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:

\[ r_t = \delta_0 + a_t g_t + b_t \pi_t + f_{t}^{\text{ext}}, \]  

(A-1)

where we specify \( f_{t}^{\text{ext}} \) 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:

\[ X_t = \mu + \Phi X_{t-1} + \Sigma \varepsilon_t, \]  

(A-2)

where \( \varepsilon_t \sim \text{IID } N(0, I) \). We parameterize \( \Phi \) as

\[ \Phi = \begin{pmatrix} \Phi_{gg} & \Phi_{g\pi} & \Phi_{ga} & 0 & \Phi_{gf} \\ \Phi_{bg} & \Phi_{\pi\pi} & 0 & \Phi_{\pi b} & \Phi_{\pi f} \\ 0 & \Phi_{a\pi} & 0 & \Phi_{ab} & 0 \\ 0 & 0 & 0 & 0 & \Phi_{ff} \end{pmatrix}, \]  

(A-3)

and set \( \Sigma \) to take the following form:

\[ \Sigma = \begin{pmatrix} \Sigma_{gg} & 0 & 0 & 0 & 0 \\ \Sigma_{bg} & \Sigma_{\pi \pi} & 0 & 0 & 0 \\ \Sigma_{ag} & \Sigma_{a\pi} & \Sigma_{aa} & 0 & 0 \\ \Sigma_{bg} & \Sigma_{b\pi} & \Sigma_{ba} & \Sigma_{bb} & 0 \\ 0 & 0 & 0 & 0 & \Sigma_{ff} \end{pmatrix}. \]  

(A-4)

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

\[ \Omega = \begin{pmatrix} 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \]  

(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, $\Phi_{bt} = 2.952$, with a posterior standard deviation of 2.334, compared to $\Phi_{bt} = 2.682$ 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 $0.007 \times 10^{-3}$ 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 most of the movements in the short rate come from changing $g_t$ and $\pi_t$ 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 $f_t^{ext}$ 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 $f_t^{ext}$ is set to be constant at its sample mean and $\text{var}(r)$ is the variance of the short rate in data. In contrast, the corresponding variance decomposition for $f_t^{bmk}$ 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 $f_t^{ext}$ factor in addition to the $a_t$ and $b_t$ variation indicates that the role of $f_t^{ext}$ 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 $f_t^{ext}$ are also very small. For example, the variance decomposition for the 20-quarter yield for $f_t^{ext}$ 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:

\[
P_{1t} = \exp(-r_t) = \exp(-\delta_0 - \delta_1^T X_t - X_t^T \Omega X_t)
\]

\[= \exp(A_1 + B_1^T X_t + X_t^T C_1 X_t), \quad (B-1)
\]

where \(A_1 = -\delta_0, B_1 = -\delta_1 = [-0 \ 0 \ 0 \ 0 \ 1]^T\) 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 \(\Omega\) 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, \(P_{nt}\), is:

\[
P_{nt} = \mathbb{E}_t^Q (\exp(-r_t) P_{n+1,t})
\]

\[= \mathbb{E}_t^Q (\exp (-r_t + A_{n-1} + B_{n-1}^T X_{n+1} + X_{n+1}^T C_{n-1} X_{n+1}))
\]

\[= \exp (-r_t + A_{n-1} + B_{n-1}^T (\mu^Q + \Phi^Q X_t) + (\mu^Q + \Phi^Q X_t)^T C_{n-1} (\mu^Q + \Phi^Q X_t))
\]

\[\times \mathbb{E}_t^Q (\exp ((B_{n-1}^T \Sigma + 2(\mu^Q + \Phi^Q X_t)^T C_{n-1} \Sigma) \epsilon_{n+1} + \epsilon_{n+1}^T \Sigma^T C_{n-1} \Sigma \epsilon_{n+1}))).
\]

