Appendix for
Monetary Policy Shifts and the Term Structure

Andrew Ang
Columbia University and NBER

Jean Boivin
Bank of Canada, CIRPÉE, CIRANO and NBER

Sen Dong
Ortus Capital Management

Rudy Loo-Kung
Columbia University

This Version: June 2010

A Extended Model

In this section we consider the possibility that there is also a linear policy shock in addition to the time-varying policy shifts in the benchmark specification in equation (2) of the main paper:

$$r_t = \delta_0 + a_t g_t + b_t \pi_t + f_t^{\text{ext}}, \tag{A-1}$$

where we specify f_t^{ext} to be orthogonal to a_t and g_t.

The state vector now becomes $X_t = [g_t \ \pi_t \ a_t \ b_t \ f_t^{\text{ext}}]^\top$, which follows the stationary VAR:

$$X_t = \mu + \Phi X_{t-1} + \Sigma \varepsilon_t, \tag{A-2}$$

where $\varepsilon_t \sim$ IID $N(0, I)$. We parameterize Φ as

$$\Phi = \begin{pmatrix} \Phi_{gg} & \Phi_{g\pi} & \Phi_{ga} & 0 & \Phi_{gf} \\ \Phi_{\pi g} & \Phi_{\pi\pi} & 0 & \Phi_{\pi b} & \Phi_{\pi f} \\ \Phi_{ag} & 0 & \Phi_{aa} & 0 & 0 \\ 0 & \Phi_{b\pi} & 0 & \Phi_{bb} & 0 \\ 0 & 0 & 0 & 0 & \Phi_{ff} \end{pmatrix}, \tag{A-3}$$

and set Σ to take the following form:

$$\Sigma = \begin{pmatrix} \Sigma_{gg} & 0 & 0 & 0 & 0 \\ \Sigma_{\pi g} & \Sigma_{\pi\pi} & 0 & 0 & 0 \\ \Sigma_{ag} & \Sigma_{a\pi} & \Sigma_{aa} & 0 & 0 \\ \Sigma_{bg} & \Sigma_{b\pi} & \Sigma_{ba} & \Sigma_{bb} & 0 \\ 0 & 0 & 0 & 0 & \Sigma_{ff} \end{pmatrix}. \tag{A-4}$$

In this specification, we obtain the same quadratic short rate form as equation (6) in the main paper with Ω now taking the form

$$\Omega = \begin{pmatrix} 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{A-5}$$

Table **??** reports the estimates of the extended model, which are largely similar to the benchmark model reported in the main paper for the common parameters. The estimates of the long-run Fed responses to the output gap and inflation are also very similar across the benchmark and extended models. For example, the sample long-run inflation response is 1.075 in the full model and 1.117 in the benchmark model. Similar to the benchmark model, we find weak evidence of Granger-causality of past inflation to next-period b_t values. In the extended model, $\Phi_{b\pi} = 2.952$, with a posterior standard deviation of 2.334, compared to $\Phi_{b\pi} = 2.682$ with a posterior standard deviation of 2.595 in the benchmark model.

The extended model has a linear latent f_t^{ext} model in addition to time-varying policy loadings. Table **??** shows that the conditional volatility of f_t^{ext} is 0.007×10^{-3} which is several orders of magnitude smaller than the conditional volatilities of the policy shift parameters a_t and b_t, which are 0.003 and 0.036, respectively. The correlation between f_t^{ext} and the short rate is low at 0.228 and thus must of the movements in the short rate come from changing g_t and π_t interacted with monetary policy shifts. Put another way, time-varying a_t and b_t in the benchmark monetary policy shock, f_t^{bmk} accounts for most of the movements of the short rate, and the extended model's remainder monetary policy effect, f_t^{ext}, plays a relatively small role in explaining short rate movements.

This is also seen in a formal variance decomposition of the short rate, where f^{ext} accounts for 4.6% of the total short rate variance, which is computed as

$$1 - \frac{\text{var}_f(r)}{\text{var}(r)},$$

where $\text{var}_f(r)$ is the variance of the short rate computed through the sample where f^{ext} is set to be constant at its sample mean and $\text{var}(r)$ is the variance of the short rate in data.[1] In contrast, the corresponding variance decomposition for f^{bmk} in the benchmark model is 0.257. This is as expected. The benchmark model already allows the output gap and inflation response to vary over time and further allowing an independent f^{ext} factor in addition to the a_t and b_t variation indicates that the role of f^{ext} is small. The estimated time-series paths of a_t and b_t are very similar across the benchmark and the extended models, which is why we concentrate on the

[1] The variance decompositions of long-term yields in the extended model in terms of f^{ext} are also very small. For example, the variance decomposition for the 20-quarter yield for f^{ext} is 0.008.

benchmark model for looking at how Fed policy shifts have changed over time in the main paper.

