Estimating a Structural Macro Finance Model of Term Structure

Taeyoung Doh

*University of Pennsylvania**

October 23, 2006

Abstract

This paper aims to estimate a structural macro finance model of term structure based on the approximate solution of a standard dynamic general equilibrium model with nominal rigidities. To capture the nonlinearity that can contribute to a more precise analysis of the dynamics of macro variables and the yield curve, we apply a second order approximation of the equilibrium conditions to solve the model. New closed-form solutions for bond yields are proposed to make estimation practically feasible. We estimate the model based on US data by Bayesian methods. Our results provide clear macroeconomic interpretations of term structure factors such as level, slope, and curvature. Also, our analysis favors the explanation that the downward trend of the level of the yield curve and term premia after the early 1980s is primarily related with changes in monetary policy. However, the impact of the Volcker disinflation policy on this decline is found to be delayed.

JEL CLASSIFICATION: C32, E43, G12

KEY WORDS: Bayesian Econometrics, DSGE Model, Term Structure of Interest Rates

*Taeyoung Doh : Department of Economics, McNeil Building, 3718 Locust Walk, Philadelphia, PA 19104-6297; email: *doht@econ.upenn.edu*

1 Introduction

Recent empirical studies on the macroeconomics of term structure(Ang and Piazzesi (2003), Diebold et al. (2006) among others) show a close link between macroeconomic variables and bond prices. They incorporate macroeconomic variables into factors that explain the entire set of yields. In spite of the inclusion of macro factors into models, they still retain unobserved term structure factors to explain the yield curve dynamics. These latent factors are often called "level,", "slope,", and "curvature,". Since latent factors themselves are not amenable to economic interpretations, many papers search for macroeconomic variables that are highly correlated with them. For example, Diebold et al. (2006) and Rudebusch and Wu (2004) associate the level with inflation or the inflation objective of the central bank. The slope is found to be linked with capacity utilization. But the corresponding macro variable for the curvature is not provided in this literature.

By contrast, clear economic interpretations of latent factors are possible when we use equilibrium bond yields implied by a structural macro model in the analysis since model implied bond yields are entirely determined by macroeconomic fundamentals. While the estimation of a dynamic stochastic general equilibrium (DSGE) model is by now a widespread practice among researchers, bond yields data are rarely used in the estimation even though the model implies equilibrium bond yields. When bond yields data are used, risk premia are restricted to be time-invariant in most cases by the nature of the popular log-linear approximation of the macro model. The log-linear approximation eliminates nonlinear terms generating time-varying risk premia. Otherwise, the underlying macro model is used to obtain only macro dynamics while bond prices are obtained from an exogenously given stochastic discount factor. Nonetheless both ways ignore nonlinear terms implied by the model. However, recent developments in the nonlinear analysis of DSGE models (An and Schorfheide (2006), Fernández-Villaverde and Rubio-Ramírez (2006) etc.) show that DSGE models solved with nonlinear methods can be estimated. This paper extends their framework and considers bond yields data as well as macro data. Specifically, we study term structure implications of a New Keynesian DSGE model solved with a second order approximation to equilibrium conditions. New closed-form expressions for bond yields are proposed to make the joint estimation practically feasible. Our solutions for bond yields are much faster than numerical solutions in Ravenna and Seppälä (2006) and more accurate than closed-form solutions in Hördahl et al. (2005b). No exogenous components accounting for time varying risk premia added to the model.

Equilibrium bond yields derived in this paper are regarded as discrete time counterparts of bond yields in a continuous time quadratic term structure model. (e.g. Ahn et al. (2002)) The no-arbitrage restriction in a quadratic term structure model does not imply constant risk premia which are implied by the affine term structure model derived from the log-linear approximation. We magnify the model's ability to generate time varying risk premia by adding an autoregressive conditional heteroskedasticity (ARCH)effect of a structural shock. Combining the approximate macro dynamics with the quadratic term structure, we can represent the whole model as a special case of nonlinear state space models.

We estimate the model based on US data and obtain estimates of unobserved macro factors. We find the level of the yield curve is strongly related to a persistent shock to the target inflation of the central bank, while the curvature of the yield curve is dominated by a monetary policy shock. The slope of the yield curve is highly related with a markup shock which is a real disturbance. The association of the slope with a real factor is in line with empirical findings in Diebold et al. (2006) and Evans and Marshall (2002) Also, the time variation of the term premium is found to be highly correlated with that of the target inflation shock. Our analysis indicates that the downward trend of the target inflation after the early 1980s drove down both the level of the yield curve and the term premium at the same time. In spite of these successes in identifying macro sources of the yield curve dynamics, the model fails to match some predictive moments. The mean term premium is underpredicted in the model compared to the actual sample moment.

The target inflation estimated by using both macro and term structure data sheds new lights on the Volcker disinflation episode from 1979:QIV to 1982:QIV.

The rapid fall of inflation after the early 1980s is regarded as evidence on the success of the Volcker disinflation policy as suggested by Clarida et al. and (2000) Sargent et al. (2006). However, the expected inflation reflected in long term bond rates declined more slowly than the actual inflation. The long term rates showed a sharp peak in the second quarter of 1984 even though the actual inflation did not increase much. Goodfriend (1993) and Goodfriend and King (2005) observed the initial Volcker disinflation policy was not so credible and the credibility was only established after fighting the inflation scare in the second quarter of 1984. In line with their observations, the estimated target inflation in our empirical analysis indicates that the sluggish responses of long term rates to the Volcker disinflation policy were related to the lingering suspicion of the disinflation policy. The finding is further confirmed by the high correlation between estimated target inflation and inflation forecasts from survey data.

Our work is closely related to the literature on term structure implications of the log-linearized New Keynesian model. Hördahl et.al. (2005a) and Rudebusch and Wu (2004) estimate the log-linearized New Keynesian model jointly with macro and term structure data. However, they do not use stochastic discount factor implied by the underlying structural model. In Bekaert et al. (2005), the stochastic discount factor derived from the log-linearized New Keynesian model is used to compute equilibrium bond yields. They find that the level of the yield curve is driven by the target inflation of the central bank while the slope and the curvature are mainly related to monetary policy shocks. However, the conclusion that all the term structure factors are dominated by nominal disturbances seems to be extreme and at odds with the conclusion of other papers relating some of the time variation of the yield curve with real factors, such asDiebold et al. (2006), Evans and Marshall (2002), and Ludvigson and Ng (2005) etc. Compared with the micro evidence on nominal rigidities, the estimate of the degree of price stickiness in Bekaert et al. (2005) seems to be bigger. Our estimate is more in line with the estimate from micro data. For the degree of nominal rigidities compatible with the micro evidence, there is a room for real disturbances to explain the yield curve dynamics.

The paper is organized as follows. Section 2 describes the model economy and presents the second order approximation to the solution of the model. Section 3 proposes a new method to construct measurement equations for bond yields under a second order approximation. Section 4 describes the econometric methodology used. Section 5 discusses the prior-posterior analysis based on the estimation of the model using US dataset. Section 6 discusses the identification of term structure factors. Section 7 provides implications of the yield curve for impacts of the Volcker disinflation policy. Section 8 concludes. Technical details are provided in the appendix.

2 Model Economy

A small scale New Keynesian model which has been widely used in business cycle and monetary policy analysis is considered in this paper. Woodford (2003) provides an excellent discussion of many variations of this kind of model. Our model closely follows the prototypical New Keynesian model studied in An and Schorfheide (2006). To improve asset pricing implications of the model, we introduce a kind of internal habit formation mechanism into the utility function. A similar specification is considered in Ravenna and Sepp*älä* (2006). Shocks to the target inflation of the central bank and the desired markups of firms are allowed as done in Bekaert et.al. (2005), for example. But unlike the existing literature, an autoregressive conditional heteroskedasticity (ARCH)effect of a shock to the desired markup of a firm is introduced as a device to amplify time variations of term premia.

2.1 Private Agents

The production sector in the economy consists of two parts. The one is the perfectly competitive final good sector and the other is the intermediate goods sector made of a continuum of monopolistically competitive firms. The final good sector combines each intermediate good indexed by $j \in [0, 1]$ using the technology

$$Y_t = (\int_0^1 Y_t(j)^{\frac{\zeta_t - 1}{\zeta_t}} dj)^{\frac{\zeta_t}{\zeta_t - 1}} \tag{1}$$

The representative firm in the final good sector maximizes its profit given output prices P_t and input prices $P_t(j)$. The resulting input demand is given by

$$Y_t(j) = (\frac{P_t(j)}{P_t})^{-\zeta_t} Y_t \tag{2}$$

Hence, $\zeta_t > 1$ represents the elasticity of demand for each intermediate good.

Production technology for intermediate good j is linear with respect to labor.

$$Y_t(j) = A_t N_t(j) \tag{3}$$

where A_t is an exogenously given common technology and $N_t(j)$ is the labor input of firm j. We assume labor market is perfectly competitive and denote the real wage as W_t. Firms in the intermediate goods sector face nominal rigidities in the form of quadratic price adjustment cost.

$$AC_t(j) = \frac{\phi}{2}(\frac{P_t(j)}{P_{t-1}(j)} - \pi_t^\star)^2 Y_t(j) \tag{4}$$

Here ϕ is a parameter governing the degree of price stickiness in this economy and $\pi_t^\star$ is the target inflation of central bank in terms of the price of the final good. Firm j decides its labor input $N_t(j)$ and the price $P_t(j)$ to maximize the present value of profit stream[1]

$$\mathrm{E}_t[\sum_{s=0}^{\infty} \beta^s \lambda_{t+s}(\frac{P_{t+s}(j)}{P_{t+s}} Y_{t+s}(j) - W_{t+s}N_{t+s}(j) - AC_{t+s}(j))] \tag{5}$$

where λ_{t+s} is the marginal utility of a final good to the representative household at time $t+s$, which is exogenous from the viewpoint of the firm.

[1]The price of the final good is given by $P_t = (\int_0^1 P_t(j)^{1-\zeta_t} dj)^{\frac{1}{1-\zeta_t}}$.

The representative household maximizes its utility by choosing consumption (C_t), real money balance($\frac{M_t}{P_t}$), and labor supply(H_t). We deflate consumption by the current technology level to ensure a balanced growth path for the economy. Also, we introduce a kind of internal habit formation into the utility function. The specification corresponds to the closed economy version of Lubik and Schorfheide (2005). The objective function of the household is given by

$$\mathrm{E}_t[\sum_{s=0}^{\infty}\beta^s(\frac{(C^a_{t+s}/A_{t+s})^{1-\tau}-1}{1-\tau}+\chi_M\ln\frac{M_{t+s}}{P_{t+s}}-\frac{H_{t+s}^{1+\frac{1}{\nu}}}{1+\frac{1}{\nu}})] \tag{6}$$

where $C^a_{t+s}=C_{t+s}-he^{u^\star_a}C_{t+s-1}$ is the consumption relative to the habit level which is determined by the previous period consumption, h is the parameter governing the magnitude of habit persistence, τ the constant relative risk aversion, ν the short-run (Frisch) labor supply elasticity. Assuming asset markets are complete, the household is subject to the following period by period budget constraint:

$$P_tC_t+\sum_{n=1}^{\infty}P_{n,t}(B_{n,t}-B_{n+1,t-1})+M_t+T_t=P_tW_tH_t+B_{1,t-1}+M_{t-1}+Q_t+\Pi_t \tag{7}$$

where $P_{n,t}$ the price of an n quarter bond, $B_{n,t}$ bond holding, T_t lump-sum tax or subsidy, Q_t the net cash flow from participating in state-contingent security markets and Π_t the aggregate profit respectively. Here the utility function is separable with respect to consumption, real money balance, and hours worked. [2]

2.2 Monetary Policy

The monetary policy of the central bank is assumed to follow forward-looking Taylor rule with interest rate inertia. The nominal gross interest rate reacts to expected inflation, and output gap in the following way:

[2]Because of this separability assumption, the demand for real money balance does not affect the dynamics of consumption and hours worked. In our estimation below, we do not use the Euler equation for real money balance nor estimate χ_m.

$$\begin{aligned}(1+i_t) &= (1+i_t^\star)^{1-\rho_i}(1+i_{t-1})^{\rho_i}e^{\eta_i\epsilon_{i,t}}\\ 1+i_t^\star &= ((1+r^\star)(\pi^\star))(\tfrac{E_t(\pi_{t+1})}{\pi_t^\star})^{\gamma_p}(\tfrac{Y_t}{Y_t^n})^{\gamma_y}\end{aligned} \tag{8}$$

where $r^\star$ is the steady state real interest rate which is $\frac{e^{u_a^\star}}{\beta}-1$, π_t the actual inflation defined by $\frac{P_t}{P_{t-1}}$, $\pi_t^\star$ the target inflation and Y_t^n the natural rate of output which will prevail in a frictionless economy where nominal rigidities and habit formation disappear. It can be defined as follows.[3]

$$Y_t^n = C_t^n = (f_t)^{-\frac{1}{\nu+\frac{1}{\tau}}}A_t \tag{9}$$

In the mode, the time varying target inflation is assumed to be exogenously given. ρ_i captures the degree of interest rate inertia.

2.3 Exogenous Processes

The model economy is subject to four structural disturbances. Technology evolves according to

$$u_{a,t} = \frac{A_t}{A_{t-1}}\ ,\ \ln u_{a,t+1} = (1-\rho_a)u_a^\star + \rho_a \ln u_{a,t} + \eta_a\epsilon_{a,t+1} \tag{10}$$

Desired markups of firms in the intermediate goods sector are affected by a persistent shock which is subject to AR(1) process with ARCH(1) structure. The ARCH effect in the disturbance can generate time varying volatilities of macro variables.[4]

[3] In Bekaert et al. (2005), the natural rate of output is defined by the output level when exogenous disturbances are at steady state values in a flexible price economy. Our definition of the natural rate of output does not impose restrictions on exogenous disturbances but makes the fluctuation of the natural rate of output isolated from nominal disturbances by assuming away frictions in the economy.

[4] Allowing ARCH effects for other disturbances is found to create either too low average term premium or too high volatility of inflation.

$$f_t = \frac{\zeta_t}{\zeta_t - 1} \ , \ \ln f_{t+1} = (1-\rho_f)\ln f^\star + \rho_f \ln f_t + \sqrt{\eta_{0,f}^2 + \eta_{1,f}(\ln f_t - \ln f^\star)^2}\epsilon_{f,t+1} \tag{11}$$

Monetary policy is exposed to a serially uncorrelated shock($\epsilon_{i,t}$) and a persistent target inflation shock.

$$\ln \pi^\star_{t+1} = (1-\rho_{\pi^\star})\ln \pi^\star + \rho_{\pi^\star}\ln \pi^\star_t + \eta_{\pi^\star}\epsilon_{\pi^\star,t+1} \tag{12}$$

All the four serially uncorrelated innovations($\epsilon_{a,t}, \epsilon_{f,t}, \epsilon_{i,t}, \epsilon_{\pi^\star,t}$) are independent of each other at all leads and lags. Each innovation is assumed to follow a standard normal distribution.

2.4 Equilibrium Conditions

Market clearing conditions for the final good market and labor market are given by

$$Y_t = C_t + AC_t \ , \ H_t = N_t \tag{13}$$

The first order conditions for firms and the represent household can be expressed as follows:

$$\begin{aligned}
\lambda_t A_t &= \lambda_t^a = (C_t^a/A_t)^{-\tau} - \beta h e^{u_a^\star} E_t((C_{t+1}^a/A_{t+1})^{-\tau} A_t/A_{t+1}) && (14)\\
1 &= \beta \mathrm{E}_t[(\frac{\lambda_{t+1}^a}{\lambda_t^a})\frac{A_t}{A_{t+1}}\frac{1+i_t}{\pi_t}] && (15)\\
1 &= \zeta_t[1 - \frac{(\frac{Y_t}{A_t})^{\frac{1}{\nu}}}{\lambda_t^a}] + \phi\pi_t(\pi_t - \pi_t^\star) - \frac{\phi}{2}\zeta_t(\pi_t - \pi_t^\star)^2 \\
&- \phi\beta\mathrm{E}_t[(\frac{\lambda_{t+1}^a}{\lambda_t^a})\frac{Y_{t+1}/A_{t+1}}{Y_t/A_t}\pi_{t+1}(\pi_{t+1} - \pi_{t+1}^\star)] && (16)
\end{aligned}$$

2.5 Model Solutions

To solve the model around the steady state, we need to eliminate the non-stationary trend of technology and make all the variables stationary, for example by setting $c_t = C_t/A_t$. And the percentage deviation from the steady state value of consumption is denoted by $\hat{c}_t = \ln(c_t/c^\star)$. Before deriving the nonlinear solution of the model, we can analyze the log-linearized system first to have some insights on the implications of the model for macro dynamics. The log-linear approximation of optimization conditions and monetary policy around the steady state yields the following system of equations.

$$\begin{aligned}
\hat{\lambda}^a_t &= E_t(\hat{\lambda}^a_{t+1}) + \hat{i}_t - E_t(\hat{\pi}_{t+1}) - E_t(\hat{u}_{a,t+1}) && (17)\\
\hat{\pi}_t - \hat{\pi}^\star_t &= \beta E_t(\hat{\pi}_{t+1} - \hat{\pi}^\star_{t+1}) + \kappa(\hat{y}_t - \hat{y}^n_t) && (18)\\
\hat{y}^n_t &= -\frac{\hat{f}_t}{\nu + \frac{1}{\tau}} && (19)\\
\text{where } \kappa &= \frac{1}{\nu\phi\pi^{\star 2}(f^\star - 1)} \\
\hat{i}_t &= \rho_i \hat{i}_{t-1} + (1-\rho_i)(\gamma_p(E_t(\hat{\pi}_{t+1}) - \hat{\pi}^\star_t) + \gamma_y(\hat{y}_t - \hat{y}^n_t)) + \epsilon_{i,t} && (20)
\end{aligned}$$

The log-linearized system provides a clear intuition for the source of economic fluctuations. Without the nominal rigidity($\phi = 0$), κ would be equal to ∞ and the actual output is always the same as the natural rate of output and nominal disturbances have no impacts on output. On the contrary, if the degree of nominal rigidity(ϕ) is high, κ would approach to 0 and nominal disturbances would have big impacts on the actual output. So when the time variation of the yield curve is traced to macro variables, the high nominal rigidity would imply nominal disturbances are dominant factors.

