A DSGE model of the term structure with regime shifts*

Gianni Amisano[†]
European Central Bank

Oreste Tristani[‡]
European Central Bank

12 October 2010

Abstract

We construct and estimate the term structure implications of a small DSGE model with nominal rigidities in which the laws of motion of the structural shocks are subject to stochastic regime shifts. We demonstrate that, to a second order approximation, switching regimes generate time-varying risk premia. We then estimate the model on US data relying on information from both macroeconomic variables and the term structure. Our results support the specification with regime-switching: heteroskedasticity is a clear feature of the model's residuals and the regimes have intuitively appealing features. The model also generates non-negligible time-variability in excess holding period returns.

JEL classification:

Keywords: DSGE models, term structure of interest rates, policy rules, regime switches, Bayesian estimation.

*The opinions expressed are personal and should not be attributed to the European Central Bank. We wish to thank Michel Juillard for a precious suggestion on how to speed up the computation of the solution of the model. We also thank Julien Matheron, Benoît Mojon, Giorgio Primiceri, Federico Ravenna, Ken Wallis, Philippe Weil and Paolo Zagaglia for useful comments and suggestions.

[†]Email: gianni.amisano@ecb.int.

[‡]Email: oreste.tristani@ecb.int.

1 Introduction

In the past twenty years, the term structure literature has developed independently of the macro-economic literature. While various arbitrage-free models have been shown to account quite well for yields dynamics – including affine, quadratic and regime switches models – these models have largely ignored the microfoundations of the stochastic discount factor used by the market to price bonds. They have also largely ignored the fact that the short term nominal interest rate, whose expected future values shape the whole yield curve, also plays an important macroeconomic role as a monetary policy instrument.

Yet there is a clearly established empirical relationship between monetary policy and the term structure of interest rates. For example, Mankiw and Miron (1986) points out that the yield curve used to behave differently – i.e. in a way more consistent with the so-called expectations hypothesis[1] – before the founding of the Fed in 1913. Cochrane (2008) highlights an even starker structural break in 1933, with the shift from the Gold standard to an interest rate targeting regime. Contrary to the recent experience, long bond yields were systematically below short rates before 1933; also, long yields were much less volatile, and short rates much more volatile, than what is the case in post-WWII data.

The aforementioned stylised facts suggest that the monetary policy should play a central role in models of the term structure of interest rates, because it affects the structural relationship betweeen developments in inflation and economic activity and the behaviour of the term structure . Conversely, the behaviour of the yield curve can provide useful information on market perceptions of the monetary policy rule followed by the central bank, insofar as the latter contributes to shape expected future short rates *and* term premia.

Calibrated versions of a new generation of DSGE models have been shown to be capable of producing roughly realistic implications for some unconditional moments of the term structure of interest rates, including slope and volatility, provided they are solved using second-order approximations or higher (see e.g. Hördahl, Tristani and Vestin, 2008; Ravenna and Seppala, 2007a, 2007b; Rudebusch and Swanson, 2009). In this paper, we take these models further and explore their ability to match conditional moments of macroeconomic and term structure data when they are estimated using full information

[1] "The expectations hypothesis, in the broadest terms, asserts that the slope of the term structure has something to do with expectations about future interest rates" (Shiller, 1990, p. 644)

methods.

One problem from this perspective is that models solved to a second-order approximation can only generate constant risk-premia, while the finance literature has highlighted the importance of allowing for time-variation in risk premia to match the conditional features of yields – see e.g. Dai and Singleton (2002). In order to allow for time-variation in risk premia, we assume heteroskedasticity in the model's structural shocks, i.e. time-variation in the "amount of risk" faced by bond-holders at any point in time.[2] We assume that heteroskedasiticy takes the specific form of regime switching. Moreover, the assumption of regime switching has already been shown to help fit yields in the finance literature – see Hamilton (1988), Naik and Lee (1997), Ang and Bekaert (2002a,b), Bansal and Zhou (2002), Bansal, Tauchen and Zhou (2004), Ang, Bekaert and Wei (2008), Dai, Singleton and Yang (2008), Bikbov and Chernov (2008) – and is also increasingly used in macroeconomics following Sims and Zha (2007).

We demonstrate analytically that, when combined with a second order approximation of the solution, this feature leads to changes in risk premia at the time of switches in regimes. More specifically, regime changes generate variations in the prices of risk, which are entirely consistent with the microfoundations of the model. While this mechanism does not explain *why* risk premia vary, it forces their variation to be consistent with changes in the volatility of macro variables.

The second novelty of our model is a generalisation of the preferences proposed by Epstein and Zin (1989) and Weil (1990) to include habit persistence. Epstein-Zin-Weil preferences are quite standard in the finance literature – see e.g. Campbell (1999) – and they have already been successfully used to model yields in a partial equilibrium model by Piazzesi and Schneider (2006) and, more recently, Bansal and Shaliastovich (2008). Gallmeyer *et al.* (2007), Backus, Routledge and Zin (2007) and Rudebusch and Swanson (2009) have used these preferences in calibrated models. Binsbergen *et al.* (2008) is the only other application that we are aware of which estimates a DSGE model with

[2] The finance literature, especially in affine term structure models, emphasises instead time-variations in risk premia due to changed in the "price of risk". Time variations in the price of risk can be produced within general equilibrium models if they are solved up to a third order approximation (or higher). This approach is pursued in Ravenna and Seppala (2007a, b), Rudebusch, Sack and Swanson (2007) and Rudebusch and Swanson (2007, 2008). However, these papers are purely theoretical: the estimation of DSGE models solved using third order approximations appears to be infeasible at this point in time.

Epstein-Zin preferences. However, this paper relies on a benchmark RBC model and is therefore not suitable to analyse the interaction between monetary policy, inflation risk and consumption risk in the determinatino of risk premia.

Our empirical results are based on US data on aggregate consumption, GDP, inflation, the short-term interest rate and yields on 3-year and 10-year yields. The sample period runs from 1966Q1 until 2009Q1.

We find considerable support for a specification with regime switches. The residuals of the model show clear signs of heteroskedasticity, which could not be accounted for in a model with homoskedastic shocks. The model with regime switching can also fit yields reasonably well.