B Bond Pricing

The price of a one-period zero-coupon bond is given by:

$$\begin{aligned} P_t^1 &= \exp(-r_t) = \exp(-\delta_0 - \delta_1^\top X_t - X_t^\top \Omega X_t) \\ &= \exp(A_1 + B_1^\top X_t + X_t^\top C_1 X_t), \end{aligned} \tag{B-1}$$

where $A_1 = -\delta_0$, $B_1 = -\delta_1 = -[0\ 0\ 0\ 0\ 1]^\top$ for the extended model in Section A, and $\delta_1 = 0$ for the benchmark model in the main paper. The matrix $C_1 = -\Omega$, with Ω given in equation (7) of the paper for the benchmark model and equation (**??**) for the extended model.

Under measure $\mathbb{Q}$, the price of a n-period zero-coupon bond, P_t^n, is:

$$\begin{aligned} P_t^n &= \mathrm{E}_t^{\mathbb{Q}}(\exp(-r_t) P_{t+1}^{n-1}) \\ &= \mathrm{E}_t^{\mathbb{Q}}(\exp\left(-r_t + A_{n-1} + B_{n-1}^\top X_{t+1} + X_{t+1}^\top C_{n-1} X_{t+1}\right)) \\ &= \exp\left(-r_t + A_{n-1} + B_{n-1}^\top(\mu^Q + \Phi^Q X_t) + (\mu^Q + \Phi^Q X_t)^\top C_{n-1}(\mu^Q + \Phi^Q X_t)\right) \\ &\quad \times \mathrm{E}_t^{\mathbb{Q}}(\exp\left((B_{n-1}^\top \Sigma + 2(\mu^Q + \Phi^Q X_t)^\top C_{n-1}\Sigma)\varepsilon_{t+1} + \varepsilon_{t+1}^\top \Sigma^\top C_{n-1}\Sigma\varepsilon_{t+1}\right)). \end{aligned}$$

To take the expectation, note that the expectation of the exponential of a quadratic Gaussian variable is given by:

$$\mathrm{E}[\exp(A^\top \epsilon + \epsilon^\top \Gamma \epsilon)] = \exp\left(-\frac{1}{2}\ln\det\left(I - 2\Psi\Gamma\right) + \frac{1}{2}A^\top(\Psi^{-1} - 2\Gamma)^{-1}A\right)$$

for $\epsilon \sim N(0, \Psi)$. This can be derived by general properties of Gaussian quadratic forms (see Mathai and Provost, 1992; Searle, 1997).

After taking the expectation and equating the terms with

$$P_t^n = \exp(A_n + B_n^\top X_t + X_t^\top C_n X_t), \tag{B-2}$$

the coefficients A_n, B_n, and C_n are given by the recursions:

$$\begin{aligned}
A_n &= -\delta_0 + A_{n-1} + B_{n-1}^\top \mu^Q + \mu^{Q\top} C_{n-1}\mu^Q - \frac{1}{2}\ln\det(I - 2\Sigma^\top C_{n-1}\Sigma) \\
&\quad + \frac{1}{2}(\Sigma^\top B_{n-1} + 2\Sigma^\top C_{n-1}\mu^Q)^\top (I - 2\Sigma^\top C_{n-1}\Sigma)^{-1}(\Sigma^\top B_{n-1} + 2\Sigma^\top C_{n-1}\mu^Q) \\
B_n^\top &= -\delta_1^\top + B_{n-1}^\top \Phi^Q + 2\mu^{Q\top} C_{n-1}\Phi^Q + 2(\Sigma^\top B_{n-1} \\
&\quad + 2\Sigma^\top C_{n-1}\mu^Q)^\top (I - 2\Sigma^\top C_{n-1}\Sigma)^{-1}\Sigma^\top C_{n-1}\Phi^Q \\
C_n &= -\Omega + \Phi^{Q\top} C_{n-1}\Phi^Q + 2(\Sigma^\top C_{n-1}\Phi^Q)^\top (I - 2\Sigma^\top C_{n-1}\Sigma)^{-1}(\Sigma^\top C_{n-1}\Phi^Q) \qquad \text{(B-3)}
\end{aligned}$$

If the model were specified in continuous time, then the recursions in equation (**??**) are versions of the ordinary differential equations derived by Ahn, Dittmar and Gallant (2002) on the bond pricing coefficients.

From equation (**??**), if we denote the yield on a zero-coupon bond with maturity n quarters as $y_t^n = -1/n \log P_t^n$, yields are quadratic functions of X_t:

$$y_t^n = a_n + b_n^\top X_t + X_t^\top c_n X_t, \qquad \text{(B-4)}$$

where $a_n = -A_n/n$, $b_n = -B_n/n$, and $c_n = -C_n/n$.

