# Interest Rate Volatility and No-Arbitrage Term Structure Models*

Scott Joslin† Anh Le‡

November 1, 2012

PRELIMINARY
COMMENTS WELCOME

**Abstract**

Forecasting volatility of interest rates remains a challenge in finance. An important aspect of any dynamic model of volatility is the fact that volatility is a positive process, not only with respect to the historical measure, but also with respect to the risk neutral measure. As a consequence, risk neutral forecasts of volatility must also always remain positive. One way for this admissibility condition to hold is for volatility to represent an autonomous process under the risk neutral measure. In contrast to the historical time series, the cross section of bond yield provides very precise information about risk neutral forecasts and therefore strong guidance to which combinations of yields are autonomous. We conclude that in the stochastic volatility setting, the no arbitrage assumption provides strong over-identifying constraints which, when the model is correctly specified, improve inference on the volatility instrument.

*We thank Caio Almeida, Francisco Barillas, Greg Duffee, Michael Gallmeyer, Kenneth Singleton, Anders Trolle and seminar participants at the Banco de España - Bank of Canada Workshop on Advances in Fixed Income Modeling, Emory Goizueta, EPFL/Lausanne, Federal Reserve Bank of San Francisco, and Gerzensee Asset Pricing Meetings (evening sessions) for helpful comments.

†University of Southern California, Marshall School of Business, sjoslin@usc.edu

‡Kenan-Flagler Business School, University of North Carolina at Chapel Hill, anh_le@unc.edu

# 1 Introduction

Forecasting volatility of interest rates remains an important challenge in finance.[1] A rich body of literature has shown that the volatility of the yield curve is, at least partially, related to the shape of the yield curve. For example, volatility of interest rates is usually high when interest rates are high and when the yield curve exhibits higher curvature (see Cox, Ingersoll, and Ross (1985), Litterman, Scheinkman, and Weiss (1991), and Longstaff and Schwartz (1992), among others). This suggests that the shape of the yield curve is a potentially useful instrument for forecasting volatility. In this paper, we assess whether standard no arbitrage affine term structure models can improve the precision of estimation of volatility instruments.

We find that instruments for a positive volatility process are strongly identified by the cross section of yields. The key to our results is that volatility must be a positive process not only under the historical measure but also under the risk neutral measure. Any linear combination of yields which is forecasted to be negative (or has a positive probability of achieving a negative value) under the risk neutral measure *cannot* be an instrument for volatility and maintain positivity. We show that since the risk neutral distribution is estimated much more precisely than the historical distribution, these conditions precisely identify which instruments maintain positive volatility. Morever, risk neutral forecasts of yields are largely invariant to any volatility considerations (see, e.g. Campbell (1986) or Joslin (2010).) This implies the striking conclusion that at a Gaussian term structure model – which has constant volatility – can reveal which instruments would be admissible for a stochastic volatility model.

To develop some intuition for our results, consider a simple time series model of the yield curve. Consider a standard vector autoregression of the principle components of the yield curve (level, slope and curvature), denoted $\mathcal{P}$:

$$\Delta \mathcal{P}_t = K_0 + K_1 \mathcal{P}_t + \epsilon_{t+1}. \tag{1}$$

In order to capture time-variation in volatility of the level, we can project squared changes in the level of interest rates onto the principal components along the lines of Campbell (1987) and Harvey (1989) through the regression

$$(\Delta \mathcal{P}_{t,1})^2 = \alpha + \beta \cdot \mathcal{P}_t + e_t. \tag{2}$$

This simple instrumentation procedure suggests that the variance of the level factor could be instrumented by $V_t \equiv \alpha + \beta \cdot \mathcal{P}_t$.[2] Alternatively, one could use other proxies for conditional variance, such as from realized volatility or time-series models.

Assuming that volatilities of all the principle components have a one-factor structure, we could then generate a fully dynamic affine model of the term structure by modifying the VAR in (1) so that the conditional covariance of the innovation, $\epsilon_{t+1}$, is affine in $\alpha + \beta \cdot \mathcal{P}_t$. However, this process may not be well-defined since volatility may become negative. Following Dai and Singleton (2000), we will refer to any condition required for a well-defined positive

[1] See Poon and Granger (2003) or Andersen and Benzoni (2008) for recent surveys of volatility forecasting.

[2] In principle, the projection in (2) could result in negative variances. Ensuring positivity here represents an additional constraint.

volatility process as an admissibility condition. A sufficient condition for volatility to be a positive process is that the variance defines an autonomous (discrete-time) Cox Ingersoll Ross (CIR) process. When there is a single volatility factor and the other risk factors are conditionally Gaussian, this condition is also necessary.[3] Among other things, this means that the conditional mean of $\Delta V_t$, $\beta'(K_0 + K_1\mathcal{P}_t)$, should depend on only $V_t$. This happens only when $\beta' K_1$ is a multiple of $\beta'$. That is, $\beta'$ is a left-eigenvector of $K_1$.

Requiring that this time series model rule out arbitrage opportunities introduces a starker version of these over-identifying restrictions. The condition of no arbitrage gives rise to the existence of a risk neutral measure $\mathbb{Q}$. Joslin, Singleton, and Zhu (2010) show that the risk-neutral distribution can be characterized through the equation

$$\Delta\mathcal{P}_t = K_0^{\mathbb{Q}} + K_1^{\mathbb{Q}}\mathcal{P}_t + \epsilon_{t+1}^{\mathbb{Q}}, \tag{3}$$

where $(K_0^{\mathbb{Q}}, K_1^{\mathbb{Q}})$ are appropriately constrained to maintain the internal consistency of no arbitrage. Admissibility requires that volatility must be autonomous under $\mathbb{Q}$. Thus $\beta'$ must also be a left eigenvector of $K_1^{\mathbb{Q}}$.

The fact that admissibility requires that volatility is an autonomous process under both the historical and risk neutral measures represents a tension between the conditional first and second moments given by (1), (2), and (3). Unless the estimate of $\beta$ is, simultaneously, a left eigenvector of $K_1$ and $K_1^{\mathbb{Q}}$, a joint estimation procedure designed to produce an admissible process will necessarily trade off the fit of these three equations. Naturally, equations with the least statistical errors will be given the most weight. As is commonly observed, $K_1^{\mathbb{Q}}$ can be estimated very precisely, particularly relative to parameters underlying (1) and (2). Intuitively, although we observe only one historical time series, each observation of the yield curve effectively represents a term structure of risk neutral expectations of $\mathcal{P}$. Given a precise estimate of $K_1^{\mathbb{Q}}$, the left eigenvectors are set and thus so too are the potential volatility instruments.

To shed further light on this trade-off, we compare the estimate of $K_1^{\mathbb{Q}}$ implied by the above model to two alternative estimates: (a) one implied by a Gaussian term structure model – where the volatility equation (2) is left out of the estimation; and (b) a model-free estimate using the regression technique in Joslin (2011b) – where both equations (1) and (2) are left out of the estimation. Strikingly, the estimates are practically identical across the three different estimation methods – and this is the case for all choices of model specifications and sample periods that we consider. These results indicate that $K_1^{\mathbb{Q}}$ are pinned down very strongly effectively by the cross-sectional information *alone* from equation (3). As a consequence, the no arbitrage restrictions allow very reliable inferences for the admissible volatility instruments. When the no arbitrage restrictions are ignored (i.e. equation (3) is dropped from the estimation) we demonstrate that the admissible volatility instruments implied by the unconstrained estimates of $K_1$ from (1) are estimated much less precisely.

[3]More precisely, this result is for the continuous time $A_1(N)$ class of models in the notation of Dai and Singleton (2000). Analogous conditions maintain in the discrete time stochastic volatility models of Bansal and Shaliastovich (2009) and Le, Singleton, and Dai (2010).

Our results help clarify the relationship between volatility instruments extracted from the cross-section of bond yields documented by several recent studies. For example, Collin-Dufresne, Goldstein, and Jones (2009) find an extracted volatility factor from the cross-section of yields through a no arbitrage model to be negatively correlated with model-free estimates. Jacobs and Karoui (2009) in contrast generally find volatility extracted from affine models are generally positively related though in some cases they also find a negative correlation. Almeida, Graveline, and Joslin (2011) also find a positive relationship. Our results show that for the $A_1(N)$ class of models, the cross-section of bonds will reveal up to $N$ linear combinations of yields (given by the left eigenvectors of $K_1^{\mathbb{Q}}$) that can serve as instruments for volatility. The no arbitrage structure then essentially implies *nothing* more for the properties of volatility beyond the assumed one factor structure and the admissibility conditions. To the extent that the instrument obtained from $K_1^{\mathbb{Q}}$ is optimal, the model implied-volatility will match the analogous results obtained from a model without the no arbitrage structure.

Our results clearly spell out aspects of model specifications that may or may not have any significant bearing on the model's volatility outputs. For example, within the $A_1(N)$ class of models, different specifications of the market prices of risks are unlikely to affect the identification of the volatility factor as long as the risk-neutral dynamics is maximally flexible. While this may seem at odds with numerous studies (e.g. Dai and Singleton (2000) and Duffee (2002)) that demonstrate the importance of the market prices of risks specification in matching yields dynamics, it can be understood in light of our results. Intuitively, since the market prices of risks serve as the linkage between the physical and risk-neutral measures, and since the risk neutral dynamics is very strongly identified, different forms of the market prices of risks are most likely to result in different estimates for the physical dynamics while leaving estimates of $K_1^{\mathbb{Q}}$ essentially intact. This in turn implies that identifications of the volatility factor are likely identical across these models. This explains the almost identical performances by the volatility estimates implied by the completely affine and essentially affine $A_1(3)$ models as reported in Jacobs and Karoui (2009).