More specifically, we find strong evidence of time variability in expected excess holding period returns. Prema tend to be higher in the eighties, and display fluctuations which can be associated with the economic cycle. Volatility in premia is mainly linked to regime switches in the variance of technology shocks.

Our model is related to a growing literature exploring empirically the term structure implications of new-Keynesian models. The closest papers to ours is Doh (2006), which also estimates a quadratic DSGE model of the term structure of interest rates with heteroskedastic shocks. However, Doh (2006) allows for additional non-structural parameters to model the unconditional slope of the yield curve, while our approach is fully theoretically consistent. Another difference between the two papers is that heteroskedasticity in Doh (2006) is modelled through ARCH shocks, while it is generated by regime switching in our case. Andreasen (2008) shows that the estimation of a richer term structure model, which includes capital accumulation, is feasible to second order. However, the model cannot generate time-variation in risk premia because shocks are homoskedastic. Bekaert, Cho and Moreno (2006) and De Graeve, Emiris and Wouters (2007) estimate the loglinearised reduced form of DSGE models using both macroeconomic and term structure data. As in Doh (2006), these papers do not impose theoretical restrictions on the unconditional slope of the yield curve. In addition, they assume at the outset that risk-premia are constant.

2 The model

We rely on a relatively standard model in the spirit of Woodford (2003). The central feature is the assumption of nominal rigidities.

We only deviate from the standard model in postulating that households' preferences can be described by the non-expected utility specification proposed by Epstein and Zin (1989) and Weil (1990). This specification is quite standard in the consumption-based asset pricing literature and it has already been employed to analyse the term structure of interest rate in a partial equilibrium model by Piazzesi and Schneider (2006). Here we extend this specification to a general equilibrium model in which we also allow for habit persistence in consumption and a labour-leisure choice – see also Backus, Routledge and Zin (2004, 2005). Rudebusch and Swanson (2009) also use non-expected utility preferences in a model similar to ours, but that paper relies on the assumption of homoskedastic shocks.

2.1 Households

We assume that each household i provides $N(i)$ hours of differentiated labor services to firms in exchange for a labour income $w_t(i)\, N_t(i)$. Each household owns an equal share of all firms j and receives profits $\int_0^1 \Pi_t(j)\mathrm{d}j$.

As in Erceg, Henderson and Levin (2000), an employment agency combines households' labor hours in the same proportions as firms would choose. The agency's demand for each household's labour is therefore equal to the sum of firms' demands. The labor index L_t has the Dixit-Stiglitz form $L_t = \left[\int_0^1 N_t(i)^{\frac{\theta_{w,t}-1}{\theta_{w,t}}} \mathrm{d}i\right]^{\frac{\theta_{w,t}}{\theta_{w,t}-1}}$ where $\theta_{w,t} > 1$ is subject to exogenous shocks. At time t, the employment minimizes the cost of producing a given amount of the aggregate labor index, taking each household's wage rate $w_t(i)$ as given and then sells units of the labor index to the production sector at the aggregate wage index $w_t = \left[\int_0^1 w(i)^{1-\theta_{w,t}} \mathrm{d}i\right]^{\frac{1}{1-\theta_{w,t}}}$. The employment agency's demand for the labor hours of household i is given by

$$N_t(i) = L_t \left(\frac{w_t(i)}{w_t}\right)^{-\theta_{w,t}} \tag{1}$$

Each household i maximizes its intertemporal utility with respect to consumption, the wage rate and holdings of contingent claims, subject to its labor demand function (1) and the budget constraint

$$P_t C_t(i) + \mathrm{E}_t Q_{t,t+1} W_{t+1}(i) \leq W_t(i) + w_t(i)\, N_t(i) + \int_0^1 \Xi_t(j)\, \mathrm{d}j \tag{2}$$

where C_t is a consumption index satisfying

$$C_t = \left(\int_0^1 C_t \left(z \right)^{\frac{\theta - 1}{\theta}} \mathrm{d}z \right)^{\frac{\theta}{\theta - 1}} \tag{3}$$

W_t denotes the beginning-of-period value of a complete portfolio of state contingent assets, $Q_{t,t+1}$ is their price, $w_t\left(i\right)$ is the nominal wage rate and $\Xi_t\left(j\right)$ are the profits received from investment in firm j. The price level P_t is defined as the minimal cost of buying one unit of C_t, hence equal to

$$P_t = \left(\int_0^1 p\left(z\right)^{1-\theta} \mathrm{d}z \right)^{\frac{1}{1-\theta}}. \tag{4}$$

Equation (2) states that each household can only consume or hold assets for amounts that must be less than or equal to its salary, the profits received from holding equity in all the existing firms and the revenues from holding a portfolio of state-contingent assets.

Households' preferences are described by the Kreps and Porteus (1978) specification proposed by Epstein and Zin (1989). In that paper, utility is defined recursively through the aggregator U such that

$$U\left[C_t, \left(\mathrm{E}_t V_{t+1}^{1-\gamma}\right)\right] = \left\{ \left(1-\beta\right) C_t^{1-\psi} + \beta \left(\mathrm{E}_t V_{t+1}^{1-\gamma}\right)^{\frac{1-\psi}{1-\gamma}} \right\}^{\frac{1}{1-\psi}}, \qquad \psi, \gamma \neq 1 \tag{5}$$

where β, ψ and γ are positive constants. Using a specification equivalent to that in equation (5), Weil (1990) shows that β is, under certainty, the subjective discount factor, but time preference is in general endogenous under uncertainty. The parameter γ is the relative risk aversion coefficient for timeless gambles. The parameter $1/\psi$ measures the elasticity of intertemporal of substitution for deterministic consumption paths.

The distinguishing feature of the Epstein-Zin-Weil preferences, compared to the standard expected utility specification, is that the coefficient of relative risk aversion can differ from the reciprocal of the intertemporal elasticity of substitution. In addition, Kreps and Porteus (1978) show that, again contrary to the expected utility specification, the timing of uncertainty is relevant in their class of preferences. The specification in equation (5) displays preferences for an early resolution of uncertainty when the aggregator is convex in its second argument, i.e. when $\gamma > \psi$. Any source of risk will be reflected in asset prices not only if it makes consumption more volatile, but also if it affects the temporal distribution of consumption volatility.