To compute conditional excess holding period returns, we use the exponential quadratic form for zero-coupon bond prices in equation (**??**) to write:

$$\begin{aligned}
xhpr_{t+1}^n &= \log \frac{P_{t+1}^{n-1}}{P_t^n} - r_t \\
&= A_{n-1} + B_{n-1}^\top X_{t+1} + X_{t+1}^\top C_{n-1} X_{t+1} - (A_n + B_n^\top X_t + X_t^\top C_n X_t) \\
&\qquad + (A_1 + B_1^\top X_t + X_t^\top C_1 X_t). \qquad \text{(B-5)}
\end{aligned}$$

Since $X_{t+1} \sim N(\mu + \Phi X_t, \Sigma\Sigma^\top)$, we can write the conditional expectation of this quadratic form as:

$$\mathrm{E}_t(X_{t+1}^\top C X_{t+1}) = tr(C\Sigma\Sigma^\top) + (\mu + \Phi X_t)^\top C(\mu + \Phi X_t).$$

This allows us to compute the expectation as:

$$\mathrm{E}_t[xhpr_{t+1}^n] = \bar{A}_n + \bar{B}_n^\top X_t + X_t^\top \bar{C}_n X_t, \qquad \text{(B-6)}$$

where

$$\begin{aligned}
\bar{A}_n &= A_{n-1} - A_n + A_1 + tr(C_{n-1}\Sigma\Sigma^\top) + \mu^\top C_{n-1}\mu + B_{n-1}^\top \mu \\
\bar{B}_n &= \Phi^\top B_{n-1} - B_n + B_1 + 2\Phi^\top C_{n-1}\mu \\
\bar{C}_n &= \Phi^\top C_{n-1}\Phi - C_n + C_1. \qquad \text{(B-7)}
\end{aligned}$$

C Estimating the Model

The model is estimated using a Bayesian Gibbs sampling algorithm. While there are several examples of these types of estimations for affine models (see, among others, Lamoureux and Witte, 2002; Johannes and Polson, 2005; Ang, Dong and Piazzesi, 2006; Dong, 2006), these cannot be directly employed to estimate the quadratic model because in an affine setting, drawing the latent factors requires a Kalman filter. The Kalman filter assumes that yields are linear functions of state variables, whereas they are non-linear functions in the quadratic model. In this appendix, we develop an acceptance-rejection algorithm to draw the latent factors without approximation.

To estimate the model, we assume that all yields, including the short rate, are measured with error. Specifically, we assume:

$$\tilde{y}_t^n = y_t^n + u_t^n, \tag{C-1}$$

where y_t^n is the model-implied yield, $\tilde{y}_t^n$ is the yield observed in data, and u_t^n IID $\sim N(0, \sigma_n^2)$, are additive measurement errors for all yields n. We report observation standard errors for the yields in Table **??** for the Constant Taylor Rule Model and the Policy-Shifts Model, which are analyzed in depth in the main paper.

For ease of notation, we group the macro variables as $M_t = [g_t\, \pi_t]^\top$ and the latent factors as $L_t = [a_t\, b_t\, f_t]^\top$ and rewrite the dynamics of $X_t = \left[M_t^\top\, L_t^\top\right]^\top$ in equation (**??**) as:

$$\begin{pmatrix} M_t \\ L_t \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix} \begin{pmatrix} M_{t-1} \\ L_{t-1} \end{pmatrix} + \begin{pmatrix} \Sigma_{11} & 0 \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{M,t} \\ \varepsilon_{L,t} \end{pmatrix}, \tag{C-2}$$

where $\varepsilon_t = (\varepsilon_{M,t}^\top\, \varepsilon_{L,t}^\top)^\top \sim$ IID $N(0, I)$ and Σ_{11} and Σ_{22} are lower triangular.

The parameters of the model are $\Theta = (\mu,\, \Phi,\, \Sigma,\, \delta_0, \delta_1,\, \Omega,\, \mu^Q,\, \Phi^Q,\, \sigma_u)$, where μ^Q and Φ^Q are parameters governing the state variable process under the risk neutral probability measure, and σ_u denotes the vector of observation error volatilities $\{\sigma_n\}$. We draw μ^Q and Φ^Q, but invert the prices of risk λ_0 and λ_1 using the relations:

$$\begin{aligned} \lambda_0 &= \Sigma^{-1}(\mu - \mu^Q) \\ \lambda_1 &= \Sigma^{-1}(\Phi - \Phi^Q). \end{aligned} \tag{C-3}$$

The latent factors $L_t = \{a_t\ b_t\ f_t\}$ are generated in each iteration of the Gibbs sampler. Note

that Ω and δ_1 are not estimated, given that they are fixed (see equation (**??**)). We also do not draw δ_0, but set this parameter to match the sample mean of the short rate in each iteration.

We simulate 500,000 observations in addition to using a burn-in period of 50,000. We sample every fifth observation to lower the serial correlation of the parameter draws. To check the adequacy of the number of simulations, we use the tests of Geweke (1992) and Raftery and Lewis (1992). For all parameters the simulation length is more than adequate except for some companion form parameters where the stationarity constraint is binding. These parameters are estimated to be always close to the unit circle no matter how many iterations are used as they capture the high persistence of the factors.