Our results also add to the recent discussion that suggests that no arbitrage restrictions are completely or nearly irrelevant for the estimation of Gaussian dynamic term structure models (DTSM). See, for example, Duffee (2011), Joslin, Singleton, and Zhu (2010), and Joslin, Le, and Singleton (2012). Left open by these studies is the question of whether the no arbitrage restrictions are useful in the estimation of DTSMs with stochastic volatility. Our results show that the answer to this question is a resounding yes! The cross-section of yields strongly identifies $K_1^{\mathbb{Q}}$ and therefore strongly identifies the admissible instruments for volatility. Key to the linkage between $K_1^{\mathbb{Q}}$ and the identification of the volatility factor is the admissibility conditions that require the volatility factor be autonomous. To the extent that the admissibility conditions hold in the true data generating process, these over-identifying restrictions will be very useful for inferring the true process of the yield curve. The admissibility conditions though are motivated not by economic considerations so this seems a potentially dubious assumption. In Section 5, we consider alternative non-affine (but nearly-affine) models which relax the admissibility constraints.

The rest of the paper is organized as follows. In Section 2, we lay out the general setup of

the term structure models with stochastic volatility that we subsequently consider. Section 3 provides intuitive examples as to why the cross-sectional information is likely to provide very strong identification for volatility instruments. Section 4 reports our empirical analysis and Section 5 concludes.

To fix notation, suppose that a DTSM is to be evaluated using a set of $J$ yields $y_t = (y_t^{m_1}, \ldots, y_t^{m_J})'$ with maturities $(m_1, \ldots, m_J)$ in periods and with $J \geq N$, where $N$ is the number of pricing factors. To be consistent with our empirical work, we fix the period length to be one month. We introduce a fixed, full-rank matrix of portfolio weights $W \in \mathbb{R}^{J \times J}$ and define the "portfolios" of yields $\mathcal{P}_t = W y_t$. The modeler's choice of $W$ will determine which portfolios of yields enter the DTSM as risk factors and which additional portfolios are used in estimation. Throughout, we assume a flat prior on the initial observed data.

# 2 Stochastic Volatility Term Structure Models

This section gives an overview of the stochastic volatility term structure models that we consider. As we elaborate, a key consideration is the admissibility conditions required to maintain a positive volatility process under both the historical and risk neutral measures. The conditions for a positive volatility process can be considered even without a formal term structure model by considering a factor time series model with stochastic volatility. This formulation is also of interest in its own right since it allows us to assess the role of the no arbitrage assumption in stochastic volatility models. For simplicity, we focus in the main text on the case of a single volatility factor under a continuous time setup; modifications for discrete time processes and more technical details are described in Appendix A.

## 2.1 General admissibility conditions in latent factor models

We first review the conditions required for a well-defined positive volatility process within a multi-factor setting. Following Dai and Singleton (2000), hereafter DS, we refer to these conditions as admissibility conditions. Recall the $N$-factor $A_1(N)$ process of DS. This process has an $N$-dimensional state variable composed of a single volatility factor, $V_t$, and $(N-1)$ conditionally Gaussian state variables, $X_t$. The state variable $Z_t = (V_t, X_t')'$ follows the Itô diffusion

$$d \begin{bmatrix} V_t \\ X_t \end{bmatrix} = \mu_{Z,t} dt + \Sigma_{Z,t} dB_t^{\mathbb{P}}, \tag{4}$$

where

$$\mu_{Z,t} = \begin{bmatrix} K_{0V} \\ K_{0X} \end{bmatrix} + \begin{bmatrix} K_{1V} & K_{1VX} \\ K_{1XV} & K_{1X} \end{bmatrix} \begin{bmatrix} V_t \\ X_t \end{bmatrix}, \text{ and } \Sigma_{Z,t} \Sigma_{Z,t}' = \Sigma_{0Z} + \Sigma_{1Z} V_t, \tag{5}$$

and $B_t^{\mathbb{P}}$ is a standard $N$-dimensional Brownian motion under the historical measure, $\mathbb{P}$. Duffie, Filipovic, and Schachermayer (2003) show that this is the most general affine process on $\mathbb{R}^+ \times \mathbb{R}^{N-1}$.

In order to ensure that the volatility factor, $V_t$, remains positive, we need that when $V_t$ is zero (a) the expected change of $V_t$ is non-negative and (b) the volatility of $V_t$ becomes zero.

Otherwise there is a positive probability that $V_t$ will become negative. Imposing additionally the Feller condition for boundary non-attainment, our admissibility conditions are then

$$K_{1VX} = 0, \;\; \Sigma_{0Z,11} = 0, \;\; \text{and} \;\; K_{0V} \;\; \geq \tfrac{1}{2}\Sigma_{1Z,11}. \tag{6}$$

A consequence of these conditions is that volatility must follow an autonomous process under $\mathbb{P}$ since the conditional mean and variance of $V_t$ depends only on $V_t$ and not on $X_t$. We now show how to embed the $A_1(N)$ specification into generic term structure models and re-interpret these admissibility constraints in terms of conditions on the volatility instruments.

## 2.2 An $A_1(N)$ model without no arbitrage restrictions

We can extend the latent factor model of (4–5) to a factor model for yields by appending the factor equation

$$y_t = A_Z + B_Z Z_t, \tag{7}$$

where $(A_Z, B_Z)$ are conformable matrices. Given the parameters of the model, we can replace the unobservable state variable with observed yields through (7).

Following Joslin, Singleton, and Zhu (2010), hereafter JSZ, we can identify the model by observing that equation (7) implies $\mathcal{P}_t \equiv W y_t = (W A_Z) + (W B_Z) Z_t$. Assuming $W B_Z$ is full rank[4], this in turn allows us to replace the latent state variable $Z_t$ with $\mathcal{P}_t$.

$$d\mathcal{P}_t = (K_0 + K_1 \mathcal{P}_t)dt + \sqrt{\Sigma_0 + \Sigma_1 V_t} dB_t^{\mathbb{P}}, \tag{8}$$

where $V_t = \alpha + \beta \cdot \mathcal{P}_t$ serves as the volatility instrument. Individual yields are then related to the yield factors through

$$y_t = A + B\mathcal{P}_t. \tag{9}$$

For identification of the parameters, we also impose that $WA = 0$ and $WB = I_N$, as in JSZ. The admissibility conditions (6) map into:

$$\beta' \Sigma_0 \beta = 0, \;\; \text{and} \;\; \beta' K_0 \geq \frac{1}{2}\beta' \Sigma_1 \beta, \;\; \text{and} \;\; \beta' K_1 = c\beta', \tag{10}$$

where $c$ is an arbitrary constant. The last admissibility condition can be restated as the requirement that $\beta'$ be a left eigenvector of $K_1$, which is required so that $V_t$ is an autonomous process under $\mathbb{P}$.

We will denote the stochastic volatility model in (8–9) by $F_1(N)$. The model is parameterized by $\Theta_F \equiv (K_0, K_1, \Sigma_0, \Sigma_1, \alpha, \beta, A, B)$ which is subject to the conditions in (10). Our development shows that the $F_1(N)$ model is the most general factor model with an underlying affine $A_1(N)$ state variable.

[4]This is slightly overidentifying. For details, see JSZ.

## 2.3 No arbitrage term structure models with stochastic volatility

The $A_1(N)$ no arbitrage short rate model of DS represents a special case of the $F_1(N)$ model. That is, when one imposes additional constraints to the parameter vector $\Theta_F$ one will obtain a model consistent with no arbitrage. In this section, we first review the standard formulation of the $A_1(N)$ no arbitrage model. We then focus on the the effect of no arbitrage on the volatility instrument through the restriction it implies on the loadings parameter $\beta$.

### The latent factor specification of the $A_1(N)$ model

We now consider affine short rate models which take a latent variable $Z_t$ with dynamics given by (4-5) and append a short rate which is affine in a latent state variable. We consider the general market prices of risk of Cheridito, Filipovic, and Kimmel (2007). Joslin (2011a) shows that any such latent state term structure model can be drift normalized under $\mathbb{Q}$ so that we have the short rate equation

$$r_t = r_\infty + \rho_V V_t + \iota \cdot X_t, \tag{11}$$

where $\iota$ denotes a vector of ones, $\rho_V$ is either +1 or -1, and the risk-neutral dynamics of $Z_t$ are given by

$$dZ_t = \left( \begin{bmatrix} K_{0V}^{\mathbb{Q}} \\ 0_{N-1\times 1} \end{bmatrix} + \begin{bmatrix} \lambda_V^{\mathbb{Q}} & 0_{1\times N-1} \\ 0_{N-1\times 1} & diag(\lambda_X^{\mathbb{Q}}) \end{bmatrix} Z_t \right) dt + \sqrt{\Sigma_{0Z} + \Sigma_{1Z} V_t} dB_t^{\mathbb{Q}}, \tag{12}$$

where $\lambda_X^{\mathbb{Q}}$ is ordered. To ensure the absence of arbitrage, we impose the Feller condition that $K_{0V}^{\mathbb{Q}} \geq \frac{1}{2}\Sigma_{1Z,11}$.

No arbitrage pricing then allows us to obtain the no arbitrage loadings that replaces the unconstrained version of (7) in the $F_1(N)$ model:

$$y_t = A_Z^{\mathbb{Q}} + B_Z^{\mathbb{Q}} Z_t, \tag{13}$$

The no arbitrage latent factor model can then be parameterized by

$$\Theta_{NA,Z} = (K_{0Z}, K_{1Z}, \Sigma_{0Z}, \Sigma_{1Z}, r_\infty, \rho_V, K_{0V}^{\mathbb{Q}}, \lambda^{\mathbb{Q}}).$$

Here, we use $\lambda^{\mathbb{Q}}$ to denote the vector $(\lambda_V^{\mathbb{Q}}, \lambda_X^{\mathbb{Q}})$.

### Implications of the no arbitrage restrictions for the factor model

Ideally, we would like to characterize the no arbitrage model as restrictions on the parameter vector $\Theta_F$ in the $F_1(N)$ model. The $\Theta_{NA,Z}$ parameterization only indirectly accomplishes this goal. In JSZ, they were able to succinctly characterize the parameter restrictions of the no arbitrage model as a special of the factor VAR model. In their case, essentially the only restriction was that the factor loadings ($B$) belongs to an $N$-parameter family characterized by the eigenvalues of the $\mathbb{Q}$ feedback matrix. In our current context of a

stochastic volatility models, such a simple characterization is not possible because changing the volatility parameters $\Sigma_{1Z}$ affects not only the volatility structure but also the loadings. This is because higher volatility implies higher convexity and thus higher bond prices or lower yields. The fact that the $\Sigma_{1Z}$ shows up both in volatility and in yields complicates a clean characterization of the restrictions on $\Theta_F$ that no arbitrage implies.