While the log-linear approximation can provide insights on macro dynamics, asset pricing implications of the model can not be studied in the pure log-linear approximation because there is no consideration of the risk in the log-linearized macro dynamics. Jermann (1998) obtains the risk premium term by combing the log-normality of disturbance with the log-linear approximation of macro dynamics. He uses this method in the context of studying on the asset pricing implications of a real business cycle model. The basic idea of this approach is to derive risk premia from the second order terms implied by the log-linear approximation. Wu (2005) applies the same approach for studying on term structure implications of a New Keynesian model. He computes equilibrium bond prices from using an affine term structure model derived from the log-linear and log-normal approach. Recent literature(An and Schorfheide (2006), Fernández-Villaverde and Rubio-Ramírez (2006)) show that there may be some significant differences between the nonlinear analysis and the linear one even for only macro variables. Since the main goal of the macro finance model is to identify term structure factors in terms of macro risks, a more accurate analysis of macro dynamics would be desirable.

We follow Schmitt-Grohé and Uribe (2004) to obtain the second order approximate solution of the DSGE model. The nonlinear solution starts from representing equilibrium conditions as a rational expectations system with respect to:
$y_t = [\widehat{Y_t/A_t}, \widehat{C_t/A_t}, \hat{\pi}_t, (\widehat{1+i_t}), \widehat{C_t^a/A_t}, \hat{\lambda}_t^a]$ and
$x_t = [\hat{u}_{a,t}, \hat{f}_t, \epsilon_{i,t}, \hat{\pi}_t^\star, (\widehat{1+i_{t-1}}), \widehat{C_{t-1}/A_{t-1}}]$.

$$E_t f(y_{t+1}, y_t, x_{t+1}, x_t, \sigma\epsilon_{t+1}) = 0 \tag{21}$$

Here $\sigma \in [0, 1]$ is a perturbation parameter which determines the distance from the deterministic steady state. $\sigma = 0$ corresponds to the non-stochastic steady state. Since $\sigma\epsilon_{t+1}$ is the only source of uncertainty, the approximation order in the perturbed system is determined by the the degree of powers of σ in the approximated system.

The exact solution of the nonlinear model is given as follows:

$$y_t = g(x_t, x_{t-1}, \sigma) \ , \ x_{t+1} = h(x_t, \sigma) + \sigma\eta(x_t)\epsilon_{t+1} \tag{22}$$

Since the previous period markup affects the volatility of the current period markup, the solution for the policy function needs to include x_{t-1} as well as x_t.

However, up to the second order approximation of g and h around the deterministic steady state $(x_t, \sigma) = (0, 0)$, it can be shown there is no need to add x_{t-1} as additional state variables to approximate policy functions.

Result 1 : Up to the second order, the coefficients on g and h are not affected by time varying volatility. Indeed they are the same as those under the homoskedastic case where $\eta(x_t)$ is replaced by $\eta(0)$.

Proof In Appendix.

The resulting approximate solutions are:[5]

$$y_t = \frac{1}{2}g_{\sigma\sigma}\sigma^2 + g_x x_t + \frac{1}{2}(I_{n_y} \otimes x_t)'(g_{xx})x_t \tag{23}$$

$$x_{t+1} = \frac{1}{2}h_{\sigma\sigma}\sigma^2 + h_x x_t + \frac{1}{2}(I_{n_x} \otimes x_t)'(h_{xx})x_t + \eta(x_t)\sigma\epsilon_{t+1} \tag{24}$$

3 Measurement Equations for Bond Yields

3.1 Construction of arbitrage-free bond prices

Measurement equations for macro variables can be constructed by the approximate model solutions. Measurement equations for bond yields require an additional work. With the second order approximation, it is much more complicated than the log-linear and log-normal case. First of all, implied bond yields are not affine with respect to the state vector because the short rate is now a quadratic function of the state vector rather than an affine one. However the discrete time counterpart of a quadratic term structure model studied in Ahn et al. (2002) can be obtained by using Euler equation for the representative household in our model. One complication in our setting is that current period consumption, therefore, marginal utility of consumption is quadratic with respect to current state variables. Mechanical forward iteration of second order approximation generates extra higher order terms

[5]Here, the representation of solutions follows Klein (2005).

in measurement equations because long term bond yields under the no-arbitrage assumption involve the conditional expectations of the future state variables. As discussed in Kim et.al. (2005), these extra higher order terms do not necessarily increase the accuracy of approximation because they do not correspond to higher order terms of the Taylor series expansion. Following their suggestion, we use the pruning scheme which effectively ignores extra higher order terms. Then, we can generate the future state variables according to the following scheme:

$$\begin{aligned} x_{t+1}^l &= h_x(x_t^l + x_t^q) + \eta_0 \sigma \epsilon_{t+1} \\ x_{t+1}^q &= \frac{1}{2} h_{\sigma\sigma} + \frac{1}{2}(I_{n_x} \otimes x_t^l)' h_{xx} x_t^l + \eta_1 (I_{n_\epsilon} \otimes x_t) \sigma \epsilon_{t+1} \\ x_{t+1} &= x_t^l + x_t^q \end{aligned} \tag{25}$$

The one-period ahead log nominal stochastic discount factor in our equilibrium model and equilibrium bond prices are expressed as follows:

$$\begin{aligned} \hat{M}_{t,t+1} &= (\hat{\lambda}_{t+1}^a - \hat{\lambda}_t^a) - \hat{u}_{a,t+1} - \hat{\pi}_{t+1} \\ e^{\hat{p}_{n,t}} &= E_t(e^{\sum_{j=0}^{n-1} \hat{m}_{t+j,t+j+1}}) \end{aligned} \tag{26}$$

After applying the pruning to the log stochastic discount factor, we can obtain the following representation.

$$\hat{M}_{t,t+1} = \epsilon_{t+1}' m_0 \epsilon_{t+1} + (m_1 + m_2 x_t^l)\epsilon_{t+1} + (m_3 + m_4(x_t^l + x_t^q) + x_t^{l\prime} m_5 x_t^l) \tag{27}$$

Another thing that we should be careful about is the ARCH(1) part of shocks to desired markups. Since the functional form of it is known, we do not approximate the ARCH(1) part in the transition equation. Even though the procedure may not create any problem for using the transition equation in the estimation, it may induce higher order terms for the law of motion of $\ln f_t$ and thus complicate measurement equations for bond yields. Therefore we consider only terms up to the second order of structural shocks ϵ_t when we construct measurement equations.[6] In the end, we obtain closed form solutions for bond prices which are accurate up to the second order of stochastic shocks.

Result 2 : The log of bond prices of maturity of n, $p_{n,t}$ under the no-arbitrage assumption has the following representation:

$$\hat{p}_{n,t} = a_n + b_n(x_t^l + x_t^q) + x_t^{l\prime} c_n x_t^l + o_p(\sigma^2)$$

Proof : In appendix.

The coefficients(a_n, b_n, c_n) in the above formula are obtained in the recursion implied by the Euler equation of the representative household. Our closed-form solutions for bond prices are different from those suggested by Hördahl et.al. (2005b). They also use the second order approximation of macro variables to derive closed form expressions for bond prices. However, they approximate $e^{\hat{p}_{n,t}}$ up to the second order of stochastic shocks, not $\hat{p}_{n,t}$. Since $e^{\hat{p}_{n,t}} = \sum_{j=0}^{\infty} \frac{\hat{p}_{n,t}^j}{j!}$, some part of second order terms in $\hat{p}_{n,t}$ correspond to higher order terms in $e^{\hat{p}_{n,t}}$. But they throw away all these higher order terms in $e^{\hat{p}_{n,t}}$. Resulting solutions are not second-order accurate regarding $\hat{p}_{n,t}$. If bond yields share the same approximation order with $e^{\hat{p}_{n,t}}$, that would not be a serious concern. However, bond yields are linear functions of $\hat{p}_{n,t}$[7] and, therefore, the second order accurate approximation of $e^{\hat{p}_{n,t}}$ would not be the valid second order approximation of bond yields. On the contrary, our solutions for bond prices are second-order accurate because we throw away only higher order terms in $\hat{p}_{n,t}$.

3.2 Term structure implications of the approximate model solution

Before taking the model to data, it would be useful to examine the properties of the term structure model implied by the model. The purpose is checking whether or

[6]The second order approximation of $\sqrt{\eta_{0,f}^2 + \eta_{1,f}(\ln f_t - \ln f^\star)^2}$ around $\ln f^\star$ is $\eta_{0,f}$ while the second order approximation of $\eta_{0,f}^2 + \eta_{1,f}(\ln f_t - \ln f^\star)^2$ is exact.

[7]Note that $\hat{y}_{n,t} = -\frac{\hat{p}_{n,t}}{n}$.

not some kind of the salient features of data can be replicated in the model. We will start the discussion from briefly discussing deficiencies in term structure implications of a standard real business cycle model pointed out by Den Haan (1995) would be repeated in our model. Consider the log real stochastic discount factor implied by a standard real business cycle model as illustrated in Den Haan (1995).

$$m^r_{t,t+1} = \ln\beta - \tau(\ln C_{t+1} - \ln C_t) \tag{28}$$

In this case, the autocorrelation of the log real stochastic discount factor is tightly related to that of consumption growth. Therefore, if consumption growth is positively autocorrelated, the uncertainty in h-period ahead stochastic discount factor($\mathrm{Var}_t(m^r_{t,t+h})$) is greater than $h \times \mathrm{Var}_t(m^r_{t,t+1})$. That means a long-term bond is a hedge against unforeseen movements in consumption and ,therefore, on average commands a negative term premium.[8] A similar argument can be applied to a nominal stochastic discount factor ($m_{t,t+1} = m^r_{t,t+1} - \pi_{t+1}$) when inflation and consumption growth are uncorrelated. Because inflation is more positively autocorrelated than consumption growth, the same observation can be made for the nominal stochastic discount factor too. It seems that positive autocorrelations of consumption growth and inflation make one of stylized facts of the yield curve- a positive term premium, on average-hard to match by a standard equilibrium business cycle model. It should be noticed that we focus on the term premium rather than the yield spread. The reason is that the unconditional mean of the term premium is easier to understand compared to that of the yield spread. Den Haan (1995)'s discussion of the unconditional mean of the yield spread is misleading because his derivation of the unconditional expectation of interest rates is incorrect. When the log consumption growth $\ln(C_{t+1}/C_t)$ follows an AR(1) process like $\ln(C_{t+1}/C_t) = \rho_c \ln(C_t/C_{t-1}) + (1-\rho_c)\bar{C} + \epsilon_{c,t+1}$, the unconditional mean of

[8] Indeed, this can be analytically proved when consumption growth is log-normally distributed. In this case, the unconditional mean of the term premium defined by the difference between the long term rate and the average of expected short rates $y_{n,t} - \frac{E_t(\sum_{j=0}^{n-1} i_{t+j})}{n}$ is $-\tau^2(\frac{E\mathrm{Var}_t(\sum_{j=1}^{n}\ln(y_{t+j}/y_{t+j-1}))}{2n} - \frac{E(\sum_{j=1}^{n}\mathrm{Var}_t(\ln(y_{t+j}/y_{t+j-1})))}{2})$.

$E_t(\ln(C_{t+2}/C_t))$ is $(1+\rho_c)\bar{C}$ instead of $2\bar{C}$. But in Den Haan (1995), $E(\ln(C_{t+2}/C_t))$ equates to $2\bar{C}$. If we correct this error, the expression for the yield spread is much more complicated.

Going back to the counterfactual implications on the term premium, we can imagine several ways out of these deficiencies as pointed out by Hördahl et.al.(2005b). First, one can sever the linear relationship between log marginal utility of consumption and log consumption growth by introducing habit formation. Second, a negative correlation between inflation and consumption growth can mitigate the variance of the long horizon stochastic discount factor and so increase the long-term bond yield because the inter-temporal insurance service value of a long term bond is reduced.[9] Our model incorporates both features. Habit formation is already introduced in the utility function of the representative household. And shocks to desired markups of firms can create a negative correlation between inflation and output growth because high prices of intermediate goods depresses the output of the final good and increase inflation rate at the same time.

4 Econometric Methodology

4.1 Construction of the likelihood by particle filtering

In the previous section, we derived the law of motions of state variables and the measurement equations for bond yields under the second order approximation. The resulting state transition equations and measurement equations can be cast into the following nonlinear state space model.

$$x_t = \Gamma_0(\vartheta) + \Gamma_1(\vartheta)x_{t-1} + (I_{n_x} \otimes x_{t-1})'\Gamma_2(\vartheta)x_{t-1} + \eta(x_{t-1})\sigma\epsilon_t \quad (29)$$

$$z_t = \alpha_0(\vartheta) + \alpha_1(\vartheta)x_t + (I_{n_z} \otimes x_t)'\alpha_2(\vartheta)x_t + H\xi_t \text{ where } \xi_t \sim \mathcal{N}(0, I_{n_z}) \quad (30)$$

[9] Piazzesi and Schneider (2006) emphasize this channel in order to get a positive inflation risk premium.

Here, z_t is a set of observed variables such as $[\ln Y_t, \ln(1+\pi_t), \ln(1+i_t), y_{4,t}, y_{8,t}, y_{12,t}, y_{16,t}, y_{20,t}]$, ξ_t a vector of measurement errors, and ϑ a vector of structural parameters. Standard deviations of measurement errors are fixed because they mitigate the evaluation of the likelihood by particle filtering convenient.[10] We allow measurement errors for all the observed variables. If the above state model is linear and Gaussian, Kalman filter is optimal and we can compute the prediction error likelihood by recursively applying it. However, once we get out of the linear Gaussian case, Kalman filter is no longer optimal. Simulation-based particle filtering is found to work well in the estimation of a DSGE model solved with nonlinear methods.(An (2005), Fernández-Villaverde and Rubio-Ramírez (2006)) The basic idea of particle filtering is to approximate unknown filtering density $p(x_t|z^t, \vartheta)$ by a large swarm of particles x_t^i $(i = 1, \cdots N)$. Arulampalam et.al. (2002) provide an excellent survey of various algorithms. We use the sequential importance resampling filter suggested by Kitagawa (1996). The same algorithm is used and discussed in An (2005) in the context of the constructing the likelihood for a DSGE model solved with the second order approximation. Detailed description of the filtering algorithm can be found in the appendix.

4.2 Posterior simulation

Once the likelihood function is obtained, we can combine that information with the prior information on structural parameters to compute the posterior density. Prior information on structural parameters can be represented by the following prior density $p(\vartheta)$. All the parameters are assumed to be independent *a priori*. The posterior kernel is the product of prior density and likelihood:

$$p(\vartheta|z^T) \propto p(\vartheta)\mathcal{L}(\vartheta|z^T) \tag{31}$$

The analytical form of the posterior density is not known but we can use simulation methods to generate draws from the posterior density. Markov Chain

[10]The detailed explanation is given in the appendix.

Monte Carlo(MCMC) methods can be used for the purpose as explained in An and Schorfheide (2006). We adopt their algorithm with some modifications to avoid the posterior density maximization step whose solution is not found reliably in our case. We initialize the MCMC chain at the point with the highest log-likelihood among 100 prior draws. Multiple MCMC chains are run starting from different points. If each chain is wandering around deeply separated region, we pick up the highest posterior area.[11] More details of our methods are available in the appendix.

4.3 Monte Carlo smoothing

The estimation of a macroeconomic model jointly with term structure data may provide valuable insights about macroeconomic dynamics which are not uncovered in the analysis of only macro data. For example, Doh (2006) finds that incorporating term structure data gives very different estimates of the federal reserve's target inflation especially during the mid 1980s. Estimates of unobserved macro state variables conditional on all the observations $E(x_t|z^T, \vartheta)$ can enhance our understanding of macro dynamics. In the linear model, we can use the classical fixed interval Kalman smoothing without resorting to simulation. However, in the nonlinear model, we have to use simulation methods to figure out latent factors. We will use an efficient way of smoothing via backward simulation as proposed in Godsill et.al. (2004) and applied in Fernández-Villaverde and Rubio-Ramírez (2006). The key idea of the simulation scheme is to regenerate a lot trajectories of resmapled state variables stored in the forward filtering. Information from all the observations is used to select particular values of state variables at each time period. The more trajectories we generate, the more accurate smoothed estimates would be. However, the computational cost of generating one trajectory is the same as that of evaluating a likelihood value which is around 6 seconds. We choose the number of trajectories

[11] In the linear analysis, two separated regions have the similar magnitude of the posterior density and the selection of the highest posterior region is difficult. Unless it is mentioned otherwise, we report results of the linear analysis from the posterior region which has a higher posterior density in the nonlinear analysis.

used in the smoothing based on the performance for the simulated dataset. The detailed information is provided in the appendix.

5 Estimation Results

5.1 Data

We estimate the model based on U.S. macro and treasury bond data. Macro variables are taken from the Federal Reserve Database (FRED)Saint Louis. The measure of output is per-capita real GDP which is obtained by dividing real GDP (GDPC1) by total population (POP). For the inflation rate, the log difference of the GDP price index (GDPCTPI) is used. The nominal interest rate is extracted from the Fama CRSP risk free rate file. We select the average quote of 3-month treasury bill rate. Five bond yields (1,2,3,4,5 year) are obtained from Fama CRSP discount bond yields files. Observations from 1983:QI to 2002:QIV are used for the estimation. To match the frequency of the yields with that of macro data, the monthly observations of the treasury bill rate and bond yields are transformed into quarterly data by averaging three monthly observations.