We generalise the utility function in equation (5) by allowing for habit formation and a labour-leisure choice. More specifically, time-t utility will not only depend on consumption

C_t but it will more generally be given by

$$u_t = (C_t - hC_{t-1}) \cdot v(N_t)$$

where $v(N_t)$ captures the disutility of labour supply and the h parameter represents the force of habits in the model: the higher h, the less utility is generated by a given amount of current consumption.[3] For $h = 0$, our preferences collapse to a special case of the class of preferences defined in Uhlig (2007). In our numerical analysis, we will adopt for the disutility of labour the formulation suggested by Trabandt and Uhligh (2009). This is given by $v(N_t) = \left(1 - \kappa(1-\psi) N_t^{1+\frac{1}{\phi}}\right)^{\frac{\psi}{1-\psi}}$ and it has the advantage of implying a constant Frisch elasticity equal to ϕ in the $h = 0$ case.

With our more general preferences specification, γ and ψ are no-longer related one-to-one to risk aversion and to the (inverse of the) elasticity of intertemporal substitution of consumption, respectively. Swanson (2009) discusses the appropriate measures of risk aversion in a dynamic setting with consumption and leisure entering the utility function. In the rest of this paper, we simply refer to γ and ψ as utility parameters.

Each households i maximises

$$U[C_t(i), N_t(i), \mathrm{E}_t V_{t+1}] = \left\{(1-\beta)\left[(C_t(i) - hC_{t-1}(i)) \cdot v(N_t(i))\right]^{1-\psi} + \beta\left(\mathrm{E}_t V_{t+1}^{1-\gamma}\right)^{\frac{1-\psi}{1-\gamma}}\right\}^{\frac{1}{1-\psi}}$$

subject to

$$P_t C_t(i) + \mathrm{E}_t Q_{t,t+1} W_{t+1}(i) \leq W_t(i) + w_t(i) N_t(i) + \int_0^1 \Xi_t(j)\,\mathrm{d}j$$

and

$$N_t(i) = L_t \left(\frac{w_t(i)}{w_t}\right)^{-\theta_{w,t}}$$

where the choice variables are $w_t(i)$ and $C_t(i)$.

To Bellman equation for this problem (abstracting from the i subscript to simplify the notation) is

$$\begin{aligned} J(W_t, C_{t-1}) \quad = \quad & \max\left\{(1-\beta)\left[(C_t - h_t C_{t-1}) v(N_t)\right]^{1-\psi} + \beta\left[\mathrm{E}_t J^{1-\gamma}(W_{t+1}, C_t)\right]^{\frac{1-\psi}{1-\gamma}}\right\}^{\frac{1}{1-\psi}} \\ & -\Lambda_t\left[P_t C_t + \mathrm{E}_t Q_{t,t+1} W_{t+1} - W_t - w_t N_t - \int_0^1 \Xi_t(i)\,di + T_t\right] \end{aligned} \tag{6}$$

[3] Guariglia and Rossi (2002) also use expected utility preferences combined with habit formation to study precautionary savings in UK consumption. Koskievic (1999) studies an intertemporal consumption-leisure model with non-expected utility.

The appendix shows that the first order conditions can be written as

$$\widetilde{w}_t = -\mu_{w,t} \frac{v'(N_t)}{v(N_t)} \frac{\left[(C_t - hC_{t-1}) v(N_t)\right]^{1-\psi}}{\widetilde{\Lambda}_t} \tag{7}$$

$$Q_{t,t+1} = \beta \frac{\widetilde{\Lambda}_{t+1}}{\widetilde{\Lambda}_t} \frac{1}{\pi_{t+1}} \left(\frac{\left[\mathrm{E}_t J_{t+1}^{1-\gamma} \right]^{\frac{1}{1-\gamma}}}{J_{t+1}} \right)^{\gamma-\psi} \tag{8}$$

$$\widetilde{\Lambda}_t = (C_t - hC_{t-1})^{-\psi} \left[v(N_t)\right]^{1-\psi} - \beta h \mathrm{E}_t (C_{t+1} - hC_t)^{-\psi} \left[v(N_{t+1})\right]^{1-\psi} \left(\frac{\left[\mathrm{E}_t J_{t+1}^{1-\gamma} \right]^{\frac{1}{1-\gamma}}}{J_{t+1}} \right)^{\gamma-\psi} \tag{9}$$

where $\widetilde{\Lambda}_t \equiv \Lambda_t P_t (1-\beta)^{-1} J_t^{-\psi}$ and $\widetilde{w}_t$ is the real wage $\widetilde{w}_t \equiv w_t / P_t$ and $\mu_{w,t} \equiv \theta_{w,t} / (\theta_{w,t} - 1)$.

The gross interest rate, I_t, equals the conditional expectation of the stochastic discount factor, i.e.

$$I_t^{-1} = \mathrm{E}_t Q_{t,t+1} \tag{10}$$

Note that we will focus on a symmetric equilibrium in which nominal wage rates are all allowed to change optimally at each point in time, so that individual nominal wages will equal the average w_t.

Equations (8)-(9) highlight how our model nests the standard power utility case, in which $\psi = \gamma$ and the maximum value function J_t disappears from the first order conditions. The same equations also demonstrate that the parameter γ only affects the dynamics of higher order approximations. To first order, the term $\left[\mathrm{E}_t J_{t+1}^{1-\gamma} \right]^{\frac{\gamma-\psi}{1-\gamma}} / J_{t+1}^{\gamma-\psi}$ in equations (9) and (10) cancels out in expectation.

2.2 Firms

We assume a continuum of monopolistically competitive firms (indexed on the unit interval by j), each of which produces a differentiated good. Demand arises from households' consumption and from government purchases G_t, which is an aggregate of differentiated goods of the same form as households' consumption. It follows that total demand for the output of firm i takes the form $Y_t(j) = \left(\frac{P_t(i)}{P_t}\right)^{-\theta} Y_t$. Y_t is an index of aggregate demand which satisfies $Y_t = C_t + G_t$.

Firms have the production function

$$Y_t(j) = A_t L_t^{\alpha}(j)$$

where L_t is the labour index L_t defined above.