For now, let us focus on a simpler question: what is the relationship between the loadings $B$ in (9) and the volatility instrument loadings $\beta$ when the no arbitrage restrictions are applied? In the general $F_1(N)$ model, both are fully flexible and independent. Surprisingly, in the no arbitrage model the loadings $B$ essentially fix the volatility instrument. We can see this in two ways.

First, observe that in (14), the loadings $B_Z^{\mathbb{Q}}$ depend only on $(\rho_V, \lambda^{\mathbb{Q}}, \Sigma_{1Z})$. Since $\rho_V$ is a normalization factor, we ignore this. $\Sigma_{1Z}$ will have an effect on how the loadings due to the stochastic convexity effects. However, these affects will generally be small and will be dominated by variation in risk neutral expectations which will be determined by $\lambda^{\mathbb{Q}}$. Thus if we approximate (14) by replacing

$$y_t \approx A_Z^{\mathbb{Q}} + \hat{B}_Z^{\mathbb{Q}}(\lambda^{\mathbb{Q}}) Z_t, \tag{14}$$

where $\hat{B}_Z^{\mathbb{Q}}(\lambda^{\mathbb{Q}})$ is a functional approximation for the true loading function that depends only on $(\rho_V, \lambda^{\mathbb{Q}}, \Sigma_{1Z})$. For example, we could take $\hat{B}_Z^{\mathbb{Q}}(\lambda^{\mathbb{Q}})$ to be the loadings from a Gaussian term structure model which does not have a stochastic volatility effect. Multiplying by $W$ we can solve this approximation to obtain that

$$Z_t \approx (W \hat{B}_Z^{\mathbb{Q}})^{-1}(\mathcal{P}_t - W A_Z^{\mathbb{Q}}). \tag{15}$$

Now, the volatility factor, $V_t$, is the first component of $Z_t$ and so we see that the volatility instrument $\beta$ must be given by the first entry of $(W \hat{B}_Z^{\mathbb{Q}}(\lambda^{\mathbb{Q}}))^{-1}$, which depends only on the parameter $\lambda^{\mathbb{Q}}$. Thus, given a set of no arbitrage loadings $B$, one could extract an associated $\lambda^{\mathbb{Q}}$ (according to this approximation) and this completely fixes the volatility loading $\beta$ (up to scaling).

A more careful analysis of these steps indicates a slightly stronger relationship. In fact, $\beta$ is completely determined by $\lambda_X^{\mathbb{Q}}$. This can be see as follows. Let us write $W_1$ for the first row of $W$ and $W_2$ for matrix consisting of rows 2 to row $N$ of $W$. Similarly, let $\mathcal{P}_t^1$ denote the first entry of $\mathcal{P}_t$ and $\mathcal{P}_t^{(2)}$ denote entries two to $N$ of $\mathcal{P}_t$. Finally, let $B_V^{\mathbb{Q}}$ denote the first column of $B_Z^{\mathbb{Q}}$ and $B_X^{\mathbb{Q}}$ the remaining columns of $B_Z^{\mathbb{Q}}$. Then we have, ignoring constants,

$$\begin{aligned}
\mathcal{P}_t^{(1)} &= W_1 B_V^{\mathbb{Q}} V_t + W_1 B_X^{\mathbb{Q}} X_t \\
\mathcal{P}_t^{(2)} &= W_2 B_V^{\mathbb{Q}} V_t + W_2 B_X^{\mathbb{Q}} X_t.
\end{aligned}$$

This gives two equations and two unknowns, so we can subtract $W_1 B_X^{\mathbb{Q}}(W_2 B_X^{\mathbb{Q}})^{-1}$ times the second equation from the first equation to eliminate $X_t$ and obtain

$$\mathcal{P}_t^{(1)} - W_1 B_X^{\mathbb{Q}}(W_2 B_X^{\mathbb{Q}})^{-1} \mathcal{P}_t^{(2)} = c V_t,$$

where $c$ is a constant. This shows directly that the no arbitrage imposes the restriction

$$\beta = (1, -W_1 B_X^{\mathbb{Q}} (W_2 B_X^{\mathbb{Q}})^{-1}).$$

We see that the volatility instrument is in fact determined entirely by $\lambda_X^{\mathbb{Q}}$.

Thus our overall conclusion is that the no arbitrage restriction on the $F_1(N)$ model imposes a very tight link between the cross-sectional loadings $B_{\mathcal{P}}$ and the volatility instrument $\beta$.

# 3 Positive volatility under the risk neutral measure

In this section, we provide intuitive examples why the no-arbitrage restrictions can strongly pin down the volatility factor, without reference to the time series properties of the data.

We have seen in Section 2.2 that in order to have a well-defined autonomous volatility process we must have the condition (10). This condition can be restated as that $\beta'$ is a left eigenvector of $K_1$. This means that $\beta \cdot \mathcal{P}_t$ is a sufficient statistic to forecast future volatility. This provides some guidance on potential volatility instruments. For example, although level is known to be related to volatility, it is also well-known (for example Campbell and Shiller (1987)), that the slope of the yield curve predicts future changes in the level of interest rates. Up to the associated uncertainty of such statistical evidence, this suggests that the slope of the yield curve predicts the level and thus also that the level of interest rates is not an autonomous process. We now show that the risk neutral measure provides strong guidance for which combinations of yields form autonomous processes, as required for a valid volatility instrument.

For illustrative purposes, we consider the case of three factors $N = 3$, but the same intuition applies to any $N$.

## 3.1 Example 1

Let's define the convexity-adjusted forward rate by:

$$f_t(n) = E_t^{\mathbb{Q}}[r_{t+n}]. \tag{16}$$

In the spirit of Collin-Dufresne, Goldstein, and Jones (2008) we can write the following risk-neutral dynamics:

$$E_t^{\mathbb{Q}} \begin{pmatrix} V_{t+1} \\ r_{t+1} \\ f_{t+1}(1) \end{pmatrix} = \text{constant} + \begin{pmatrix} a_1 & 0 & 0 \\ 0 & 0 & 1 \\ a_2 & a_3 & a_4 \end{pmatrix} \begin{pmatrix} V_t \\ r_t \\ f_t(1) \end{pmatrix}. \tag{17}$$

The first row is due to the autonomous nature of $V_t$. The second row is the definition of the forward rate in (16). The last row is obtained from the fact that in a three factor affine model, $(V_t,\, r_t,\, f_t(1))$ are informationally equivalent to the three underlying states at time $t$. From the last row and by applying the law of iterated expectation to (16), we have:

$$a_2 V_t = \text{constant} + f_t(2) - a_3 r_t - a_4 f_t(1). \tag{18}$$

Using (18) to substitute $V_t$ out from (17), we have:

$$E_t^{\mathbb{Q}} \begin{pmatrix} r_{t+1} \\ f_{t+1}(1) \\ f_{t+1}(2) \end{pmatrix} = \text{constant} + \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 \end{pmatrix} \begin{pmatrix} r_t \\ f_t(1) \\ f_t(2) \end{pmatrix}. \tag{19}$$

where $\alpha_1 = -a_1 a_3$, $\alpha_2 = a_3 - a_1 a_4$, and $\alpha_3 = a_4 + a_1$. It follows from the last row of (19) that:

$$f_t(3) = \text{constant} + \alpha_1 r_t + \alpha_2 f_t(1) + \alpha_3 f_t(2). \tag{20}$$

Equation (20) reveals that if the short rate as well as the forward rates can be empirically observed, the loadings $\alpha$ can in principle be pinned down simply by regressing $f_t(3)$ on $r_t$, $f_t(1)$, and $f_t(2)$. Based on the mappings from $(a_1, a_3, a_4)$ to $\alpha$, it follows that the regression implied by (20) will also identify all the $a$ coefficients, except for $a_2$. In the context of equation (18), it means that the volatility factor is already determined up to a translation and scaling effect.

Repeated iterations of the above steps allow us to write any forward rate $f_t(n)$ as a linear function of $(r_t, f_t(1), f_t(2))$. Suppose that we use $J$ forwards in $(f_t(1), \ldots f_t(J))$ in estimation, then:

$$\begin{pmatrix} r_t \\ f_t(1) \\ f_t(2) \\ f_t(3) \\ f_t(4) \\ \vdots \\ f_t(J) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 \\ & g_4(\alpha) & \\ & \cdots & \\ & g_J(\alpha) & \end{pmatrix} \begin{pmatrix} r_t \\ f_t(1) \\ f_t(2) \end{pmatrix}$$

where $(g_4, \ldots, g_J)$ represent the cross-sectional restrictions of no-arbitrage. So we can think of no-arbitrage as having two facets. First, it imposes a cross-section to time series link through the fact that fixing $\alpha$ constrains what the volatility factor must look like, through $a_3$ and $a_4$. Second, it induces cross-sectional restrictions on the loadings $(g_4, \ldots g_J)$, just as is seen with pure gaussian term structure models.

## 3.2 Example 2

We can also use the approach of Joslin (2011b) to consider arbitrary linear combinations of yields as factors. The insight again is that risk-neutral expectations are, up to convexity, observed as forward rates. The $m$-term forward rate the begins in one period, $f_t^{1,m}$ is given by

$$f_t^{1,m} = \frac{1}{m}((m+1)y_t^{m+1} - r_t). \tag{21}$$

Thus we can use (21) to approximate $E_t^{\mathbb{Q}}[y_{t+1}^m]$ whereby we simply ignore any convexity term. Notice that since our primary interest is not in the level of expected-risk neutral

changes but in their variation (i.e. $K_1^{\mathbb{Q}}$), it is only stochastic convexity effects that will violate this approximation. Thus to the extent that changes in convexity effects are small this approximation will be valid for inference on $K_1^{\mathbb{Q}}$.