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

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

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

\[
P_{nt} = \exp(A_n + B_n^T X_t + X_t^T C_n X_t), \quad (B-2)
\]
the coefficients $A_n$, $B_n$, and $C_n$ are given by the recursions:

$$A_n = -\delta_n + A_{n-1} + B_{n-1}^T \mu^Q + \mu^Q C_{n-1} \mu^Q - \frac{1}{2} \ln \det(I - 2\Sigma^\top C_{n-1} \Sigma)$$

$$B_n = -\delta_n^T + B_{n-1}^T \Phi^Q + 2\mu^Q C_{n-1} \Phi^Q + 2(\Sigma^T B_{n-1})$$

$$C_n = -\Omega + \Phi^Q C_{n-1} \Phi^Q + 2(\Sigma^T C_{n-1} \Phi^Q) (I - 2\Sigma^T C_{n-1} \Sigma)^{-1} (\Sigma^T C_{n-1} \Phi^Q)$$

(B-3)

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 $y^n_t = -1/n \log P^n_t$, yields are quadratic functions of $X_t$:

$$y^n_t = a_n + b_n^T X_t + X_t^T c_n X_t,$$  

(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:

$$x_{hpr}^{n+1} = \log \frac{P^{n+1}_t}{P^n_t} - r_t$$

$$= A_{n+1} + B_{n+1}^T X_{t+1} + X_{t+1}^T C_{n+1} X_{t+1} - (A_n + B_n^T X_t + X_t^T C_n X_t)$$

$$+ (A_1 + B_1^T X_t + X_t^T C_1 X_t).$$  

(B-5)

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

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

This allows us to compute the expectation as:

$$E_t[x_{hpr}^{n+1}] = \bar{A}_n + \bar{B}_n^T X_t + X_t^T \bar{C}_n X_t,$$  

(B-6)

where

$$\bar{A}_n = A_{n+1} - A_n + A_1 + tr(C_{n+1} \Sigma \Sigma^\top) + \mu^T C_{n+1} \mu + B_{n+1}^T \mu$$

$$\bar{B}_n = \Phi^T B_{n+1} - B_n + B_1 + 2\Phi^T C_{n+1} \mu$$

$$\bar{C}_n = \Phi^T C_{n+1} \Phi - C_n + C_1.$$  

(B-7)
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:

\[ \tilde{y}_n^t = y_n^t + u_n^t, \]  

where \( y_n^t \) is the model-implied yield, \( \tilde{y}_n^t \) is the yield observed in data, and \( u_n^t \) 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 \( X_t = [M_t^{\top} \, L_t^{\top}]^{\top} \) in equation (??) as:

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

where \( \varepsilon_t = (\varepsilon_{M,t}^{\top} \, \varepsilon_{L,t}^{\top})^{\top} \) 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 \( X_t = [M_t^{\top} \, L_t^{\top}]^{\top} \) in equation (??) as:

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

where \( \varepsilon_t = (\varepsilon_{M,t}^{\top} \, \varepsilon_{L,t}^{\top})^{\top} \) 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.

The latent factors \( L_t = \{a_t \, b_t \, f_t\} \) are generated in each iteration of the Gibbs sampler. Note
that $\Omega$ and $\delta_1$ are not estimated, given that they are fixed (see equation (??)). We also do not draw $\delta_0$, 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 $\tilde{Y} = \{\tilde{y}_t\}$. Note that the model-implied yields $Y = \{y_t\}$ differ from the yields in data, $\tilde{Y}$, by observation error. By definition, $\tilde{Y} = Y + u$, where $u = \{u_t\}$ 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:

$$P(L_t|L_{-t}, \tilde{Y}, M, \Theta) \propto P(L_t|L_{t-1}, M, \Theta)P(\tilde{Y}_t|L_t, M, \Theta)P(L_{t+1}|L_t, M, \Theta).$$  \hfill (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 $\Theta$ are treated as known, we can write the dynamics for $L_t$ in equation (??) as

$$L_t = \mu_2 + \Sigma_{12}\varepsilon_{M,t} + \Phi_{21}M_{t-1} + \Phi_{22}L_{t-1} + \Sigma_{22}\varepsilon_{L,t}$$

$$= \mu_{L,t} + \Phi_L L_{t-1} + \Sigma_L\varepsilon_{L,t},$$  \hfill (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 $\Theta$ as fixed, $\mu_{L,t}$ is known at time $t$.