Once aggregate demand is realised, the firm demands the labour necessary to satisfy it $L_t(j) = (Y_t(j)/A_t)^{\frac{1}{\alpha}}$ so that the total nominal cost function will be given by

$$TC_t(j) = w_t \left(\frac{Y_t(j)}{A_t}\right)^{\frac{1}{\alpha}}$$

As a result, real marginal costs will be

$$mc_t(j) = \frac{1}{\alpha}\frac{\widetilde{w}_t}{A_t}\left(\frac{Y_t(j)}{A_t}\right)^{\frac{1-\alpha}{\alpha}}$$

where nominal costs are deflated using the aggregate price level.

As in Rotemberg (1982), we assume the firms face quadratic costs in adjusting their prices. This assumption is also adopted, for example, by Schmitt-Grohé and Uribe (2004) and Ireland (1997). It is well-known to yield first-order inflation dynamics equivalent to those arising from the assumption of Calvo pricing.[4] From our viewpoint, it has the advantage of greater computational simplicity, as it allows us to avoid having to include an additional state variable in the model, i.e. the cross-sectional dispersion of prices across firms.

The specific assumption we adopt is that firm j faces a quadratic cost when changing its prices in period t, compared to period $t-1$. Consistently with what is typically done in the Calvo literature, we modify the original Rotemberg (1982) formulation for partial indexation of prices to lagged inflation. More specifically, we assume that

$$\frac{\zeta}{2}\left(\frac{P_t^j}{P_{t-1}^j} - (\Pi_t^*)^{1-\iota}\,\Pi_{t-1}^{\iota}\right)^2 Y_t$$

where Π_t^* is the inflation objective.

Firms maximise their real profits

$$\max_{P_t^j} \mathrm{E}_t \sum_{s=t}^{\infty} Q_{t,s}\left[\frac{P_s^j Y_s^j\left(P_s^j\right)}{P_s} - \frac{TC_s\left(Y_s^j\left(P_s^j\right)\right)}{P_s} - \frac{\zeta}{2}\left(\frac{P_s^j}{P_{s-1}^j} - (\Pi_s^*)^{1-\iota}\,\Pi_{s-1}^{\iota}\right)^2 Y_s\right]$$

subject to

$$Y_t(j) = \left(\frac{P_t(j)}{P_t}\right)^{-\theta} Y_t$$

and to

$$Y_t(j) = A_t L_t^{\alpha}(j)$$

[4] The two pricing models, however, have in general different welfare implications – see Lombardo and Vestin (2008).

Focusing on a symmetric equilibrium in which all firms adjust their price at the same time, the first order condition for price setting can be written as

$$(\theta-1)\,Y_t+\zeta\left(\Pi_t-\left(\Pi_t^*\right)^{1-\iota}\Pi_{t-1}^{\iota}\right)Y_t\Pi_t=\frac{\theta}{\alpha}\widetilde{w}_t\left(\frac{Y_t}{A_t}\right)^{\frac{1}{\alpha}}+\mathrm{E}_tQ_{t,t+1}\zeta\left(\Pi_{t+1}-\left(\Pi_{t+1}^*\right)^{1-\iota}\Pi_t^{\iota}\right)Y_{t+1}\Pi_{t+1}$$

2.3 Monetary policy

We close the model with the simple Taylor-type policy rule

$$I_t=\left(\frac{\Pi^*}{\beta}\right)^{1-\rho_I}\left(\frac{\Pi_t}{\Pi^*}\right)^{\psi_\Pi}\left(\frac{Y_t}{Y}\right)^{\psi_Y}I_{t-1}^{\rho_I}e^{\eta_{t+1}} \tag{11}$$

where Y_t is aggregate output, Π^* is the inflation target and η_{t+1} is a policy shock.

2.4 Market clearing

Market clearing in the goods market requires

$$Y_t=C_t+G_t$$

In the labour market, labour demand will have to equal labour supply. In addition, the total demand for hours worked in the economy must equal the sum of the hours worked by all individuals. Taking into account that at any point in time the nominal wage rate is identical across all labor markets because all wages are allowed to change optimally, individual wages will equal the average w_t. As a result, all households will chose to supply the same amount of labour and labour market equilibrium will require that

$$L_t=\left(\frac{Y_t}{A_t}\right)^{\frac{1}{\alpha}}$$

2.5 Exogenous shocks

In macroeconomic applications, exogenous shocks are almost always assumed to be (log-)normal, partly because models are typically log-linearised and researchers are mainly interested in characterising conditional means. However, Hamilton (2008) argues that a correct modelling of conditional variances is always necessary, for example because inference on conditional means can be inappropriately influenced by outliers and high-variance episodes. The need for an appropriate treatment of heteroskedasticity becomes even more compelling when models are solved nonlinearly, because conditional variances have a direct impact on conditional means.

In this paper, we assume that variances are subject to stochastic regime switches for shocks other than the inflation target. More specifically, we assume a deterministic trend in technology growth

$$\begin{aligned} A_t &= Z_t B_t \\ B_t &= B_{t-1}\overline{\Xi} \\ Z_t &= Z_{t-1}^{\rho_z} e^{\varepsilon_t^z}, \qquad \varepsilon_{t+1}^z \approx N\left(0, \sigma_{z,s_{z,t}}\right) \end{aligned}$$

where $\overline{\Xi}$ is the long run productivity growth rate. We specify the exogenous government spending process in deviation from trend, so that

$$\frac{G_t}{B_t} = \left(\frac{gY}{B}\right)^{1-\rho_g} \left(\frac{G_{t-1}}{B_{t-1}}\right)^{\rho_g} e^{\varepsilon_t^g} \qquad \varepsilon_{t+1}^G \approx N\left(0, \sigma_{G,s_{G,t}}\right)$$

where the long run level g is specified in percent of output, so that $g \equiv G/Y$. Finally, for monetary policy and mark-up shocks we assume

$$\begin{aligned} \eta_{t+1} &= e^{\varepsilon_{t+1}^\eta}, \qquad \varepsilon_{t+1}^\eta \tilde{\ } N\left(0, \sigma_{\eta,s_{\eta,t}}\right) \\ \mu_{w,t+1} &= \mu_w^{1-\rho_\Pi} \left(\mu_{w,t}\right)^{\rho\mu} e^{\varepsilon_{t+1}^\mu}, \qquad \varepsilon_{t+1}^\mu \approx N\left(0, \sigma_\mu\right) \end{aligned}$$