From this method, we extract observations on $E_t^{\mathbb{Q}}[y_{t+1}^m]$ from forward rates which we can then convert into estimates of $E_t^{\mathbb{Q}}[\mathcal{P}_{t+1}]$ using the weighting matrix $W$. We denote this approximation of $E_t^{\mathbb{Q}}[\mathcal{P}_{t+1}]$ by $\mathcal{P}_t^f$. We can consider the regression of the form

$$\mathcal{P}_t^f = \text{constant} + K_1^{\mathbb{Q}}\mathcal{P}_t + u_t. \tag{22}$$

Given the validity of our convexity approach and repeated cross-sections of yields, we should be able to obtain estimates of $K_{1\mathcal{P}}^{\mathbb{Q}}$ even without estimating a term structure model.

## 3.3 Example 3

Consider the $J \times 1$ vector of yields $y_t$ used in estimation:

$$y_t = A + BX_t, \tag{23}$$

where $X_t$ denotes the underlying states. To have affine bond pricing, the essential requirement is that the risk-neutral dynamics of $X_t$ be affine. Focusing on the conditional mean, and applying standard rotations to $X_t$, we can always write:[5]

$$E_t^{\mathbb{Q}}[X_{t+1}] = diag(\lambda^{\mathbb{Q}})X_t. \tag{24}$$

In the spirit of Joslin, Singleton, and Zhu (2010), we can rotate $X_t$ to any $N$ yields portfolios: $\mathcal{P}_t = Wy_t$. Rewriting (23) and (24), replacing $X_t$ with $\mathcal{P}_t$, we have:

$$y_t = constant + B(WB)^{-1}\mathcal{P}_t, \tag{25}$$

$$E_t^{\mathbb{Q}}[\mathcal{P}_{t+1}] = constant + \underbrace{(WB)diag(\lambda^{\mathbb{Q}})(WB)^{-1}}_{K_{1\mathcal{P}}^{\mathbb{Q}}}\mathcal{P}_t. \tag{26}$$

Assuming the effect of convexity on the loadings $B$ is minimal (which we confirm subsequently in our empirical exercises), to the first-order approximation, the loadings $B$ are only dependent on $\lambda^{\mathbb{Q}}$.[6] Consequently, equation (25) implies that we could estimate $\lambda^{\mathbb{Q}}$ by regressing $y_t$ on $\mathcal{P}_t$, forcing the loadings to be of the form $B(WB)^{-1}$, which is only dependent on $\lambda^{\mathbb{Q}}$. Once $\lambda^{\mathbb{Q}}$ is identified, the risk-neutral feedback matrix $K_{1\mathcal{P}}^{\mathbb{Q}}$ in (26) is in turn pinned down. Notably, this must be true of all canonical affine term structure models regardless of whether a volatility factor is included or not.

[5] Suppose that we start from $E_t^{\mathbb{Q}}[X_{t+1}] = K_0 + K_1X_t$. The appropriate rotation is $\tilde{X}_t = C + DX_t$ where $D$ is the inversion of the eigenvector matrix of $K_1$ such that $DK_1D^{-1}$ is diagonal. And $C = (DK_1D^{-1} - I)^{-1}DK_0$. Here for ease of exposition, we assume that the eigenvalues of $K_1$ are distinct and real-valued.

[6] Specifically, for the loading on the $i^{th}$ factor for yield $y_t^n$ with $n$ periods to maturity is approximately $\frac{1+\lambda^Q(i)+...+\lambda^Q(i)^{n-1}}{n\Delta t}$.

Focusing on term structure models with one stochastic volatility factor $V_t$, affine bond pricing implies that we can write:

$$V_t = constant + b'\mathcal{P}_t \tag{27}$$

for some $b$. It follows that:

$$E_t^{\mathbb{Q}}[V_{t+1}] = constant + b' K_{1\mathcal{P}}^{\mathbb{Q}} \mathcal{P}_t. \tag{28}$$

Since $V_{t+1}$ must be autonomous, its conditional mean can only depend on its lagged value. Consequently, we must have:

$$b' K_{1\mathcal{P}}^{\mathbb{Q}} = cb' \tag{29}$$

for some scaler $c$. In other words, $b'$ must be a left eigenvector of $K_{1\mathcal{P}}^{\mathbb{Q}}$.

# 4 Empirical Results

In this section, we empirically evaluate the implications of our theory, using the US Treasury data spanning the period from March 1984 until December 2006. The sample is chosen such that our analysis is not materially affected by either the Fed experiment regime or the current financial crisis. We use the monthly unsmoothed Fama Bliss zero yields and the following seven maturities in our estimation: 3–month, 6–month, one– out to five–year. We first provide evidence that for our sample period, conditional volatilities of yields can be captured to a large extent by linear combinations of yields. Next, we examine the extent to which the requirement of autonomous volatility can impact the identification of volatility under both the physical and risk-neutral measures. For all of our exercises, we consider models with three factors.

## 4.1 Conditional volatility from linear projections

In this section, we investigate whether interest rate volatility can be approximately spanned by yields. In doing this, we focus on the conditional volatility of the level factor – the most important determinant of yields for all maturities. We first construct the first three principal principal components of yields, $\mathcal{P}_t$, for our data and fit these to a VAR:

$$\mathcal{P}_{t+1} = K_0 + K_1 \mathcal{P}_t + \epsilon_{t+1}. \tag{30}$$

To estimate the conditional volatility of the level factor $\mathcal{P}_{t+1,1}$, we project the fitted residuals of the level factor, $\hat{\epsilon}_{t+1,1}$, from (30) onto $\mathcal{P}_t$:

$$\hat{\epsilon}_{t+1,1}^2 = \alpha + \beta \mathcal{P}_t + e_{t+1}. \tag{31}$$

The square root of the predicted component of (31) is a measure of the conditional volatility of the level factor implied by the affine structure (of the conditional means and

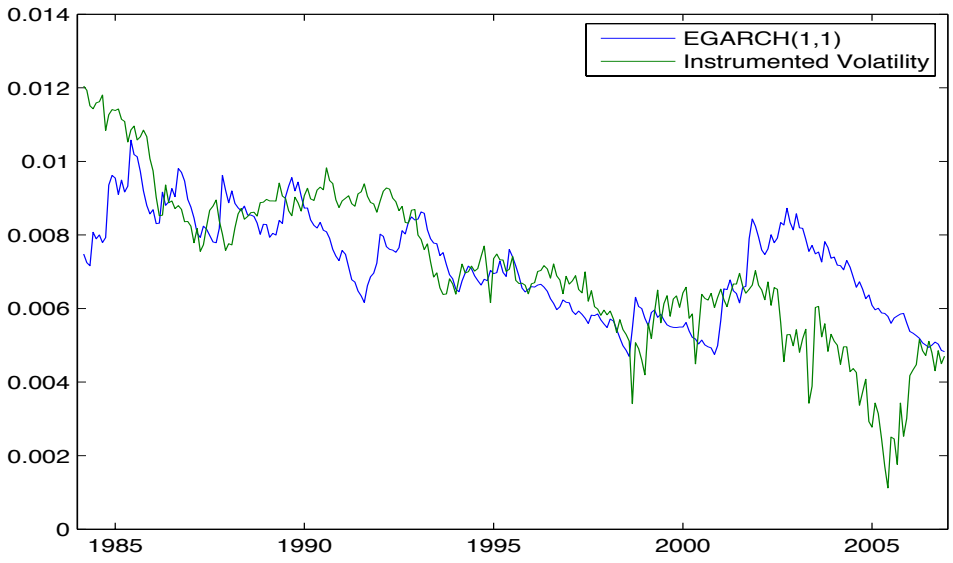


Figure 1: Volatility of the level factor estimated from the EGARCH(1,1) model and a two-stage regression.

variances) of standard term structure models. What we are primarily interested in is whether this spanned measure of volatility, *free of admissibility constraints*, can reasonably capture interest rate volatility. As a result, in estimating (30) and (31), we do not require that the volatility factor be autonomous. While the true volatility process is not known, we follow standard practices and use volatility estimates from the EGARCH(1,1) model as a benchmark. Specifically, using the fitted residuals $\hat{\epsilon}_{t+1,1}$ from (30) as input, we obtain and plot the EGARCH(1,1) estimates in Figure 1 side by side the predicted component of (31), labeled "Instrumented Volatility." It is visually clear that the two series track each other quite well with a sample correlation of 66%. Comparing volatility estimates of a $A_1(3)$ model with the EGARCH(1,1) estimates over a longer sample period, Jacobs and Karoui (2009) find a very similar correlation.

Additionally, using a multi-period version of (31),[7] Le and Singleton (2011) show that spanned volatility is considerably rich of information relevant for predicting future excess returns. Together, these suggest that volatilities instrumented by yields can potentially reflect realistic fluctuations in yields as well as meaningful information about priced risks. As we overlay the estimation of (31) with the requirement that volatility be autonomous, there is a possibility that the estimates of the loading $\beta$ can be altered in such a way that the information content of the resulting volatility estimates can be quite different. We take on these issues at greater depth in the subsequent sections. Here with the aim of providing some initial descriptive evidence of the impact of the autonomous requirement, we perform the

[7] In particular, they stack $\hat{\epsilon}^2_{t+h}$ over multiple horizons and use the technique of rank reduced regression to estimate $\alpha$ and $\beta$.

following simple test. Using the estimated loading $\hat{\beta}$ from (31), we regress $\hat{\beta}\mathcal{P}_{t+1}$ on $\mathcal{P}_t$:

$$\hat{\beta}\mathcal{P}_{t+1} = \text{constant} + \gamma\mathcal{P}_t + noises. \tag{32}$$

Under the null hypothesis that the volatility factor $\alpha+\beta\mathcal{P}_t$ is autonomous, it must be true that $\gamma$ is a scaled version of $\beta$. This amounts to a $\chi^2$ test with two degrees of freedom. Using twelve lags to construct the robust asymptotic covariance matrix of the estimates, accounting for estimation errors with sequential inferences, we obtain a $\chi^2(2)$ statistics of 11.2, statistically significant at the 1% level. This evidence shows that while there exists a linear combination of yields that can give a reasonable account of the time-varying conditional variances of yields, such a combination does not give rise to a volatility factor that is autonomous under $\mathbb{P}$. This suggests the potential tension in fit when autonomy is imposed in estimation. In the next two subsections, we separately consider the impact of the autonomous requirement under $\mathbb{P}$ and $\mathbb{Q}$.