5.2 Prior information

5.2.1 Prior distributions of parameters

The specification of prior distributions is shown in Table 1. The prior mean of the average technological progress ($u_a^\star$) is set to be 0.5% at the quarterly frequency to imply the steady state growth rate of 2% per year which is roughly the average growth rate of per capita real GDP during the pre-sample period (1960:QI to 1982:QIV). The prior mean of the discount factor (β) is chosen to set the steady state real interest rate to 2.8% in annualized percentage.[12] The elasticity of labor

[12]Since the average real interest rate during the pre-sample period was less than 2%, we can set $u_a^\star$ and β to match the real interest rate and output growth at the same time. The real interest rate is about 3.2% during the sample period and we target the steady state real interest rate is slightly

supply(ν) is fixed at 0.5 which is roughly the posterior mean in Chang et al. (2006). The average markup of 1.1 used in Ravenna and Walsh (2004) motivates our choice of 0.1 as the prior mean of the steady state log-markup. The prior mean of $\ln A_0$ is set in order to match the detrended output of 1982:QIII with the steady state level implied by prior mean values of parameters. Measurement errors of output, inflation rate, and bond yields are fixed at about 20% of the sample standard deviations of output growth, inflation rate, and the nominal interest rate. The prior distribution ARCH parameter $\eta_{1,f}$ is assumed to be distributed uniformly over the interval where the existence of the stationary distribution of $\ln f_t$ is guaranteed.[13] The prior standard deviation of the monetary policy reaction to expected inflation is set to be a bit smaller than that in other literature. This is because a loose prior generates a low value of the reaction coefficient which creates a numerical problem in the computation of the likelihood. It turns out the likelihood is extremely sensitive to the reaction coefficient when the parameter gets closer to 1. For other parameters, we set pretty loose priors to enhance the model's ex-ante explanatory power. Table 1 summarizes the prior information for all the parameters.

5.2.2 Prior predictive checks

Before estimating the model, it is useful to evaluate the model's ability to replicate salient features of data. If the model can not match salient features of data, it may not be worth estimating the model. Here, we generate 80 observations from 100 prior draws and compute sample moments of macro variables and term structure variables such as the standard deviation of inflation, mean term premium and the standard deviation of term premium. We also compute sample statistics for empirical proxies for the level($\frac{y_{20,t}+y_{8,t}+y_{1,t}}{3}$), the slope($y_{1,t} - y_{20,t}$), and the curvature($2y_{8,t} - y_{1,t} - y_{20,t}$) of the yield curve. To detect the differences across models, we generate data from (i)the linear model, (ii)the nonlinear model without ARCH effect, and (iii)the

lower than that.

[13] Borkovec and Klüppelberg (2001) derive the interval under the condition that ρ_f belongs to $[0, 1]$. As we decrease the possible range of ρ_f, the boundary where the strict stationarity of $\ln f_t$ is satisfied.

nonlinear model with ARCH effect. We assume no measurement error in predictive checks to isolated the differences across models transparent. The resulting 100 sample moments for three models are plotted in Figure 1 - 2. Several things are noticeable. First, ARCH effect gets mean term premium closer to the actual sample moment. The average of 100 means of the term premium increases 30 basis points with ARCH effect than other models.[14] At the same time, the standard deviation of the term premium increases and gets closer to the actual sample moment in the quadratic model with ARCH effect. Time-varying volatility induced by ARCH effect seems to have the potential to explain time variation of the term premium better. However, the term structure of volatilities is not well matched even with ARCH effect. Volatilities of bond yields decrease rapidly over maturities. This explains the underprediction of the volatility of the level together with the overprediction of volatilities of the slope and the curvature. The same problem is documented in other literature using equilibrium bond yields implied by a macro model.(Den Haan (1995), Piazzesi and Schneider (2006), Ravenna and Seppälä (2006) etc.)[15] Making structural shocks very persistent increase volatilities of long term rates but creates the poor fit for macro data which are not so persistent at least during our sample period.[16] Prior checks indicate that ARCH effect may not be useful for solving this problem. Second, the three models tend to overpredict the volatility of inflation a bit. Assuming less persistent or less volatile shocks may help us in getting the moment closer to the actual sample moment. However, by doing so, term structure implications of the model deteriorate seriously because long term rates are much more persistent and volatile in data than in the model. Our prior distribution is a compromise between macro and term structure implications of the model. Third, consistent with the previous discussion of term structure implications of habit formation, all the models can account for a positive autocorrelation of

[14] Nonlinear terms without ARCH effect do not change mean term premium much.

[15] The exception is Wachter (2006). She introduces a complicated nonlinear habit shock which creates a time-varying risk aversion by construction. However, the structural interpretation of habit shock is not clear and for that reason, we avoid introducing the channel.

[16] The AR(1) coefficient of inflation during the sample period is just 0.6 while AR(1) coefficients of bond yields are around 0.95.

output growth and a positive mean term premium at the same time.[17] Finally, the markup shock generates the negative correlation of output growth and inflation. Since ARCH effect amplifies the volatility of a markup shock, the impact is more pronounced in the case of the quadratic model with ARCH effect.

5.3 Posterior analysis

5.3.1 Posterior predictive checks

We compute predictive moments from the posterior distribution by simulating data from posterior draws. Since posterior draws obtained by MCMC methods are serially correlated, we take every 500th draw from 50,000 posterior draws to attenuate the dependence of parameter draws. Comparing bi-dimensional scatter plots of posterior predictive moments in Figure 3 - 4 with plots in Figure 1 - 2, we can extract additional information from data. Data move posterior moments of macro variables such as the autocorrelation of output growth and the standard deviation of inflation rate closer to actual sample moments than prior moments even though the unconditional volatility of inflation is higher than the actual sample moment. And the nonlinear model does better at matching the standard deviation of the inflation rate. On the other hand, the posterior analysis fails to match the first moments of term structure data. The mean term premium and the mean slope are still significantly away from sample counterparts of actual data. Here, the nonlinear terms including ARCH effect do not improve term structure implications much. The mean term premium does not show much difference across models. As mentioned before, the channel through which ARCH effect increases mean term premium is the increased volatility of a markup shock. In our model, the markup shock affect inflation via Phillips curve relation created by nominal rigidities.[18] Hence to match

[17] We generate data from the linear model without habit formation too. As expected, moments from simulated data can not match the positive autocorrelation of output growth and the positive mean term premium at the same time.

[18] Without nominal rigidities, a markup shock induces one-time adjustment of the price and would not affect inflation.

the relatively less volatile inflation series during the sample period, the volatility of a markup shock is constrained. In prior predictive checks, we can select the ARCH effect parameter either in the direction of the better fit for first moments or that for second moments. The lesson from the posterior predictive analysis is that the latter aspect is more important in data.

5.3.2 Posterior distribution of parameters

Additional information from data about the model also can be gleaned from contrasting prior distributions of parameters with the posterior counterparts. Figure 5 - 6 depict the draws from prior and posterior distribution of each parameter. For most parameters, the posterior draws are more concentrated as expected. Also, there are clear shifts of the mean values for some parameters. The risk aversion parameter(τ) and the degree of habit formation(h) are higher in the posterior distribution for both linear and nonlinear models. The steady state of target inflation($\ln \pi^{\star}$) is higher in the nonlinear model. In the steady state target inflation is not the same as the unconditional average inflation in the nonlinear model due to the impacts of quadratic terms while the two concepts coincide in the linear model. The difference can be illustrated in the following equation.

$$E(\ln \pi_t) - \ln \pi^{\star} = \frac{1}{2} g^{\pi}_{\sigma\sigma} + g^{\pi}_{i_{t-1}} E(\hat{i}_{t-1}) + g^{\pi}_{c_{t-1}} E(\hat{c}_{t-1}) + tr(g^{\pi}_{xx} E(x_t x_t')) \qquad (32)$$

We can call this term the quadratic adjustment term. The magnitude of this term can be quantified by simulating a long time series and replacing the expectation by the average. In our case, the term is negative reflecting the fact the estimated steady state target inflation is higher in the nonlinear analysis. The above finding is consistent with An and Schorfheide (2006) who call this term stochastic steady state effect. The reaction of the central bank to expected inflation is stronger in the posterior distribution. Interestingly, the nonlinear model implies a bit more stronger response of the central bank. This is due to the fact that the likelihood in the quadratic model decreases rapidly when the reaction parameter gets closer to the indeterminacy region. The ARCH effect parameter $\eta_{1,f}$ is much smaller in the posterior distribution.[19] The finding is consistent with discussions made regarding predictive moments. Boudoukh (1993) finds that in a standard real business cycle model, the stochastic volatilities do not generate enough average term premium while it can improves the model fit for the standard deviation of the term premium. Our finding is in line with his observation. Still, there can be other channels that we can introduce into DSGE models in order to amplify the average term premium. One lesson from our analysis is that the channel should not induce a high volatility of inflation to explain that.

Even though term structure implications in terms of posterior predictive moments are not much different across models, the nonlinear analysis provides useful information which can not be obtained in the linear analysis. Figure 7 provides the posterior contours of parameters around posterior mean values of $(\rho_f\ ,\ \rho_{\pi^\star})$ computed from different MCMC chains. It turns out the nonlinear analysis clearly identifies the highest posterior region of parameters while the linear analysis can not do that. Also, the nonlinear analysis can give us smaller posterior intervals of parameters than the linear analysis. The posterior draws of risk aversion parameter (τ), the reaction of the central bank to expected inflation (γ_p), the average technological progress $(u_a^\star)$, and the degree of habit formation (h) in Figure 5 - 6 are more concentrated in the nonlinear version. This finding is consistent with other literature on the nonlinear analysis of DSGE models.(An and Schorfheide (2006),Fernández-Villaverde and Rubio-Ramírez (2006)) On the other hand, the precautionary saving effect which is related to the third order derivative of the utility function is considered only in the nonlinear analysis. As the curvature of the utility function increases, the precautionary saving increase and that would push down the real interest rate. The bi-variable scatter plots of the risk aversion parameter(τ) and the steady state real interest rate in Figure 8 show that the two variables are slightly negatively correlated in the nonlinear model as the theory predicts. However, the linear model

[19] If $\eta_{1,f}$ is greater than $1-\rho_f^2$, the second moment of $\ln f_t$ does not exist. This would make the unconditional distribution of macro variables and bond yields fairly heavy-tailed.

implies a positive correlation. In terms of the overall fit, the nonlinear model is a bit better though the difference is not great. Log marginal data densities in Table 3 supports the conclusion.[20] Another measure of assessing the model's ability to fit the data is to check for ex-post measurement errors. If standard deviations of measurement errors were estimated, we could evaluate the model in terms of the estimated standard deviations of measurement errors. Since these are fixed in our estimation for the evaluation of the likelihood function by particle filtering, we directly extract smoothed estimates of measurement errors and compute means and standard deviations of them. Results in Table 4 show that means of ex-post measurement errors are very close to zero. Also standard deviations of measurement errors are quite small. For example, standard deviations of the measurement errors of bond yields are fixed at 44 basis points but the corresponding values of ex-post measurement errors are at most around 25 basis points. Brandt and Yaron (2002) report that the average mean absolute pricing error of bond yields is around 9 basis points when the exogenous stochastic discount factor with macro variables are constructed to fit the cross-sectional data at each time period. The corresponding value for our model is 15 basis points. Given restrictions on our stochastic discount factor imposed by the structural model, the fit does not seem to be much worse.

6 Identification of Term Structure Factors

6.1 Macro explanation for empirical counterparts of term structure factors

The posterior analysis of our model gives us two lessons. First, the predictive performance of the model is limited. Second, the model's ability to capture the dynamics of the yield curve evaluated by ex-post measurement errors is not so bad. Therefore, even though the DSGE model itself may not be the best tool to deliver good forecasts of bond yields, the structural model has the potential to provide the clear economic

[20] Geweke (1999)'s modified harmonic mean estimator is used to compute the marginal data density.

interpretation of the yield curve dynamics captured in other flexible models. Here we try to identify term structure factors in terms of macro factors appearing in the DSGE model. It is well known that the statistical decomposition of the yield curve into the level, the slope, and the curvature factors can explain most(up to 98%) of the time variation of bond yields(e.g.) Litterman and Scheinkman (1991).). These latent factors are constructed by rotating the observed bond yields and do not have clear economic meanings by themselves. There are many papers which try to figure out macro forces of term structure factors by linking the empirical counterparts of latent term structure factors with observed macroeconomic variables (e.g.)Diebold et al. (2006)), identified macroeconomic shocks from structural vector autoregressions (VAR)(e.g.)Evans and Marshall (2002)), or macro shocks in DSGE models(Bekaert et al. (2005), and Wu (2005) etc.).

We link smoothed estimates of macro shocks in the model with empirical counterparts of latent term structure factors. Empirical proxies of term structure factors in sample data are regressed on smoothed estimates of macro shocks as follows:

$$\begin{aligned} Dep_{i,t} &= c_{i,0} + c_{i,1}E(\hat{u}_{a,t}|z^T) + c_{i,2}E(\hat{\ln f}_t|z^T) \\ &\quad + c_{i,3}E(\epsilon_{i,t}|z^T) + c_{i,4}E(\hat{\ln \pi}^{\star}_t|z^T) + \upsilon_t \qquad (33) \\ \text{where} &\quad , \quad i = (\text{level,slope,curvature}) \end{aligned}$$

Here Dep stands for a term structure factor. Each macro factor is normalized so that the standard deviation of it is equal to one. The results in Table 5 show that macro factors capture time variations of latent factors remarkably well. R^2 is 0.997 for the level, 0.794 for the slope, and 0.565 for the curvature. It seems that there is a big discrepancy between the evidence from predictive checks implying the poor fit of the DSGE model and strong explanatory power of macro factors in the above regressions. The difference comes from the fact that constant terms in these regressions are free parameters. Fitting average values of term structure factors is much easier in the above regressions than the structural model where the constant term is restricted by structural parameters. In predictive checks, the model is relatively better at generating the magnitude of the time variation of term structure factors while it misses first moments of those factors. Also, estimated macro factors are constructed by using information from all the observations and the above regressions can not be predictive regressions in a strict sense. Our analysis associates the level with the target inflation of the central bank, the slope with the time variation of the desired markup, and the curvature with the transitory monetary policy shock. Linking the level of the yield curve with the target inflation of the central bank is common in other literature. (Bekaert et al. (2005), Rudebusch and Wu (2004) among others) No-arbitrage arguments imply that long term rates reflect expected future short term rates after adjusting risk premia. Hence, the target inflation of the central bank which can affect the future behavior of the short term rate is important in determining the level of the yield curve. There is some disagreement about the interpretation of the slope. Diebold et al. (2006) and Rudebusch and Wu (2004) link the slope with the cyclical variation of a real factor like the capacity utilization rate. On the contrary, Bekaert et al. (2005) find only limited roles for real factors. They conclude that the transitory monetary policy shock drives the slope of the yield curve. The difference of our finding from Bekaert et al. (2005) comes from the fact that the degree of price stickiness estimated in our model is lower than that of Bekaert et al. (2005).[21] As mentioned before, the higher the degree of price stickiness, the bigger portion of economic fluctuations are attributed to nominal disturbances. The implied value of the Phillips curve parameter in the log-linearized model κ is 0.064 in their estimation while it is 0.261 at the posterior mean in our estimation. And the implied mean price duration according to the estimates of Bekaert et al. (2005) is longer than 1 year which is in conflict with micro evidence from Bils and Klenow (2004) reporting much more frequent price adjustments. Also, the estimate of the reaction of the monetary policy is very small and not significantly different from zero in Bekaert et al. (2005) which

[21] Another possible reason is that their model includes the dynamic indexation of prices unlike ours. The inclusion of the previous period inflation into the Phillips curve due to the dynamic indexation of prices can mitigate the role of real disturbances in generating the persistence of the observed inflation as documented in Del Negro and Schorfheide (2006).

dampens the response of the systematic monetary policy to real disturbances. This is in conflict with estimates reported in Rudebusch and Wu (2004) and Ang et al. (2006) who report the significant positive response of the monetary policy to real disturbances. Weighing in all the evidences, we conclude the identification of the slope with the shock to the desired markup is more plausible. Our identification of the curvature with transitory monetary policy shocks is in line with Bekaert et al. (2005) and Cogley (2005). Transitory policy shocks are found to be more strongly loaded on the short term rate than medium or long term rates. Dynamic responses of the yield curve to various macro shocks shown in Figure 9 and time series plots of smoothed estimates of macro shocks together with empirical counterparts of term structure factors in Figure 10 strengthen our conclusion.

6.2 Macro factors and the term premium

The decomposition of the yield curve into three latent factors can be useful for describing the overall variation of the yield curve. However, from the perspective of economics, the decomposition of long term rates into expected future short term rates and term premium is more interesting. Here, we seek a macro explanation for the time variation of the term premium. It should be noticed that the empirical proxy of the level factor is highly correlated with ex-post term premium. The co-movement of two variables is clear in Figure 11.[22] Hence, macro factors explaining the downward trend of the level factor can also account for a similar downward trend of term premium. Our identification of term structure factors implies the downward trend of term premium can be related to the change in the target inflation. According to this evidence, the term premium is mainly driven by a persistent policy shock.

[22]Cochrane and Piazzesi (2006) claim the time variation of the term premium defined by the expected excess return of the long term bond is well explained by the return forecasting factor which is a linear combination of forward rates and orthogonal to the level, the slope, and the curvature of the yield curve. Our definition of risk premium is different from theirs.