Technology and monetary policy shocks have regime-switching variances, namely

$$\begin{aligned} \sigma_{z,s_{z,t}} &= \sigma_{z,L} s_{z,t} + \sigma_{z,H}\left(1 - s_{z,t}\right) \\ \sigma_{\eta,s_{\eta,t}} &= \sigma_{\eta,L} s_{\eta,t} + \sigma_{\eta,H}\left(1 - s_{\eta,t}\right) \end{aligned}$$

and the variables $s_{z,t}$ and $s_{\eta,t}$ can assume the discrete values 0 and 1. For each variable $s_{j,t}$ ($j = z, \eta$), the probabilities of remaining in states 0 and 1 are constant and equal to $p_{j,0}$ and $p_{j,1}$, respectively.[5]

We assume regime switches in these particular variances for the following reasons. The literature on the "Great moderation" (see e.g. McDonnell and Perez-Quiros, 2000) has emphasised the reduction in the volatility of real aggregate variables starting in the second half of the 1980s. We conjecture that this phenomenon could be captured by a reduction in the volatility of technology shocks in our structural setting. The heteroskedasticity in policy shocks aims to capture the large increase in interest rate volatility in the early 1980s, the time of the so-called "monetarist experiment" of the Federal Reserve.[6]

[5] In previous versions of the paper we have allowed for regime-switching also in the variance of government spending and mark-up shocks. These additional dimensions of regime switching receive little support from the data.

[6] A similar assumption in made in Schorfeide (2005).

2.6 Solution method

To solve the model, we exploit the recursive nature of bonds in equilibrium. We first solve for all macroeconomic variables and then construct the prices of bonds of various maturities.

2.6.1 Solving the macroeconomic system

We approximate the system around a deterministic steady state in which all variables are detrended. Detrend variables are denoted by a tilde and defined as a ratio to B_t (with the exception is $\widetilde{\Lambda}_t$, which is detrended by $B_t^{-\psi}$). For example, detrended output is $\widetilde{Y}_t \equiv Y_t/B_t$. In the solution, we expand variables around their natural logarithms, which are denoted by lower-case letters. The logarithm of a variable in deviation from its non-stochastic steady state is denoted by a hat. For example, approximate (detrended) output is denoted by $\widehat{\widetilde{y}}_t$.

For the solution, we collect all predetermined variables (including both lagged endogenous predetermined variables and exogenous states) in a vector $\mathbf{x}_t$ and all the non-predetermined variables in a vector $\mathbf{y}_t$ (note that $\mathbf{y}_t$ is different from output y_t).

The macroeconomic system can thus be written in compact form as

$$\mathbf{y}_t = g\left(\mathbf{x}_t, \widetilde{\sigma}, \mathbf{s}_t\right) \tag{12}$$

$$\mathbf{x}_{t+1} = h\left(\mathbf{x}_t, \widetilde{\sigma}, \mathbf{s}_t\right) + \widetilde{\sigma}\Sigma\left(\mathbf{s}_t\right)\mathbf{u}_{t+1} \tag{13}$$

$$\mathbf{s}_{t+1} = \boldsymbol{\kappa}_0 + \boldsymbol{\kappa}_1\mathbf{s}_t + \boldsymbol{\nu}_{t+1} \tag{14}$$

for matrix functions $g(\cdot)$, $h(\cdot)$, and $\Sigma(\cdot)$, a vector $\mathbf{s}_t$ including the state variables that index the discrete regimes, and a vector of innovations $\mathbf{u}_t$. In the above system, $\widetilde{\sigma}$ is a perturbation parameter. Following Hamilton (1994), we can write the law of motion of the discrete processes $\mathbf{s}_t$ in the form implied in equation (14) for a vector $\boldsymbol{\kappa}_0$ and a matrix $\boldsymbol{\kappa}_1$. The law of motion of state $s_{z,t}$, for example, is written as $s_{z,t+1} = (1-p_{z,0}) + (-1+p_{z,1}+p_{z,0})\, s_{z,t} + \nu_{z,t+1}$, where $\nu_{z,t+1}$ is an innovation with mean zero and heteroskedastic variance.

We seek a second-order approximation to the functions $g\left(\mathbf{x}_t, \widetilde{\sigma}, \mathbf{s}_t\right)$ and $h\left(\mathbf{x}_t, \widetilde{\sigma}, \mathbf{s}_t\right)$ around the non-stochastic steady state, namely the point where $\mathbf{x}_t = \overline{\mathbf{x}}$ and $\widetilde{\sigma} = 0$. Due to the presence of the discrete regimes in the system, both the steady state and the coefficients of the second order approximation could potentially depend on $\mathbf{s}_t$. Since the discrete states only affect the variance of the shocks, however, they disappear when $\widetilde{\sigma} = 0$ so that the non-stochastic steady state is not regime-dependent. In a companion paper (Amisano and Tristani, 2009b), we show that the second order approximation to the solution can be written as

$$g(\mathbf{x}_t, \widetilde{\sigma}, \mathbf{s}_t) = F\widehat{\mathbf{x}}_t + \frac{1}{2}\left(I_{n_y} \otimes \widehat{\mathbf{x}}_t'\right) E\widehat{\mathbf{x}}_t + k_{y,s_t}\widetilde{\sigma}^2$$

and

$$h(\mathbf{x}_t, \widetilde{\sigma}, \mathbf{s}_t) = P\widehat{\mathbf{x}}_t + \frac{1}{2}\left(I_{n_x} \otimes \widehat{\mathbf{x}}_t'\right) G\widehat{\mathbf{x}}_t + k_{x,s_t}\widetilde{\sigma}^2$$

where F, E, P and G are constant variances and only the vectors k_{y,s_t} and k_{x,s_t} are regime dependent. As a result, regime switching plays no role to a first order approximation. It only affects the means of endogenous variables.

2.6.2 Pricing bonds

Once the solution of the macroeconomic model is available, bond yields can be solved for analytically.