Each conditional distribution of the RHS of equation (C-9) is known. From equation (C-9) we have

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

Similarly, from the VAR in equation (C-7) we can write:

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

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

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

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 (C-6), and $\sigma_n^2$ is the observation error variance of the yield of maturity $n$, which is assumed to be normally distributed.

We can combine equations (C-6)-(C-8) and complete the square to obtain:

$$P(L_t | L_{-t}) \propto P(\tilde{Y}_t | L_t) \exp \left(-\frac{1}{2} \left[ L_t^\top (\Phi_{L}^\top (\Sigma_L^{-1}) + (\Sigma_L^{-1}) L_t 
- 2(L_{t+1}^\top (\Sigma_L^{-1}) - \mu_{L,t+1}^\top (\Sigma_L^{-1}) - \mu_{L,t}^\top (\Sigma_L^{-1}) L_t 
+ \mu_{L,t} (\Sigma_L^{-1}) L_t) \right] \right) \quad (C-9)$$

where

$$\Sigma_t^* = (\Phi_{L}^\top (\Sigma_L^{-1}) + (\Sigma_L^{-1})^\top)^{-1} \quad (C-9)$$

and

$$\mu_t^* = \Sigma_t^* (L_{t+1}^\top (\Sigma_L^{-1}) - \mu_{L,t+1}^\top (\Sigma_L^{-1}) - \mu_{L,t}^\top (\Sigma_L^{-1}) L_t 
+ \mu_{L,t} (\Sigma_L^{-1}) L_t)^\top. \quad (C-9)$$

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 $P(\tilde{Y}_t | L_t)$. 7
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 $[k_{p,t}^p - \sigma_{k,t}^p ; k_{p,t}^p + \sigma_{k,t}^p]$ for $k = a, b$; where $k_{p,t}^p$ 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 $\mu$ and $\Phi$

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 $\mu$ and $\Phi$. We note that the posterior of $\mu$ and $\Phi$ conditional on $X$, $\bar{Y}$ and the other parameters is:

$$P(\mu, \Phi|\Theta_-, X, \bar{Y}) \propto P(\bar{Y}|\Theta, X)P(X|\mu, \Phi, \Sigma)P(\mu, \Phi)$$
$$\propto P(\bar{Y}|\Sigma, \delta_0, \delta_1, \mu^Q, \Phi^Q, \sigma_q, X)P(X|\mu, \Phi, \Sigma)P(\mu, \Phi)$$
$$\propto P(X|\mu, \Phi, \Sigma)P(\mu, \Phi),$$

where $\Theta_-$ denotes the set of all parameters except $\mu$ and $\Phi$, 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 $\mu^Q$ and $\Phi^Q$, the bond price is independent of $\mu$ and $\Phi$. We specify the prior $P(\mu, \Phi)$ to be $N(0, 1000)$, which effectively represents an uninformative prior. We draw $\mu$ and $\Phi$ separately for each equation in the VAR system (??). Given that we impose the restriction that $f_t$ is mean zero for identification, we set $\mu_f$ 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:

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

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

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

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 $\alpha$, where

$$\alpha = \min \left\{ \frac{P((\Sigma \Sigma^\top)^{m+1} | \Theta_-, X, \tilde{Y}) q((\Sigma \Sigma^\top)^m)}{P((\Sigma \Sigma^\top)^m | \Theta_-, X, Y) q((\Sigma \Sigma^\top)^{m+1})}, 1 \right\},$$

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 (??), $\alpha$ is just the ratio of the likelihoods of the new draw of $\Sigma \Sigma^\top$ relative to the old draw.

