroll growth is particularly influential among various data releases; in fact, most of the days of the highest realized volatility in recent years have been the nonfarm payroll announcement days. Thus traders have come to expect a particularly high volatility on these dates. This effect is strong enough that if a one-week bond option expires after the nonfarm payroll announcement day, it has a notably elevated price. Figure 3 shows the daily time series of the implied volatility from a 1-week option on the five-year Treasury note, from which a nonfarm payroll announcement-induced seasonality can be clearly seen. Obviously, these "seasonal" features cannot be captured with time-homogeneous no-arbitrage models like affine and QG models.

3.2. **Time Scales and Time Series.** The complexity of high-frequency volatility dynamics discussed above can create difficulties even if one wishes to focus on the more fundamental variation in interest rate uncertainty. For instance, in LZ (2006)'s USV study with the QG model, their use of *daily* data may lead to distortions when the estimation tries to fit the "unfittable": the largely anticipated rise in volatility on certain days would be treated as a purely random (unanticipated) rise in volatility.

One response to this problem is to build a more elaborate model that accommodates anticipated jumps and inhomogeneities (by identifying all dates of important macroeconomic data releases). However, that would be too ambitious an undertaking for the present paper, especially in view of the fact that still too little is agreed about the proper modeling of volatility in the term structure modeling context.

Another response – something of a "poor man's solution" – is to sample less frequently (weekly or monthly, instead of daily). The idea is this: if term structure data were sampled less frequently, the data would appear more homogeneous, and the dynamics would look more like a diffusion. A visual comparison of 10 years' monthly data and 6 months' daily data would immediately illustrate this point (jumps and inhomogeneities are harder to see in the former), but one can also see this from the comparison of the kurtosis of 1-day, 1-week, and 1-month changes $(\Delta_d, \Delta_w, \Delta_m)$ in interest rates. For example, computing the kurtosis k for the 2-year swap rate from 1995 to 2007 gives

$$k(\Delta_d) = 5.98, \quad k(\Delta_w) = 4.90, \quad k(\Delta_m) = 4.19. \tag{13}$$

Note that the kurtosis of the daily change is the largest of the three, as jump-like effects make the distribution more heavy-tailed (recall that the normal distribution has a kurtosis of 3), and this heavy-tail effect declines with increasing time scale. Thus, we could think of time-homogeneous diffusion models as an approximate, "effective" description of the weekly- or monthly-frequency data, even though the "true" process is a far more complicated one.

From the classical econometric point of view (asymptotic theory), it may appear unwise to sample intentionally less frequently, when the availability of data and computational resources is not an issue. However, the point is that any model is necessarily an imperfect approximation, and a model that is designed to best capture the behavior at certain time scale (e.g., high-frequency features) might not necessarily be the best model for addressing questions about an effect that occurs at another characteristic time scale.

While the idea of there being no such thing as "one model fits all" may appear pessimistic to time-series econometricians, physicists are comfortable with this idea and accept that different kinds of physical laws (or "effective models") are needed to describe systems at different time scales, length scales, or energy scales.[13] To give a simple example, in order to describe the behavior of a steel ball thrown into the air, we should use the laws of classical mechanics (Newtonian mechanics); quantum mechanics (a more advanced and modern theory) would be unnecessary and distracting here.[14] Another example, which I believe is due to Benoit Mandelbrot, is the fact that a piece of paper appears zero-dimensional (a point) if seen from a distance; sufficiently close, it would look two-dimensional (a plane); and even closer on a microscope, it would look one-dimensional (fibers).

In sum, although time-homogeneous diffusion models may "break down" at very high frequencies, it is at least an open question whether such models would be so bad an approximation of the weekly or monthly data.[15] Lastly, note that, while sampling at lower

[13] Philip Anderson (1972)'s influential essay, a critique of the so-called reductionist philosophy of science, gives a deeper discussion of this issue.

[14] Quantum mechanics would be needed to describe the iron atoms inside the ball, but for this example we can treat the ball as a point mass rather than a system of many atoms.

[15] Admittedly the problem under consideration here (dynamic term structure estimation) is much more complicated than the steel ball example or the piece-of-paper example: the qualitative distinction

frequencies is likely to improve the prospects for time-homogeneous diffusion models, it does involve information loss for some aspects of the data. In particular, it is well known that although the "drift" part of a diffusion model is not much better estimated by a more frequent sampling, the "diffusion" part is better estimated with a more frequent sampling (e.g., Merton (1980)). Later in the paper we shall discuss some ways of addressing this problem.

3.3. **Term Structure Behavior in the Past Few Decades.** In this paper, I focus mainly on the US term structure behavior in the 1990s and 2000s, as this is the period that is most representative of the current interest rate environment, and as most other studies of relative pricing between bonds and bond options have focused on this period (due to the ready availability of the options data).

As mentioned in the Introduction, in the earlier periods (especially in the late 70s and early 80s), there seemed to be a positive relation between the level of interest rates and the amount of interest rate uncertainty, thus a model like the CIR model,

$$dr_t = \kappa(\mu - r_t)dt + \sigma\sqrt{r_t}dB_t, \tag{14}$$

whose short rate volatility is proportional to $\sqrt{r}$, seemed sensible. However, the relation between the level of interest rates and interest rate uncertainty is much less clear-cut in the more recent period (1990s and 2000s). For instance, there is little in Figure 1 (shown earlier in Sec. 2.2) that jumps out to the eye to indicate a positive relation between the short rate and the uncertainty. This might be consistent with the USV models like CDG (2002), which have a volatility variable that is independent of bond yields.

However, there is another possibility, namely that the relation between volatility (interest rate uncertainty) and the term structure has become more complex; in other words, there is still possibility for multi-factor term structure models without the USV feature to be consistent with the observed behavior of volatility and term structure. Models like JKS (2003) and LZ (2006) may look especially promising, in view of their potentially rich volatility dynamics (having more than one factor to describe volatility variation). However, these studies make a normalization choice that is problematic for the past few decades' data.

I now illustrate this point with the multi-factor CIR model used by JKS (2003). Their 3-factor model is specified as

$$\begin{aligned} r_t &= \phi + x_{1t} + x_{2t} + x_{3t}, \\ dx_{it} &= \kappa_i(\mu_i - x_{it})dt + \sigma_i\sqrt{x_{it}}dB_t. \end{aligned} \tag{15}$$

By a re-scaling of the x_{it}'s, one can show that the specification (15) can be written in the form

$$\begin{aligned} r_t &= \phi + \rho_1\, x_{1t} + \rho_2\, x_{2t} + \rho_3\, x_{3t}, \\ dx_{it} &= \kappa_i(\mu_i - x_{it})dt + \sqrt{x_{it}}dB_t. \end{aligned} \tag{16}$$

between daily and monthly behaviors of the yield curve is not as clear-cut as the fibers-plane distinction; furthermore, the term structure effects that occur at different time-scales are likely correlated.

Hence, having the coefficient of 1 on x_{it} for the short rate in eq. (15) is not really a restriction but a "normalization." That said, the two normalization choices are *not* equivalent (if the signs of the ρ_i's in eq. (16) are unrestricted): eq. (16) implies eq. (15), but not vice versa. For example, the specification

$$\begin{aligned} r_t &= \phi + x_{1t} + x_{2t} - x_{3t}, \\ dx_{it} &= \kappa_i(\mu_i - x_{it})dt + \sigma_i\sqrt{x_{it}}dB_t. \end{aligned} \tag{17}$$

is covered by the normalization (16) but not by the normalization (15).

The normalization (15) has become quite standard, and almost all studies of the multi-factor CIR models have used it (e.g., JKS (2003) and Duffie and Singleton (1997)). Besides tradition, the likely reason for the prevalence of the normalization (15) is that such a normalization is consistent with the positive-definiteness of the nominal short rate. For this reason, I shall refer to this normalization as the "positive-definite normalization". However, the positive-definiteness is not as compelling a reason as it might appear. The short rate process (15) is indeed bounded at zero (if $\phi = 0$), but it would poorly describe the short rate behavior if the short rate was indeed at or near the zero boundary. The Japanese experience in the 2000s clearly shows that if the nominal short rate hits the zero bound, it can stay there for quite some time before rising again. Such a behavior can be captured by Black (1995)'s "interest rates as options" model, in which the short rate takes the form

$$r_t = \max[r_t^*, 0], \tag{18}$$

where r_t^* is a shadow-rate process that can go below 0. By contrast, the process described by eq. (15) does not spend much time at the boundary.[16] Recall also that the empirical studies that use the normalization (15) without imposing a restriction on ϕ often find that ϕ is negative (e.g., Duffie and Singleton (1997) and JKS (2003)), thus the positive-definiteness is anyway not satisfied by the model with positive-definite normalization.

The US experience of the 2001-2003, during which the Fed's target rate was lowered from 6.5% to the historical low of 1%, is also a revealing episode. According to the one-factor CIR model (14), the fourfold reduction in the short rate in 2001 should have reduced the interest rate volatility by twofold. By contrast, interest rate uncertainty as measured from eurodollar futures options tended to *rise* in 2001, as can be seen in Figure 1. Note that under the normalization (15) even the multi-factor version of the CIR model would have problems with this episode: because x_{it}'s are positive processes, r_t $(= \phi + \sum_i x_{it})$ being low means that x_{it}'s have to be small, thus the volatility of the short rate would also tend to be low. The "non-positive-definite" normalized models,

[16] Note that the one-factor CIR model is a *reflecting boundary process*. In the case of the multi-factor model (15), *all* factors $(x_{1t}, x_{2t}, ...)$ have to be zero for the process to be at the boundary, thus it is even less likely to spend time at or near the boundary.

on the other hand, would have easier time in this regard.[17] For example, one can obtain a declining level and rising volatility when the "$-x_{3t}$" term in eq. (17) declines.

Another problem with the normalization (15) is that it has difficulty capturing certain interest rate distributions that arise in practice. Note that options on eurodollar futures with different strikes can provide information on the whole (risk-neutral) distribution, not just the means and variances of the distribution. We know from these data that the distribution of future short rate can display a positive, negative, or zero skewness. For example, the risk-neutral probability density function based on the options data (Figure 4) on March 20, 2007 is clearly asymmetric and shows a negative skew. There are also times during which the opposite (positive) skew is observed.