7 The Yield Curve and Monetary Policy

7.1 Term structure information on the Volcker disinflation policy

So far, focus has been concentrated on the identification of macro factors behind the movement of bond yields. However, we can extract useful information about macroeconomy from bond yields themselves. The rapid fall of inflation after the Volcker disinflation period of the early 1980s has been paid much attention in macroeconomics.(Lubik and Schorfheide (2004), Primiceri (2005), and Sargent et al. (2006) etc.) However, whether or not that decline can be also interpreted as the decrease of the long run expected inflation deserves a more careful study. Information from the yield curve may shed some lights on this issue because long term rates reflect expected future inflation rates. Recent studies incorporating term structure information for the monetary policy analysis dispute the immediate impacts and magnitude of the Volcker disinflation policy. For example, Ang et al. (2006) show the variability of the monetary policy shock decreases much when term structure data are included in the estimation of the monetary policy rule and conclude that the Volcker experience was not as a big surprise as macro data suggested. Doh (2006) takes the non-stationary target inflation of the central bank as the level factor of the yield curve and estimates a log-linearized New Keynesian model with term structure data. Smoothed estimates of the target inflation stayed high during the mid 1980s when the actual inflation stayed low after the Volcker disinflation period.

Empirical findings in the above literature can be related to the delayed and gradual responses of bond yields to the Volcker disinflation policy. If bond yields had reacted to the Volcker disinflation much like macro data indicated, we would have observed the large policy shock and the rapid fall of the target inflation during this period. The lack of these observations indicate that there might have been lingering suspicions of the disinflation policy even after the Volcker period. The conjecture is consistent with the observation in Goodfriend (1993) and Goodfriend and King (2005). They argue that the credibility of the central bank to low inflation was not obtained immediately after the Volcker experiment. Instead, the credibility was established after increasing the short term rate during the third quarter of 1984 in response to the sharp rise of the long term rate. The pattern of smoothed estimates of the target inflation in this paper supports the observation. The level factor of the yield curve and smoothed estimates of the target inflation in Figure 11 both peaked in the second quarter of 1984 and declined subsequently. In Doh (2006), the target inflation is assumed to be the level factor of bond yields. Here, we allow competing macroeconomic sources for the level factor of the yield curve. In the end, the target inflation is selected as determining the level of the yield curve. In this sense, our evidence for the movement of the target inflation is stronger than Doh (2006). On the other hand, our results are distinguished from Ang et al. (2006) in that no explicit role is given to latent term structure factors. In Ang et al. (2006), the dynamics of the latent factor is modeled together with macro variables. In our setting, there is no independent latent factor except for macro state variables. Their conclusion that the Volcker disinflation policy was not a big surprise may be reversed if we decompose the latent factor into macro fundamentals. We agree with macroeconomic studies emphasizing large monetary policy changes during the Volcker disinflation period.(Clarida et al. (2000), Lubik and Schorfheide (2004), Primiceri (2005), Sargent et al. (2006), and Schorfheide (2005) etc.) But we find that rapid fall of inflation rate did not induce the immediate swift decline of expected inflation. One-year ahead inflation forecasts shown in Figure 12 began a downward movement after the third quarter of 1984. Estimates of target inflation from the estimation using only macro data tracks the actual inflation better than those from the estimation with term structure data. The correlation of actual inflation and estimates of target inflation from the macro estimation is 0.749 while the corresponding number is 0.616 for estimates of target inflation in the joint estimation. However, estimated target inflation from the macro estimation is poorly correlated with expected inflation from the survey data. The correlation is just 0.258. On the contrary, the correlation is 0.910 for the estimated target inflation from the joint estimation. Estimated target inflation from the joint estimation in Figure 12 seems to be consistently higher than actual inflation. As mentioned before, in the quadratic model the steady inflation rate is not the same as the unconditional mean of the actual inflation. To quantify the impacts of quadratic terms, we simulate 5,000 observations from the model at the posterior mean of parameters. It turns out the average inflation in simulated data is 1.1% lower than the steady state inflation.[23] If this factor is considered, the estimated target inflation tracks the survey data of expected inflation well. To sum it up, while the Volcker disinflation policy was a large policy shock, the impacts on expected inflation was somewhat delayed unlike the actual inflation. The additional information from the yield curve provides us estimates of target inflation which are more consistent with survey data of inflation forecasts.

One possible objection to our analysis is that high long term rates might have reflected high inflation risk premia rather than high expected inflation. This is a perfectly plausible scenario but begs the following question. What drove up inflation risk premia? If people expected the long run expected inflation would decline with little uncertainty, why did inflation risk premia go up even though actual inflation rapidly declined and stayed low? Maybe it was related to the high uncertainty about future monetary policy changes. There are some evidences for this hypothesis. Cogley (2005) points out that the fluctuation of the target short term rate is closely related to the yield spread. He argues that the uncertainty about the target rate at the time of the Volcker disinflation was high and this fact can explain high risk premia of long term bonds during this period. Also, Buraschi and Jiltsov (2006) mention that high inflation risk premia during this period is related to a lack of confidence by the market in the central bank's effectiveness in pursuing the low inflation target. The basic message of our analysis that there was lingering suspicions of the disinflation policy is consistent with findings in the above literature.

7.2 Lessons from the joint estimation

Our analysis of the impacts of the Volcker disinflation policy provides some caution against estimating macro dynamics by only macro data and plug those estimates into

[23] For the posterior mean of parameters from the macro estimation, the impact is just 0.2%.

equilibrium bond prices implied by the structural model.(e.g.)Piazzesi and Schneider (2006), Wachter (2006)) The practice hides the interesting implications of bond yields for macro dynamics. If we used only macro data, our model could not detect sluggish responses of bond yields to the Volcker disinflation policy are related to the expectations of future monetary policies. Also, the significant portion of the term premium variation may not be linked with macro factors by doing so. For example, Duffee (2006) argues there is only weak evidence that macro variables like inflation, output growth, and the short term interest rate are related to term premia. The argument is based on the observation that the steady decline of the term premium from more than 4 percent in 1985:QI to about 1 percent in 2004:QIV could not be explained by actual inflation series even though the survey data of expected inflation can explain some part of the decline. However, when changes in the monetary policy are decomposed into the persistent part(target inflation) and the transitory part(temporary policy shock), the downward trend of the term premium can be well explained by the persistent part according to our analysis. The analysis is consistent with Kozicki and Tinsley (2005) who report sizeable movements in bond yields can be explained by permanent policy shocks. It should be emphasized that the estimation using macro data can not reveal this point. The intricate relationship between macro dynamics and the yield curve can be better captured in the joint estimation of the macro equilibrium model with both macro variables and bond yields.

8 Conclusion

We estimate a small-scale New Keynesian model solved with a second-order approximation to the equilibrium conditions. Both macro and term structure data of the United States from 1983:QI to 2002:QIV are used in the estimation. To magnify the time variation of the term premium, we introduce an ARCH(1) effect to a markup shock to an otherwise standard model. New closed form solutions of bond prices are proposed to make the estimation practically feasible. Main empirical findings are as follows. First, the model can provide clear economic interpretations of empirical counterparts of latent term structure factors. The level, the slope, and the curvature of the yield curve are found to be closely related to the persistent part of monetary policy, the real disturbance to the desired markup of a firm, and the transitory monetary policy shock, respectively. Also, the time variation of the term premium is found to be driven by the persistent monetary policy shock. Second, from the viewpoint of macroeconomics, our analysis sheds new lights on the interpretation of the Volcker disinflation policy. Smoothed estimates of the target inflation from the joint estimation closely track the inflation forecasts of the survey data while those from the macro estimation do not. Sluggish responses of long term rates reveal the lingering suspicion of the disinflation policy in spite of the relatively low inflation rate. Even though a relatively standard macro model is used, our analysis provide interesting implications for both macroeconomics and finance. However, the model misses some predictive moments of bond yields and implies a bit more volatile inflation than the actual data. Dynamic macro models with richer structures-asymmetric information between the central bank and private agents, for example- can improve term structure implications in terms of predictive moments. Our analysis indicates that the joint estimation of such a model with macro and term structure data is essential to fully investigate the issue. That remains a challenging but promising task.

9 Appendix

9.1 Proof of result 1

We adopt notations used in Schmitt-Grohé and Uribe (2004). The nonlinear rational expectations system is represented by

$$E_t(y', y, x', x, \sigma\epsilon') = 0 \tag{34}$$

Let $s_t = [x_t, x_{t-1}]$. First, we would establish the fact that first order terms like (g_s and h_s) are not affected by the heteroskedasticity. Consider the derivative of F with respect to s_t and σ.

$$
\begin{aligned}
[F_s(0,0)]^i_j &= E_t([f_{y'}]^i_\alpha [g_s]^\alpha_\beta [h_s + \sigma\eta'\epsilon']^\beta_j + [f_y]^i_\alpha [g_s]^\alpha_j + [f_{s'}]^i_\beta [h_s]^\beta_j + [f_s]^i_j) \\
&= ([f_{y'}]^i_\alpha [g_s]^\alpha_\beta [h_s]^\beta_j + [f_y]^i_\alpha [g_s]^\alpha_j + [f_{s'}]^i_\beta [h_s]^\beta_j + [f_s]^i_j) \\
&= 0 \qquad (35)
\end{aligned}
$$

$$
\begin{aligned}
[F_\sigma(0,0)]^i &= E_t([f_{y'}]^i_\alpha [g_s]^\alpha_\beta [h_\sigma]^\beta + [f_{y'}]^i_\alpha [g_s]^\alpha_\beta [\eta]^\beta_\phi [\epsilon']^\phi + [f_{y'}]^i_\alpha [g_\sigma]^\alpha + [f_y]^i_\alpha [g_{\sigma_\epsilon}]^\alpha \\
&+ [f_{s'}]^i_\beta [h_\sigma]^\beta + [f_{s'}]^i_\beta [\eta]^\beta_\phi [\epsilon']^\phi) \\
&= [f_{y'}]^i_\alpha [g_s]^\alpha_\beta [h_\sigma]^\beta + [f_{y'}]^i_\alpha [g_\sigma]^\alpha + [f_y]^i_\alpha [g_\sigma]^\alpha + [f_{s'}]^i_\beta [h_\sigma]^\beta \\
&= 0 \qquad (36)
\end{aligned}
$$

So in the end, we are back to the same equations as those in Schmitt-Grohé and Uribe (2004) derived under the homoskedasticity assumption. Therefore, $h_\sigma = g_\sigma = 0$.

Second, we need to take a look at the second order terms too. Although the algebra is complicated, the conclusion is simple.

$$
\begin{aligned}
[F_{ss}(0,0)]^i_{jk} &= E_t\{([f_{y'y'}]^i_{\alpha\gamma}[g_x]^\gamma_\delta[h_s+\sigma\eta'\epsilon']^\delta_k + [f_{y'y}]^i_{\alpha\gamma}[g_s]^\gamma_k \\
&+ [f_{y's'}]^i_{\alpha\delta}[h_s+\sigma\eta'\epsilon']^\delta_k + [f_{y's}]^i_{\alpha k})[g_s]^\alpha_\beta[h_s+\sigma\eta'\epsilon']^\beta_j \\
&+ [f_{y'}]^i_\alpha[g_{ss}]^\alpha_{\beta\delta}[h_s+\sigma\eta'\epsilon']^\delta_k[h_s+\sigma\eta'\epsilon']^\beta_j \\
&+ [f_{y'}]^i_\alpha[g_s]^\alpha_\beta[h_{ss}+\sigma\eta''\epsilon']^\beta_{jk} \\
&+ ([f_{yy'}]^i_{\alpha\gamma}[g_s]^\gamma_\delta[h_s+\sigma\eta'\epsilon']^\delta_k + [f_{yy}]^i_{\alpha\gamma}[g_s]^\gamma_k + [f_{ys'}]^i_{\alpha\delta}[h_s+\sigma\eta'\epsilon']^\delta_k + [f_{ys}]^i_{\alpha k})[g_s]^\alpha_j \\
&+ [f_y]^i_\alpha[g_{ss}]^\alpha_{jk} \\
&+ ([f_{s'y'}]^i_{\beta\gamma}[g_s]^\gamma_\delta[h_s+\sigma\eta'\epsilon']^\delta_k + [f_{s'y}]^i_{\beta\gamma}[g_s]^\gamma_k + [f_{s's'}]^i_{\beta\delta}[h_s+\sigma\eta'\epsilon']^\delta_k + [f_{s's}]^i_{\beta k})[h_s+\sigma\eta'\epsilon']^\beta_j \\
&+ [f_{s'}]^i_\beta[h_{ss}+\sigma\eta''\epsilon']^\beta_{jk} \\
&+ [f_{sy'}]^i_{j\gamma}[g_s]^\gamma_\delta[h_s+\sigma\eta'\epsilon']^\delta_k + [f_{sy}]^i_{j\gamma}[g_s]^\gamma_k + [f_{ss'}]^i_{j\delta}[h_s+\sigma\eta'\epsilon']^\delta_k + [f_{ss}]^i_{jk}\} \\
&= 0 \qquad (37)
\end{aligned}
$$

Once again, the fact we evaluate the above expression at $\sigma = 0$ and $E_t(\epsilon')$ is equal to 0 gives us the same equation as that in Schmitt-Grohé and Uribe (2004).

$$
\begin{aligned}
[F_{\sigma\sigma}(0,0)]^i &= E_t\{[f_{y'}]^i[g_x]^\alpha_\beta[h_{\sigma\sigma}]^\beta \\
&+ [f_{y'y'}]^i_{\alpha\gamma}[g_s]^\gamma_\delta[h_\sigma+\eta\epsilon']^\delta_\zeta[g_s]^\alpha_\beta[h_\sigma+\eta\epsilon']^\beta_\phi \\
&+ [f_{y's'}]^i_{\alpha\delta}[h_\sigma+\eta\epsilon']^\delta_\zeta[g_s]^\alpha_\beta[h_\sigma+\eta\epsilon']^\beta_\phi \\
&+ [f_{y'}]^i_\alpha[g_{ss}]^\alpha_{\beta\delta}[h_\sigma+\eta\epsilon']^\delta_\zeta[h_\sigma+\eta\epsilon']^\beta_\phi \\
&+ [f_{y'}]^i_\alpha[g_{\sigma\sigma}]^\alpha \\
&+ [f_y]^i_\alpha[g_{\sigma\sigma}]^\alpha \\
&+ [f_{s'}]^i_\beta[h_{\sigma\sigma}]^\beta \\
&+ [f_{s'y'}]^i_{\beta\gamma}[g_s]^\gamma_\delta[h_\sigma+\eta\epsilon']^\delta_\zeta[h_\sigma+\eta\epsilon']^\beta_\phi \\
&+ [f_{s's'}]^i_{\beta\delta}[h_\sigma+\eta\epsilon']^\delta_\zeta[h_\sigma+\eta\epsilon']^\beta_\phi\} \\
&= 0 \qquad (38)
\end{aligned}
$$

From the result on first order terms, we know $g_\sigma = h_\sigma = 0$. Then we get the same expression as the one in Schmitt-Grohé and Uribe (2004) except that here η is a constant matrix evaluated at $(x = 0, \sigma = 0)$.

The cross derivatives can be obtained in a similar way.

$$\begin{aligned}
[F_{\sigma s}(0,0)]^i_j &= E_t\{[f_{y'}]^i_\alpha [g_s]^\alpha_\beta [h_{\sigma x} + \eta' \epsilon']^\beta_j + [f_{y'}]^i_\alpha [g_{\sigma x}]^\alpha_\gamma [h_s]^\gamma_j \\
&+ [f_y]^i_\alpha [g_{\sigma s}]^\alpha_j + [f_{s'}]^j_\beta [h_{\sigma s} + \eta' \epsilon']^\beta_j\} \\
&= 0 \qquad (39)
\end{aligned}$$

The system of equations is homogenous in the unknown $g_{\sigma s}$ and $h_{\sigma s}$. Therefore, both of them are equal to 0.

Therefore, it is shown that up to the second order the heteroskedasticity does not affect the approximation of g, h. This implies that we can safely replace s_t by x_t in the approximation of g, h.