We now detail the procedure for drawing each of these variables. We denote the factors $X = \{X_t\}$ and the set of yields for all maturities in data as $\tilde{Y} = \{\tilde{y}_t^n\}$. Note that the model-implied yields $Y = \{y_t^n\}$ differ from the yields in data, Y, by observation error. By definition, $\tilde{Y} = Y + u$, where $u = \{u_t^n\}$ is the set of all observation errors for all yields. This notation also implies that the short rate in data, $\tilde{r}_t$, is the same as $\tilde{y}_t^1$.

C.1 Drawing the Latent Factors

We use a single-move algorithm based on Jacquier, Polson and Rossi (1994, 2004) adapted to our model. We derive a draw from the distribution $P(L_t|\tilde{Y}, L_{-t}, M)$, where L_t is the t-th observation of the latent factors, L_{-t} denotes all the latent factors except the t-th observation, and $\tilde{Y}$ and M are the complete time-series of yields and macro variables, respectively. We use the notation $\tilde{Y}_t$ and M_t to denote the t-th observation of the set of yields and macro variables. We draw the latent factors L_t conditional on the macro factors, yields, and other parameters.

From the Markov structure of the model, we can write:

$$P(L_t|L_{-t}, \tilde{Y}, M, \Theta) \propto P(L_t|L_{t-1}, M, \Theta)P(\tilde{Y}_t|L_t, M, \Theta)P(L_{t+1}|L_t, M, \Theta). \tag{C-4}$$

To keep the notation to a minimum, we write this as:

$$P(L_t|L_{-t}) \propto P(L_t|L_{t-1})P(\tilde{Y}_t|L_t)P(L_{t+1}|L_t).$$

Since M and Θ are treated as known, we can write the dynamics for L_t in equation (**??**) as

$$\begin{aligned} L_t &= \mu_2 + \Sigma_{12}\varepsilon_{M,t} + \Phi_{21}M_{t-1} + \Phi_{22}L_{t-1} + \Sigma_{22}\varepsilon_{L,t} \\ &= \mu_{L,t} + \Phi_L L_{t-1} + \Sigma_L \varepsilon_{L,t}, \end{aligned} \tag{C-5}$$

where $\mu_{L,t} = \mu_2 + \Sigma_{12}\varepsilon_{M,t}$, $\Phi_L = \Phi_{22}$, and $\Sigma_L = \Sigma_{22}$. Since M is observable and we hold Θ as fixed, $\mu_{L,t}$ is known at time t.

Each conditional distribution of the RHS of equation (**??**) is known. From equation (**??**) we have

$$P(L_t|L_{t-1}) \propto \exp\left(-\frac{1}{2}(L_t - \mu_{L,t} - \Phi_L L_{t-1})^\top (\Sigma_L \Sigma_L^\top)^{-1}(L_t - \mu_{L,t} - \Phi_L L_{t-1})\right). \quad \text{(C-6)}$$

Similarly, from the VAR in equation (**??**) we can write:

$$P(L_{t+1}|L_t) \propto \exp\left(-\frac{1}{2}(L_{t+1} - \mu_{L,t+1} - \Phi_L L_t)^\top (\Sigma_L \Sigma_L^\top)^{-1}(L_{t+1} - \mu_{L,t+1} - \Phi_L L_t)\right). \quad \text{(C-7)}$$

Finally, the likelihood of bond yields, $P(\tilde{Y}_t|L_t)$, is given by:

$$P(\tilde{Y}_t|L_t) \propto \exp\left(-\frac{1}{2}\sum_n \left[\frac{(\tilde{y}_t^n - (a_n + b_n^\top X_t + X_t^\top c_n X_t))^2}{\sigma_n^2}\right]\right), \quad \text{(C-8)}$$

where $X_t = [L_t^\top \; M_t^\top]$ and the summation is taken over yield maturities n. In the likelihood, the model-implied yield, $y_t^n = a_n + b_n^\top X_t + X_t^\top c_n X_t$, is given in equation (**??**), and σ_n^2 is the observation error variance of the yield of maturity n, which is assumed to be normally distributed.