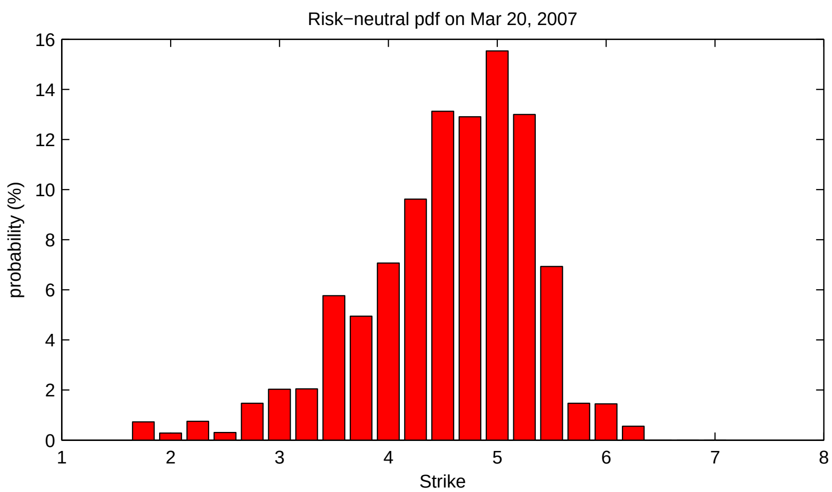

Figure 4: Risk-neutral probability density function of the short rate 1-year-ahead, based on the eurodollar futures options (Source: internal data, Division of Monetary Affairs, Federal Reserve Board).

It is well known that the unconditional distribution of the CIR process has a positive skew. The conditional distribution of the CIR process $f(r_{t+\tau}|r_t)$ is not easy to characterize in a simple manner, as its shape depends on r_t and the parameters of the model. Typically though, the conditional distribution has a positive skew or is close to symmetric. Therefore, the kind of distribution seen in Figure 4 is difficult to capture with the CIR model.[18] The multi-factor CIR model with normalization (15) would face similar

[17]Besides the present paper, the only other paper (that I am aware of) that explores the empirical consequences of non-positive-definite normalization is Backus et al (2001). Their main focus is different (they focus on predictability, rather than volatility), but they also find that the non-positive-definite normalization (their models D,E) describes data better than the positive-definite normalization (their model C); their model E is the eq. (17) in the present paper. Incidentally, note that the "canonical specification" of Dai and Singleton (2000) corresponds to the normalization (16) in this paper, rather than the normalization (15).

[18]The $A_1(n)$ USV models of CDG (2002) and CDGJ (2004) also have problems in this regard, as they can produce neither the positive skew nor the negative skew. Thus they might encounter difficulties in pricing out-of-money options, whose prices tend to be sensitive to distributional assumptions.

difficulties, as it is a sum of "positive-skewed" processes. By contrast, non-positive-definite normalization like (16) or (17) can potentially accommodate different cases of skewness (+/0/−) observed in the options market.

The QG model has also been traditionally specified with an analogous "positive-definite normalization." In particular, LZ (2006)' 3-factor QG model study uses a normalization (same as Ahn, Dittmar, and Gallant (2002)) that sets ρ in eq (7) to zero and the diagonal elements of the Ψ matrix to 1.[19] A quadratic form $r_t = \phi + x_t'\Psi x_t$, where Ψ is a positive-definite matrix (thus has a Cholesky decomposition $\Psi = ZZ'$), can be written as $\phi + \tilde{x}_{1t}^2 + \tilde{x}_{2t}^2 + ... + \tilde{x}_{nt}^2$ (where $\tilde{x}_t = Z'x_t$). Note that this form is reminiscent of the positive-definite normalization (15) of the CIR model, hence we may expect similar difficulties as those explained above for the multi-factor CIR model.[20] Therefore, also for the QG model, a non-positive-definite normalization may be more promising than the "traditional" (positive-definite) normalization for the past few decades' data.

3.4. **Potential Problems with Conventional Estimation.** As discussed briefly in the Introduction, the empirical evidence on USV from no-arbitrage term structure model-based studies is somewhat mixed. Part of the problem is that it is difficult to tell how much of the conclusion in some of these studies are due to the *richness* of the model, as opposed to the *correctness* of the model.

Consider, for instance, the study of CDGJ (2004). They find that their $A_1(4)$ USV model describes data better than an $A_1(3)$ non-USV model. Though the $A_1(3)$ non-USV model is not nested by the $A_1(4)$ USV model, the latter is a richer model in the usual sense, having one more factor and more parameters. Hence there is a concern that the outcome reflects the richness of the model instead of the presence of USV.

There may be a similar concern in the case of Joslin (2007), who concludes against the presence of USV. He finds that $A_1(4)$ and $A_2(4)$ non-USV models capture bond option prices quite well and generate smaller option pricing errors than the USV counterparts. However, these are very rich models with many parameters. Especially in view of the fact that bond option prices were also used in the estimation (i.e., the estimation explicitly tries to minimize option pricing errors), some caution is warranted in interpreting the result.

A well-known result in latent-factor term structure modeling may serve as a useful reminder: estimations of 3-factor latent-factor models can fit the last 15 years' yield data quite well (typical fitting error for the 10-year yield being 5 basis points or less), but this does not mean that the model captures well the important features of the data *that are not explicitly fitted* (such as the term premia and volatility variation), as the

[19] Fixing the diagonal elements of Ψ at 1 by itself does not guarantee the positive definiteness of the interest rates. (Depending on ϕ or the off-diagonal elements of the matrix Ψ, it can be non-positive-definite.) However, I shall still refer to this normalization as "positive-definite normalization" to emphasize the motivation behind the normalization.

[20] However, the problem might be less severe for the QG model (than the multi-factor CIR model), since the factors are allowed to have a general correlation.

literature has found on many occasions. The flexibility inherent in latent-factor models is thus a double-edged sword; in some cases the model might fit "too well" the aspects of the data that are explicitly asked to be fit.[21]

In principle, this might not be a problem if one has long enough data. In practice, however, commonly used estimation techniques in the literature might not have enough discipline to guard against the possibility of "obtaining conclusions by design." There are at least two problems with conventional estimation techniques in a typical setting:

(I) As emphasized in Kim and Orphanides (2005), there is a severe small sample problem. For example, a standard ML estimation of the 3-factor affine-Gaussian model with a 1990-2003 sample generates term premia in forward rates that most market participants would find nonsensical (even though the bond pricing errors are small). The problem is that the time span of the samples typically used in the term structure estimation is too short to pin down some of the key parameters that determine the behavior of the expectations (in the physical measure).

(II) A second problem is that the basic premise of the classical estimation, namely having the "true model" (or testing if a model is true), is untenable for the present context. As the discussion in Sec. 3.1 makes it clear, time-homogeneous diffusion models (like affine models and QG models) are at best an approximation of a far more complicated model (or reality); it is a foregone conclusion that these models (and even the jump-diffusion models) would be rejected if one focuses on the features at short enough time scales. Even aside from the issue of high-frequency dynamics, it is clear that reduced-form latent-factor models are just a "statistical representation" of data, and hence there is no assurance that the classical criterion function (likelihood function in the case of the ML estimation) of the model should have a unique "meaningful" maximum: it is possible that there are different local maxima that fit various aspects of the data with differing degree of emphasis, e.g., one maximum might have a relatively good fit of volatility dynamics, another maximum might have a particularly small bond pricing errors, and so on.

These two problems raise questions about the relevance of the asymptotic concept that underlies much of the classical econometrics. Some papers including CDGJ (2004) proposed to use Bayesian techniques instead. However, they have mostly used *uninformative (flat) priors* (hence have not introduced extra information or constraints to address the problems), so it is not clear whether and how they can overcome above problems.

4. REEXAMINATION OF RELATIVE PRICING

The remainder of this paper explores whether a more encouraging result would be obtained in bonds/options relative pricing once the potential problems discussed above are taken care of in some way.

[21]Kim (2007) discusses an example in which long-horizon survey inflation forecasts are fit "too well" in a latent-factor model of inflation and nominal term structure.

4.1. **More on the QG Model.** For this investigation, I shall restrict attention to the quadratic-Gaussian class of models. The principal motivation for focusing on the QG model is to address several questions that naturally arise from the recent 3-factor QG model study of LZ (2006), which is arguably the most sophisticated model so far explored in the bonds/bond-options relative pricing. The rather disappointing relative pricing performance that they find is somewhat at odds with Ahn et al (2002) and Kim (2004), who used earlier and longer samples and concluded fairly positively about the QG model's empirical performance, especially as regards its ability to capture the volatility and term premium variations. Do LZ (2006)'s findings indicate that the QG model works less well for recent years? Are they really evidence for strong USV? In other words, beside its inability to capture the high-frequency features discussed in Sec. 3.1 (which is shared by other models like affine models), does the QG model have difficulty also in capturing the lower-frequency variations in volatility and bond option prices? Beyond these immediate questions, the QG model merits more study also because it has not been explored as extensively as affine models despite its many attractive features.

Reduced-form latent-factor models like affine models and QG models are implicitly a projection of various information about interest rates onto a dynamical basis (statistical representation of data). The QG model is particularly appealing in this regard, as it can be viewed as a multivariate Taylor series representation (reminiscent of the polynomial regression (series estimator) in nonparametric statistics; see, e.g., Andrews (1991)), i.e., bond yield of an arbitrary maturity could be represented as

$$y_{t,\tau} = a_\tau + \sum_i b_{\tau,i} x_{it} + \sum_{ij} C_{\tau,ij} x_{it} x_{jt} + \ldots \tag{19}$$

for some set of variables x_t that follow the multivariate Gaussian (Ornstein-Uhlenbeck) process. Note that if one had to treat the coefficients $a_\tau, b_{\tau,i}, C_{\tau,ij}$ separately for different maturities, that would be not only inefficient (since the coefficient for adjacent τ's are expected to be similar) but also negligent of the no-arbitrage principle. Specifying the QG model in the form of eq. (7), i.e., via the specification of the short rate and the market price of risk, is a convenient way of writing down a second-order polynomial form that is arbitrage-free and, at the same time, general and tractable.

If the expansion (19) is truncated at the first order, one obtains the affine-Gaussian model.[22] While the affine-Gaussian model has been useful in many contexts, it gives up the possibility of capturing the time-varying volatility of interest rates.[23] The affine

[22]In contrast to the affine-Gaussian and QG models, it is difficult to view models like multi-factor CIR models as a series expansion. No matter how many factors they have, it is hard to rationalize that the one can project information about interest rates onto a set "square-root" factors such that the factors are all independent. Indeed, estimated CIR models typically lead to the result that the model-implied state variables are correlated (e.g., Duffie and Singleton (1997)).