9.2 Proof of result 2

The coefficients on $\hat{p}_{1,t}$ are obtained after solving the policy function for nominal interest rate because $\hat{p}_{1,t} = -\hat{i}_t$.

$$\begin{aligned}
a_1 &= -g^i_{\sigma\sigma}\,,\ b_1 = -g^i_x\,,\ c_1 = -g^i_{xx} \\
\hat{M}_{t,t+1} &= (\hat{\lambda}^a_{t+1} - \hat{\lambda}^a_t) - \hat{u}_{a,t+1} - \hat{\pi}_{t+1} \\
\hat{M}_{t,t+1} + \hat{p}_{n-1,t+1} &= x'_{t+1}(\frac{1}{2} g^{\lambda^a}_{xx} - \frac{1}{2} g^{\pi}_{xx} + c_{n-1}) x_{t+1} + (g^{\lambda^a}_x - b_a - g^{\pi}_x + b_{n-1}) x_{t+1} \\
&+ x'_t(-\frac{1}{2} g^{\lambda^a}_{xx}) x_t - g^{\lambda^a}_x x_t - \frac{1}{2} g^{\pi}_{\sigma\sigma} + a_{n-1} \\
&= x'_{t+1} \Omega_0 x_{t+1} + \Omega_1 x_{t+1} + \Omega_2 \qquad (40)
\end{aligned}$$

where $b_a = [1, 0, \cdots, 0]$

If we apply the pruning scheme to $\hat{M}_{t,t+1} + \hat{p}_{n-1,t+1}$, then we would obtain:

$$
\begin{aligned}
x'_{t+1}\Omega_0 x_{t+1} &= (\Gamma_1 x_t + \eta_0\epsilon_{t+1})'\Omega_0(\Gamma_1 x_t + \eta_0\epsilon_{t+1}) + o_p(\sigma^2) \\
&\approx x'_t(\Gamma'_1\Omega_0\Gamma_1)x_t + 2x'_t\Gamma'_1\Omega_0\eta_0\epsilon_{t+1} + \epsilon'_{t+1}\eta'_0\Omega_0\eta_0\epsilon_{t+1} \\
\Omega_1 x_{t+1} &= \Omega_1(\Gamma_0 + \Gamma_1 x_t + (I_{n_x}\otimes x_t)'\Gamma_2 x_t + (\eta_0 + \eta_1(I_{n_\epsilon}\otimes x_t))\epsilon_{t+1}) + o_p(\sigma^2) \\
&\approx \Omega_1(\Gamma_0 + \Gamma_1 x_t + (I_{n_x}\otimes x_t)'\Gamma_2 x_t + (\eta_0 + \eta_1(I_{n_\epsilon}\otimes x_t))\epsilon_{t+1}) \\
\Omega_2 &= x'_t(-\frac{1}{2}g^{\lambda^a}_{xx})x_t - g^{\lambda^a}_x x_t - \frac{1}{2}g^{\pi}_{\sigma\sigma} + a_{n-1} \qquad (41)
\end{aligned}
$$

Rearranging the approximated $\hat{M}_{t,t+1} + \hat{p}_{n-1,t+1}$ by the order of ϵ_{t+1} leads into:

$$
\begin{aligned}
\hat{M}_{t,t+1} + \hat{p}_{n-1,t+1} &\approx \epsilon'_{t+1}C_{n-1}\epsilon_{t+1} + B_{n-1}\epsilon_{t+1} + A_{n-1} \\
\text{In the end, } C_{n-1} &= \eta'_0\Omega_0\eta_0\ ,\ B_{n-1} = 2x'_t\Gamma'_1\Omega_0\eta_0 + \Omega_1(\eta_0 + \eta_1(I_{n_\epsilon}\otimes x_t)) \\
\text{where },\ A_{n-1} &= x'_t(\Gamma'_1\Omega_0\Gamma_1 - \frac{1}{2}g^{\lambda^a}_{xx})x_t + \Omega_1(I_{n_x}\otimes x_t)'\Gamma_2 x_t \\
&\quad + \Omega_1(\Gamma_0 + \Gamma_1 x_t) - g^{\lambda}_x x_t - \frac{1}{2}g^{\pi}_{\sigma\sigma} + a_{n-1} \qquad (42)
\end{aligned}
$$

Now, we can use the multi-normality of ϵ_{t+1} to derive:

$$
\begin{aligned}
& E_t(e^{\epsilon'_{t+1}C_{n-1}\epsilon_{t+1} + B_{n-1}\epsilon_{t+1} + A_{n-1}}) \\
= & \int |2\pi I|^{-\frac{1}{2}} e^{(-\frac{1}{2}\epsilon_{t+1}{}'(I-2C_{n-1})\epsilon_{t+1} + B_{n-1}\epsilon_{t+1} + A_{n-1})} d\epsilon_{t+1} \\
= & \int |2\pi(I-2C_{n-1})^{-1}|^{-\frac{1}{2}} e^{(-\frac{1}{2}\epsilon'_{t+1}(I-2C_{n-1})\epsilon_{t+1})} \times e^{(B_{n-1}\epsilon_{t+1})}|I-2C_{n-1}|^{-\frac{1}{2}} e^{A_{n-1}} d\epsilon_{t+1} \\
= & \int |2\pi(I-2C_{n-1})^{-1}|^{-\frac{1}{2}} e^{(-\frac{1}{2}(\epsilon_{t+1} - (I-2C_{n-1})^{-1}B'_{n-1})'(I-2C_{n-1})(\epsilon_{t+1} - (I-2C_{n-1})^{-1}B'_{n-1}))} \\
& \times e^{(B_{n-1}(I-2C_{n-1})^{-1}B'_{n-1})}|I-2C_{n-1}|^{-\frac{1}{2}} e^{A_{n-1}} d\epsilon_{t+1} \\
= & \ e^{\frac{1}{2}B_{n-1}(I-2C_{n-1})^{-1}B'_{n-1}}|I-2C_{n-1}|^{-\frac{1}{2}} e^{A_{n-1}}\ ,\ \epsilon_{t+1} \sim \mathcal{N}(0, I) \qquad (43)
\end{aligned}
$$

As long as $I - 2C_{n-1}$ is positive definite, this is well defined. Practically, C_{n-1} is pretty small relative to I and this condition is satisfied. For all the parameter draws from prior distribution or MCMC chains, this condition was satisfied. By matching

$\frac{1}{2}B_{n-1}(I-2C_{n-1})^{-1}B'_{n-1} - \frac{1}{2}\ln|I-2C_{n-1}| + A_{n-1}$ with $a_n + b_n x_t + x'_t c_n x_t$, we obtain the following recursion formula for the coefficients.

$$
\begin{aligned}
a_n &= a_{n-1} + \Omega_1\Gamma_0 + \frac{1}{2}\Omega_1\eta_0(I-2C_{n-1})^{-1}\eta'_0\Omega'_1 - \frac{1}{2}g^{\pi}_{\sigma\sigma} - \frac{1}{2}\ln|I-2C_{n-1}| \\
&\quad -\frac{1}{2}g^{\pi}_{\sigma\sigma} \\
b_n &= -g^{\lambda^a}_x + \Omega_1\Gamma_1 + 2\Omega_1\eta_0(I-2C_{n-1})^{-1}\eta'_0\Omega_1 \\
c_n &= 2\Gamma'_1\Omega_0\eta_0(I-2C_{n-1})^{-1}\eta'_0\Gamma_1 + \Gamma'_1\Omega_0\Gamma_1 - \frac{1}{2}g^{\lambda^a}_{xx} \\
&\quad +[(g^{\lambda^a}_x - b_a - g^{\pi}_x + b_{n-1})]_{i_{t-1}}[\Gamma_2]_{i_{t-1}} \\
&\quad +[(g^{\lambda^a}_x - b_a - g^{\pi}_x + b_{n-1})]_{c_{t-1}}[\Gamma_2]_{c_{t-1}} + \frac{1}{2}\Omega^{2,2}_1(I-2C_{n-1})^{-1}\eta_{1,f} \qquad (44)
\end{aligned}
$$

Here the subscript (i_{t-1} , c_{t-1}) denote the element or the matrix related to these variables and $\Omega^{2,2}_1$ means (2,2)th element of Ω_1. Notice that $\eta_{1,f}$ is missed in the approximation of the model solution but captured in the measurement equation for bond yields. The above recursion is the extension of recursion formula in the affine term structure model. If we consider only constant and first order terms and ignore the second order terms in the approximate model solution, we are back to the following affine recursion formula much discussed in Wu (2005), for example.

$$
\begin{aligned}
a_n &= a_{n-1} + \frac{1}{2}\Omega_1\eta_0\eta'_0\Omega'_1 \\
b_n &= -g^{\lambda^a}_x + \Omega_1\Gamma_1 \\
\text{where } a_0 &= 0\ ,\ b_0 = [0,\cdots,0] \qquad (45)
\end{aligned}
$$

9.3 Particle filtering algorithm

The predictive likelihood $p(z_t|z^{t-1},\vartheta)$ can be evaluated by $\int\int p(z_t|x_t,\vartheta)p(x_t|x_{t-1},\vartheta)p(x_{t-1}|z^{t-1},\vartheta)dx_{t-1}dx_t$. In the linear and Gaussian world, the Kalman filter provides the analytical solution for the integral. In the general nonlinear model, that is no longer the case but we can approximate the integral by

Monte Carlo methods. Particle filtering belongs to these methods. The algorithm can be described as follows.

- **Step 1 Initialization** : Draw N particles $\{x_0^i\}_{i=1}^N$ by using the initial distribution of $p(x_0|\vartheta)$. In period t, we are given with the particles $\{x_{t-1}^i\}_{i=1}^N$ which are randomly sampled from the discrete approximation of the true filtering density $p(x_{t-1}|z^{t-1}, \vartheta)$.

- **Step 2 Prediction** : Draw one-step ahead particles $\{\hat{x_t^i}\}_{i=1}^N$ by generating one draw from $p(x_t|x_{t-1}^i, \vartheta)$ for each i. Thus, the distribution of $\{\hat{x_t^i}\}_{i=1}^N$ approximates that of $p(x_t|z^{t-1}, \vartheta)$ by

$$p(x_t|z^{t-1}, \vartheta) \approx \frac{1}{N}\sum_{i=1}^{N} p(x_t|x_{t-1}^i, \vartheta).$$

- **Step 3 Updating** : The true filtering density $p(x_t|z^t, \vartheta)$ is proportional to $p(z_t|x_t, \vartheta)p(x_t|z^{t-1}, \vartheta)$. Since $\{\hat{x_t^i}\}_{i=1}^N$ are generated from the approximated $p(x_t|z^{t-1}, \vartheta)$, the approximation of the filtering density reduces to adjusting the probability weights assigned to $\hat{x_t^i}_{i=1}^N$ according to $\hat{\pi_t^i} = p(z_t|\hat{x_t^i}, \vartheta)$. Here, the role of measurement errors is crucial. Without measurement errors, $\hat{\pi_t^i}$ would have 0 or 1 point mass. With continuously distributed measurement errors, $\hat{\pi_t^i}$ would be positive for all i. That essentially makes the likelihood function which is evaluated as the average of $\hat{\pi_t^i}$, $(i = 1, \cdots, N)$ less rough. We normalize $\{\hat{\pi_t^i}\}_{i=1}^N$ as follows:

$$\pi_t^i = \frac{\hat{\pi_t^i}}{\sum_{j=1}^N \hat{\pi_t^j}}$$

  The resulting sampler $\{\hat{x_t^i}, \pi_t^i\}_{i=1}^N$ approximates the true filtering density $p(x_t|z^t, \vartheta)$.

- **Step 4 Resampling** : The above samplers are undesirable because after a few iterations, most particles will have negligible weights and the accuracy of Monte Carlo approximation of the integral in **Step 2** and **Step 3** would deteriorate. To overcome this problem, we generate a new swarm of particles $\{x_t^i\}_{i=1}^N$ such that

$$Pr(x_t^i = \hat{x_t^i}) = \pi_t^i \ , \ i = 1, \cdots, N$$

The resulting sample is indeed a random sample form the discrete approximation of the filtering density $p(x_t|z^t, \vartheta)$, and hence is equally weighted.[24]

- **Step 5 Likelihood Evaluation** : The log-likelihood can be approximated by using the average of unnormalized weights.

$$\ln \mathcal{L}(\vartheta|z^T) \approx \sum_{t=1}^{T} \ln(\frac{1}{N} \sum_{i=1}^{N} \hat{\pi_t^i})$$

Several remarks are necessary for the actual implementation of the above algorithm. First, since our state variables include pre-determined endogenous variables as well as structural shocks which follow linear processes, it is not obvious to get initial values for them. As in An (2005), we draw the initial structural shocks from their unconditional distributions and generate the initial values of pre-determined endogenous state variables from putting the previous period's values to their steady state values. Second, we choose the number of particles based on the evaluation of the log-likelihood across 40 different random seeds. The standard deviation of the likelihood values changes only by a small amount after 60,000 particles. That motivates our choice of 60,000 particles.

9.4 MCMC algorithm

The random walk Metropolis-Hastings algorithm widely used in the estimation of DSGE models usually starts from the posterior mode. However, in our case, finding

[24]By the nature of the discrete approximation of the filtering density, the likelihood evaluated by particle filters may not be continuous with respect to parameters. This can be a serious problem for the numerical maximization of the likelihood. Our Markov Chain Monte Carlo methods are less susceptible for the lack of the continuity but still the problem is a concern for the posterior inference.

the posterior mode by numerical optimization routines does not work well. But as Chernozuhkov and Hong (2003) shows, MCMC algorithm itself can be used as an optimization tool. Indeed, we find even after a small number of draws we get to the point with a higher posterior density than the one reached by the simplex method.

- **Step 1 Selection of the Starting Point** : Compute the log-likelihood for 100 draws from prior distribution. Select one point which gives the highest log-likelihood value $\vartheta^{\star}$.

- **Step 2 Proposal** : Starting from $\vartheta^{\star}$, generate a new draw by the following random-walk proposal density. The scaling matrix c is chosen by multiplying a small positive real number to the prior covariance matrix.

$$\hat{\vartheta}^{j+1} = \vartheta^{j} + c\mathcal{N}(0, I) \ , \ j = 0, \cdots, R-1$$

- **Step 3 Accept/Reject** Compute the acceptance rate $\alpha = \min\{\frac{p(\hat{\vartheta}^{j+1}|z^T)}{p(\vartheta^{j}|z^T)}, 1\}$ and accept or reject $\hat{\vartheta}^{j+1}$ according to the value of u which is drawn from the uniform distribution over the unit interval $[0, 1]$.

$$\begin{aligned} \vartheta^{j+1} &= \hat{\vartheta}^{j+1} \ , \ \text{if } u < \alpha \\ & \quad \vartheta^{j} \ , \ \text{otherwise} \end{aligned}$$

- **Step 4 Burn-In** For the purpose of the posterior inference, burn in the initial B draws and use the remaining draws.

We try multiple MCMC chains and select the highest posterior region. Unfortunately, different MCMC chains are wandering around deeply separated regions of the parameter space even after significant draws (200,000 for the nonlinear model, 1,000,000 for the linear model). In the quadratic model, the highest posterior region is associated with the MCMC chain starting from the $\vartheta^{\star}$. However, in the linear model, the difference of the posterior density across different chains is small. Although Hoogerheide et al. (2006) suggest a sophisticated proposal density which can explore deeply separated regions of the parameter space, the algorithm depends on the numerical optimization which does not work well for our case. We left this issue as a future agenda. In the paper, we report results from the MCMC chain whose starting point is $\vartheta^{\star}$.

9.5 Monte Carlo smoothing algorithm

The key steps of Monte Carlo smoothing can be described as follows:

- **Step 1** : Store resampled state variables x_t^i at each time for $i = 1, \cdots N$
- **Step 2** : For $\{x_T^i\}_{i=1}^M$, calculate $w_{T-1|T}^{i,j} \propto p(x_T^i | x_{T-1}^j)$ where $M <= N$.
- **Step 3** : Choose $x_{T-1}^k = x_{T-1}^j$ with probability $w_{T-1|T}^{i,j}$ for $k = 1, \cdots N$.
- **Step 4** : Repeat **Step 2** and **Step 3** until we get M trajectories of smoothed states $x_{1:T}$ conditional on z^T

We follow Godsill et.al. (2004) and generate a bunch of trajectories of state variables based on the resampled state variables in the forward filtering. Fernández-Villaverde and Rubio-Ramírez (2006) make the number of trajectories equal to that of particles. However, since the computation time of one trajectory amount to that of the evaluation of the likelihood, this is computationally costly. In a simulation study, 6,000 trajectories are enough to make the mean estimates close to true values. Smoothed estimates of macro factors in the paper are based on 6,000 trajectories of state variables. Figure 13 shows the accuracy of smoothed estimates for the data simulated from our model. Mean smoothed estimates are found to track true states reasonably well.

References

Ahn,D.,R.F. Dittmar, and A.R. Gallant (2002): “Quadratic Term Structure Models:Theory and Evidence,” *Review of Financial Studies*, **15**(1), 243-288.

An, S. (2005): "Bayesian Estimation of DSGE Models:Lessons from Second-order Approximations," *Manusript*, University of Pennsylvania.

An, S. and F. Schorfheide (2006): "Bayesian Analysis of DSGE Models," *Manuscript*, University of Pennsylvania.

Ang, A. and M. Piazzesi (2003): "A no-arbitrage vector-autoregression of term structure dynamics with macroeconomic and latent variables," *Journal of Monetary Economics*, **50**, 745-787.

Ang, A.,S. Dong, and M. Piazzesi (2005): "No-Arbitrage Taylor Rules," *Manuscript*, University of Chicago.

Arulampalam, M.S.,S.Maskell,N.Gordon, and T. Clapp (2002): "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, **50**(2), 173-188.

Bekaert, G.,S. Cho, and A. Moreno (2005): "New-Keynesian Macroeconomics and the Term Structure," *Working Paper*, National Bureau of Economic Research.

Bils, M. and P. Klenow (2004): "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, **112**(5), 383-398.

Borkovec, M. and C. Klüppelberg (2001): "The Tail of the Stationary Distribution of an Autoregressive Process with ARCH(1) Errors," *Annals of Applied Probability*, **11**, 1220-1241.

Boudoukh, J. (1993): "An Equilibrium Model of Nominal Bond Prices with Inflation-Output Correlation and Stochastic Volatility," *Journal of Money, Credit, and Banking*, **25**(3), 636-665.

Brandt, M. and A. Yaron (2002): "Time-Consistent No-Arbitrage Models of the Term Structure," *Manuscript*, University of Pennsylvania.

Buraschi, A. and A. Jiltsov (2005): "Inflation risk premia and the expectations hypothesis," *Journal of Financial Economics*, **75**, 429-490.

Chang, Y.,T. Doh, and F. Schorfheide (2006): "Non-stationary Hours in a DSGE Model," *Journal of Money, Credit, and Banking*, forthcoming.

Chernozuhkov, V. and H. Hong (2003): "An MCMC Approach to Classical Estimation,", *Journal of Econometrics*, **115**(2), 293-346.

Clarida, R.,J.Galí, and M. Gertler (2000): "Monetary Policy Rules and Macroeconomic Stability:Evidence and Some Theory," *Quarterly Journal of Economics*, **115**(1), 147-180.

Cochrane,J. and M.Piazzesi (2006): "Decomposing the Yield Curve," *Manuscript*, University of Chicago.

Cogley, T. (2005): "Changing Beliefs and the Term Structure of Interest Rates:Cross-equation Restrictions with Drifting Parameters," *Review of Economic Dynamics*, **8**, 420-451.

Del Negro, M. and F. Schorfheide (2006): "Forming Priors for DSGE Models(and How It Affects the Assessment of Nominal Rigidities," *Manuscript*, University of Pennsylvania.

Den Haan W. (1995): "The Term Structure of Interest Rates in Real and Monetary Economies," *Journal of Economic Dynamics and Control*, **19**, 909-940.

Diebold, F.X.,G.D. Rudebusch, and B. Aruoba (2006): "The Macroeconomy and the Yield Curve:A Dynamic Latent Factor Approach," *Journal of Econometrics*,**131**(1-2), 309-338.