We can combine equations (**??**)-(**??**) and complete the square to obtain:

$$\begin{aligned} P(L_t|L_{-t}) \;&\propto\; P(\tilde{Y}_t|L_t)\exp\Big(-\frac{1}{2}\Big[L_t^\top(\Phi_L'^\top(\Sigma_L\Sigma_L^\top)^{-1}\Phi_L + (\Sigma_L\Sigma_L^\top)^{-1})L_t \\ &\quad -2(L_{t+1}^\top(\Sigma_L\Sigma_L^\top)^{-1}\Phi_L + L_{t-1}^\top\Phi_L(\Sigma_L\Sigma_L^\top)^{-1} - \mu_{L,t+1}(\Sigma_L\Sigma_L^\top)^{-1}\Phi_L \\ &\quad +\mu_{L,t}(\Sigma_L\Sigma_L^\top)^{-1})L_t\Big]\Big) \\ &\propto\; P(\tilde{Y}_t|L_t)\exp\left(-\frac{1}{2}(L_t - \mu_t^*)^\top(\Sigma_t^*)^{-1}(L_t - \mu_t^*)\right) \end{aligned} \quad \text{(C-9)}$$

where

$$\begin{aligned} \Sigma_t^* &= \left(\Phi_L'^\top(\Sigma_L\Sigma_L^\top)^{-1}\Phi_L + (\Sigma_L\Sigma_L^\top)^{-1}\right)^{-1} \\ \mu_t^* &= \Sigma_t^*(L_{t+1}^\top(\Sigma_L\Sigma_L^\top)^{-1}\Phi_L + L_{t-1}^\top\Phi_L(\Sigma_L\Sigma_L^\top)^{-1} - \mu_{L,t+1}(\Sigma_L\Sigma_L^\top)^{-1}\Phi_L \\ &\qquad +\mu_{L,t}(\Sigma_L\Sigma_L^\top)^{-1})^\top. \end{aligned}$$

Since this distribution is not recognizable, we use a Metropolis draw. We draw a proposal from the distribution $N(\mu_t^*, \Sigma_t^*)$ and then the acceptance probability is based on the likelihood of $P(\tilde{Y}_t|L_t)$.

In the three-factor constant Taylor rule model, yields are linear functions of the factors and there is no need for the single-move algorithm. In this case, we employ the more efficient Carter and Kohn (1994) forward-backward algorithm to first run a Kalman filter forward and then sample f_t backwards. When the single-move algorithm is employed, it produces parameter values and posterior sample paths of f_t that are almost identical to those produced by the forward-backward algorithm. Since we specify the mean of f_t to be zero for identification, we set each generated draw of this factor to have a mean of zero.

In the benchmark four-factor specification, we additionally require that at each point in time both a_t and b_t are non-negative for purposes of identification.

In the extended five-factor model, we impose a prior for the draw of a_t and b_t period by period. Specifically, the prior used is given by the uniform distribution on the interval [k_t^p - $\sigma_{k,t}^p$; k_t^p + $\sigma_{k,t}^p$] for $k = a, b$; where k_t^p and $\sigma_{k,t}^p$ represent the posterior mean and standard deviation of factor k in period t from the estimated benchmark model. The motivation for imposing this prior is that we want the latent factor f_t^{ext} in the extended model to capture only for short rate and term structure movements not accounted for by the four factors $[g_t \ \pi_t \ a_t \ b_t]^\top$ since the model specifies f_t^{ext} as a factor orthogonal to $[g_t \ \pi_t \ a_t \ b_t]^\top$.

C.2 Drawing μ and Φ

We follow Johannes and Polson (2005) and explicitly differentiate between $\{\mu, \Phi\}$ under the real measure and $\{\mu^Q, \Phi^Q\}$ under the risk-neutral measure. As X_t follows a VAR in equation (**??**), we follow standard Gibbs sampling and use conjugate normal priors and posteriors for the draw of μ and Φ. We note that the posterior of μ and Φ conditional on X, $\tilde{Y}$ and the other parameters is:

$$\begin{aligned} P(\mu, \Phi|\Theta_-, X, \tilde{Y}) &\propto P(\tilde{Y}|\Theta, X)P(X|\mu, \Phi, \Sigma)P(\mu, \Phi) \\ &\propto P(\tilde{Y}|\Sigma, \delta_0, \delta_1, \mu^Q, \Phi^Q, \sigma_\eta, X)P(X|\mu, \Phi, \Sigma)P(\mu, \Phi) \\ &\propto P(X|\mu, \Phi, \Sigma)P(\mu, \Phi), \end{aligned} \tag{C-10}$$

where Θ_- denotes the set of all parameters except μ and Φ, and $P(X|\mu, \Phi, \Sigma)$ is the likelihood function of the VAR, which is normally distributed from the assumption of normality for the errors in the VAR. The validity of going from the first line to the second line is ensured by the

bond recursion in equation (**??**): given μ^Q and Φ^Q, the bond price is independent of μ and Φ. We specify the prior $P(\mu, \Phi)$ to be $N(0, 1000)$, which effectively represents an uninformative prior. We draw μ and Φ separately for each equation in the VAR system (**??**). Given that we impose the restriction that f_t is mean zero for identification, we set μ_f to zero.