## 4.2 Autonomy under $\mathbb{P}$

We first consider the requirement by both the $F_1(N)$ and $A_1(N)$ models that volatility forms an autonomous process under $\mathbb{P}$. This requires (10) which states that the volatility instrument, $\beta'$ be a left eignevector of $K_1$. To the extent that the conditional mean is strongly identified by the time-series, this condition will pin down the admissible volatility instruments up to a sign choice and the choice of which of the $N$ left eigenvectors instruments volatility. However, in general even with a moderately long time series, such as our 22 year sample, inference on the conditional mean is not very precise. At the same time however, this condition does provide some guidance as we have just seen that there is statistical evidence that the linear projection of the squared residuals does not seem to follow an autonomous process.

To gauge how strongly identified the volatility instrument is by the autonomous requirement under $\mathbb{P}$, we implement the following exercise. First we estimate an unconstrained VAR on the first three principal factors, $\mathcal{P}_t$. Ignoring the intercepts, the estimates for our sample period are:

$$\mathcal{P}_{t+1} = \underbrace{\begin{pmatrix} 0.9805 & -0.0190 & -0.5530 \\ -0.0027 & 0.9581 & 0.3928 \\ 0.0014 & -0.0005 & 0.8433 \end{pmatrix}}_{K_1} \mathcal{P}_t + noises \tag{33}$$

Then, for *each* potential volatility instrument $\beta\mathcal{P}_t$ (as $\beta$ roaming over $\mathbb{R}^3$), we re-estimate the VAR under the constraint that $\beta$ is a left eigenvector of $K_1$. The VAR is easily estimated under this constraint after a change of variables so the the eigenvector constraint becomes a zero constraint (compare the constraints in (6) and (10)). We then conduct a likelihood ratio test of the unconstrained versus the constrained alternative and compute the associated p-value. A p-value close to 1 indicates that the evidence is consistent with such an instrument being

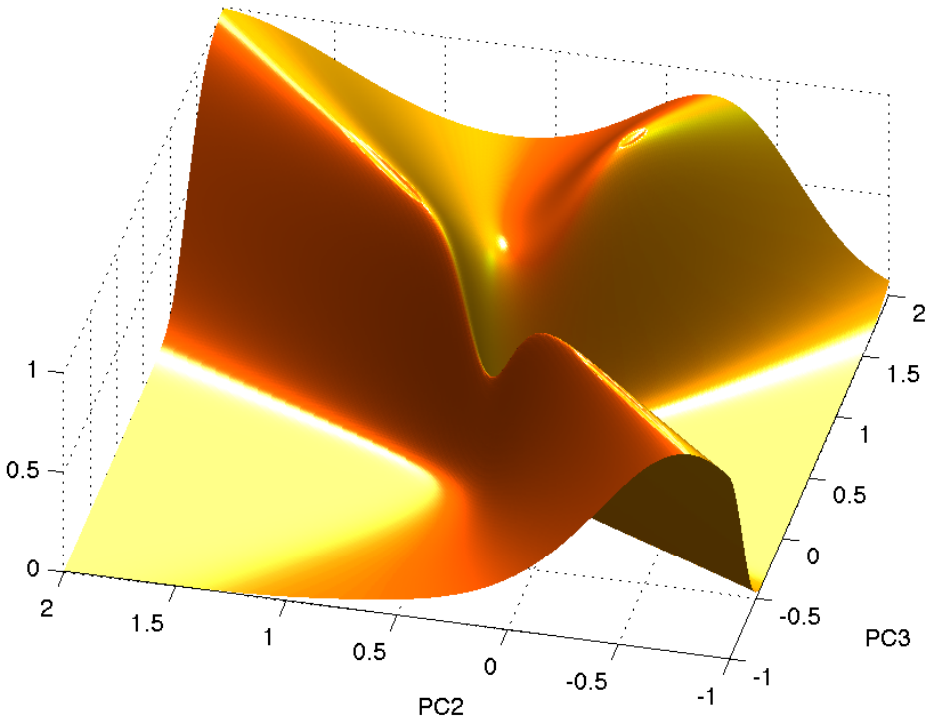


Figure 2: Likelihood Ratio Tests of $\mathbb{P}$-Autonomous Restrictions. This figure reports the p-values of the likelihood ratio test of whether a particular linear combination of yields, $\beta\mathcal{P}_t$, is autonomous under $\mathbb{P}$, plotted against the loadings of PC2 and PC3. The loading of PC1 is one minus the loadings on PC2 and PC3 ($\beta(1) = 1 - \beta(2) - \beta(3)$). PC1, PC2, and PC3 are scaled to have in-sample variances of one.

an autonomous process while a p-value close to 0 indicates evidence against the instrument being an autonomous process.[8]

Since $\beta\mathcal{P}_t$ and its scaled version, $c\beta\mathcal{P}_t$, for any constant $c$, effectively give the same volatility factor (and hence deliver the same p-values in our exercise), we scale so that all elements of $\beta$ sum up to one. We plot the p-values against the corresponding pairs of loadings on PC2 and PC3 (the loading on PC1 $\beta(1) = 1 - \beta(2) - \beta(3)$) in Figure 2. For ease of presentation, in this graph the three PCs are scaled to have in-sample variances of one.

We see that there are three peaks which correspond to the three left eigenvectors of the maximum likelihood estimate of $K_1$. When $\beta$ is equal to one of these left eigenvectors, the likelihood ratio test statistic must be zero and hence the corresponding p-value must be one, by construction. As our intuition suggests, many, though not all, instruments appear to

[8]We view this test as an approximation since it assumes volatility is constant. However, exploratory computations of p-values, accounting for heteroskedasticity of the errors, deliver very similar results.

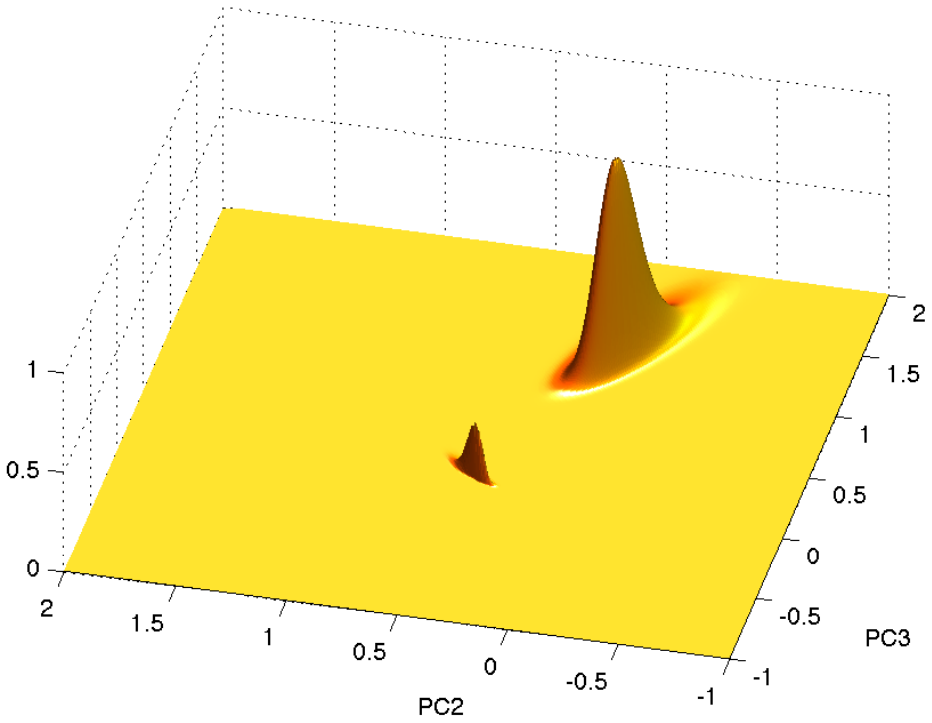


Figure 3: Likelihood Ratio Tests of $\mathbb{Q}$-Autonomous Restrictions. This figure reports the p-values of the likelihood ratio test of whether a particular linear combination of yields, $\beta\mathcal{P}_t$, is autonomous under $\mathbb{Q}$, plotted against the loadings of PC2 and PC3. The loading of PC1 is one minus the loadings on PC2 and PC3 ($\beta(1) = 1 - \beta(2) - \beta(3)$). PC1, PC2, and PC3 are scaled to have in-sample variances of one.

potentially satisfy the requirement of (10) according to the metric that we are considering. Thus we conclude that the autonomy requirement under the $\mathbb{P}$ measure in general still leaves a great deal of flexibility in forming the volatility instrument.

## 4.3 Autonomy under $\mathbb{Q}$

Having seen the precision with which the historical measure guides the set of admissible volatility instruments, we now turn to the information gained from the risk-neutral measure. We follow the model-free approach of Joslin (2011b) as in Section 3.2 to study this restriction in a similar manner to our previous exercise under the historical measure. Specifically, we consider, ignoring constants, the regression

$$E_t^{\mathbb{Q}}[\mathcal{P}_{t+1}] = K_1^{\mathbb{Q}}\mathcal{P}_t + u_t,$$

where, as before, we proxy for $E_t^{\mathbb{Q}}[\mathcal{P}_{t+1}]$ using the linear combination of forwards $\mathcal{P}_t^f$. As in our analysis of the autonomy constraints under the historical measure, we consider both the unconstrained version of this regression and the same regression with the constraint that a particular $\beta$ is a left eigenvector of $K_1^{\mathbb{Q}}$. Intuitively, although we observe only a single time series under the historical measure with which to draw inferences, we observe repeated term structures of risk-neutral expectations every month and this allows us to draw much more precise inferences.