Note that the stochastic discount factor can be rewritten in terms of detrended variables as

$$Q_{t,t+1} = \beta \frac{\widetilde{\widetilde{\Lambda}}_{t+1}}{\widetilde{\widetilde{\Lambda}}_t} \frac{1}{\Pi_{t+1}} \left(\frac{\left[\mathrm{E}_t \widetilde{J}_{t+1}^{1-\gamma}\right]^{\frac{1}{1-\gamma}}}{\widetilde{J}_{t+1}} \right)^{\gamma-\psi}$$

This expression can be written more simply as

$$Q_{t,t+1} = \beta \frac{\overleftrightarrow{\Lambda}_{t+1}}{\overleftrightarrow{\Lambda}_t} \frac{1}{\overleftrightarrow{\Pi}_{t+1}} \tag{15}$$

for

$$\begin{aligned} \overleftrightarrow{\Pi}_t &\equiv \Pi_t \widetilde{J}_t^{\gamma-\psi} \left[\mathrm{E}_t \widetilde{J}_{t+1}^{1-\gamma}\right]^{\frac{\psi-\gamma}{1-\gamma}} \\ \overleftrightarrow{\Lambda}_t &\equiv \widetilde{\widetilde{\Lambda}}_t \left[\mathrm{E}_t \widetilde{J}_{t+1}^{1-\gamma}\right]^{\frac{\psi-\gamma}{1-\gamma}} \end{aligned}$$

Since these relationships are all loglinear, the law of motion for $\overleftrightarrow{\Pi}_t$ and $\overleftrightarrow{\Lambda}_t$ can immediately be derived from those of Π_t, $\widetilde{\Lambda}_t$, $\widetilde{J}_t$ and $D_t \equiv \mathrm{E}_t J_{t+1}^{1-\gamma}$. It follows that

$$\begin{aligned} \widehat{\overleftrightarrow{\lambda}}_t &= F_\lambda \widehat{\mathbf{x}}_t + \frac{1}{2}\widehat{\mathbf{x}}_t' E_\lambda \widehat{\mathbf{x}}_t + k_{\lambda,s_t}\widetilde{\sigma}^2 \\ \widehat{\overleftrightarrow{\pi}}_t &= F_\pi \widehat{\mathbf{x}}_t + \frac{1}{2}\widehat{\mathbf{x}}_t' E_\pi \widehat{\mathbf{x}}_t + k_{\pi,s_t}\widetilde{\sigma}^2 \end{aligned}$$

where F_λ and F_π are row vectors, and E_λ and E_π are matrices. We can now compute bond prices using the method in Hördahl, Tristani and Vestin (2008). The appendix shows that, in log-deviation from its deterministic steady state, the approximate price of a bond of maturity n, $\widehat{b}_{t,n}$, can be written as

$$\widehat{b}_{t,n} = F_{B_n}\widehat{\mathbf{x}}_t + \frac{1}{2}\widehat{\mathbf{x}}_t' E_{B_n}\widehat{\mathbf{x}}_t + k_{B_n,s_t}\widetilde{\sigma}^2 \tag{16}$$

where F_{B_n}, E_{B_n} and k_{B_n,s_t} are defined through a recursion. Note that k_{B_n,s_t} changes depending on the realisation of the discrete states, but matrices F_{B_n} and E_{B_n} are state-independent.

3 Some properties of the model

3.1 The stochastic discount factor

The appendix shows that, to a second order approximation, the stochastic discount factor in equation (8) can be written as

$$\begin{aligned} \widehat{q}_{t,t+1} &= \Delta\widehat{\widehat{\lambda}}_{t+1} - \widehat{\pi}_{t+1} \\ &\quad - (\gamma-\psi)\left(\widehat{j}_{t+1} - \mathrm{E}_t\widehat{\widehat{j}}_{t+1}\right) - \frac{1}{2}(\gamma-\psi)(\gamma-1)\,\mathrm{Var}_t\left[\widehat{\widehat{j}}_{t+1}\right] \end{aligned} \tag{17}$$

When: (a) $\psi = 1$; (b) temporary utility depends on consumption only and $\Delta\widehat{\widehat{\lambda}}_{t+1} = -\Delta\widehat{c}_{t+1}$, equation (17) boils down to

$$\widehat{q}_{t,t+1} = -\Delta\widehat{c}_{t+1} - \widehat{\pi}_{t+1} - (\gamma-1)\left(\widehat{j}_{t+1} - \mathrm{E}_t\left[\widehat{j}_{t+1}\right]\right) - \frac{1}{2}(\gamma-1)^2\,\mathrm{var}_t\left[\widehat{j}_{t+1}\right]$$

which corresponds to the case considered by Piazzesi and Schneider (2006).

In the expected utility case, only the risk of unpredictable changes in future consumer prices, in future detrended marginal utility, or in technology growth matter for the investor.

With Epstein-Zin preferences, the whole temporal distribution of future risks to detrended marginal utility and technology growth becomes relevant. In our case, similarly to Uhlig (2007), detrended marginal utility is affected both by future detrended consumption growth and by future levels of labour supply.

Following Restoy and Weil (2010) and Piazzesi and Schneider (2006), we can solve out the value function as an infinite sum of future expected utility. The appendix derives this representation. In general, the value function will be affected by expected future productivity growth and the discounted future growth rates of consumption and of labour supply. The expected future change in consumption growth also matters because of habit formation.