C.3 Drawing $\Sigma\Sigma^\top$

To draw $\Sigma\Sigma^\top$, we note that the posterior of $\Sigma\Sigma^\top$ conditional on X, $\tilde{Y}$ and the other parameters is:

$$P(\Sigma\Sigma^\top|\Theta_-, X, \tilde{Y}) \propto P(\tilde{Y}|\Theta, X)P(X|\mu, \Phi, \Sigma)P(\Sigma\Sigma^\top), \tag{C-11}$$

where Θ_- denotes the set of all parameters except Σ. This posterior suggests an Independence Metropolis draw. We draw $\Sigma\Sigma^\top$ from the proposal density

$$q(\Sigma\Sigma^\top) = P(X \mid \mu, \Phi, \Sigma)P(\Sigma\Sigma^\top),$$

which is an Inverse Wishart (IW) distribution if we specify the prior $P(\Sigma\Sigma^\top)$ to be IW, so that $q(\Sigma\Sigma^\top)$ is an IW natural conjugate. The proposal draw $(\Sigma\Sigma^\top)^{m+1}$ for the $(m+1)$th draw is then accepted with probability α, where

$$\begin{aligned} \alpha &= \min\left\{ \frac{P((\Sigma\Sigma^\top)^{m+1} \mid \Theta_-, X, \tilde{Y})}{P((\Sigma\Sigma^\top)^{m} \mid \Theta_-, X, \tilde{Y})} \frac{q((\Sigma\Sigma^\top)^{m})}{q((\Sigma\Sigma^\top)^{m+1})}, 1 \right\} \\ &= \min\left\{ \frac{P(\tilde{Y} \mid (\Sigma\Sigma^\top)^{m+1}, \Theta_-, X)}{P(\tilde{Y} \mid (\Sigma\Sigma^\top)^{m}, \Theta_-, X)}, 1 \right\}, \end{aligned} \tag{C-12}$$

where $P(\tilde{Y}|\mu, \Phi, \Theta_-, X)$ is the likelihood function of all yields, including the short rate, which is normally distributed from the assumption of normality for the observation errors. From equation (**??**), α is just the ratio of the likelihoods of the new draw of $\Sigma\Sigma^\top$ relative to the old draw.

C.4 Drawing μ^Q and Φ^Q

We draw μ^Q and Φ^Q with a Random Walk Metropolis algorithm assuming a flat prior. We draw each parameter separately in μ^Q, and each row in Φ^Q. The accept/reject probability for the draws of μ^Q and Φ^Q is the ratio of the likelihood of bond yields based on candidate and last

draw of μ^Q and Φ^Q:

$$\begin{aligned}\alpha &= \min\left\{\frac{P((\mu^Q,\Phi^Q)^{m+1} \mid \Theta_-, X, \tilde{Y})}{P((\mu^Q,\Phi^Q)^{m} \mid \Theta_-, X, \tilde{Y})}\frac{q((\mu^Q,\Phi^Q)^{m})}{q((\mu^Q,\Phi^Q)^{m+1})}, 1\right\} \\ &= \min\left\{\frac{P(\tilde{Y} \mid (\mu^Q,\Phi^Q)^{m+1}, \Theta_-, X)}{P(\tilde{Y} \mid (\mu^Q,\Phi^Q)^{m}, \Theta_-, X)}, 1\right\}, \end{aligned} \tag{C-13}$$

In each iteration, we invert λ_0 and λ_1 and report the estimates of the prices of risk instead of μ^Q and Φ^Q. We discard non-stationary draws of Φ^Q.

C.5 Drawing σ_u

Drawing the variance of the observation errors, σ_u^2, is straightforward, because we can view the observation errors η as regression residuals from equation (**??**). We draw the observation variance $(\sigma_\eta^n)^2$ separately from each yield. We specify a conjugate prior $IG(0, 0.00001)$, so that the posterior distribution of σ_η^2 is a natural conjugate Inverse Gamma distribution. The prior information roughly translates into a 30bp bid-ask spread in Treasury securities, which is consistent with studies on the liquidity of spot Treasury market yields (see, for example, Fleming, 2003).

D Short Rate Variance Decomposition

For the short rate variance decomposition presented in Section 4.2 of the main paper, we write the short rate as

$$r_t = \delta_0 + \delta_1^\top X_t + X_t^\top \Omega^1 X_t + X_t^\top \Omega^2 X_t, \tag{D-1}$$

where the matrices Ω^1 and Ω^2 have elements $\Omega^1_{ga} = \Omega^1_{ag} = \Omega^2_{\pi b} = \Omega^2_{b\pi} = 0.5$ and zeros elsewhere. Then, the unconditional variance of the short rate can be decomposed as:

$$\begin{aligned}\text{var}(r_t) &= \text{var}(a_t g_t) + \text{var}(b_t \pi_t) + 2\text{cov}(a_t g_t,\, b_t \pi_t) \\ &= \text{var}(X_t^\top \Omega^1 X_t) + \text{var}(X_t^\top \Omega^2 X_t) + 2\text{cov}(X_t^\top \Omega^1 X_t,\, X_t^\top \Omega^2 X_t), \end{aligned} \tag{D-2}$$