[23]In the affine-Gaussian model, the instantaneous excess return (bond risk premium) $e_{t,\tau}$ takes the form $e_{t,\tau} = -\tau b_\tau' \Sigma \lambda_t$. (See, for example, eq. (7) of Duffee (2002)). Thus, all of the variation in $e_{t,\tau}$ comes from the variation in the market price of risk λ_t, and nothing from Σ, which is constant in the affine-Gaussian model. If we think of Σ as variable, this leads to second-order effects of the form $x_{it} x_{jt}$;

literature has dealt with this problem by modifying the factor dynamics (x_t-process) from the Gaussian process to a non-Gaussian process that has the time-varying volatility feature. This, however, introduces restrictions on the factor space and the factor dynamics. For example, if an affine model factor x_{1t} is a volatility variable, it has to be nonnegative, and this restricts the feedback and correlation structure of x_{1t} with other variables. As noted by Dai and Singleton (2000), there is a trade-off between the richness of volatility dynamics and the flexibility of factor correlation in the affine model, the latter aspect decreasing in the order $A_0(n), A_1(n), ..., A_n(n)$. By contrast, the QG model does not face this trade-off, and there is no need to classify the general n-factor model into subclasses as in affine models. It is also particularly attractive for the present problem (bond option pricing) that multiple factors can affect the volatility dynamics in the QG model, as opposed to having just a single volatility factor as in the $A_1(3)$ or $A_1(4)$ model considered by CDG (2002) and CDGJ (2004).[24]

4.2. **Overall Plan.** In this paper I estimate two versions (normalizations) of the maximally flexible 2-factor QG model. The use of a 2-factor model here can be viewed as *complementary* to the studies of Joslin (2007) and CDGJ (2004) using much richer (4-factor affine) models. Recall from the discussion in Sec. 3.4 that with rich models there is a risk that a desired conclusion is unwittingly obtained "by design." Therefore, while it *is* important to develop realistic models that can fit bond option prices well, it is also worth investigating whether some of the fundamental questions can be addressed using simple models.

In this paper, I intentionally make it challenging for the model to match the observed bond option prices in two ways: first, using a relatively parsimonious model in the form of a 2-factor model (so that there would be little concern of overfitting), and second, by restricting attention to purely relative pricing (i.e., without using bond option data in the estimation). Of course, in this case there is a risk that the model is "too challenged," i.e., it might fare too poorly to be of interest. We shall see, however, that the 2-factor QG model has a surprisingly rich empirical content, and has much to say about the USV and the relative pricing puzzle. The added benefit is that due to the relative simplicity of the model it is easy to pinpoint the sources of problem when something doesn't work well;

---

this can be viewed as another motivation for the QG model. (This is however only heuristic, since b_τ also depends on Σ and the expression was derived under the assumption of constant Σ.) In the affine models with stochastic volatility factors, the excess return takes the form $e_{t,\tau} = c_0 + c_1\nu_t + c_2x_{2t} + ...$, i.e., affine in the volatility factor ν_t and other factors; higher-order effects like ν_t^2 or $\nu_t x_{2t}$ are ruled out in affine models. One might view some of the factors x_{it} in the affine model as compensating for the "omitted variables" effect, but the factors might be too constrained by other requirements to play that role in a relatively low-dimensional model.

[24]The QG model also turns out to have an elegant mathematical structure. In particular, the differential equation (Riccati equation) for the QG model admits an *exact* solution, first derived in Kim (2004). Beyond potential practical benefits, there may be a deeper significance to this, as the exact solvability of the differential equation is connected to the existence of a hidden symmetry (underlying Lie group) in the general QG model.

we shall see that the cases in which rather poor results are obtained are also interesting, as they may shed light on *why* some of the results in the existing studies are poor.

4.3. **Normalizations.** I explore two different normalizations in light of the discussion in Sec. 3.3 about the potential limitations of the positive-definite normalization. Especially in view of the "multivariate Taylor series approximation" motivation of the QG model, it befits to let the *data* decide the normalization.

Specifically, the following two specifications are estimated. One version, denoted $QG2^{++}$, uses the "positive-definite normalization," the same as in LZ (2006). In the notation of eqs. (2) and (7), it is:

$$\rho = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ \Psi = \begin{bmatrix} 1 & \Psi_{21} \\ \Psi_{21} & 1 \end{bmatrix}, \ \mathcal{K} = \begin{bmatrix} \mathcal{K}_{11} & 0 \\ \mathcal{K}_{21} & \mathcal{K}_{22} \end{bmatrix}, \ \Sigma = \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}, \tag{20}$$

variables not written out $(\phi, \mu, \lambda_a, \Lambda_b)$ being unrestricted.

The second version, denoted $QG2^{+-}$, is the same as $QG2^{++}$, except that the Ψ matrix has one diagonal element normalized to 1, and the other diagonal element normalized to -1, i.e.,

$$\Psi = \begin{bmatrix} 1 & \Psi_{21} \\ \Psi_{21} & -1 \end{bmatrix}. \tag{21}$$

Thus the $QG2^{+-}$ model is obviously not positive-definite.[25]

As discussed in connection with the CIR model, a more promising model of the short rate for situations where the non-negativity of interest rates matters would be a Black (1995)-type model,

$$r_t = \max[\phi + \rho' x_t + x_t' \Psi x_t, \, 0]. \tag{22}$$

The very-near zero-boundary behavior of the short rate is simply too singular to be described by analytic functions like (7), regardless of the positive-definiteness constraint. Unfortunately, computations with the model (22) would be far more complicated, as yields would be no longer quadratic in state variables. We shall thus simply assume that the specification in eq. (7) is a good approximation to eq. (22) for the US data under consideration (though obviously it would not be so for the Japanese data). Recall also that other models that are popular in the literature, such as the $A_{i<n}(n)$ affine models, are also non-positive-definite, but this reason by itself has not discouraged researchers from using them.

[25]The following normalization

$$\mathcal{K} = \begin{bmatrix} \mathcal{K}_{11} & 0 \\ \mathcal{K}_{21} & \mathcal{K}_{22} \end{bmatrix}, \ \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ \mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

with $\phi, \rho, \Psi, \lambda_a, \Lambda_b$ being unrestricted, is another way to introduce (allow for) non-positive-definiteness; this may be more convenient especially in the case of higher-dimensional models. Leippold and Wu (2003) used this normalization, but they imposed the positive-definiteness condition in the estimation, which makes it equivalent to the $QG2^{++}$ model.

4.4. **Empirical Setup.** The models are estimated with weekly-sampled (every Wednesday) LIBOR/swap-based zero-coupon yields (at maturities of 3-month, 6-month, 1-, 2-, 4-, 7-, 10-year) from January 1995 to June 2007.[26] Weekly data are used here (as opposed to the daily estimation of LZ (2006)) to alleviate the concerns about the inhomogeneity of the high-frequency volatility dynamics, as discussed in Sec 3.2. As emphasized earlier, this paper examines only the pure relative pricing: options data are not used in the estimation; instead they are reserved for comparison with the values implied by the estimated model.

For estimating the model, in addition to the standard ML method based on the extended-Kalman filter (linearizing the observation equation),[27] I explore the use of certain "regularized ML" estimations. Since the QG model, like affine models, is an "obviously false" model, there is little point in testing whether the model is rejected. Instead, the relevant question here is how well the model works *as an approximation*, in particular, whether it can capture key features of the bond option price variation. For such an inquiry, it is not clear that the standard ML estimation (inference based on the global optimum of the likelihood function) tells all that we need to know about the model. Due to the informational insufficiency (small sample problem) and the inherently approximate nature of the models (discussed in Sec. 3.4), the estimation problem may be less "well-posed" than is often presumed.

Therefore, I propose to approach the problem from the perspective of the so-called *ill-posed inverse problems*. There is an extensive literature on this sort of problems in physical sciences and engineering, but they have received relatively little attention in econometrics.[28] Some examples in statistics do belong to this class of problems, but are not often labeled as such. One example is the "mixture of normal distributions" model: Suppose the data $z_1, z_2, ..., z_N$ are generated from the distribution

$$\begin{aligned} z &\sim N(\mu_1, \sigma_1^2) \qquad \text{with prob } \alpha, \\ &\sim N(\mu_2, \sigma_2^2) \qquad \text{with prob } 1-\alpha. \end{aligned} \tag{23}$$

The likelihood function of this problem does not have a global maximum: it diverges when $\mu_1 = z_1$ (or any other z_i) and $\sigma_1 \to 0$; hence searching for the global maximum is pointless and misleading. This example is thus a useful reminder that a problem that *looks* well-posed may not be so in reality. To solve the problem, some additional

[26] The zero-coupon yields data are constructed by fitting a flexible parametric form to the $3m$, $6m$, $1y$ LIBOR rates and available swap rates.

[27] See LZ (2006, Sec. II.B) for relevant formulae for the standard extended-Kalman-filter-based maximum likelihood estimation of the QG model and Harvey (1989, Sec. 3.7.2) for a more general discussion of the extended-Kalman filter.

[28] The strict definition of a "well-posed" problem according to Hadamard is that (1) the solution exists and is unique and (2) the solution depends continuously on data. Ill-posed problems in the narrow sense can be defined as the problems that are not "well-posed". However, the techniques developed in the ill-posed problems literature are also useful for nearly-ill-posed problems (e.g., when the likelihood function is nearly singular or almost flat). O'Sullivan (1986) provides a survey of ill-posed problems in which randomness plays a role.

information (or constraint) has to be provided to render the problem "better posed"; such a procedure is generically called *regularization*.[29]

For the term structure estimation problem at hand, I explore two ideas for regularization. One idea is to supplement the standard estimation with information (or priors) about the expected short-term interest rate from survey forecasts. This is intended to address the first of the two problems discussed in Sec. 3.4, namely the small sample problem. Because the source of the problem is that the parameters of the model related to the interest rate expectations in the physical measure are often only weakly identified due to the relatively short time span of data, one can expect that providing information about the short rate expectation in the form of a noisy proxy (survey forecast) would help.