Doh,T. (2006): "What does the Yield Curve Tell us about the Fed's Target Inflation: A Structural Approach?," *Manuscript*, University of Pennsylvania.

Duffee, G. (2006):"Are Variations in Term Premia Related to the Macroeconomy?," *Manuscript*, University of Berkeley.

Evans, C.L. and D. Marshall (2002): "Economic Determinants of the Nominal Treasury Yield Curve," *Working Paper*, Federal Reserve Bank of Chicago.

Fernàndez-Villaverde, J. and J. Rubio-Ramírez (2006): "Estimating Macroeconomic Models:A Likelihood Approach," *Manuscript*, University of Pennsylvania.

Geweke, J. (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication," *Econometric Reviews*, **18**(1), 1-126.

Godsill, S.J.,A. Doucet, and M.West (2004): "Monte Carlo Smoothing for Nonlinear Time Series," *Journal of the American Statistical Association*, **99**(465), 156-168.

Goodfriend, M. (1993): "Interest Rate Policy and Inflation Scare Problem:1979-1992," *Federal Reserve Bank of Richmond Economic Quarterly*, **79**(1), 1-23.

Goodfriend, M. and R. King (2005):"The Incredible Volcker Disinflation," *Journal of Monetary Economics*, **52**(5), 981-1015.

Hoogerheide, L.F.,J.F.Kaashoek, and H.K. van Dijk (2006): "On the Shape of Posterior Densities and Credible Sets in Instrumental Variable Regression Models with Reduced Rank:An Application of Flexible Sampling Methods Using Neural Networks," *Journal of Econometrics*, forthcoming.

Hördahl, P.,O. Tristani, and D. Vestin (2005a): "A Joint Econometric Model of Macroeconomic and Term Structure Dynamics," *Journal of Econometrics*, forthcoming.

Hördahl, P.,O. Tristani, and D. Vestin (2005b): "The Yield Curve and Macroeconomic Dynamics," *Manuscript*, European Central Bank.

Jermann, U. (1998): "Asset Pricing in Production Economy," *Journal of Monetary Economics*, **41**, 257-275.

Kim, J.,S.H. Kim,E.Schaumburg, and C.A. Sims (2005): "Calculating and Using Second Order Accurate Solutions of Discrete Time Dynamic Equilibrium Models," *Manuscript*, Princeton University.

Kitagawa, G. (1996): "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear Stats Space Models," *Journal of Computational and Graphical Statistics*, **5**(1), 1-25.

Klein, P. (2005): "Second-order Approximation of Dynamic Models without the Use of Tensors," *Manuscript*, University of Western Ontario.

Kozicki, S. and P.A. Tinsley(2005): "Permanent and Transitory Policy Shocks in an Empirical Macro Model with Asymmetric Information," *Journal of Economic Dynamics and Control*, **29**(11), 1985-2015.

Litterman, R. and J.A. Scheinkman (1991): "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, **1**, 51-61.

Ludvigson, S. and S. Ng (2005): "Macro Factors in Bond Risk Premia," *Working Paper*, National Bureau of Economic Research.

Lubik, T. and F. Schorfheide (2004): "Testing for Indeterminacy:An Application to U.S. Monetary Policy," *American Economic Review*, **94**(1), 190-217.

Lubik, T. and F. Schorfheide (2005): "A Bayesian Look at New Open Economy Macroeconomics" *NBER Macroeconomics Annual*, forthcoming.

Piazzesi, M. and M. Schneider (2006): "Equilibrium Bond Yields," *NBER Macroeconomics Annual*, forthcoming.

Primiceri, G. (2005): "Why Inflation Rose and Fell:Policymakers' Beliefs and US Postwar Stabilization Policy," *Quarterly Journal of Economics*, forthcoming.

Ravena, F. and C.E. Walsh (2004): "Optimal Monetary Policy with the Cost Channel," *Manuscript*, University of California Santa Cruz.

Rudebusch, G.D. and T. Wu (2004): "A Macro-Finance Model of the Term Structure, Monetary Policy, and the Economy," *Staff Report*, Federal Reserve Bank of San Francisco.

Sargent, T.,N. Williams, and T. Zha (2006): "Shocks and Government Beliefs: The Rise and Fall of American Inflation," *American Economic Review*, **96**(4), 1193-1224.

Schmitt-Grohé,S.,and M. Uribe (2004): "Solving Dynamic General Equilibrium Models Using an Second-order Approximation to the Policy Function," *Journal of Economic Dynamics and Control*, **28**(4), 755-775.

Schorfheide, F. (2005): "Learning and Monetary Policy Shift," *Review of Economic Dynamics*, **8**(2), 392-419.

Wachter, J. (2006): "A Consumption-based Model of the Term Structure of Interest Rates," *Journal of Financial Economics*, **79**, 365-399.

Woodford, M. (2003): "Interests and Prices: Foundations of a Theory of Monetary Policy,"Princeton University Press.

Wu, T. (2005): "Macro Factors and the Affine Term Structure of Interest Rates," *Journal of Money, Credit, and Banking*, forthcoming.

Table 1: Prior Distribution

Parameters	Domain	Density	Para(1)	Para(2)
τ	$\mathbb{R}^+$	Gamma	2	0.5
β	$[0, 1)$	Beta	0.998	0.001
$\ln f^\star$	$\mathbb{R}^+$	Gamma	0.11	0.05
ϕ	$\mathbb{R}^+$	Gamma	100	40
$u_a^\star$	$\mathbb{R}^+$	Gamma	0.005	0.002
γ_p	$\mathbb{R}^+$	Gamma	2	0.2
γ_y	$\mathbb{R}^+$	Gamma	0.5	0.2
ρ_a	[0,1)	Beta	0.3	0.1
ρ_f	[0,1)	Beta	0.8	0.1
ρ_i	[0,1)	Beta	0.5	0.2
$\rho_{\pi^\star}$	[0,1)	Beta	0.8	0.1
η_a	$\mathbb{R}^+$	Inverse Gamma	0.004	4
$\eta_{0,f}$	$\mathbb{R}^+$	Inverse Gamma	0.010	4
η_i	$\mathbb{R}^+$	Inverse Gamma	0.003	4
$\eta_{\pi^\star}$	$\mathbb{R}^+$	Inverse Gamma	0.002	4
$\ln A_0$	$\mathbb{R}$	Normal	9.951	0.2
$\ln \pi_0^\star$	$\mathbb{R}^+$	Gamma	0.01	0.002
h	[0,1)	Beta	0.3	0.1
$\eta_{1,f}$	[0,2.42)	Uniform	0	2.42

Notes: Para (1) and Para (2) list the means and the standard deviations for Beta, Gamma, and Normal distributions; s and ν for the Inverse Gamma distribution, where $p_{\mathcal{IG}}(\sigma|\nu, s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$, a and b for the Uniform distribution from a to b. The elasticity of labor supply ν is fixed at 0.5 which is roughly the posterior mean of the parameter in Chang et al.(2006). Standard deviations of measurement errors are also fixed as explained in the paper.

Table 2: Posterior Distribution

	Prior	Posterior(Linear)		Posterior(Quadratic)	
	90% Interval	Mean	90% Interval	Mean	90% Interval
τ	[1.18,2.79]	2.47	[1.71,3.27]	2.72	[2.11,3.36]
β	[0.997,0.999]	0.997	[0.995,0.998]	0.997	[0.996,0.999]
$\ln f^{\star}$	[0.033,0.185]	0.055	[0.021,0.092]	0.055	[0.029,0.088]
ϕ	[36.0,160]	58.2	[18.5,93.9]	65.6	[30.9,98.8]
$400u_a^{\star}$	[0.8,3.2]	2.4	[1.6,2.8]	2.0	[1.6,2.4]
γ_p	[1.68,2.33]	2.10	[1.76,2.44]	2.18	[1.89,2.53]
γ_y	[0.19,0.81]	0.17	[0.06,0.28]	0.15	[0.06,0.23]
ρ_a	[0.135,0.459]	0.145	[0.074,0.219]	0.162	[0.073,0.239]
ρ_f	[0.650,0.961]	0.822	[0.776,0.867]	0.801	[0.718,0.855]
ρ_i	[0.171,0.825]	0.627	[0.488,0.786]	0.642	[0.541,0.758]
$\rho_{\pi^{\star}}$	[0.651,0.960]	0.990	[0.983,0.997]	0.986	[0.979,0.994]
$400\eta_a$	[0.8,3.2]	2.8	[2.4,3.2]	2.8	[2.4,3.2]
$100\eta_{0,f}$	[0.5,2.0]	1.3	[0.9,1.7]	1.0	[0.7,1.2]
$400\eta_i$	[0.64,2.36]	0.8	[0.68,0.96]	0.8	[0.68,0.96]
$400\eta_{\pi^{\star}}$	[0.4,1.6]	0.36	[0.32,0.44]	0.4	[0.32,0.44]
$\ln A_0$	[9.62,10.28]	9.82	[9.75,9.90]	9.81	[9.75,9.87]
$400\ln\pi^{\star}$	[2.8,5.2]	5.2	[4.0,6.0]	6.4	[4.4,7.6]
h	[0.14,0.46]	0.42	[0.30,0.54]	0.41	[0.34,0.47]
$\eta_{1,f}$	[0.223,2.399]	0.948	[0.001,1.717]	0.040	[0.002,0.086]

Table 3: Log Marginal Data Densities

	Linear	Quadratic
Log marginal data density	3149.7	3151.3

Table 4: Smoothed Estimates of Measurement Errors

	π_t	$y_{1,t}$	$y_{4,t}$	$y_{8,t}$	$y_{12,t}$	$y_{16,t}$	$y_{20,t}$
mean	-0.009	-0.036	-0.005	-0.0248	-0.0116	0.0358	0.0362
standard deviation	0.044	0.221	0.252	0.214	0.152	0.148	0.182
mean absolute values	0.033	0.180	0.197	0.166	0.117	0.123	0.143

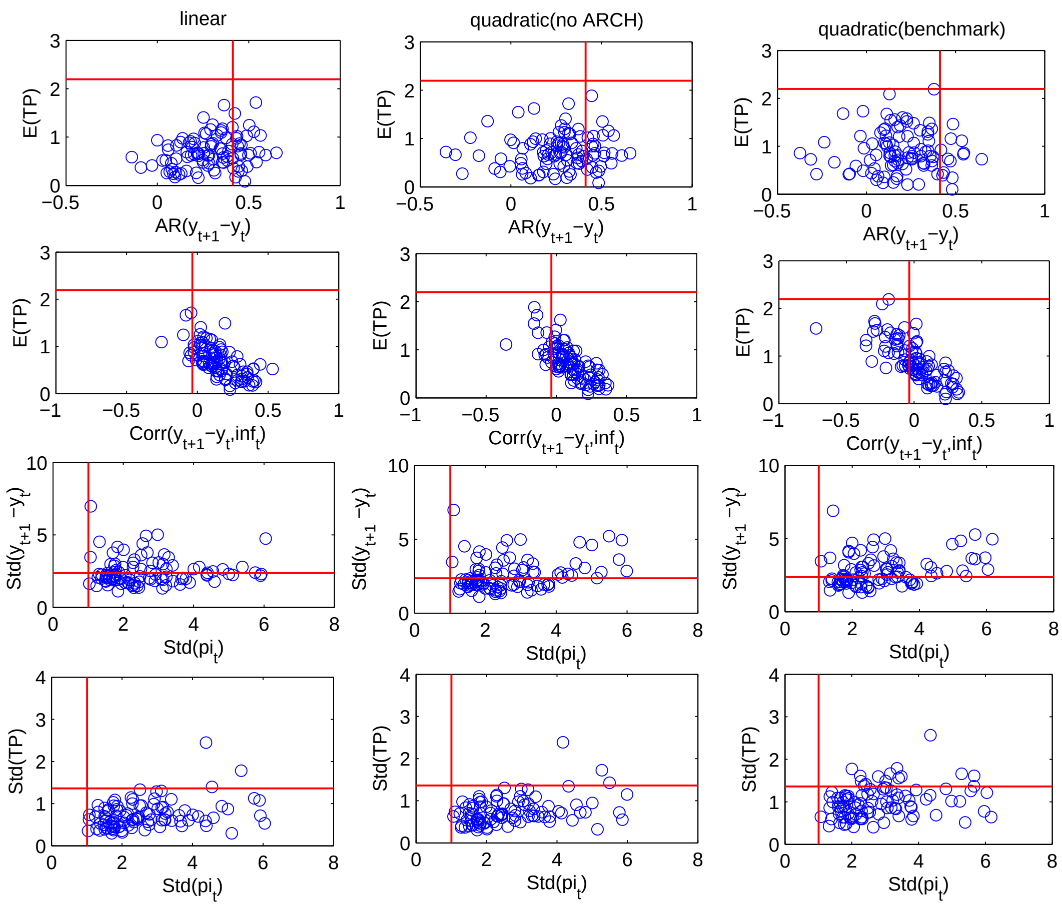
Notes: All the estimates are in terms of the annualized percentage. They are evaluated at the posterior mean values of parameters.

Table 5: Regression of Empirical Counterparts of Term Structure Factors and term premium

Regression	level	slope	curvature	term premium
constant	9.354 (0.059)	0.565 (0.202)	0.068 (0.157)	3.527 (0.469)
$\hat{u}_{a,t}$	-0.130 (0.015)	-0.214 (0.052)	0.096 (0.040)	-0.272 (0.104)
$\hat{\ln f_t}$	0.653 (0.026)	1.116 (0.088)	-0.120 (0.068)	0.331 (0.202)
$\epsilon_{i,t}$	1.483 (0.145)	1.177 (0.495)	1.952 (0.384)	0.135 (0.189)
$\hat{\ln \pi_t^\star}$	2.440 (0.033)	0.563 (0.111)	-0.032 (0.086)	1.522 (0.228)
R^2	0.997	0.794	0.565	0.728

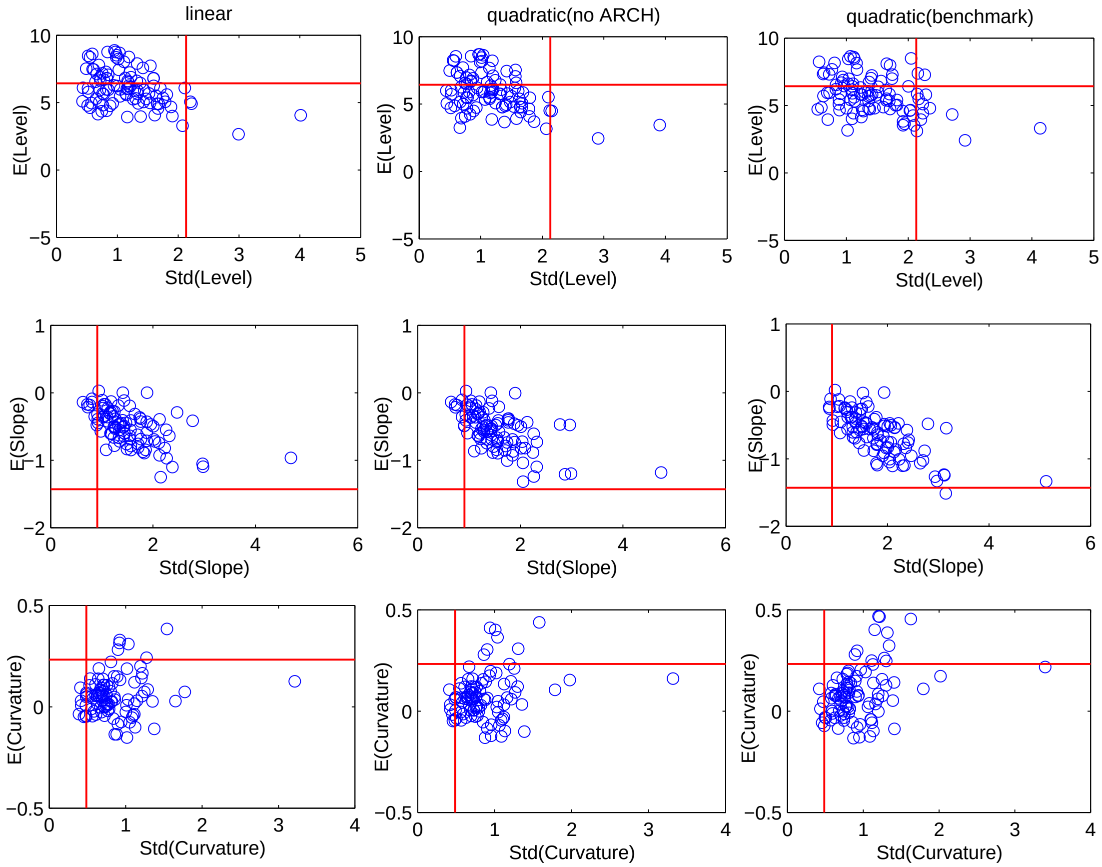
Notes: The level, slope, and curvature of the yield curve are defined by $\frac{y_{20,t}+y_{8,t}+y_{1,t}}{3}$, $y_{1,t}-y_{20,t}$, and $2y_{8,t}-y_{20,t}-y_{1,t}$. Term premium is defined by $y_{20,t}-\frac{\sum_{j=0}^{19} y_{1,t+j}}{20}$. Regressors are smoothed estimates of macro factors(expressed in terms of log deviations from steady state values) at the posterior mean of the quadratic model. Each regressor is normalized so that the variance of it is equal to 1. Numbers in ( ) are standard errors.