where

$$\begin{aligned}
\text{var}(X_t^\top \Omega^1 X_t) &= 2tr(\Omega^1 \Sigma_X \Sigma_X^\top)^2 + 4\mu_X^\top \Omega^1 \Sigma_X \Sigma_X^\top \Omega^1 \mu_X \\
\text{var}(X_t^\top \Omega^2 X_t) &= 2tr(\Omega^2 \Sigma_X \Sigma_X^\top)^2 + 4\mu_X^\top \Omega^1 \Sigma_X \Sigma_X^\top \Omega^2 \mu_X \\
2\text{cov}(X_t^\top \Omega^1 X_t,\, X_t^\top \Omega^2 X_t) &= 4tr(\Omega^1 \Sigma_X \Sigma_X^\top \Omega^2 \Sigma_X \Sigma_X^\top) + 8\mu_X^\top \Omega^1 \Sigma_X \Sigma_X^\top \Omega^2 \mu_X,
\end{aligned}$$

and Σ_X is the unconditional covariance matrix of X_t implied by the VAR in equation (**??**).

E Impulse Responses

Since the yields are non-linear, we follow Gallant, Rossi and Tauchen (1993) and Koop, Pesaran and Potter (1996), among others, and compute the impulse response functions using simulation. We start with the sample series of data (g_t and π_t) and the posterior means of the latent factors (a_t and b_t) at each observation t. We term these points X_t^*. From the VAR in equation (**??**), we construct an orthogonalized error term ν_t by taking the Cholesky of $\Sigma\Sigma^\top$. To construct the impulse response for the jth variable of X_t, we first draw a shock v_t that represents a shock only to variable j from the error term distribution ν_t. From the points X_t^*, we construct a new series where each observation has been shocked by v_t, which we denote as $X_t^v = X_t^* + v_t$.

The impulse response functions are taken as the difference between the averaged response of the yields to the evolution of X_t^* without shocks to the evolution of the shocked X_t^v series:

$$\mathrm{E}(y_{t+k}^n | X_t^v) - \mathrm{E}(y_{t+k}^n | X_t^*).$$

Using the VAR in equation (**??**), we simulate out the value of X_{t+k}^v from X_t^v and the value of X_{t+k}^* from X_t^*. This is done at each observation t. Then, we construct the yields, y_{t+k}^n, from equation (**??**) corresponding to the state vectors X_{t+k}^v and X_{t+k}^*. We take values of $k = 1 \ldots 60$ quarters.

The impulse responses are computed at each observation by taking the average of the sample paths of y_{t+k}^n computed using X_{t+k}^v minus the average of the sample paths of y_{t+k}^n computed using X_{t+k}^*. We report the average of the impulse responses across all observations t. This procedure results in impulse responses that are identical to impulse responses computed for traditional VAR systems for large numbers of observations.

References

[1] Ahn, D. H., R. F. Dittmar, and A. R. Gallant, 2002, "Quadratic Term Structure Models: Theory and Evidence," *Review of Financial Studies*, 15, 243-288.

[2] Ang, A., S. Dong, and M. Piazzesi, 2006, "No-Arbitrage Taylor Rules," working paper, Columbia University.

[3] Carter, C. K., and R. Kohn, 1994, "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541-553.

[4] Dong, S., 2006, "Macro Variables Do Drive Exchange Rate Movements: Evidence from a No-Arbitrage Model," unpublished dissertation, Columbia University.

[5] Fleming, M. J., 2003, "Measuring Teasury Market Liquidity," *Federal Reserve Bank of New York Economic Policy Review*, 9, 83-108.

[6] Gallant, A. R., P. E. Rossi, and G. Tauchen, 1993, "Nonlinear Dynamic Structures," *Econometrica*, 61, 871-908.

[7] Geweke, J., 1992, "Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments" in J. M. Bernardo, J. Berger, A. P. David, and A. F. M. Smith, ed., *Bayesian Statistics 4*, Oxford University Press, Oxford, pp. 169-193.

[8] Jacquier, E., N. G. Polson, and P. E. Rossi, 1994, "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 371-417.

[9] Jacquier, E., N. G. Polson, and P. E. Rossi, 2004, "Bayesian Analysis of Fat-Tailed Stochastic Volatility Models with Correlated Errors," *Journal of Econometrics*, 122, 185-212.

[10] Johannes, M., and N. Polson, 2005, "MCMC Methods for Financial Econometrics," working paper, Columbia University.

[11] Koop, G., M. H. Pesaran, and S. M. Potter, 1996, "Impulse Response Analysis in Nonlinear Multivariate Models," *Journal of Econometrics*, 74, 119-147.

[12] Lamoureux, C., and D. Witte, 2002, "Empirical Analysis of the Yield Curve: The Information in the Data Viewed through the Window of Cox, Ingersoll, and Ross," *Journal of Finance*, 57, 1479-1520.

[13] Mathai, A. M., and S. B. Provost, 1992, *Quadratic Forms in Random Variables: Theory and Applications*, Marcel Dekker, New York.