This can be done by including the condition

$$E_t^{survey}(y_{t+u,3m}) = E_t(y_{t+u,3m}) + \epsilon_{t,u}^F, \qquad \epsilon_{t,u}^F \sim N(0, \sigma_{F,u}^2) \tag{24}$$

in the estimation. Note that in the QG model $E_t(y_{t+u,3m})$ is quadratic in the state variables, as in the case of bond yields. For $E_t^{survey}(y_{t+u,3m})$, I use the 6-month and 12-month horizon forecasts of the 3-month LIBOR rate from the *Blue Chip Financial Forecasts* (BCFF) survey, and the long-horizon BCFF forecast, which I approximate as a 7.5-year horizon forecast.[30] The variables $\sigma_{F,u}$ ($u = 6m, 12m, 7.5y$) can be viewed as the size of the assumed measurement errors, and will be discussed more below.

One could also introduce an additional regularization, in the form of some information (or prior) about the time-variation in interest rate volatility. The idea is to compensate a little bit for the volatility information lost due to a less frequent sampling (necessitated by the "break down" of diffusion models at high frequencies); recall the discussion at the end of Sec. 3.2. In this regard, one could include the following condition for instantaneous volatility,

$$v_{t,\tau}^{proxy} = v_{t,\tau} + \epsilon_{t,\tau}^V, \qquad \epsilon_{t,\tau}^V \sim N(0, \sigma_{V,\tau}^2), \tag{25}$$

where $v_{t,\tau}^{proxy}$ denotes some proxy for the instantaneous volatility and $v_{t,\tau}$ denotes the model-implied instantaneous volatility, i.e., eq. (9).[31] In this paper, I take $\tau = 2$-year, i.e., the instantaneous volatility for the 2-year yield.

[29]In the case of the "normal mixtures" problem, one could add a condition that $\sigma_1/\sigma_2 > c$ for some number c; see, for example, Hathaway (1985). Another example in statistics that can be viewed as a regularization is the ridge-regression for nearly-multicollinear problems.

[30]The BCFF forecasts for the next 5 or 6 quarters are available every month, while the long-horizon BCFF forecasts are available twice a year. Note that these are less frequent observations than yields data, but the Kalman filter setup can easily handle such (missing data) situation; see, e.g., Kim and Orphanides (2005).

[31]I use $v_{t,\tau}$ rather than $v_{t,\tau}^2$ for regularization in order to avoid the possibility of the occasions of large $v_{t,\tau}^2$ having an excessive influence. Note that the formula for $v_{t,\tau}$ is not quadratic in x_t, but it can be still straightforwardly implemented in the standard extended-Kalman filter.

Note that instantaneous yield volatility is unobserved (and in fact not well-defined),[32] but still one could come up with simple but useful proxies. This paper uses the rolling standard deviation of daily changes of swap yields (centered at t).[33] Of course, there is some ambiguity about the length of the window for computing the rolling standard deviation. If the window was too short, the proxy would be very jagged and heavily influenced by large realized change in the swap yield on certain dates. On the other hand, if the window was too long, much of the information about the variation of volatility would be washed out. As a compromise, I use the 6-month window, which is shown in Figure 5. To give an idea about the short-window case, Figure 5 also shows the proxy based on the 1-month window.

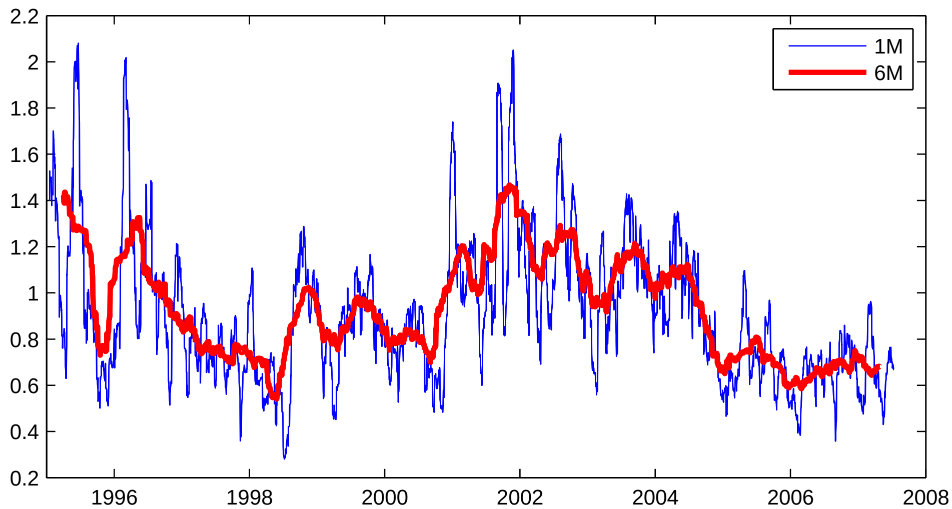

Figure 5: Rolling standard deviation of daily change in 2-year swap yields (for 1-month and 6-month windows).

In both regularization proposals (eqs. (24) and (25)), the true "market variables" ($E_t(y_{t+u,3m})$ and $v_{t,\tau}$) are *unobserved*, and the proxies are necessarily imprecise. Nonetheless, the proxy information can be introduced "fairly weakly" by fixing the measurement errors at a sizeable value, but not "too weakly" so as to make it irrelevant. In this spirit, I set $\sigma_{V,2y}$ =40 bp, which is almost twice the standard derivation of the proxy $v_{t,2y}^{proxy}$ (based on the 6-month rolling window). Similarly, I take the survey forecast measurement errors to be $\sigma_{F,6m} = 30$bp, $\sigma_{F,12m} = 40$bp, and $\sigma_{F,7.5y} = 100$bp. While survey

[32]Because jump-like events that cause very high realized volatility on certain dates are not "in the model," even if one could measure volatility at very high frequencies, the result would not lead to a suitable empirical proxy for the $v_{t,\tau}$ for the model.

[33]Other proxies such as GARCH-based estimate of volatility can be also used. Note, however, that standard GARCH models (GARCH(1,1), EGARCH(1,1), etc.) may be affected by strong data announcement effects (discussed in Sec. 3.1).

forecasts might not capture the "true" expectations in the market, most practitioners would find it also strange if the market expectations deviated by a large amount (e.g., 1 percentage point) from the survey forecasts.

While the conditions (24) and (25) can be handled most conveniently by including them in the observation vector of the extended-Kalman-filter-based ML estimation,[34] one can also pose the estimation problem as

$$\max \Bigg[(const) \log L^s - \sum_{u,t} \frac{1}{\sigma^2_{F,u}} (E_t^{survey}(y_{t+u,3m}) - E_t(y_{t+u,3m}))^2 - \sum_{\tau,t} \frac{1}{\sigma^2_{V,\tau}} (v^{proxy}_{t,\tau} - v_{t,\tau})^2 \Bigg], \tag{26}$$

where L^s is the standard ML likelihood function. This is the form that is more familiar in the regularization literature (see, for example, Press et al (2007, Sec. 19.4)); we can interpret the fixed parameters ($1/\sigma^2_{F,u}$ and $1/\sigma^2_{V,\tau}$) as "regularization coefficients". Note that the standard (non-regularized) estimation is recovered in the limit $\sigma_F, \sigma_V \to \infty$. The regularized estimations (with fixed σ_F, σ_V) can be also viewed in the Bayesian spirit, the conditions (24) and (25) being interpreted as *informative* priors.[35]

In sum, I consider two regularization implementations, one with the survey-forecasts (which I shall denote R_F), and the other with both the survey-forecasts and the volatility proxy (denoted $\mathrm{R}_{F,V}$). Together with the standard estimation with no regularization (denoted R_{none}), there are three options for the estimation of the model. With two different normalizations to be explored, this leads to six estimations, denoted $\mathrm{QG2}^{++}\mathrm{R}_{none}$, $\mathrm{QG2}^{+-}\mathrm{R}_{none}$, $\mathrm{QG2}^{++}\mathrm{R}_F$, $\mathrm{QG2}^{+-}\mathrm{R}_F$, $\mathrm{QG2}^{++}\mathrm{R}_{F,V}$, and $\mathrm{QG2}^{+-}\mathrm{R}_{F,V}$.

4.5. **Estimation Results.** The parameter estimates for the six estimations are given in Table 1. These represent the global optimum of the respective (standard or regularized) likelihood function. In all these estimations, longer-term yields are fitted quite well, while the 3-month yield is not fitted as well, a typical 3-month yield measurement error size being about 30 bp (δ_{3m} in Table 1).

Let us first examine the empirical content of the $\mathrm{QG2}^{++}\mathrm{R}_{none}$ and $\mathrm{QG2}^{+-}\mathrm{R}_{none}$ estimates (i.e., the standard ML estimation of the $\mathrm{QG2}^{++}$ and $\mathrm{QG2}^{+-}$ models). Of immediate interest is how these models perform in bond option pricing. Table 2 summarizes the results for the 1-, 2-, and 5-year at-the-money (ATM) cap pricing.[36] In

[34] This is indeed the way the "regularized estimations" are performed in the present paper.

[35] Kim and Orphanides (2005) treat eq. (24) in a more classical spirit, allowing $\sigma_{F,6m}$ and $\sigma_{F,12m}$ to be optimized in the estimation. I have also experimented with having $\sigma_{F,6m}$, $\sigma_{F,12m}$ as free parameters; in this case I obtained *smaller* values for $\sigma_{F,6m}$, $\sigma_{F,12m}$ than the numbers used here.