Figure 3 plots the p-values for this test of the restrictions of various instruments to be autonomous under $\mathbb{Q}$. In stark contrast to Figure 2 and in accordance with our intuition, we see that the risk-neutral measure provides very strong evidence for which instruments are able to be valid volatility instruments. Most potential volatility instruments are strongly ruled out with p-values essentially at zero. Bond prices essentially pin down $K_1^{\mathbb{Q}}$ and the only degrees of freedom are the sign choice and choosing which of the left eigenvectors is the volatility instrument.

## 4.4 Comparison of Gaussian and stochastic volatility models

The analysis in the previous subsections provides suggestive evidence that the cross-sectional information, in combination with the autonomy requirement, is likely to be very helpful in identifying the potential volatility instrument. What are left out of these exercises, however, are, among other things, the constraints imposed from variance equations such as (2). The informal evidence produced in Section 4.1 hints at the potential tension between fitting the variance equations and satisfying the autonomy requirements.

To shed light on how this tension is resolved, we estimate the Gaussian term structure model $A_0(3)$ with constant volatility and compare it to the $A_1(3)$ model. This comparison is informative since the main difference between the two models is simply the variance equations. In particular, we are interested in the estimates of $K_1^{\mathbb{Q}}$ implied by these two models since the volatility loading from the $A_1(3)$ model must be a left eigenvector of this matrix. Estimates of $K_1^{\mathbb{Q}}$ for the two models are reported in the first two panels of Table 1. The two estimates are strikingly close: except for the (1,2) entry, for every other entry, the two estimates are essentially identical up to the third decimal place. This evidence suggests that the identification by the cross-sectional information (and possibly other moments shared between the $A_0(3)$ and $A_1(3)$ models) for the parameter $K_1^{\mathbb{Q}}$ seems overwhelmingly stronger than the restrictions coming from the variance equations. This leads us to the surprising conclusion that the $A_0(N)$ model with *constant volatility* allows us to essentially identify (up to choice of which eigenvector) the volatility instrument in the $A_1(N)$ model.

Moreover, we have argued that variation in the risk-neutral expectations, as determined by $K_1^{\mathbb{Q}}$, is well approximated by the regression based estimate of (22). Importantly, this model free estimate of $K_1^{\mathbb{Q}}$ is independent of any physical dynamics. As can be seen from the last panel of Table 1, this estimate is also very similar to ones implied by the $A_0(3)$ and $A_1(3)$ models. This evidence suggests that the cross-sectional information *alone* is sufficient to pin down the risk-neutral feedback matrix, and this identification is so strong that information from other constraints imposed by the models seem irrelevant.

| | | | |
|---|---|---|---|
| | 1.0084 | 0.1088 | -0.4111 |
| $A_0(3)$ | -0.0105 | 0.9786 | 0.4491 |
| | 0.0042 | -0.0098 | 0.7851 |
| | 1.0085 | 0.1102 | -0.4084 |
| $A_1(3)$ | -0.0104 | 0.9781 | 0.4479 |
| | 0.0042 | -0.0088 | 0.7870 |
| | 1.0066 | 0.1230 | -0.4078 |
| Regression | -0.0082 | 0.9659 | 0.4510 |
| | 0.0020 | -0.0008 | 0.7589 |

Table 1: This table contains the estimates of $K_1^{\mathbb{Q}}$ computed from the $A_0(3)$ and $A_1(3)$ models, along with the estimates from the model-free regression of (22).

This observation has a number of implications. First, as stated previously, this allows us to pin down the potential volatility instruments using the cross-section of yields due to the admissibility constraint. Essentially the volatility instrument is free in terms of the sign but must be one of the left eigenvectors of $K_1^{\mathbb{Q}}$ which can be computed accurately from either the cross-sectional regression or from estimation of the $A_0(N)$ model which has constant volatility and can be estimated quite quickly as shown in JSZ.

This observation also shows that in some regards, the estimation of the no arbitrage $A_1(N)$ model is more tractable than estimate of the $\mathrm{F}_1(N)$ model. In the case of the Gaussian models the opposite holds: the factor model is trivial to estimate as it amounts to a set of ordinary least squares regressions while the no arbitrage model is slightly more difficult to estimate due to the non-linear constraints in the factor loadings. In the stochastic volatility models, the admissibility conditions require a number of non-linear constraints in order to ensure that volatility remains positive. The no arbitrage model essentially determines the volatility instrument up to sign and choice of eigenvector. This actually simplifies the computational burden in estimation since it reduces the set of non-linear constraints that need to be imposed.

## 4.5 Comparison of estimates of volatility

The strong identification power of the cross-sectional information for the potential volatility instruments (to the extent that the variance equations do not seem to matter) begs the important question of whether the volatility outputs by these models make sense. With this in mind, we now turn to see the estimation results for volatility in both the general $F_1(N)$ factor model and the $A_1(N)$ model. The time series of model-implied volatilities are plotted in Figure 4. We see that in general both the factor model (which imposes only admissibility without restrictions on the cross-sectional loadings) and the no arbitrage model produce a volatility time series for the level of interest rates which is generally consistent with the

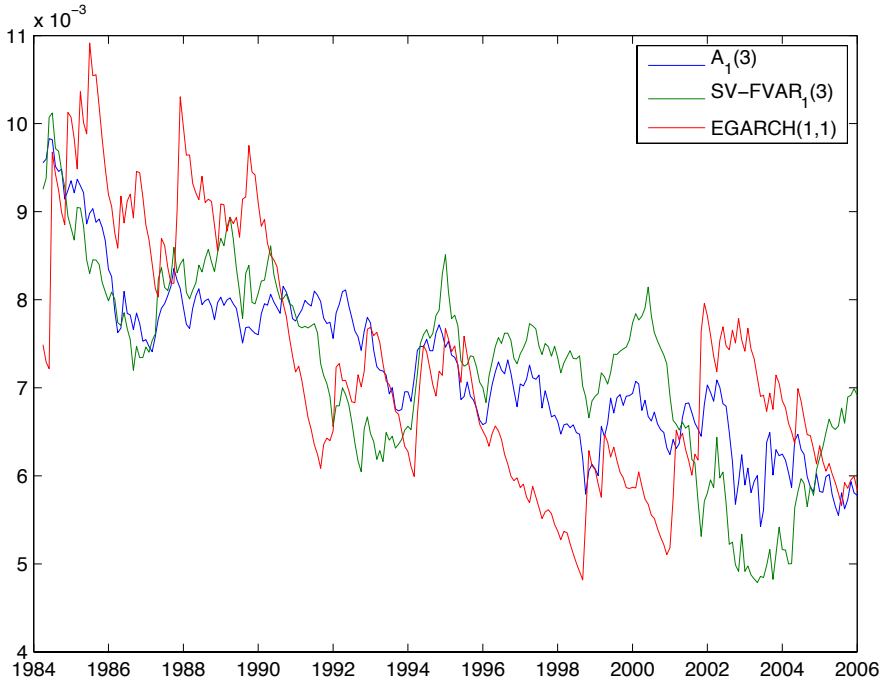


Figure 4: This figure plots the model estimates of the volatility from the estimated $A_1(N)$ and $F_1(N)$ with the EGARCH(1,1) estimates.

EGARCH(1,1) estimates. The correlations between the $F_1(N)$ and $A_1(N)$ volatility estimates and the EGARCH(1,1) estimates are 0.72 and 0.45, respectively. We note again that we should anticipate a positive correlation since both specifications are free to change the sign of the volatility instrument so to the extent that the maximum likelihood criterion function is maximized with a positive sign, this positive correlation should generally occur.

# 5 Conclusion

In this paper, we show that there is a strong link in a no arbitrage affine term structure model between the cross-section of bond yields and the set of positive processes under the risk neutral measure. Since volatility must define a positive process under any measure, this provides a tight link between the cross section of bond yields and the set of valid volatility instruments. To the extent that the no arbitrage model is correctly specified, no arbitrage therefore provides evidence for identifying volatlity instruments. Moreover, the information about volatility instruments in the cross section of bond yield can also be approximately obtained either through model-free regressions or by analyzing Gaussian models with constant volatilities. Empirically, we show that broadly the no arbitrage models appear to match

volatility with this condition in our sample. Of separate interest, we also provide a general canonical form for affine models with stochastic volatility (the $A_1(N)$ class) where nearly all of the parameters directly relate to observable variables.

Our primary results are driven by the fact that an affine drift requires a number of constraints in order to assure that volatility stays positive. A number of alternative models could be considered. First, one could consider a model with unspanned or nearly unspanned volatility. This, however, can only partially counteract our results in the sense that the projection of volatility onto yields must still mathematically be a positive process. So several of our insights maintain. Another possible model to consider is a model with non-linear drift. That is, we can suppose that there is a latent state variable $Z_t$ with the drift of $Z_t$ linear in $Z_t$ without any constraints provided that volatility (or its instrument) is far from the boundary. Near the zero boundary, the drift of the volatility may be non-linear in such a way as to maintain positivity. Provided that the probability of entering this non-linear region is small (under $\mathbb{Q}$), similar pricing equation will be obtained as in the standard affine setting.

# A A Canonical Form for Discrete-Time Term Structure with One Stochastic Volatility Factor

In this section, we lay out canonical forms for discrete-time affine term structure models with one stochastic volatility factor. Our construction draws on the discrete volatility dynamics of Le, Singleton, and Dai (2010) and the canonical setup of Joslin, Singleton, and Zhu (2010) (JSZ) for Gaussian DTSMs. The continuous time limit of our construction nets the $A_1(N)$ family of models considered in Dai and Singleton (2000).

Recall that we denote the $N$ latent risk factors by $(V_t, X_t')'$ where $V_t$ governs the conditional volatility of the states and $X_t$ represent the remaining $N-1$ conditionally gaussian risk factors. We first write down the conditional dynamics of states that: (a) delivers affine bond pricing; (b) ensures econometric identification. Next, we provide explicit bond pricing formulas. Finally, we show how to reparameterize the model in terms of observable yield portfolios and a way to generate sensible starting points for ML estimation.