Figure 1: Prior Predictive Checks I



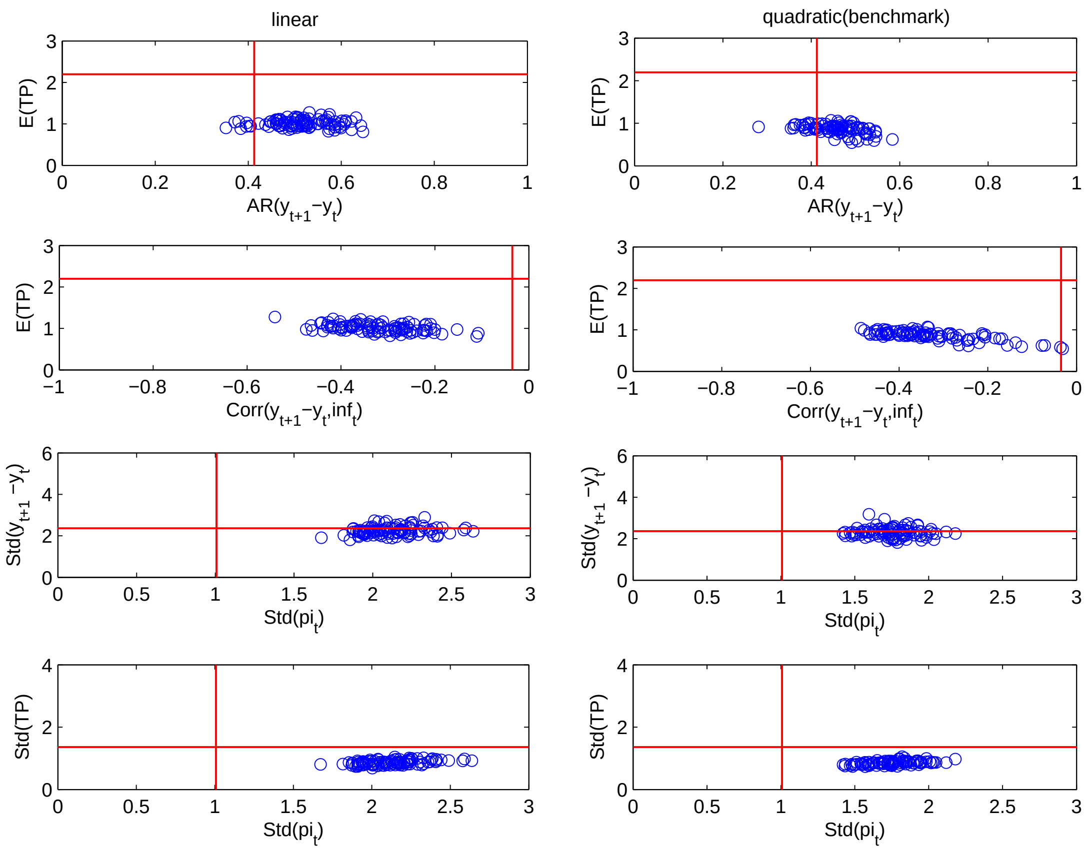
Notes: TP stands for the term premium defined by $y_{20,t} - \frac{(\sum_{j=0}^{19} i_{t+j})}{20}$. 1,000 draws from the prior distribution are used for simulation. Red lines represent sample moments from actual data.

Figure 2: Prior Predictive Checks II



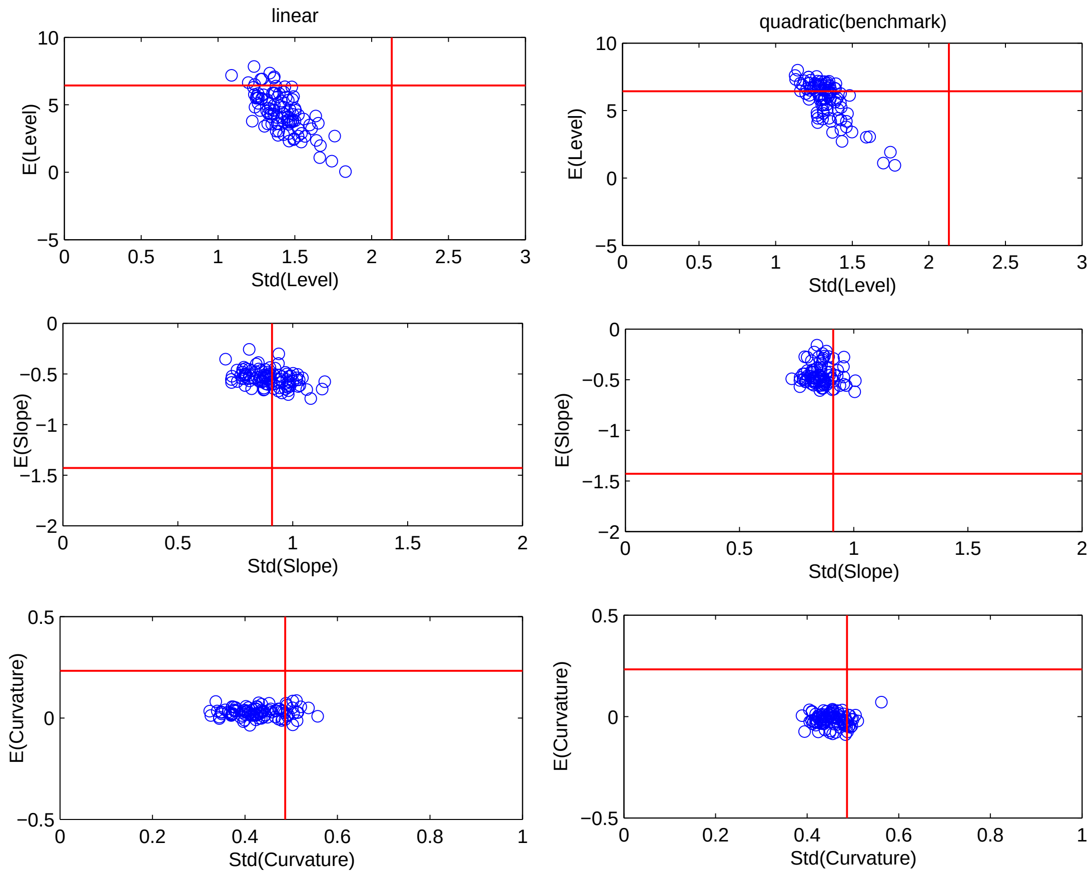
Notes: TP stands for the term premium defined by $y_{20,t} - \frac{(\sum_{j=0}^{19} i_{t+j})}{20}$. 1,000 draws from the prior distribution are used for simulation. Red lines represent sample moments from actual data.

Figure 3: Posterior Predictive Checks I



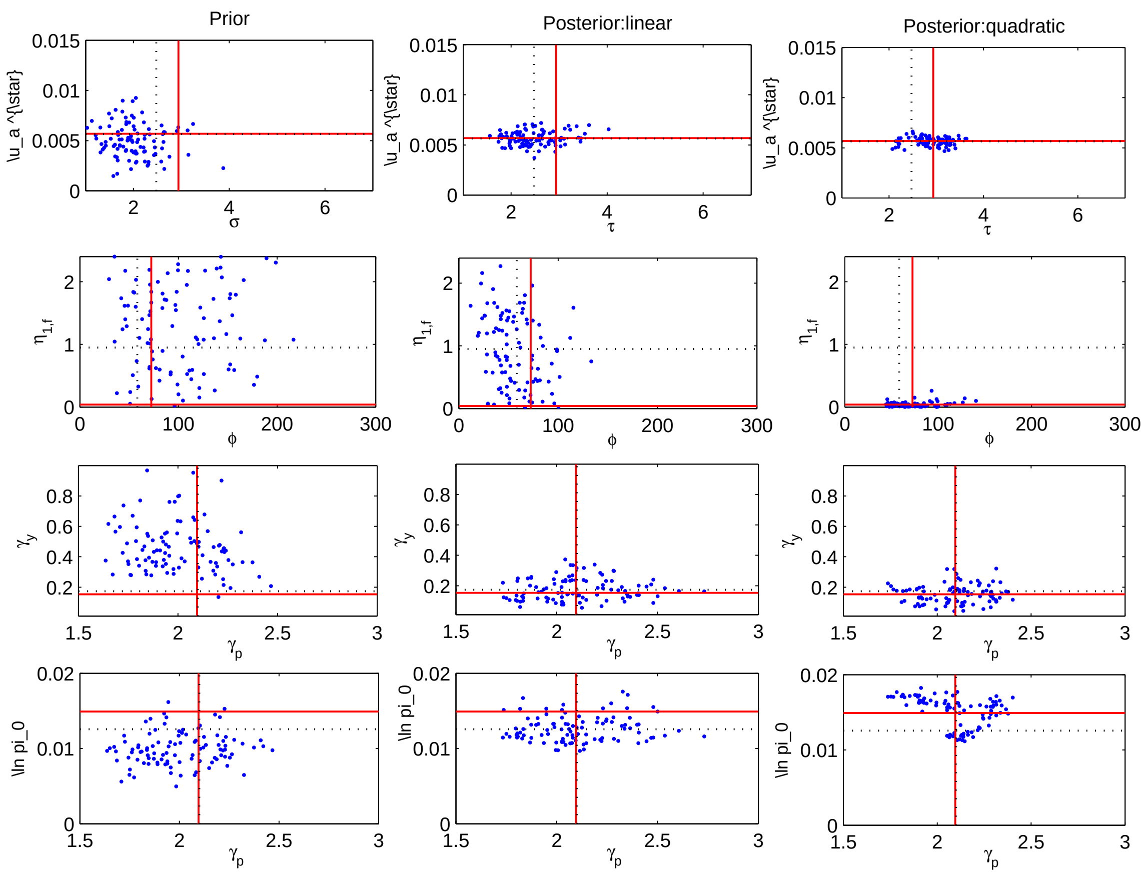
Notes: TP stands for the term premium defined by $y_{20,t} - \frac{(\sum_{j=0}^{19} i_{t+j})}{20}$. Every 500 draw of parameters from 50,000 posterior draws is used for simulation. Red lines represent sample moments from actual data.

Figure 4: Posterior Predictive Checks II



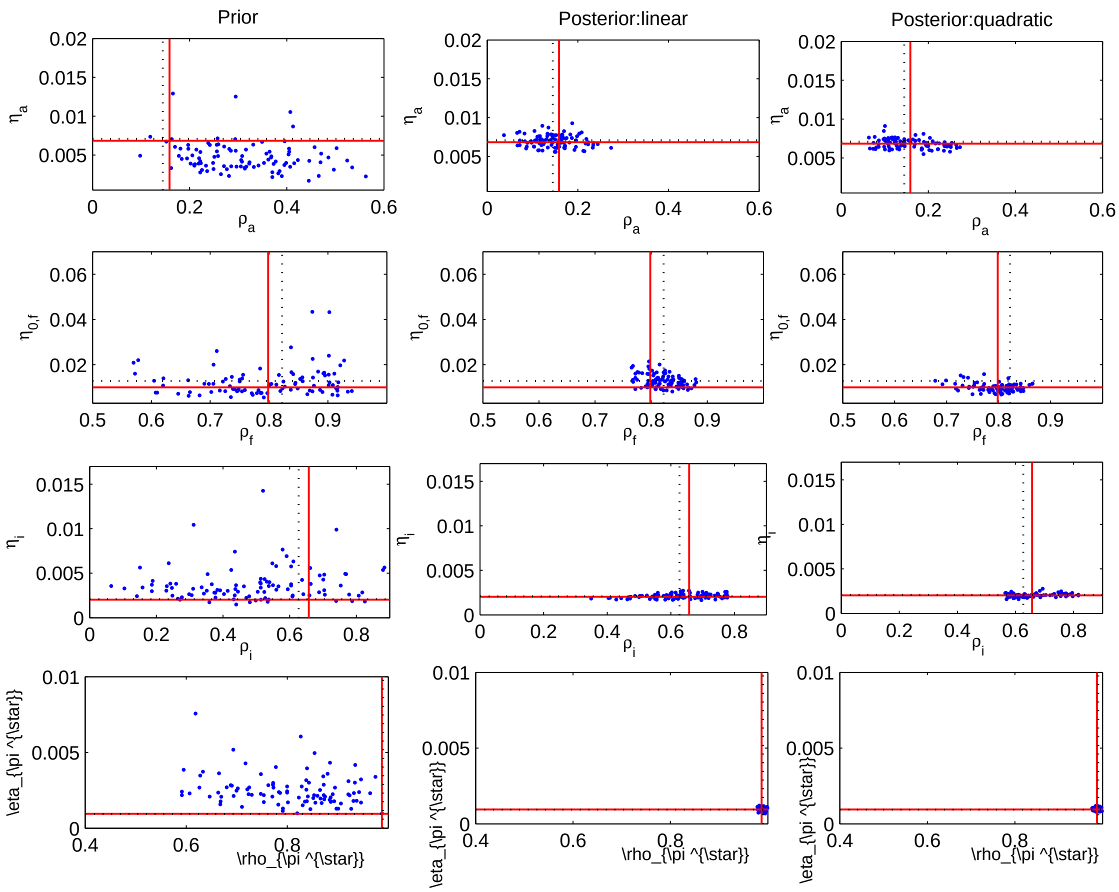
Notes: TP stands for the term premium defined by $y_{20,t} - \frac{(\sum_{j=0}^{19} i_{t+j})}{20}$. Every 500 draw of parameters from 50,000 posterior draws is used for simulation. Red lines represent sample moments from actual data.

Figure 5: Prior-Posterior Draws I



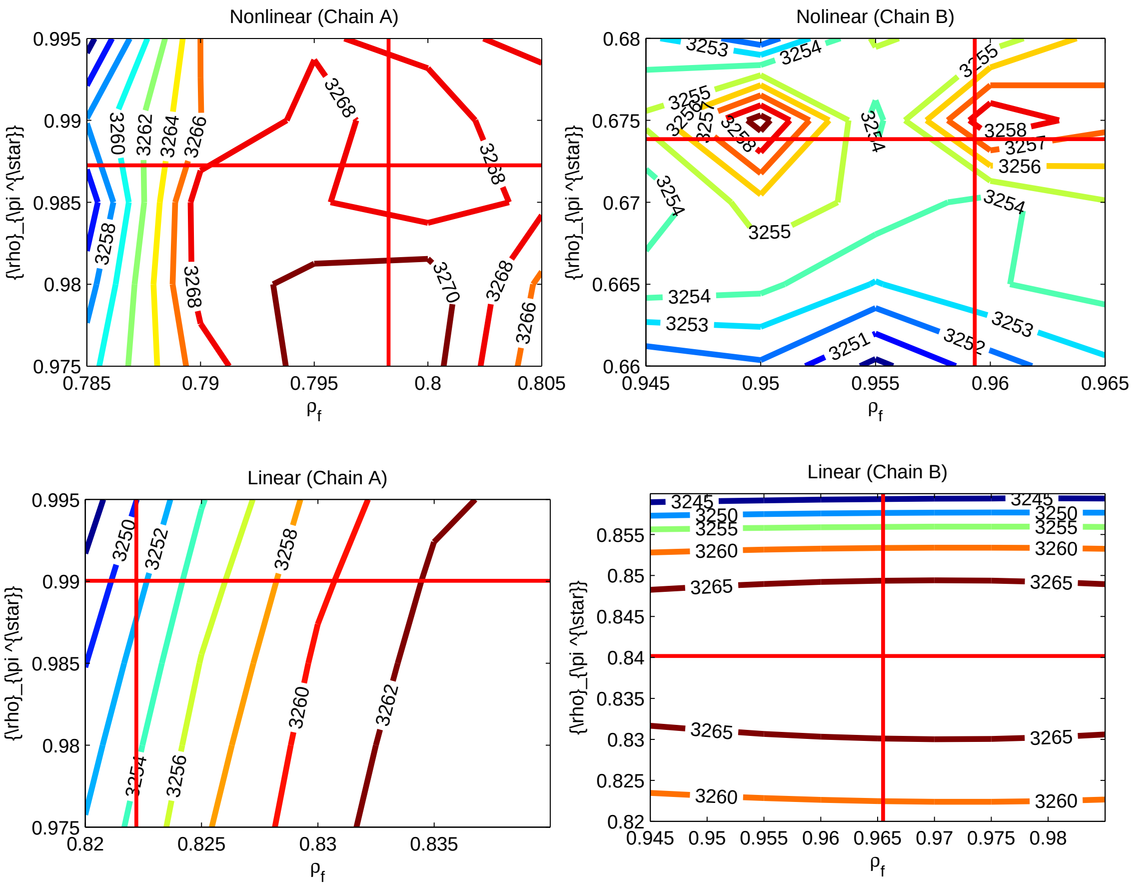
Notes: Every 500 draw of parameters from 50,000 posterior draws is plotted. 100 posterior draws are contrasted with 100 prior draws. The intersection of solid lines represent the posterior mean from the quadratic model while the intersection of dot lines represent the posterior mean from the linear model.