[14] Raftery, A. E., and S. Lewis, 1992, "How Many Iterations in the Gibbs Sampler?" in J. M. Bernardo, J. Berger, A. P. David, and A. F. M. Smith, ed., *Bayesian Statistics 4*, Oxford University Press, Oxford, pp. 763-773.

[15] Searle, S. R., 1997, *Linear Models*, John Wiley & Sons.

Table 1: Extended Model Parameter Estimates

Short Rate Parameters

	$\bar{a}$		$\bar{b}$	
δ_0	Sample	VAR	Sample	VAR
0.003	0.374	0.388	1.075	1.101
(0.000)	(0.018)	(0.184)	(0.031)	(0.516)

VAR Parameters

		Φ				
	$\mu \times 1000$	g	π	a	b	f^{ext}
g	0.355	0.907	-0.080	0.001	0	-0.058
	(0.740)	(0.025)	(0.037)	(0.001)	–	(0.048)
π	0.116	0.064	0.989	0	0.000	-0.014
	(0.165)	(0.011)	(0.007)	–	(0.000)	(0.015)
a	21.23	-1.267	0	0.945	0	0
	(11.07)	(0.693)	–	(0.027)	–	–
b	57.74	0	2.952	0	0.925	0
	(36.87)	–	(2.334)	–	(0.026)	–
f^{ext}	0	0	0	0	0	0.672
	–	–	–	–	–	(0.107)
		Volatility $\times$1000/Correlation				
g		0.004	0.024	0.006	-0.009	0
		(0.000)	(0.068)	(0.082)	(0.078)	–
π		0.024	0.001	-0.359	-0.185	0
		(0.068)	(0.000)	(0.068)	(0.086)	–
a		0.006	-0.359	3.413	0.854	0
		(0.082)	(0.068)	(0.708)	(0.068)	–
b		-0.009	-0.185	0.854	35.818	0
		(0.078)	(0.086)	(0.068)	(6.437)	–
f^{ext}		0	0	0	0	0.007
		–	–	–	–	(0.001)

Table **??** Continued

Risk Premia Parameters

		λ_1				
	λ_0	g	π	a	b	f^{ext}
g	0.711	18.63	65.50	3.130	0	-28.27
	(0.925)	(11.97)	(18.12)	(0.707)	–	(22.72)
g	-3.154	30.15	164.15	-0.076	-0.390	-15.49
	(0.570)	(11.94)	(12.26)	(0.220)	(0.156)	(15.94)
g	-0.896	-8.92	63.33	-0.422	-0.152	-6.031
	(0.391)	(15.15)	(14.84)	(0.537)	(0.071)	(7.302)
g	0.360	27.04	-14.59	0.878	-0.712	3.899
	(0.609)	(24.48)	(30.25)	(0.997)	(0.335)	(6.053)
f^{ext}	0.063	0	0	0	0	79.33
	(0.072)	–	–	–	–	(52.67)

Observation Error Standard Deviation

	$n=1$	$n=4$	$n=8$	$n=12$	$n=16$	$n=20$
σ_u^n	0.081	0.038	0.024	0.018	0.022	0.024
	(0.024)	(0.004)	(0.002)	(0.001)	(0.002)	(0.002)

The table lists parameter values for the extended model for the factors $X_t = [g_t\, \pi_t\, a_t\, b_t\, f_t^{\text{ext}}]^\top$ Any parameters without standard errors are not estimated. We report the posterior mean and posterior standard deviation (in parentheses) of each parameter. For the short rate parameters, we report two estimated long-run means $\bar{a}$ and $\bar{b}$ for a_t and b_t, respectively. The "sample" mean is the posterior mean of the latent factors averaged across the sample. For the "population" mean we compute the population mean of the latent factors implied by the VAR parameters in each iteration and report the posterior average. In the Volatility/Correlation matrix, we report standard deviations of each factor along the diagonal multiplied by 1000 and correlations between the factors on the off-diagonal elements. The zero entries in the λ_1 matrix result from the companion form Φ taking the form of equation (**??**) under both the risk neutral and the real measure. The sample period is June 1952 to December 2007 and the data frequency is quarterly.

Table 2: Observation Error Standard Deviations σ_u^n

$n = 1$	$n = 4$	$n = 8$	$n = 12$	$n = 16$	$n = 20$
Constant Taylor Rule Model					
0.177	0.106	0.052	0.032	0.044	0.058
(0.009)	(0.006)	(0.003)	(0.002)	(0.003)	(0.003)
Policy-Shifts Model					
0.168	0.069	0.030	0.021	0.024	0.026
(0.043)	(0.026)	(0.010)	(0.003)	(0.003)	(0.002)

The table lists the observation error standard deviations (see equation (**??**)) for the Constant Taylor Rule Model and the Policy-Shifts Model, which have other parameters reported in Tables 2 and 3, respectively, of the main paper.