C.4 Drawing $\mu^Q$ and $\Phi^Q$

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

$$
\alpha = \min \left\{ \frac{P((\mu^Q, \Phi^Q)^{m+1} | \Theta_-, X, \hat{Y})}{P((\mu^Q, \Phi^Q)^{m} | \Theta_-, X, \hat{Y})} \cdot q((\mu^Q, \Phi^Q)^{m+1}), 1 \right\}
$$

$$
= \min \left\{ \frac{P(\hat{Y} | (\mu^Q, \Phi^Q)^{m+1}, \Theta_-, X)}{P(\hat{Y} | (\mu^Q, \Phi^Q)^{m}, \Theta_-, X)} \cdot q, 1 \right\},
$$

(C-13)

In each iteration, we invert $\lambda_0$ and $\lambda_1$ and report the estimates of the prices of risk instead of $\mu^Q$ and $\Phi^Q$. We discard non-stationary draws of $\Phi^Q$.

C.5 Drawing $\sigma_u$

Drawing the variance of the observation errors, $\sigma_u^2$, is straightforward, because we can view the observation errors $\eta$ as regression residuals from equation (??). We draw the observation variance $(\sigma_\eta^2)^2$ separately from each yield. We specify a conjugate prior $IG(0, 0.00001)$, so that the posterior distribution of $\sigma_\eta^2$ 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

$$
r_t = \delta_0 + \delta_1^T X_t + X_t^T \Omega^1 X_t + X_t^T \Omega^2 X_t,
$$

(D-1)

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

$$
\text{var}(r_t) = \text{var}(a_t g_t) + \text{var}(b_t \pi_t) + 2 \text{cov}(a_t g_t, b_t \pi_t)
$$

$$
= \text{var}(X_t^T \Omega^1 X_t) + \text{var}(X_t^T \Omega^2 X_t) + 2 \text{cov}(X_t^T \Omega^1 X_t, X_t^T \Omega^2 X_t),
$$

(D-2)
where

\[
\text{var}(X_t^\top \Omega^1 X_t) = 2 \text{tr}(\Omega^1 \Sigma_X \Sigma_X^\top)^2 + 4 \mu_X^\top \Omega^1 \Sigma_X \Sigma_X^\top \Omega^1 \mu_X
\]

\[
\text{var}(X_t^\top \Omega^2 X_t) = 2 \text{tr}(\Omega^2 \Sigma_X \Sigma_X^\top)^2 + 4 \mu_X^\top \Omega^1 \Sigma_X \Sigma_X^\top \Omega^2 \mu_X
\]

\[
2 \text{cov}(X_t^\top \Omega^1 X_t, X_t^\top \Omega^2 X_t) = 4 \text{tr}(\Omega^1 \Sigma_X \Sigma_X^\top \Omega^2 \Sigma_X \Sigma_X^\top) + 8 \mu_X^\top \Omega^1 \Sigma_X \Sigma_X^\top \Omega^2 \mu_X,
\]

and \(\Sigma_X\) 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 \(\pi_t\)) 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 \(\nu_t\) by taking the Cholesky of \(\Sigma \Sigma^\top\). To construct the impulse response for the \(j\)th variable of \(X_t\), we first draw a shock \(v_t\) that represents a shock only to variable \(j\) from the error term distribution \(\nu_t\). 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:

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

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


Table 1: Extended Model Parameter Estimates

Short Rate Parameters

<table>
<thead>
<tr>
<th></th>
<th>Sample</th>
<th>VAR</th>
<th>Sample</th>
<th>VAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_0$</td>
<td>0.003</td>
<td>0.374</td>
<td>0.388</td>
<td>1.075</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.018)</td>
<td>(0.184)</td>
<td>(0.031)</td>
</tr>
</tbody>
</table>

VAR Parameters

<table>
<thead>
<tr>
<th></th>
<th>$\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu \times 1000$</td>
<td>$g$</td>
</tr>
<tr>
<td>$g$</td>
<td>0.355</td>
</tr>
<tr>
<td></td>
<td>(0.740)</td>
</tr>
<tr>
<td>$\pi$</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>(0.165)</td>
</tr>
<tr>
<td>$a$</td>
<td>21.23</td>
</tr>
<tr>
<td></td>
<td>(11.07)</td>
</tr>
<tr>
<td>$b$</td>
<td>57.74</td>
</tr>
<tr>
<td></td>
<td>(36.87)</td>
</tr>
<tr>
<td>$f^{ext}$</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>–</td>
</tr>
</tbody>
</table>