To have some intuition for the implications of the model in terms of bond pricing, we can derive the short-term interest rate as $\widehat{i}_t = -\mathrm{E}_t \widehat{q}_{t,t+1} - \frac{1}{2}\mathrm{Var}_t \widehat{q}_{t,t+1}$. We obtain

$$\widehat{i}_t = \widehat{r}_t + \mathrm{E}_t \widehat{\pi}_{t+1} - \frac{1}{2}\mathrm{Var}_t \widehat{\pi}_{t+1} + \mathrm{Cov}_t\left[\widehat{\mu}_{t+1}, \widehat{\pi}_{t+1}\right] - (\gamma - \psi)\,\mathrm{Cov}_t\left[\widehat{\pi}_{t+1}, \Psi_{t+1}\right] \tag{18}$$

where $\widehat{\mu}_{t+1} \equiv \Delta\widehat{\widehat{\lambda}}_t$ is the change in the marginal utility of consumption,

$$\Psi_{t+1} \equiv \mathrm{E}_{t+1} \sum_{i=0}^{\infty} \left(\beta \Xi^{1-\psi}\right)^i \Delta \overleftrightarrow{u}_{t+1+i}$$

denotes expected discounted future changes in utility – i.e. the sum of expected discounted future consumption growth and expected future changes in the disutility of labour – and $\widehat{r}_t$ is the real interest rate

$$\widehat{r}_t = -\mathrm{E}_t \widehat{\mu}_{t+1} - \frac{1}{2}\mathrm{Var}_t \widehat{\mu}_{t+1} + (\gamma - \psi)\,\mathrm{Cov}_t\left[\widehat{\mu}_{t+1}, \Psi_{t+1}\right] + \frac{1}{2}(\psi - 1)(\gamma - \psi)\,\mathrm{Var}_t \Psi_{t+1} \tag{19}$$

According to expression (18), the nominal interest rate includes three components in excess of the real rate and expected inflation. The first component is the variance of inflation, which is a Jensen's inequality term. The second component, the covariance between inflation and the marginal utility of consumption at date $t+1$, is a standard inflation risk premium term. The return on a nominal 1-period bond will be higher if inflation at $t+1$ tends to be high when the marginal utility at $t+1$ is also high. Note that this term will be larger under habit persistence, which tends to boost the sensitivity of marginal utility to changes in consumption. The last term in equation (18), which is proportional to the covariance between inflation at $t+1$ and expected future changes in utility, is also an inflation risk premium and it can be understood along the lines suggested by Restoy and Weil (2010). When the covariance is negative, inflation at $t+1$ tends to be high, and the real return on nominal bonds tends to be low, in case of bad news about expected future utility. Nominal bonds are therefore risky, and require a premium given by the term $\gamma\mathrm{Cov}_t\left[\widehat{\mu}_{t+1}, \Psi_{t+1}\right]$. At the same time, the bad news about the future induces households to increase their savings, so as to smooth consumption over time. From this viewpoint, nominal bonds are desirable assets and command a discount $-\psi\mathrm{Cov}_t\left[\widehat{\mu}_{t+1}, \Psi_{t+1}\right]$. With expected utility preferences, these effects cancel out and news about future utility growth become irrelevant for asset pricing.

The real interest rate in equation (19) also includes three additional components compared to the linearised case. The first one, the variance of the marginal utility of consumption, is a standard precautionary savings motive. The larger the volatility of future marginal utility growth, the higher the incentive to hold precautionary savings and the lower the return on real bonds. Epstein-Zin-Weil utility with $\psi = 1$ – the case analysed in Piazzesi and Schneider (2006) – introduces a second source of risk premia, namely the covariance term in equation (19). The real interest rate includes a premium when bad news about future utility growth over the indefinite future tend to be associated with high marginal utility at $t+1$. The premium is higher if agents are unwilling to adjust their level of utility across states (γ is high) and lower the more averse they are to adjust utility across time. Finally, a third source of risk premium arises under Epstein-Zin-Weil utility in the $\psi \neq 1$ case, namely a premium related to the variance in revisions of expected future utility.

3.2 Regime switching and the variability of risk premia

The state-dependence of $\widehat{b}_{t,n}$ in equation (16) implies that bond risk premia will also become state-dependent. In order to show this, it is useful to derive expected excess holding period returns, i.e. the expected return from holding a n-period bond for 1 period in excess of the return on a 1-period bond. To a second order approximation, the expected excess holding period return on an n-period bond can be written as

$$\widehat{hpr}_{t,n} - \widehat{i}_t = \mathrm{Cov}_t\left[\widehat{\overleftrightarrow{\pi}}_{t+1}, \widehat{b}_{t+1,n-1}\right] - \mathrm{Cov}_t\left[\Delta\widehat{\overleftrightarrow{\lambda}}_{t+1}, \widehat{b}_{t+1,n-1}\right]$$

This expression can be evaluated using the model solution to obtain

$$\widehat{hpr}_{t,n} - \widehat{i}_t = \widetilde{\sigma}^2 F_{B_{n-1}} \Sigma_t \Sigma_t' \left(F_\pi' - F_\lambda'\right) \tag{20}$$

where $\Sigma_t \equiv \Sigma(\mathbf{s}_t)$ and $\Sigma_t \Sigma_t'$ is the conditional variance-covariance matrix depending on vector $\mathbf{s}_t$.

Equation (20) demonstrates that excess holding period returns change when there is a switch in any of the discrete state variables. Since the conditional variance of the price of a bond of maturity n can be written, to a second order approximation, as $\mathrm{E}_t\left[\widehat{b}_{t+1,n-1}\widehat{b}_{t+1,n-1}'\right] = \widetilde{\sigma}^2 F_{B_{n-1}} \Sigma_t \Sigma_t' F_{B_{n-1}}'$, it follows that we can define the (micro-founded) price of risk for unit of volatility, or the "market prices of risk" ω_t, as

$$\omega_t \equiv \widetilde{\sigma} \Sigma_t' \left(F_\pi' - F_\lambda'\right) \tag{21}$$

Since F_π and F_λ are vectors of constants, all terms in equation (21) would be constant in a world with heteroskedastic shocks, in which Σ_t would also be constant. They becomes time-varying in our model due to the possibility of regime switches, because the variance-covariance matrix $\Sigma_t \Sigma_t'$ is regime-dependent.

In the empirical finance literature, the market prices of risk are often postulated exogenously using slightly different specifications. For example, Naik and Lee (1997), Bansal and Zhou (2002) and Ang, Bekaert and Wei (2008) assume that the market prices of risk are regime dependent, but the risk of a regime-change is not priced. On the contrary, regime-switching risk is priced in Dai, Singleton and Yang (2008).

In our model, prices of risk are only associated with variables with continuous support. These prices change across regimes. If, for example, technological risk were not diversifiable, then the price of risk associated with technology shocks would be higher in a high-variance regime for technology shocks (and lower in a low-variance regime). This is the regime-dependence of market prices of risk which is present in all the aforementioned finance models. In our set-up, however, the prices of risk are additionally derived from the model's microfoundation, rather than allowed to vary as affine functions of the continuous state variables of the model.