[36] For comparison with model-implied cap prices, I use monthly data (last Wednesday of the month) on market cap prices, constructed from the cap implied-volatility data from Bloomberg (available from January 1997) and the zero-coupon LIBOR-swap yield curve described earlier. The formula for the cap prices in the QG model can be found in Leippold and Wu (2002) and LZ (2006). To compute

	$QG2^{++}R_{none}$	$QG2^{++}R_F$	$QG2^{++}R_{F,V}$
$\mathcal{K}_{11}$	0.0275(0.0917)	0.3205(0.0501)	0.3529(0.0550)
$\mathcal{K}_{21}$	0.0281(0.0359)	0.1108(0.0684)	0.0781(0.0304)
$\mathcal{K}_{22}$	0.0438(0.0350)	0.3000(0.0572)	0.2549(0.0675)
Σ_{11}	0.0228(0.0023)	0.0110(0.0051)	0.0160(0.0032)
Σ_{22}	0.0116(0.0046)	0.0038(0.0006)	0.0048(0.0014)
μ_1	-0.0230(0.1575)	0.0773(0.0385)	0.1364(0.0323)
μ_2	0.2019(0.1482)	0.0124(0.0042)	0.0121(0.0027)
ϕ	0.0076(0.0028)	0.0110(0.0001)	0.0114(0.0002)
Ψ_{12}	0.8993(0.1265)	18.0680(6.7975)	6.5476(3.5520)
λ_{a1}	3.1801(8.9583)	-2.2975(0.3498)	-2.7064(0.5033)
λ_{a2}	0.2471(0.4986)	1.0029(0.3544)	0.3125(0.3737)
$[\Sigma\Lambda_b]_{11}$	0.6718(0.1314)	0.3919(0.0518)	0.4318(0.0566)
$[\Sigma\Lambda_b]_{21}$	-0.0135(0.0736)	-0.0061(0.0198)	0.0170(0.0212)
$[\Sigma\Lambda_b]_{12}$	-0.9537(0.7545)	-0.1928(0.2231)	-0.7344(0.2771)
$[\Sigma\Lambda_b]_{22}$	-0.0219(0.0766)	-0.2846(0.0535)	-0.3043(0.0657)
$\tilde{\delta}_{3m}$	0.0026(0.0004)	0.0027(0.0004)	0.0027(0.0004)
$\tilde{\delta}_{6m}$	0.0013(0.0002)	0.0013(0.0003)	0.0013(0.0003)
$\tilde{\delta}_{1y}$	0.0002(0.0000)	0.0002(0.0000)	0.0002(0.0000)
$\tilde{\delta}_{2y}$	0.0008(0.0001)	0.0007(0.0001)	0.0007(0.0001)
$\tilde{\delta}_{4y}$	0.0006(0.0001)	0.0006(0.0001)	0.0006(0.0001)
$\tilde{\delta}_{7y}$	0(0.0000)	0(0.0000)	0(0.0000)
$\tilde{\delta}_{10y}$	0.0005(0.0001)	0.0005(0.0001)	0.0005(0.0001)
	$QG2^{+-}R_{none}$	$QG2^{+-}R_F$	$QG2^{+-}R_{F,V}$
$\mathcal{K}_{11}$	0.0071(0.0252)	0.0166(0.0425)	0.0454(0.0767)
$\mathcal{K}_{21}$	-0.3257(0.1225)	0.0273(0.1755)	0.0797(0.1166)
$\mathcal{K}_{22}$	0.6553(0.3189)	0.3084(0.0485)	0.3167(0.0504)
Σ_{11}	0.0256(0.0010)	0.0145(0.0034)	0.0231(0.0033)
Σ_{22}	0.0358(0.0023)	0.0236(0.0019)	0.0261(0.0014)
μ_1	0.3657(0.0560)	0.1690(0.0457)	0.1530(0.0445)
μ_2	0.1419(0.0540)	-0.0501(0.0758)	-0.1099(0.0499)
ϕ	0.0115(0.0001)	0.0444(0.0067)	0.0517(0.0053)
Ψ_{12}	0.1663(0.2518)	0.4412(0.6188)	0.0912(0.5008)
λ_{a1}	-5.3861(0.4586)	0.5606(1.7735)	0.4846(1.2583)
λ_{a2}	-2.2603(0.7341)	6.8238(1.9649)	3.6010(1.3626)
$[\Sigma\Lambda_b]_{11}$	0.4935(0.0995)	-0.0707(0.1685)	-0.0869(0.1851)
$[\Sigma\Lambda_b]_{21}$	0.5254(0.1429)	-1.2386(0.3280)	-0.8383(0.1677)
$[\Sigma\Lambda_b]_{12}$	0.4517(0.0417)	0.0361(0.0724)	0.0534(0.1058)
$[\Sigma\Lambda_b]_{22}$	-0.4143(0.3608)	0.1292(0.1388)	0.1180(0.1356)
$\tilde{\delta}_{3m}$	0.0026(0.0004)	0.0031(0.0006)	0.0032(0.0006)
$\tilde{\delta}_{6m}$	0.0013(0.0002)	0.0015(0.0004)	0.0015(0.0004)
$\tilde{\delta}_{1y}$	0.0002(0.0000)	0(0.0000)	0(0.0000)
$\tilde{\delta}_{2y}$	0.0007(0.0001)	0.0008(0.0001)	0.0008(0.0001)
$\tilde{\delta}_{4y}$	0.0005(0.0001)	0.0007(0.0001)	0.0007(0.0001)
$\tilde{\delta}_{7y}$	0(0.0000)	0(0.0000)	0(0.0000)
$\tilde{\delta}_{10y}$	0.0005(0.0001)	0.0005(0.0001)	0.0005(0.0001)

Table 1: Parameter estimates for the $QG2^{++}$ and $QG2^{+-}$ models. The fixed (normalized) parameters are as follows: $\mathcal{K}_{12} = 0$, $\Sigma_{12} = \Sigma_{21} = 0$, $\rho_1 = \rho_2 = 0$, $\Psi_{11} = 1$, and $\Psi_{22} = 1$ for $QG2^{++}$ and $= -1$ for $QG2^{+-}$. The measurement errors for yields δ_τ are parameterized as $\delta_\tau^2 = \tilde{\delta}_\tau^2 + c^2$, where I have included a small constant c (=3 bp) to avoid the case of a zero measurement error. Standard errors (based on the BHHH formula) are given in parentheses.

		mean	std	MAE	corr
market	1-year	15.5	4.2		
	2-year	67.8	20.7		
	5-year	301.7	86.6		
$QG2^{++}$–R_{none}	1-year	17.3	4.0	0.42	-0.15
	2-year	61.8	6.7	0.21	0.46
	5-year	264.3	39.5	0.14	0.91
$QG2^{+-}$–R_{none}	1-year	19.2	4.4	0.53	-0.37
	2-year	67.9	4.5	0.27	0.07
	5-year	286.5	39.3	0.16	0.86
$QG2^{++}$–R_F	1-year	20.0	4.7	0.57	-0.32
	2-year	69.3	4.8	0.28	0.06
	5-year	285.7	33.7	0.15	0.94
$QG2^{+-}$–R_F	1-year	14.9	4.5	0.19	0.64
	2-year	52.8	16.7	0.22	0.91
	5-year	233.6	66.4	0.22	0.92
$QG2^{++}$–$R_{F,V}$	1-year	18.1	3.9	0.46	-0.25
	2-year	64.5	4.9	0.23	0.44
	5-year	276.0	36.4	0.14	0.95
$QG2^{+-}$–$R_{F,V}$	1-year	17.5	4.7	0.21	0.68
	2-year	60.6	17.2	0.12	0.92
	5-year	262.6	66.7	0.13	0.94

Table 2: Summary measures of relative pricing performance. "MAE" is the mean absolute errors for the percentage pricing errors. "corr" is the simple correlation between the model-implied and market cap prices.

short, the relative pricing performance is poor. The pricing errors (computed as $\epsilon = (P^{model} - P^{market})/P^{market}$) are large, as indicated by the size of the MAE (mean-absolute errors). The correlation between the model-implied value and the market value is negative for the 1-year cap for both the $QG2^{++}R_{none}$ and the $QG2^{+-}R_{none}$ estimates, and the variability (sample standard deviation) of the model-implied cap prices is much smaller than that of the market cap prices.[37]

the model-implied ATM cap prices, I use the model-implied ATM cap strike rate (based on the fitted yields) for internal consistency.

[37]These results are reminiscent of those for the multi-factor CIR models in JKS (2003), who also report much smaller standard deviation of the model-implied cap prices than those of the market prices

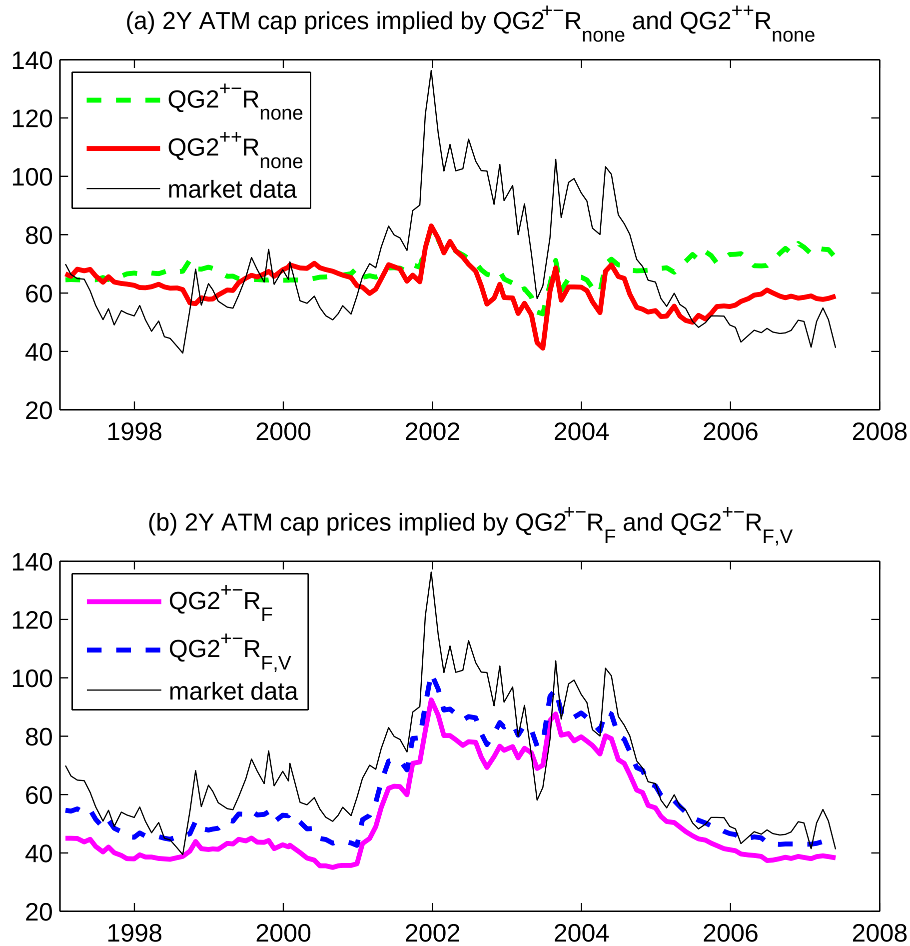

Figure 6: Two-year ATM cap prices implied by the QG2^{++}R$_{none}$, QG2^{+-}R$_{none}$, QG2^{+-}R$_{F}$, QG2^{+-}R$_{F,V}$ estimates.