## A.1 $\mathbb{P}$ and $\mathbb{Q}$ dynamics of the latent states in discrete time

Under $\mathbb{Q}$, the states follow:

$$V_{t+1}|V_t \sim \text{univariate } CAR(\rho^{\mathbb{Q}}, c^{\mathbb{Q}}, \nu^{\mathbb{Q}}), \tag{34}$$

$$X_{t+1} \sim N(K_{1Z}^{\mathbb{Q}} V_t + diag(\lambda^{\mathbb{Q}}) X_t, \Sigma_{0X} + \Sigma_{1X} V_t), \text{ independent of } V_{t+1} \tag{35}$$

$$r_t = r_\infty + \rho_V V_t + \iota \cdot X_t. \tag{36}$$

Here we have applied the rotation techniques similar to those used by DS and JSZ to obtain econometric identification for the parameters related to $X_t$. CAR refers to a compound autoregressive gamma process, characterized by three strictly positive parameters $(\nu^{\mathbb{Q}}, \rho^{\mathbb{Q}}, c^{\mathbb{Q}})$ such that:[9]

$$E_t^{\mathbb{Q}}[V_{t+1}] = \nu^{\mathbb{Q}} c^{\mathbb{Q}} + \rho^{\mathbb{Q}} V_t, \text{ and } Var_t^{\mathbb{Q}}[V_{t+1}] = \nu^{\mathbb{Q}} c^{\mathbb{Q}^2} + 2\rho^{\mathbb{Q}} c V_t. \tag{37}$$

Here, $\rho^{\mathbb{Q}}$, $\nu^{\mathbb{Q}}$, and $c^{\mathbb{Q}}$ modulate the (risk-neutral) persistence, mean and scale of $V_t$, respectively. To see the continuous time limit of CAR, let $\rho^{\mathbb{Q}} = \kappa \Delta t$, $c^{\mathbb{Q}} = \frac{1}{2}\sigma^2 \Delta t$, and $v^{\mathbb{Q}} = \frac{2\kappa\theta}{\sigma^2}$, we can see that

$$E_t^{\mathbb{Q}}[\Delta V_{t+1}] = \kappa(\theta - V_t)\Delta t \text{ and } Var_t^{\mathbb{Q}}[\Delta V_{t+1}] = \sigma^2 V_t \Delta t + o(\Delta t). \tag{38}$$

Clearly, a CAR process approaches the CIR process in the continuous time limit as $\Delta t$ approaches 0.

To avoid $V_t$ being absorbed at the zero boundary, the discrete–time counterpart to the Feller condition ($\kappa\theta \geq \frac{1}{2}\sigma^2$) requires $\nu^{\mathbb{Q}}$ be greater than one. One attractive property of the CAR process is that $V_t$ is always strictly positive, yet its conditional Laplace transform is

[9] For further details of the CAR process, such as its associated conditional density and Laplace transform, see Le, Singleton, and Dai (2010).

exponentially affine in $V_t$,[10] which is essential to obtain affine bond pricing. It is important to note that $V_t$ must be autonomous – that is, the gaussian states $X_t$ must not directly affect the conditional dynamics of $V_t$. The presence of $X_t$ on the right hand side of either equation in (37) would give a strictly positive probability that either the conditional mean or the conditional variance of $V_{t+1}$ be negative.

Note that we can still scale $V_t$ up and down by a positive constant.[11] To obtain econometric identification, we will fix the scaling of $V_t$ by normalizing $\rho_V$ the details of which will be described below.

Under $\mathbb{P}$, the states follow an analogous dynamics:

$$V_{t+1}|V_t \sim \text{univariate } CAR(\rho^{\mathbb{P}}, c^{\mathbb{P}}, \nu^{\mathbb{P}}), \tag{39}$$

$$X_{t+1} \sim N(\mathcal{L}(V_t, X_t), \Sigma_{0X} + \Sigma_{1X} V_t), \text{ independent of } V_{t+1} \tag{40}$$

where $\mathcal{L}$ denotes a linear operator. Again, non-attainment under $\mathbb{P}$ requires the Feller condition: $\nu^{\mathbb{P}} \geq 1$.

We now discuss two technical issues related to this parameterization. First, consider the market prices of variance risk:

$$\frac{E_t^{\mathbb{P}}[V_{t+1}] - E_t^{\mathbb{Q}}[V_{t+1}]}{Var_t^{\mathbb{P}}[V_{t+1}]}.$$

As discussed by Cheridito, Filipovic, and Kimmel (2007), when $Var_t^{\mathbb{P}}[V_{t+1}]$ approaches zero, there is the issue of exploding market prices of risks unless the intercept terms of $E_t^{\mathbb{P}}[V_{t+1}]$ and $E_t^{\mathbb{Q}}[V_{t+1}]$ are the same (hence the numerator too approaches zero at the same rate as the denominator). Nevertheless, in our discrete time setup, as long as $\nu^{\mathbb{P}}$ and $c^{\mathbb{P}}$ are strictly positive, $Var_t^{\mathbb{P}}[V_{t+1}]$ is bounded strictly away from zero. As a result, we don't have to directly deal with this issue. If one wishes to avoid this issue even in the continuous time limit, then a sufficient restriction on the parameters is:

$$v^{\mathbb{P}} c^{\mathbb{P}} = v^{\mathbb{Q}} c^{\mathbb{Q}}.$$

Finally, the scale parameters ($c^{\mathbb{P}}$ and $c^{\mathbb{Q}}$) in principle can be any pair of positive numbers in our discrete time setup. Nevertheless, the diffusion invariance property of the CIR process requires that these two parameters have the same continuous time limit ($\frac{1}{2}\sigma^2 dt$). To be consistent with diffusion invariance of $V_t$ in the continuous time limit, then a sufficient restriction on the parameters is:

$$c^{\mathbb{P}} = c^{\mathbb{Q}}.$$

## A.2 Bond pricing

Note that the Laplace transform for a univariate CAR process is:

$$E_t[e^{uV_{t+1}}] = e^{a(u)+b(u)V_t} \text{ where } a(u) = -\nu^{\mathbb{Q}} log(1 - uc^{\mathbb{Q}}), \; b(u) = \frac{\rho^{\mathbb{Q}} u}{1 - uc^{\mathbb{Q}}}.$$

[10] That is, $E_t^{\mathbb{Q}}[e^{uV_{t+1}}] = e^{a(u)+b(u)V_t}$.

[11] If we scale $V_t$ up by a positive constant $c$ then $cV_{t+1}|cV_t \sim CAR(\rho^{\mathbb{Q}}, cc^{\mathbb{Q}}, \nu^{\mathbb{Q}})$.

Therefore, bond prices are exponentially affine $logP_{n,t} = -A_n - B_{V,n}V_t - B_{X,n}X_t$ with loadings given by:

$$B_{X,n} = \iota' + B_{X,n-1}diag(\lambda^{\mathbb{Q}}), \tag{41}$$

$$B_{V,n} = \rho_Z + \frac{\rho^{\mathbb{Q}}B_{V,n-1}}{1 + B_{V,n-1}c^{\mathbb{Q}}} + B_{X,n-1}K_{1Z}^{\mathbb{Q}} - \frac{1}{2}B_{X,n-1}\Sigma_{1X}B'_{X,n-1}, \tag{42}$$

$$A_n = r_\infty + A_{n-1} + \nu^{\mathbb{Q}}log(1 + B_{V,n-1}c^{\mathbb{Q}}) - \frac{1}{2}B_{X,n-1}\Sigma_{0X}B'_{X,n-1}, \tag{43}$$

starting from: $A_0 = B_{V,0} = B_{X,0} \equiv 0$.

If we ignore convexity, then the loadings of yields on the factors satisfy:

$$B^c_{X,n} = \iota' + B^c_{X,n-1}diag(\lambda^{\mathbb{Q}}), \tag{44}$$

$$B^c_{V,n} = \rho_Z + \rho^{\mathbb{Q}}B^c_{V,n-1} + B^c_{X,n-1}K_{1Z}^{\mathbb{Q}}, \tag{45}$$

starting from $B^c_{V,0} = B^c_{X,0} \equiv 0$. Note that $B^c_X$ and $B_X$ are the same.

To obtain loadings on annualized yields, we need to scale these loadings appropriately by the corresponding maturities.

## A.3 Parameterization with yields portfolios as factors

In estimation, we assume that $\mathcal{N}$ portfolios of yields, $\mathcal{P}^{\mathcal{N}}_t = W^{\mathcal{N}}y_t$ are observed without errors. Let's denote the first row of $W^{\mathcal{N}}$ by $W_V$ and the last $\mathcal{N} - 1$ rows by $W_X$, with the corresponding portfolios denoted by $P_{V,t}$ and $P_{X,t}$. If we use convexity-adjusted yields instead, we add superscript $^c$ and denote these portfolios by $P^c_t$, $P^c_{V,t}$, and $P^c_{X,t}$.

Applying $W_X$ to both sides of the yield pricing equation $y_t = A + B_V V_t + B_X X_t$, we obtain

$$P_{X,t} = W_X A + W_X B_V V_t + W_X B_X X_t.$$

Thus under the physical measures:

$$P_{X,t+1} - W_X B_V V_{t+1} = N\Big(\mathcal{L}(V_t, P_{X,t}), \Sigma_0 + \Sigma_1 V_t\Big)$$

where

$$\Sigma_0 = W_X B_X \Sigma_{0X}(W_X B_X)', \text{ and } \Sigma_1 = W_X B_X \Sigma_{1X}(W_X B_X)'.$$

Now, pricing the $P_{V,t}$ portfolio, we obtain:

$$P_{V,t} = W_V \tau A + W_V \tau B_V V_t + W_V B_X (W_X B_X)^{-1} P_{X,t}$$

where $\tau = I - B_X(W_X B_X)^{-1}W_X$.