Figure 6: Prior-Posterior Draws II



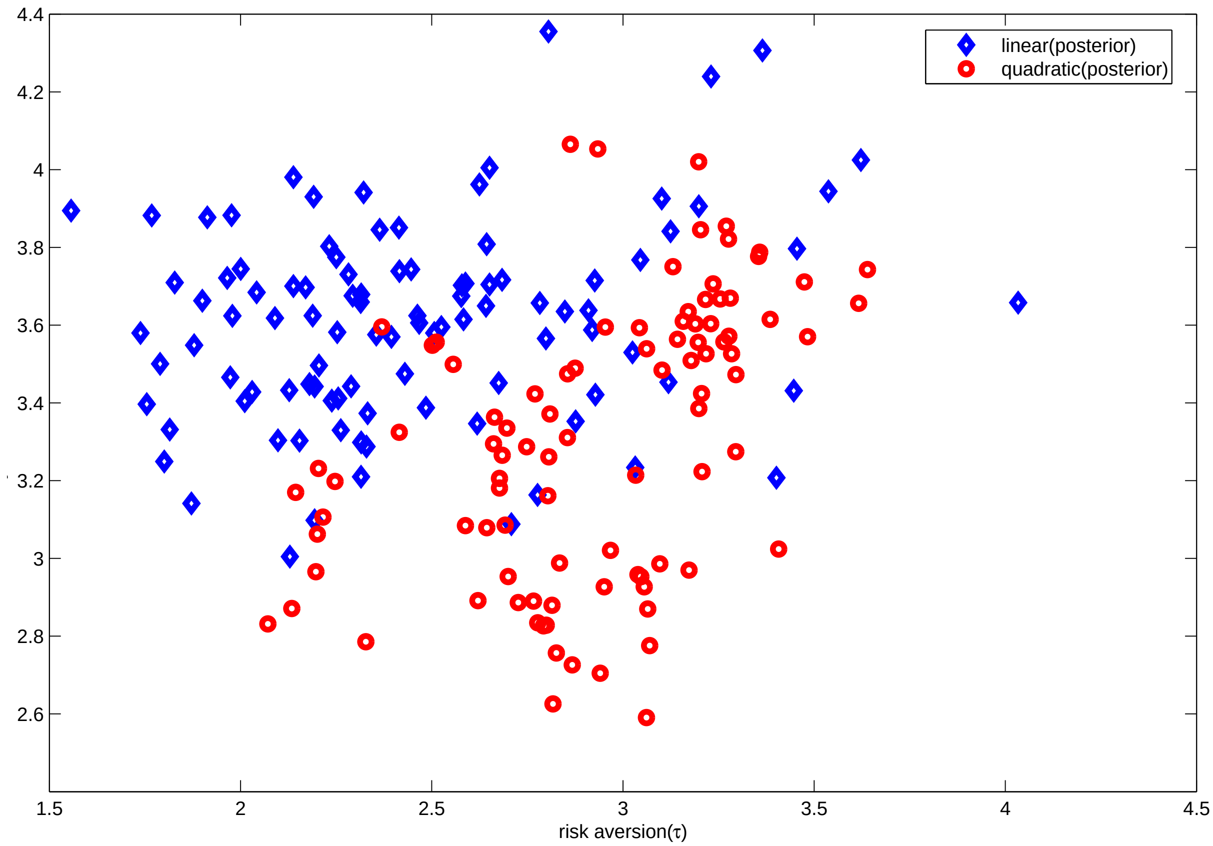
Notes: Every 500 draw of parameters from 50,000 posterior draws is plotted. 100 posterior draws are contrasted with 100 prior draws. The intersection of solid lines represent the posterior mean from the quadratic model while the intersection of dot lines represent the posterior mean from the linear model.

Figure 7: Posterior Contour



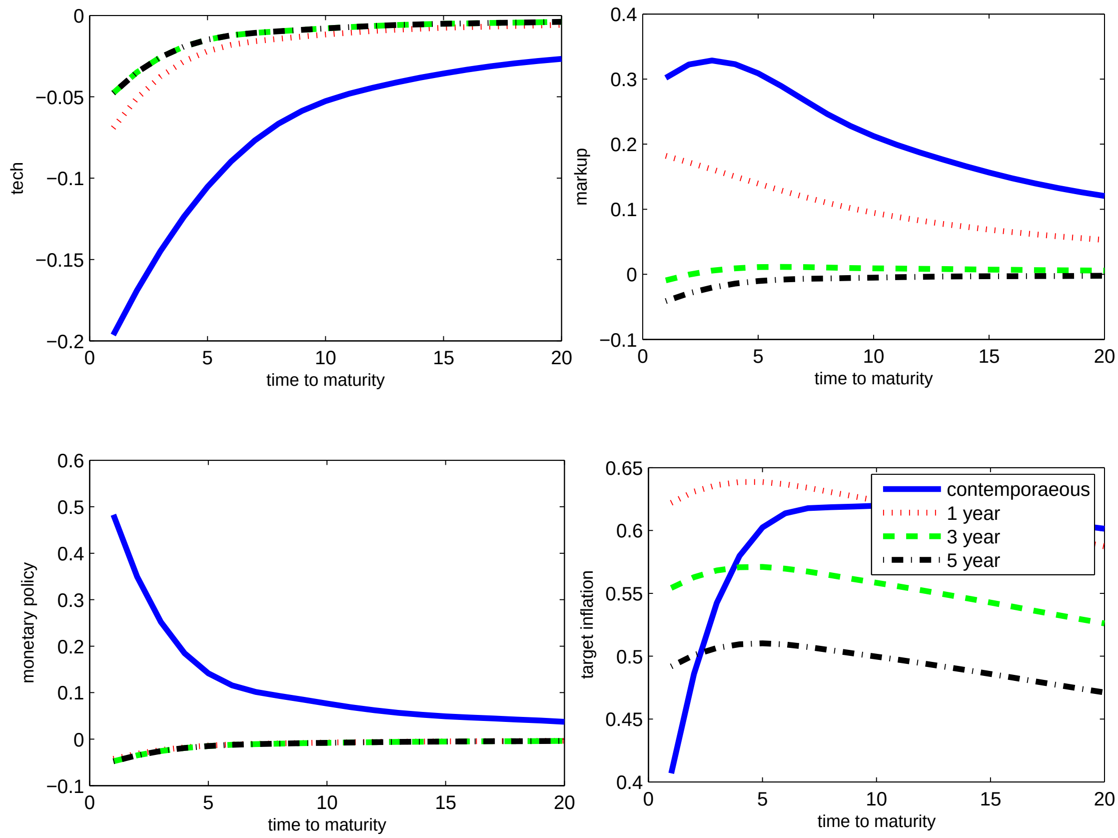
Notes: The intersection of red lines represents the posterior mean computed from the results of each chain. All the posterior inferences in this paper are based on results from chain A.

Figure 8: Steady State Real Interest Rate and Risk Aversion



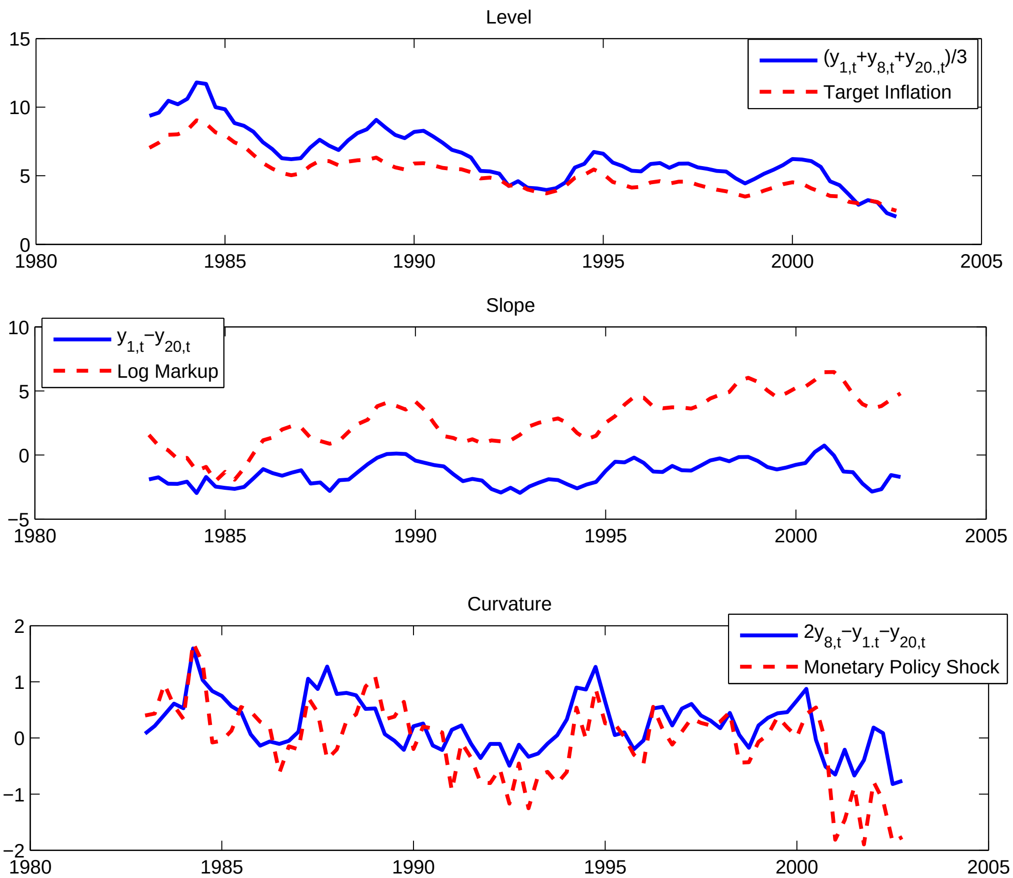
Notes: Every 500 draw of parameters from 50,000 posterior draws is used.

Figure 9: Dynamic Responses of Bond Yields



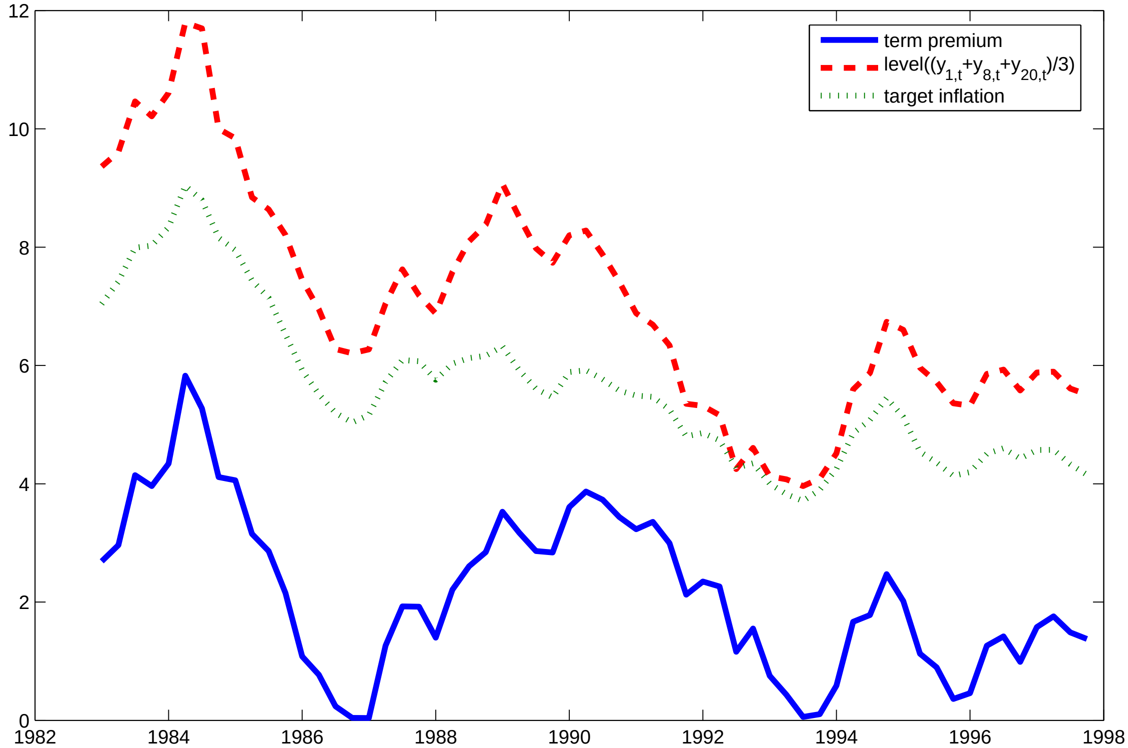
*Notes:*Dynamic responses of bond yields are conditional on the one standard deviation shock in the current period and no shocks in the future. The economy is assumed to be at the steady state initially.

Figure 10: Macro Factors and Empirical Counterparts of Term Structure Factors



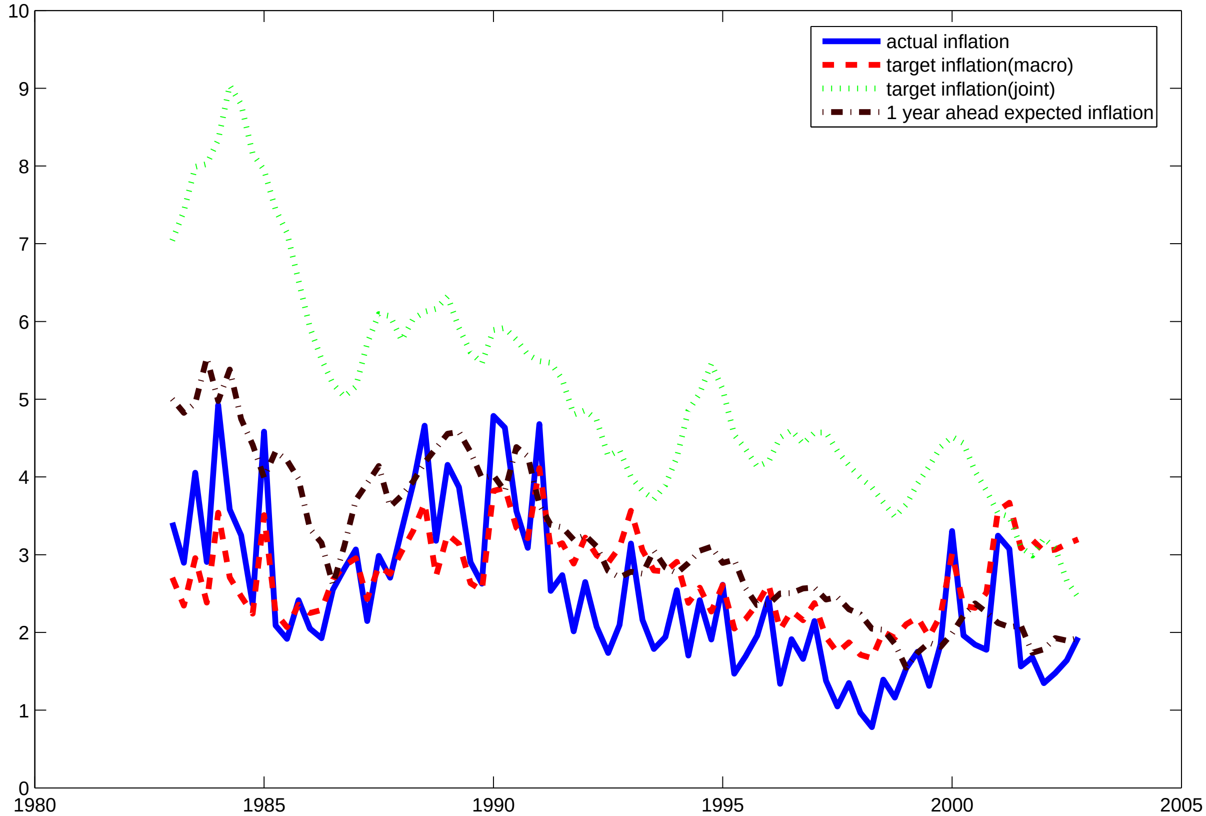
*Notes:*Smoothed estimates of macro factors obtained at the posterior mean of the quadratic model are plotted.

Figure 11: Time Variation of Term Premium



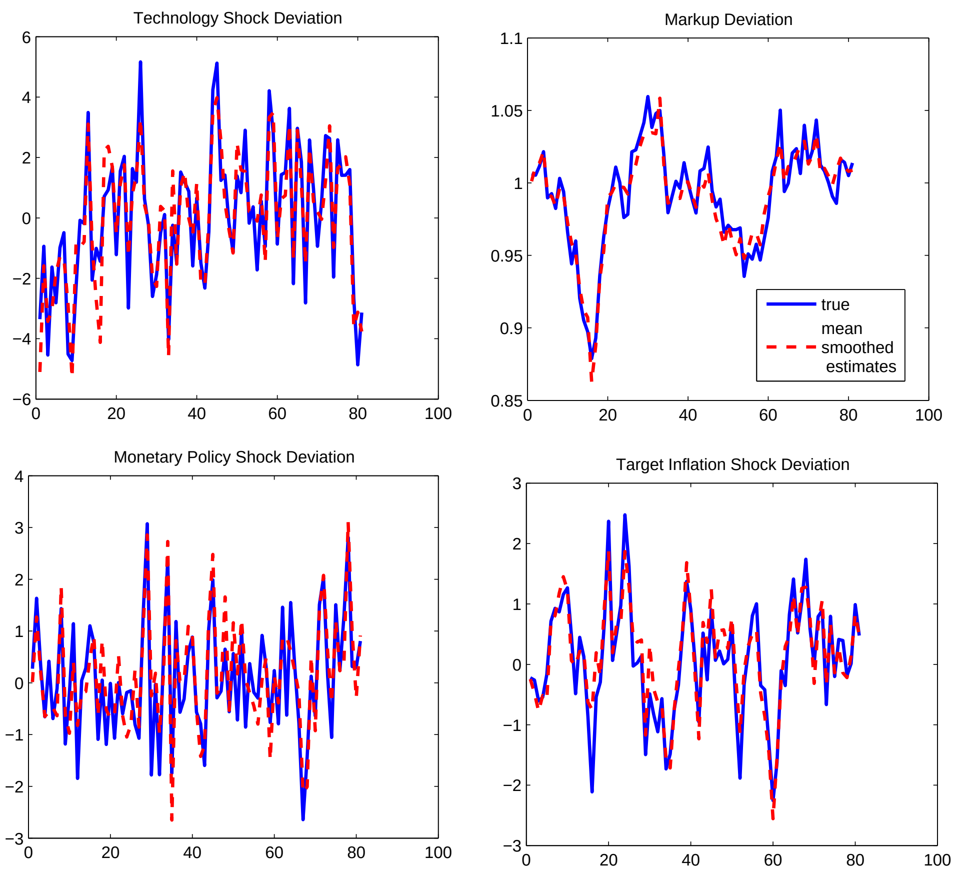
Notes: Term premium is defined by $y_{20,t} - \frac{\sum_{j=0}^{19} y_{1,t+j}}{20}$. Smoothed estimates of target inflation from the posterior mean of the quadratic model are plotted.

Figure 12: Inflation, Expected Inflation, and Target Inflation



Notes: Expected inflation is the mean of one year ahead inflation forecasts from the survey of the professional forecasters provided by the federal reserve bank of Philadelphia.

Figure 13: Accuracy of Smoothed Estimates



Notes: Simulated data from the benchmark model are used for the estimation.