Figure 6 compares the market price of the 2-year ATM cap with the QG2^{++}R$_{none}$-based and the QG2^{+-}R$_{none}$-based prices. The QG2^{+-}R$_{none}$ result is especially poor, being largely flat over the entire 1997-2007 period, but the QG2^{++}R$_{none}$ result is also very poor, showing little resemblance to the market data in the 1997-2000 period, and failing to capture much of the notable rise in 2001 and the decline in 2004-2006. These

(their Table 6). For example, the 3-factor CIR model-implied standard deviations for the 2-year and 5-year caps are 5.01 bp and 8.52 bp, respectively, while those of the market 2-year and 5-year caps are 15.8 bp and 48.88 bp, respectively.

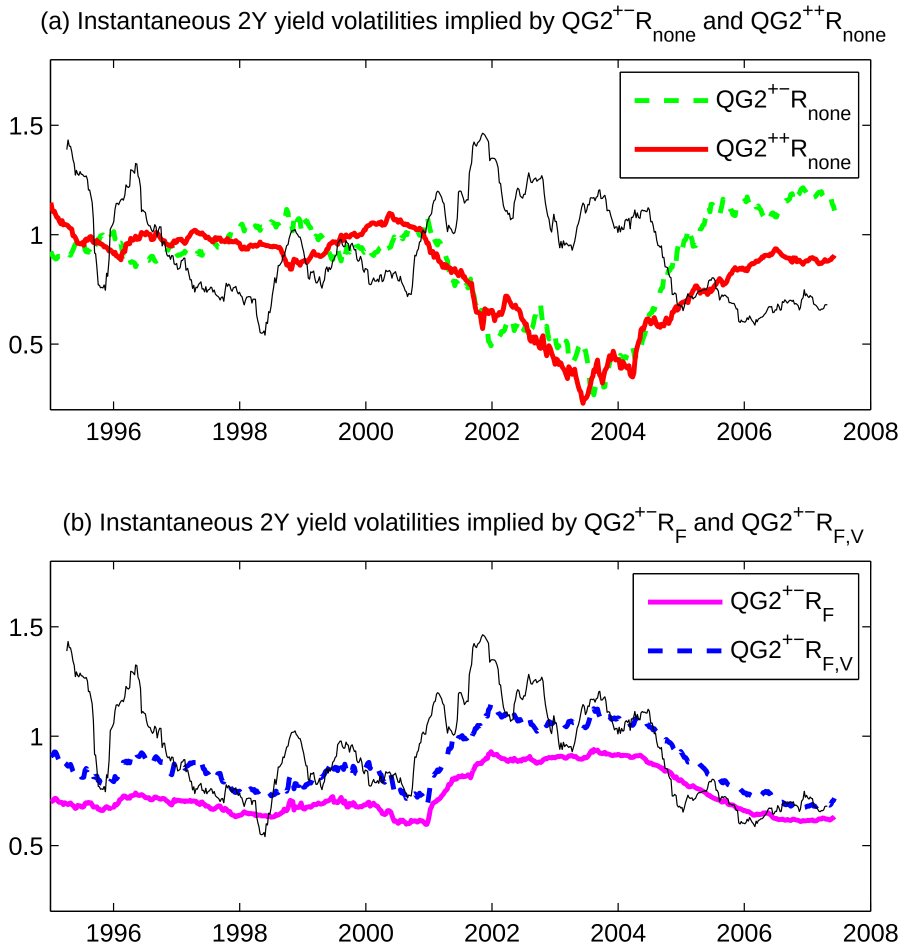

Figure 7: Instantaneous volatility of the 2-year yield ($v_{t,2y}$) implied by the QG2^{++}R$_{none}$, QG2^{+-}R$_{none}$, QG2^{+-}R$_F$, QG2^{+-}R$_{F,V}$ estimates. The empirical proxy based on the 6-month window rolling standard deviation is also shown for comparison (thin solid line).

results indicate a shortcoming in the estimated models that is far more basic than their inability to capture high-frequency variations in option prices.

Figure 7a, which plots the model-implied instantaneous volatility of the 2-year yield (i.e., $v_{t,2y}$) for QG2^{++}R$_{none}$ and QG2^{+-}R$_{none}$, explain *why* the relative pricing performance of these estimates is poor: the model-implied $v_{t,2y}$'s show little resemblance to the empirical proxy based on the rolling standard deviation. Indeed, in the 2001-2005 episode during which the policy rate was aggressively lowered and then raised back, the model-implied volatility and the empirical proxy behave in an almost opposite manner.

Even granting the imperfection of the empirical proxy, this clearly points to a very poor volatility modeling.

The QG2^{++}R$_{none}$-based and QG2^{+-}R$_{none}$-based $v_{t,2y}$'s in the 2001-05 period show a positive comovement with the policy rate. I have noted in Sec. 3.3 that this kind of behavior is expected for the CIR model and, to some extent, also for the QG model with positive-definite normalization. Interestingly, the estimated QG2^{+-} model also implies a poor volatility modeling, indicating that the non-positive-definite normalization did not help.[38] It turns out, in the QG2^{+-}R$_{none}$ estimation there is another local maximum, whose likelihood value is a little bit lower than the global maximum but has more reasonable properties.[39] This other maximum develops into the global maximum in the regularized ML estimations.

The regularized ML estimation of the QG2^{+-} model leads to a surprisingly good relative pricing performance. As can be seen in Table 2, the pricing errors (MAEs) for the 1-year cap are much smaller for the regularized estimations (QG2^{+-}R$_F$, QG2^{+-}R$_{F,V}$) than the standard estimation (QG2^{+-}R$_{none}$). MAEs, however, do not tell the whole story: if the market price and the model price had the same time variation but differed by a constant, one could still have significant MAEs. The variability and correlation numbers in Table 2 are thus also useful: the sample standard deviations of the cap prices implied by the QG2^{+-}R$_F$ and the QG2^{+-}R$_{F,V}$ estimates are much larger than those of the QG2^{+-}R$_{none}$ estimate and more in line with those of the market data. Furthermore, the correlation between the model and the market data for the 1-year cap, which was negative for QG2^{+-}R$_{none}$, has become positive; in the case of the 2-year and 5-year caps, it exceeds 90%.

Figure 6b illustrates the dramatic improvement in bond option pricing for the case of the 2-year ATM cap. The QG2^{+-} model (both R$_F$ and R$_{F,V}$) still misses a substantial part of the month-to-month variation in the market cap price and has difficulty in capturing the full extent of several of the sharp moves (e.g., in late 2001 and in mid 2003).[40] But it now does quite well in capturing the broader and slower aspects of cap price movements (e.g., decline in 1997, decline in 2000, rise in 2001, decline from mid-2004). The QG2^{+-}R$_F$-based 2-year cap price tends to be lower than the market price, but their variations are quite similar. Thus, one already obtains quite encouraging results with the R$_F$ regularization, i.e., even without the "volatility regularization" (25).

[38] I have also tried an alternative quasi-ML estimation based on the augmented state space method in Kim (2004) without the regularization conditions, but the results were qualitatively similar to the QG2^{+-}R$_{none}$ estimate reported here.

[39] The global optimum (in QG2^{+-}R$_{none}$ estimation) has a slightly smaller fitting errors for the 3-month yield than the other optimum (26 bp vs 31 bp)

[40] This failure to capture the high-frequency effects is also reflected in the different outcomes in a CDG(2002)-type regression with model-implied/market values: the regression of monthly change in the 2-year ATM cap price onto monthly change in yields using the QG2^{+-}R$_{F,V}$-model-implied values (cap price and yields) gives an R^2 of 0.99, while the same regression using market values (cap price and yields) gives a much lower R^2 of 0.55.

Underlying this more successful relative pricing is the fact that the regularized estimations of the QG2^{+-} model capture the volatility dynamics better. As can be seen in Figure 7b, the model-implied 2-year instantaneous yield volatility based on the QG2^{+-}R$_F$ estimate shows a variation similar to the empirical proxy (in its lower-frequency movements), while the QG2^{+-}R$_{F,V}$ estimate produces an even closer match.

Interestingly, even if the regularized estimations are used, the model with positive-definite normalization (QG2^{++}) still leads to a poor volatility modeling. This in turn results in a poor cap pricing performance: summary measures in Table 2 show that the cap pricing performance of the QG2^{++}R$_F$ and QG2^{++}R$_{F,V}$ estimates is not much better than the QG2^{++}R$_{none}$ results discussed earlier.

	QG2^{++}R$_{none}$	QG2^{++}R$_F$	QG2^{++}R$_{F,V}$	QG2^{+-}R$_{none}$	QG2^{+-}R$_F$	QG2^{+-}R$_{F,V}$
3-month	0.97	0.97	0.96	0.98	0.98	0.98
4-year	0.99	1.17	1.10	0.86	1.01	1.01
10-year	1.03	1.28	1.18	0.85	1.00	1.00

Table 3: Root-mean-square-errors (RMSE) of the 1-year-horizon yield forecasts implied by the model, divided by the RMSE of the 1-year-horizon yield forecasts based on the random walk hypothesis (i.e., $E_t(y_{t+u,\tau}) = y_{t,\tau}$).

The QG2^{++} model also seems to produce worse term premium estimates. For example, as can be seen in Table 3, the 1-year-ahead in-sample root-mean-square errors for yield forecasts based on the QG2^{++} model tend to be larger than those based on the QG2^{+-} model. The forecasting errors for longer-term yields are particularly large for the regularized estimations of the QG2^{++} model; for example, their 10-year yield forecast errors are substantially larger than a key benchmark based on the "random-walk" model. In fact, the excess return forecasting errors for the regularized QG2^{++} estimates (e.g., 1-year holding period return on 5-year bond over the 1-year bond) turn out to be even larger than those based on the expectations hypothesis. These results suggest that the QG2^{++} model is so poor that it cannot meet reasonable regularization conditions without generating problems elsewhere.[41]

As can be also seen in Table 3, the standard estimation of the QG2^{+-} model generates smaller yield forecasting errors than the regularized estimations of the QG2^{+-} model. It may be tempting to interpret this as indicating a superior *term premium modeling* in the standard estimation. However, even if one were interested only in term premia (and not in volatility), the QG2^{+-}R$_{none}$ estimate would be a problematic one. There are several different definitions of term premia, which are nonetheless related ("yield premium," "forward premium," and "return premium"). The version that is particularly

[41]It is also interesting to note that the estimates of Ψ_{12} for QG2^{++}R$_F$ and QG2^{++}R$_{F,V}$ have absolute values larger than 1. Thus the positive-definite condition is not satisfied in these estimations (reminiscent of the finding of $\phi < 0$ for multi-factor CIR models with the normalization (15) as in JKS (2003)). In the QG2^{++}R$_{none}$ estimation, $|\Psi_{12}|$ is smaller than 1, but it is close to 1.