Volatility $\times 1000$/Correlation

<table>
<thead>
<tr>
<th></th>
<th>$\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g$</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
</tr>
<tr>
<td>$\pi$</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
</tr>
<tr>
<td>$a$</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>(0.082)</td>
</tr>
<tr>
<td>$b$</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(0.078)</td>
</tr>
<tr>
<td>$f^{ext}$</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>–</td>
</tr>
</tbody>
</table>
Table ?? Continued

Risk Premia Parameters

<table>
<thead>
<tr>
<th></th>
<th>$\lambda_0$</th>
<th>$g$</th>
<th>$\pi$</th>
<th>$a$</th>
<th>$b$</th>
<th>$f^{\text{ext}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g$</td>
<td>0.711</td>
<td>18.63</td>
<td>65.50</td>
<td>3.130</td>
<td>0</td>
<td>-28.27</td>
</tr>
<tr>
<td></td>
<td>(0.925)</td>
<td>(11.97)</td>
<td>(18.12)</td>
<td>(0.707)</td>
<td>–</td>
<td>(22.72)</td>
</tr>
<tr>
<td>$g$</td>
<td>-3.154</td>
<td>30.15</td>
<td>164.15</td>
<td>-0.076</td>
<td>-0.390</td>
<td>-15.49</td>
</tr>
<tr>
<td></td>
<td>(0.570)</td>
<td>(11.94)</td>
<td>(12.26)</td>
<td>(0.220)</td>
<td>(0.156)</td>
<td>(15.94)</td>
</tr>
<tr>
<td>$g$</td>
<td>-0.896</td>
<td>-8.92</td>
<td>63.33</td>
<td>-0.422</td>
<td>-0.152</td>
<td>-6.031</td>
</tr>
<tr>
<td></td>
<td>(0.391)</td>
<td>(15.15)</td>
<td>(14.84)</td>
<td>(0.537)</td>
<td>(0.071)</td>
<td>(7.302)</td>
</tr>
<tr>
<td>$g$</td>
<td>0.360</td>
<td>27.04</td>
<td>-14.59</td>
<td>0.878</td>
<td>-0.712</td>
<td>3.899</td>
</tr>
<tr>
<td></td>
<td>(0.609)</td>
<td>(24.48)</td>
<td>(30.25)</td>
<td>(0.997)</td>
<td>(0.335)</td>
<td>(6.053)</td>
</tr>
<tr>
<td>$f^{\text{ext}}$</td>
<td>0.063</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>79.33</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>(52.67)</td>
</tr>
</tbody>
</table>

Observation Error Standard Deviation

<table>
<thead>
<tr>
<th></th>
<th>$n = 1$</th>
<th>$n = 4$</th>
<th>$n = 8$</th>
<th>$n = 12$</th>
<th>$n = 16$</th>
<th>$n = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_u^n$</td>
<td>0.081</td>
<td>0.038</td>
<td>0.024</td>
<td>0.018</td>
<td>0.022</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
</tbody>
</table>

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 $\lambda_1$ matrix result from the companion form $\Phi$ 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 $\sigma_u^n$

<table>
<thead>
<tr>
<th></th>
<th>$n = 1$</th>
<th>$n = 4$</th>
<th>$n = 8$</th>
<th>$n = 12$</th>
<th>$n = 16$</th>
<th>$n = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant Taylor Rule Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.177</td>
<td>0.106</td>
<td>0.052</td>
<td>0.032</td>
<td>0.044</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.006)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td><strong>Policy-Shifts Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.168</td>
<td>0.069</td>
<td>0.030</td>
<td>0.021</td>
<td>0.024</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.026)</td>
<td>(0.010)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
</tbody>
</table>

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.