We choose to fix the scaling of $V_t$ to the observable portfolio $P_{V,t}$. Specifically, we choose $\rho_Z$ such that:

$$W_V \tau B_V = s_V \text{ where } s_V = 1 \text{ or } s_V = -1.$$

Let $\beta = (s_V, -s_V W_V B_X (W_X B_X)^{-1})$, we can write:[12]

$$V_t = \beta \mathcal{P}_t - s_V W_V \tau A = \beta \mathcal{P}_t - \beta W^{\mathcal{N}} A.$$

Now to make sure that $V_t$ be strictly positive, it is convenient to write instead:

$$V_t = \beta \mathcal{P}_t - min(\beta \mathcal{P}_t) + \tilde{\alpha}.$$

Obviously, as long as $\tilde{\alpha}$ is positive, $V_t$ is positive. To this extent, it is more convenient to use $\tilde{\alpha}$ as a primitive parameter. For this, we will give up our degree of freedom in $r_\infty$. Matching:

$$\beta W^{\mathcal{N}} A = min(\beta \mathcal{P}_t) - \tilde{\alpha},$$

noting that (1) $r_\infty$ does not enter into the calculations of the loadings $B_X$ and $B_V$; and (2) $r_\infty$ contributes to $A$ linearly: $A = h r_\infty + \tilde{A}$ where $h$ is a constant vector and $\tilde{A}$ is the component of $A$ that is not dependent on $r_\infty$, we can write:

$$r_\infty = \frac{min(\beta \mathcal{P}_t) - \tilde{\alpha} - \beta W^{\mathcal{N}} \tilde{A}}{\beta W^{\mathcal{N}} h}.$$

The full set of $\mathbb{Q}$ parameters include $\{\rho^{\mathbb{Q}}, c^{\mathbb{Q}}, \nu^{\mathbb{Q}}, K_{1V}^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, s_V, \tilde{\alpha}, \Sigma_0, \Sigma_1\}$. The $\mathbb{P}$ parameters include $\{\rho^{\mathbb{P}}, c^{\mathbb{P}}, \nu^{\mathbb{P}}\}$ and other parameters $\{K_0, K_{1V}, K_{1\mathcal{P}}\}$ that determine the conditional means of $\mathcal{P}_{X,t}$.

$$E_t[\mathcal{P}_{X,t+1} - W_X B_V V_{t+1}] = \mathcal{L}(V_t, P_{X,t}) = K_0 + K_{1V} V_t + K_{1\mathcal{P}} \mathcal{P}_{X,t}.$$

## A.4 Starting points

### A.4.1 $\lambda^{\mathbb{Q}}$

We guess that the eigenvalues of the risk-neutral feedback matrix, ($\rho^{\mathbb{Q}}$, $\lambda^{\mathbb{Q}}$), are similar to those obtained in estimating the corresponding $A_0(N)$ model. Since there is no ordering between $\rho^{\mathbb{Q}}$ and $\lambda^{\mathbb{Q}}$, given the N risk-neutral eigenvalues implied by the $A_0(N)$ model, there are $N$ possibilities as to which one corresponds to $\rho^{\mathbb{Q}}$. The remaining eigenvalues are then ordered from high to low and form into $\lambda^{\mathbb{Q}}$. Although we can make a guess of $\rho^{\mathbb{Q}}$ here (to be the remaining risk-neutral eigenvalue), we will defer this until later by choosing a value for $\rho^{\mathbb{Q}}$ that best fits the cross-section of yields.

Given a guess of $\lambda^{\mathbb{Q}}$, we can compute $B_X$ and $\beta = (s_V, -s_V W_V B_X (W_X B_X)^{-1})$. For $\beta$, we need to know $s_V$. We will consider both possibilities $s_V = 1$ and $s_V = -1$.

### A.4.2 $\tilde{\alpha}$, $\Sigma_0$, and $\Sigma_1$

Given each $\beta$ from the previous subsection, we fit the following physical dynamics:

$$V_{t+1} = \beta \mathcal{P}_t - min(\beta \mathcal{P}_t) + \tilde{\alpha} \sim CAR(\rho^{\mathbb{P}}, c^{\mathbb{P}}, \nu^{\mathbb{P}}), \tag{46}$$

$$\mathcal{P}_{X,t+1} \sim N(K_0 + K_{1V} V_t + K_{1\mathcal{P}} \mathcal{P}_{X,t} + \phi_V V_{t+1}, \Sigma_0 + \Sigma_1 V_t), \tag{47}$$

[12] It is straightforward to show that $\beta W^{\mathbb{N}} = s_V W_V \tau$.

subject to $0 < \rho^{\mathbb{P}} < 1$, $\nu^{\mathbb{P}} >= 1$, and $c^{\mathbb{P}} > 0$. Note that for each value of $\tilde{\alpha}$, the triplet $\{\rho^{\mathbb{P}}, c^{\mathbb{P}}, \nu^{\mathbb{P}}\}$ can be obtained relatively efficiently by maximizing the likelihood implied by the CAR dynamics. Additionally, for each value of $\tilde{\alpha}$ and $\Sigma_0$ and $\Sigma_1$, the mean parameters in the second equation ($\{K_0, K_{1V}, K_{1\mathcal{P}}, \phi_V\}$) can be obtained analytically. Note that $\phi_V$ corresponds to $W_X B_V$.

### A.4.3 The remaining risk neutral parameters

The remaining risk neutral parameters are $c^{\mathbb{Q}}$, $\nu^{\mathbb{Q}}$, $\rho^{\mathbb{Q}}$, and $K_{1V}^{\mathbb{Q}}$. For $\nu^{\mathbb{Q}}$ and $c^{\mathbb{Q}}$, we will use their physical counterparts, $\nu^{\mathbb{P}}$ and $c^{\mathbb{P}}$, from the previous subsection. To guess $\rho^{\mathbb{Q}}$ and $K_{1V}^{Q}$ we follow the following steps:

1. Regressing $y_t - B_X(W_X B_X)^{-1} P_{X,t}$ on $\beta \mathcal{P}_t$ to obtain a loading vector $\theta$; Note that by construction, $W_X \theta = 0$, and $W_V \theta = s_V$.

2. Let $\tilde{B}_V = \theta + B_X(W_X B_X)^{-1}\phi_V$. It can be checked that $W_X \tilde{B}_V = \phi_V$, and $W_V \tau \tilde{B}_V = s_V$.

3. Fit a cubic spline through $B_V$ to obtain a series of $\tilde{B}_{V,n}$. Assuming that $B_{V,n}$'s are close to $B_{V,n}^c$ therefore satisfying equation (45).

4. Regressing $\tilde{B}_{V,n}$ on $\tilde{B}_{V,n-1}$ and $B_{X,n-1}$. The loadings obtained from this regression correspond to $\rho^{\mathbb{Q}}$ and $K_{1V}^{\mathbb{Q}}$.

# References


Almeida, C., J. J. Graveline, and S. Joslin, 2011, "Do interest rate options contain information about excess returns?," *Journal of Econometrics*.

Andersen, T. G., and L. Benzoni, 2008, "Stochastic Volatility," Discussion paper, Northwestern University.

Bansal, R., and I. Shaliastovich, 2009, "Risk and Return in Bond, Currency and Equity Markets," Discussion paper, Duke University.

Campbell, J., 1986, "A defense of the traditional hypotheses about the term structure of interest rates," *Journal of Finance*.

Campbell, J., 1987, "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18, 373–399.

Campbell, J., and R. Shiller, 1987, "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95, 1062–1087.

Cheridito, R., D. Filipovic, and R. Kimmel, 2007, "Market Price of Risk Specifications for Affine Models: Theory and Evidence," *Journal of Financial Economics*, 83, 123 – 170.

Collin-Dufresne, P., R. Goldstein, and C. Jones, 2008, "Identification of Maximal Affine Term Structure Models," *Journal of Finance*, LXIII, 743–795.

Collin-Dufresne, P., R. Goldstein, and C. Jones, 2009, "Can Interest Rate Volatility Be Extracted From the Cross Section of Bond Yields?," *Journal of Financial Economics*, 94, 47–66.

Cox, J., J. Ingersoll, and S. Ross, 1985, "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 53, 363–384.

Dai, Q., and K. Singleton, 2000, "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55, 1943–1978.

Duffee, G., 2002, "Term Premia and Interest Rates Forecasts in Affine Models," *Journal of Finance*, 57, 405–443.

Duffee, G., 2011, "Forecasting with the Term Structure: the Role of No-Arbitrage," Discussion paper, Johns Hopkins University.

Duffie, D., D. Filipovic, and W. Schachermayer, 2003, "Affine Processes and Applications in Finance," *Annals of Applied Probability*, 13, 984–1053.

Harvey, C. R., 1989, "Time-varying conditional covariances in tests of asset pricing models," *Journal of Financial Economics*, 24, 289–317.

Jacobs, K., and L. Karoui, 2009, "Conditional volatility in affine term-structure models: Evidence from Treasury and swap markets," *Journal of Financial Economics*, 91, 288–318.

Joslin, S., 2010, "Pricing and Hedging Volatility in Fixed Income Markets," Discussion paper, Working Paper, MIT.

Joslin, S., 2011a, "Can Unspanned Stochastic Volatility Models Explain the Cross Section of Bond Volatilities?," Discussion paper, USC.

Joslin, S., 2011b, "Risk Neutral Dynamics of the Term Structure," Discussion paper, MIT.

Joslin, S., A. Le, and K. Singleton, 2012, "Why Gaussian Macro-Finance Term Structure Models Are (Nearly) Unconstrained Factor-VARs," *Journal of Financial Economics, forthcoming*.

Joslin, S., K. Singleton, and H. Zhu, 2010, "A New Perspective on Gaussian DTSMs," Discussion paper, forthcoming, *Review of Financial Studies*.

Le, A., and K. Singleton, 2011, "A Robust Analysis of the Risk-Structure of Equilibrium Term Structures of Bond Yields," Discussion paper, University of North Carolina.

Le, A., K. Singleton, and J. Dai, 2010, "Discrete-Time Affine$^Q$ Term Structure Models with Generalized Market Prices of Risk," *Review of Financial Studies*, 23, 2184–2227.

Litterman, R., J. Scheinkman, and L. Weiss, 1991, "Volatility and the Yield Curve," *Journal of Fixed Income*, 1, 49–53.

Longstaff, F. A., and E. S. Schwartz, 1992, "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model," *Journal of Finance*, 47, 1259–1282.

Poon, S.-H., and C. W. Granger, 2003, "Forecasting Financial Market Volatility: A Review," *Journal of Economic Literature*, 41, 478–539.