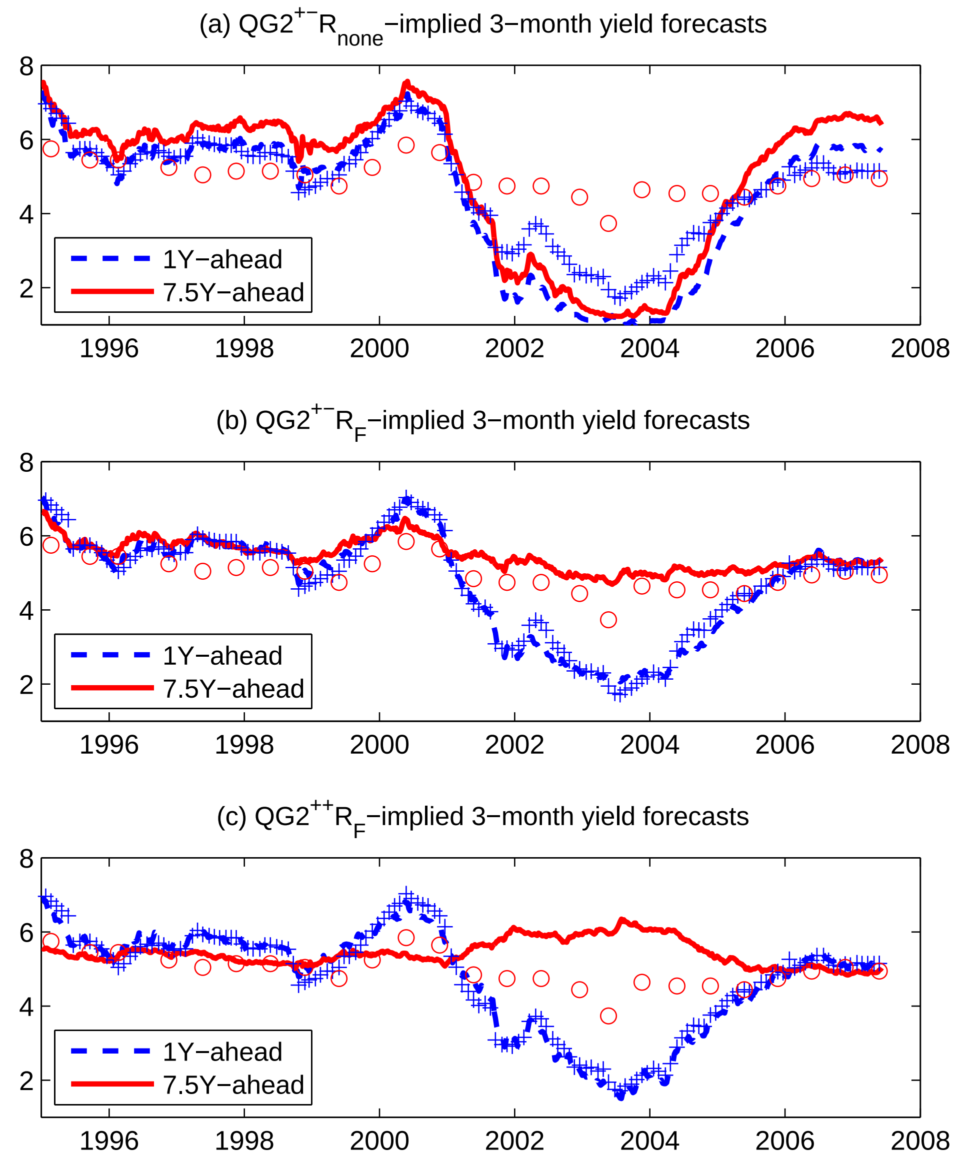

Figure 8: The 1-year-ahead and 7.5-year-ahead expectations of the 3-month yield, implied by the QG2^{+-}R$_{none}$, QG2^{+-}R$_F$, QG2^{++}R$_F$ estimates. The 1-year-ahead and the "long horizon" BCFF survey forecasts are also shown as "+" and "o" signs, respectively.

relevant to policy discussions is the forward premium. Because the forward premium is the forward rate minus the expected spot rate, estimating the forward premium is equivalent to estimating the expected spot rate across various horizons. Figure 8a shows the expected 3-month LIBOR rate for the "long" horizon (7.5-years) implied by the QG2^{+-}R$_{none}$ estimate. Note that when the short-term rate is at the historical lowest point in 2003, the model-implied 7.5-year-ahead expectation is also extremely low, being barely above the 1-year-ahead expectation. One can safely say that such an outcome is outside everyone's priors, i.e., the forward premium implied by QG2^{+-}R$_{none}$ is not sensible. Comparison with the long-horizon survey forecast (also shown in the same figure) helps illustrate this point. Unfortunately, this problem is hard to detect from the usual metrics on forecasting performance (e.g., Table 3), since such metrics are necessarily limited to relatively short horizons.[42]

Figure 8b plots the model-implied long-horizon expectation from the regularized estimation QG2^{+-}R$_F$, which can be seen to be more reasonable. This is partly by design, because the estimation was done with the condition (24) included. I've added the qualifier "partly", since the result also hinges on the model. Figure 8c shows the result based on the QG2^{++} model. Although the QG2^{+-}R$_F$ estimation and the QG2^{++}R$_F$ estimation have used the same regularization condition (24), the behaviors of the model-implied $E_t(y_{t+7.5y,3m})$ are quite different. For instance, in 2001-2005 the long-horizon forecast implied by the QG2^{++}R$_F$ estimate is elevated, contrary to intuition.

4.6. **Summary of Implications.** The key implications of various results presented so far can be summarized as follows:

(A) *The main reason for the poor relative pricing in the extant literature is likely the poor volatility modeling.* The implausible volatility dynamics implied by the QG2^{++}R$_{none}$ and the QG2^{+-}R$_{none}$ estimates (as seen in Figure 7a) suggests that the likely explanation of the relative pricing puzzle is a more basic one: problems in volatility modeling (rather than USV). USV might facilitate volatility modeling (by decoupling the volatility dynamics from the term structure dynamics), but that might not necessarily be the only way or the correct way to model volatility.

(B) *The good relative pricing performance of the 2-factor model casts doubts on the strong USV scenario.* QG2^{+-} is a relatively simple non-USV model, but it is capable of capturing key features of the bond option (interest rate cap) price variation, apart from the "high-frequency effects" which it misses. This result, along with the results based on much richer models like Joslin (2007), makes a strong case against the strong USV scenario. Thus, imposing the USV condition as in the $A_1(n)$ USV models of CDG (2002) and CDGJ (2004) does not seem advisable for investigating the issues discussed in Sec. 2.2 (possible relationship between term premium and interest rate uncertainty).

(C) *Normalization matters.* None of the estimations with the QG2^{++} model led to results that are comparable to those of the regularized QG2^{+-} model estimations. In

[42]Incidentally, Kim and Orphanides (2007) argue that searching for a model that generates the smallest in-sample yield forecasting errors may be misleading, due to the look-ahead bias and other problems.

short, positive-definite normalization (demanding the positive-definiteness of the model) is not promising for the past few decades' data.

(D) *Econometric technique matters.* Even with non-positive-definite normalization ($QG2^{+-}$ model), sensible results could not be obtained without some regularization to address the problems of the standard estimation. Had we used the standard ML estimation only, we would have ended up with discouraging conclusions about the relative pricing performance of the 2-factor QG model, similar to those of JKS (2003) about the multi-factor CIR model.

In the 3-factor QG model of LZ (2006), the points (C) and (D) may be harder to see, because the effects can be diluted amid the greater flexibility of the model and because it can be more difficult to separate out various sources of problems. However, that does not mean that the problems disappeared. In fact, the problem with the conventional estimation may be even more serious in higher-dimensional models, as the greater dimensionality means more room for things to go wrong.

4.7. **What the Present Paper Does Not Imply.** It may also help to discuss what this paper does *not* imply:

(A) *That two factors are sufficient.* As noted above, the 2-factor model ($QG2^{+-}R_F$, $QG2^{+-}R_{F,V}$) does miss some of the sharper movements in the ATM cap prices. The cross-sectional fit and term premia also leave room for improvement.[43]

It is a tall order for any model with only two sources of shocks to capture the complexity of the term structure and options data. Basically, two factors, x_{1t} and x_{2t}, are being asked to play many roles at the same time: current rates, expectation of future rates, volatilities, market price of the volatility risk, market price of other risks, and so on. This means that there is a lot of "tension" in the model in the sense of Duffee (2002).[44] However, tension should not necessarily be viewed as bad. After all, tension is the other side of the coin called "structure," and models are a means to impose structure on the data. One can always use a large model to fit data well, or introduce separate factors to explain different effects. But there is an accompanying danger of losing discipline or losing the predictive content of the model (due to the neglect of possible interrelationship between various effects). Too little tension may be as problematic as too much tension.

Incidentally, it is not clear that there *should* exist a 3-factor model that captures the full intricacy of the term structure dynamics (be it QG, affine, or other models). A frequently used argument for the 3-factor model is based on the finding that 3 factors in factor analysis (FA) or principal components analysis (PCA) explain something like 99% of the term structure movements. However, this is not convincing evidence, since the

[43]In addition, it does not do well in capturing the time variation of the convexity bias. It is straightforward to compute the convexity bias implied by the estimated QG model, but I shall not discuss detailed results on the convexity premium in this paper (to save space).

[44]Duffee (2002) makes the point that certain affine models impose too tight a relation between expected excess bond returns and volatility, i.e., there is too much "tension" between the two.

ability to reproduce the overall variability (unconditional variance) does not guarantee that the model can capture the "correct" internal dynamics.[45]

In short, the low-dimensional models will always have room for improvements; the key point is that there is no compelling reason to believe that not having the (strong) USV feature is the source of the deficiency of these models.

(B) *That the maximally flexible QG model is the best model.* To address basic questions raised by the USV debate and the relative pricing puzzle, this paper has taken an exploratory approach, searching for robust implications rather than the best possible fit of bond option prices. Accordingly, I have remained in the simple and familiar framework of the time-homogeneous diffusion models, viewing them as "effective models" for "not-too-high"-frequency data. Clearly, a more advanced model that accommodates features like jumps and inhomogeneities can be constructed, although a considerable care may be required in the process.[46]

Being an approximate statistical representation, the maximally flexible QG model can still be consistent with many other kinds of models. While it may be suitable for an exploratory study like this one, in some other applications (especially in the higher-dimensional case), other (more parsimonious) models might be more convenient.

(C) *That the regularization used here is the best possible one.* The regularization techniques used here should be also viewed in the light of the exploratory nature of this paper. Clearly, one can further develop the regularization procedure. For example, (if one is not restricting attention to the pure relative pricing) option prices can be also utilized in the estimation, as in Joslin (2007), who developed a fast method for computing bond option prices in affine models. A simple way to introduce the options information would be to use the relatively short-dated swaption-implied volatility as a proxy for instantaneous volatility in eq. (25). While very short option maturities (like 1-week maturity) have the problems that we have seen in Figure 3, the implied-volatility from swaptions with option maturity of 3-month is fairly smooth, and hence may be a better proxy for instantaneous volatility than the rolling standard deviation used in the present paper.[47]

(D) *That this paper solves all mysteries about USV.* The results in this paper clearly indicate that there is *a lot less* unspanned volatility than is presumed in the specific affine

[45]There is an interesting parallel in fluid dynamics/climate modeling, where the complexity of the problem (the presence of many degrees of freedom) often forces one to build low-dimensional factor models to approximate the reality: Aubry et al (1993) provides an example (a problem in fluid modeling) in which a low-dimensional representation of a more complicated model can capture 99.9995% of the variance but cannot reproduce the correct dynamics, as discussed in Crommelin and Majda (2004).

[46]The effects of anticipated jumps can be different from unanticipated jumps; the jump intensity may also depend on the general level of uncertainty.

[47]Because swaption volatilities are usually quoted as a relative (percentage) volatility based on Black's formula, one may have to convert them to an "absolute" volatility. Note that near-term option-implied volatilities tend to be somewhat higher than the realized volatilities empirically, possibly due to a volatility risk premium. This can be accommodated by adding a constant term to the right-hand side of eq. (25).

USV models of CDG (2002) and CDGJ (2004). That said, there *is* a relatively high-frequency unspanned variation in bond option prices, as indicated by CDG (2002)'s regression. This paper is silent about the sources of this variation: how much it has to do with the complexity of high-frequency yield volatility dynamics (highlighted by AB (2006)'s regression),[48] how much is due to the limitation of low-dimensional term structure models (like the 2-factor QG model), and whether there are special mechanisms in option markets (special hedging demand, option market microstructure effects, etc.) that would explain it.

5. Concluding remarks

Collin-Dufresne and Goldstein (2002)'s proposal of the unspanned stochastic volatility for the bond market brought into sharp focus a long-standing difficulty in time-consistent term structure modeling, namely the poor relative pricing between bonds and bond options: attempts to use no-arbitrage term structure models to price bond options with bond yields data have not reported much success in reproducing the behavior of the market option prices.

The USV scenario leads to an interesting take on this issue: relative pricing fails because it is meant to fail. (The bond market is incomplete; hence yields information is not sufficient to determine bond option prices.) This answer is not complete, however, as it opens up the following questions: *How incomplete* is the bond market? *To what extent* should relative pricing fail?

To address these questions, in this paper I have introduced a distinction between the "weak USV" and the "strong USV" scenarios, and reexamined existing empirical evidence and also presented new evidence. I have argued that although some kind of weak USV effect seems to be present, the evidence for a more fundamental USV effect (strong USV) is lacking. I find that a 2-factor QG model that gives up the requirement of the positive-definiteness of nominal interest rates (i.e., model with non-positive-definite normalization) can capture the variation of bond option prices fairly well, largely obviating the need for a special USV factor to explain their (non-high-frequency) variation. I have also presented much poorer results from the model with positive-definite normalization; they may help explain the poor relative pricing results in some of the existing studies, and may have useful implications for the specification (normalization) of affine and QG models in future studies.

The present paper is also of methodological interest. The existing literature has largely assumed that conventional econometric techniques work well in term structure estimation settings and that the global optimum of the criterion function (likelihood function) tells all that one needs to know about the model. This paper has shown that even at the level of the 2-factor model the empirical analysis can be quite complicated, and

[48]Inhomogeneities in volatility dynamics due to scheduled macro data releases might be important for short-dated options but likely not for caps and swaptions with long maturities.

it provides a concrete example in which the conventional estimation leads to misleading results. One can only guess how much more challenging the empirical analysis of higher-dimensional models would be. I have proposed to address these problems from the perspective of "ill-posed inverse problems" (introducing regularization conditions for the estimation), but more progress on the implementation front is clearly desirable.

References

[1] Ahn, D.-H., R. Dittmar, and A. R. Gallant (2002), "Quadratic term structure models: theory and evidence", *Review of Financial Studies*, 15, pp243-88.

[2] Almeida, C., J. Graveline, and S. Joslin (2006), "Do options contain information about excess bond returns?", working paper.

[3] Andersen, T. and L. Benzoni (2006), "Do bonds span volatility risk in the U.S. Treasury market? A specification test for affine term structure models", working paper.

[4] Anderson, P. W. (1972), "More is different", *Science*, 177, pp393-6.

[5] Andrews, D. W. K. (1991), "Asymptotic normality of series estimators for nonparameteric and semiparametric regression models," *Econometrica*, 59, p307-45.

[6] Aubry, N., W.-Y. Lian, and E. S. Titi (1993), "Preserving symmetries in the proper orthogonal decomposition", *SIAM Journal of Scientific Computation*, 14, pp483-505.

[7] Backus, D., S. Foresi, A. Mozumdar, and L. Wu (2001), "Predictable changes in yields and forward rates", *Journal of Financial Economics*, 59, p281.

[8] Backus, D. and J. Wright (2007), "Cracking the conundrum," *Brookings Papers on Economic Activity*, 1, pp293-329.

[9] Bikbov, R. and M. Chernov (2005), "Term structure and volatility: lessons from the eurodollar futures and options", working paper.

[10] Black, F. (1995), "Interest rates as options", *Journal of Finance*, 50, 1371-76.

[11] Burghardt, G. and W. Hoskins (1995), "A question of bias", *RISK*, March issue.

[12] Cheridito, P., D. Filipovic, and R. L. Kimmel (2007), "Market price of risk specification for affine models: theory and evidence", *Journal of Financial Economics*, 83, pp123-70.

[13] Collin-Dufresne, P. and R. Goldstein (2002), "Do bonds span the fixed income markets?: Theory and evidence for unspanned stochastic volatility", *Journal of Finance*57, p1685-1729.

[14] Collin-Dufresne, R. Goldstein, and C. Jones (2004), "Can interest rate volatility be extracted from the cross section of bond yields? An investigation of unspanned stochastic volatility", NBER working paper.

[15] Crommelin, D. T. and A. J. Majda (2004), "Strategies for model reduction: Comparing different optimal bases", *Journal of Atmospheric Sciences*, 61, pp2206-17.

[16] Cutler, D., J. M. Poterba, and L. H. Summers (1989), "What moves stock prices", *Journal of Portfolio Management*, 15, pp4-12.

[17] Dai, Q. and K. J. Singleton (2000), "Specification analysis of affine term structure models", *Journal of Finance*, 55, pp1943-78.

[18] Duffee, G. R. (2002), "Term premia and interest rate forecasts in affine models," *Journal of Finance*, 57, pp405-443.

[19] Duffie, D. and K. Singleton (1997), "An econometric model of the term structure of interest-rate swap yields", *Journal of Finance*, 52, 1287-1321.

[20] Fan, R., A. Gupta, and P. Ritchken (2003), "Hedging in the possible presence of unspanned volatility: Evidence from swaption markets," *Journal of Finance*, 58, pp2219-48.

[21] Fleming, M. J., and E. Remolona (1997), "What moves the bond market", *Federal Reserve Bank of New York Economic Policy Review*, 3, pp31-50.
[22] Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.
[23] Hathaway, R. J. (1985), "A constrained maximum likelihood estimation for normal-mixture distributions", *Annals of Statistics*, 13, pp785-800.
[24] Jagannathan, R., A. Kaplin and S. Sun (2003), "An evaluation of multi-factor CIR model using LIBOR, swap rates, and cap and swaption prices," *Journal of Econometrics*, 116, pp113-46.
[25] Jones, C. M., O. Lamont, and R. L. Lumsdaine (1998), "Macroeconomic news and bond market volatility", *Journal of Financial Economics*, 47, p315-37.
[26] Joslin, S. (2007), "Pricing and hedging volatility risk in fixed income markets", working paper.
[27] Kim, D. H. (2004), "Time-varying risk and return in the quadratic-Gaussian model of the term structure", working paper.
[28] Kim, D. H. (2007), "Challenges in Macro-Finance Modeling", working paper.
[29] Kim, D. H. and A. Orphanides (2005), "Term structure estimation with survey data on interest rate forecasts", FEDS working paper.
[30] Kim, D. H. and A. Orphanides (2007), "Bond market term premium: what is it, and how can we measure it?" *BIS Quarterly Review*, June issue.
[31] Leippold, M. and L. Wu (2002), "Asset pricing under the quadratic class", *Journal of Financial and Quantitative Analysis*, 37, pp271-297.
[32] Leippold, M. and L. Wu (2003), "Design and estimation of quadratic term structure models", *European Finance Review*, 7, 47-73.
[33] Li, H. and F. Zhao (2006), "Unspanned stochastic volatility: evidence from hedging interest rate derivatives", *Journal of Finance*, 61, 341-378.
[34] Merton, R. (1980), "On estimating the expected return on the market: an exploratory analysis", *Journal of Financial Economics*, 8, pp323-361.
[35] Moessner, R., and W. Nelson (2007), "Central bank policy rate guidance and financial market functioning", working paper.
[36] O'Sullivan, F. (1986), "A statistical perspective on ill-posed inverse problems," *Statistical Science*, 1, pp502-18.
[37] Press, W., S. Teukolsky, W. Vetterling, B. Flannery (2007), *Numerical Recipes: the art of scientific computing* (3rd ed.), Cambridge University Press.
[38] Roll, R. (1988) "R^2", *Journal of Finance*, 43, pp541-66.