The risk of regime-switches is not priced because the possibility of changes in regime does not have any impact on the F_π and F_λ vectors. Regime switching risk would only be priced if it affected some structural parameters. For example, one could thing of allowing for changes in the parameters of the (11) as in Bikbov and Chernov (2008). We leave this extension to future research.

4 Empirical results

4.1 Estimation methodology

The system of equations (12) and (13) can be re-written as

$$\mathbf{y}_{t+1}^{o} = \mathbf{c}_j + \mathbf{C}_1 \mathbf{x}_{t+1} + \mathbf{C}_2 vech(\mathbf{x}_{t+1}\mathbf{x}_{t+1}') + \mathbf{D}\mathbf{v}_{t+1} \quad (22)$$

$$\mathbf{x}_{t+1} = \mathbf{a}_i + \mathbf{A}_1 \mathbf{x}_t + \mathbf{A}_2 vech(\mathbf{x}_t \mathbf{x}_t') + \mathbf{B}_i \mathbf{w}_{t+1} \quad (23)$$

$$\mathbf{s}_t \backsim \text{Markov switching} \quad (24)$$

where the vector $\mathbf{y}_t^o$ includes all observable variables, and $\mathbf{v}_{t+1}$ and $\mathbf{w}_{t+1}$ are measurement and structural shocks, respectively. In this representation, the regime switching variables affect the system by changing the intercepts $\mathbf{a}_i$ and $\mathbf{c}_j$, and the loadings of the structural innovations $\mathbf{B}_i$(we indicate here with i the value of the discrete state variables at t and with j the value of the discrete state variables at $t+1$).

If a linear approximation were used, we would then have a linear state space model with Markov switching (see Kim, 1994, Kim and Nelson, 1999, and Schorfheide, 2005).

Focusing on the case in which the number of continuous shocks (measurement and structural) is equal to the number of observables, and there are no unobserved predetermined variables, the continuous latent variables could be obtained via inversion of the observation equation (22). The system could then be written as a Markov Switching VAR in the observable variables and the likelihood could be obtained using the Kitagawa-Hamilton filter i.e. by integrating out the discrete latent variables.

In the quadratic case, however, the likelihood cannot in general be obtained in closed form. One possible approach to compute the likelihood is to rely on Sequential Monte Carlo techniques (henceforth SMC, see Amisano and Tristani, 2010a, for an application of these techniques in a DSGE setting with homoskedastic shocks). The convergence of these methods, however, can be very slow in a case, such as the one of our model, in which both nonlinearities and non-Gaussianity of the shocks characterise the economy.

In the special case mentioned above, i.e. when when the number of continuous shocks is equal to the number of observables and there are no unobserved predetermined variables, we can use exactly the same approach used in the linear case, i.e. doing filtering with respect to the continuous latent variables by inversion, and filtering out the discrete variables by integrating over their discrete domain. See Amisano and Tristani (2010b), for further details. The problem of filtering the continuous latent variables through an inversion of the quadratic observation equation (22) is that the inversion is not unique. At each point in time, multiple values of the latent variables are consistent with the observation vector $\mathbf{y}_t^o$. In a scalar case and in the absence of measurement errors, for example, we would obtain the two solutions $x_t^{(1),(2)} = 1/2\left(-C_1 \pm \sqrt{C_1^2 - 4C_2\left(c_j - y_t^o\right)}\right)/C_2$. In the more general case of our model, we are going to use six series in the estimation process. As a result, at each point in time we will have up to 16 solutions for each of the latent variables.[7]

[7] Given that four structural shocks enter the model, we are going to assume that two series are observed with measurement error. The inversion for the measurement error is unique, because the measurement error enters the model linearly.

Rather than choosing arbitrarily a particular solution at each point in time, we compute the likelihood taking all (real) solutions into account. We simply exploit the property that, while equally likely based on the sole observation vector $\mathbf{y}_t^o$, the different solutions for $\mathbf{x}_t$ have different probability (or likelihood) conditional on $\mathbf{x}_{t-1}$. In a homoskedastic model, the filtered values of our latent variables in t would simply be a weighted average of all $\mathbf{x}_t$ solutions, with weights given by their conditional probabilities. In our model with regime switching, solutions for $\mathbf{x}_t$ must be found for each of the regimes in $\mathbf{s}_t$, and then weighed by the probability of each regime.

Provided *all* solutions of the observation equation are found, this procedure produces the exact likelihood of the quadratic system (22)-(23). To find all solutions, we rely on homotopy continuation methods – see e.g. Judd (1998) and Morgan (1987). More specifically, we rely on the PHCpack solver described in Verschelde (1999) and its Matlab interface PHClab presented in Guan and Verschelde (2008). We discard all complex solutions (which tend to be the majority) and compute the likelihood using the real ones. If no real solutions are available at any point in time for a certain value of the parameter vector, we impute to the likelihood a large negative value.

In the results presented in this paper, we use another viable route to estimation, which can be applied to any context, irrespective of the number of shocks and of the presence or absence of unobservable predetermined variables. Our proposed estimation routine works as follows.

Using the extended Kalman filter framework, at each point in time we linearise state and measurement equations around the conditional mean of the continuous state variables

$$
\begin{aligned}
\mathbf{y}^o_{t+1} &= \widetilde{\mathbf{c}}_{ij,t+1} + \widetilde{\mathbf{C}}_{1,t+1}\mathbf{x}_{t+1} + \mathbf{D}\mathbf{v}_{t+1} && (25)\\
\mathbf{x}_{t+1} &= \widetilde{\mathbf{a}}_{it} + \widetilde{\mathbf{A}}_{1,t}\mathbf{x}_t + \mathbf{B}_i\mathbf{w}_{t+1} && (26)\\
\widetilde{\mathbf{c}}_{jt+1} &= \mathbf{c}_j + \mathbf{C}_2\left[vech(\mathbf{x}^{(i)}_{t+1|t}\mathbf{x}^{(i)\prime}_{t+1|t}) - \boldsymbol{\Delta}_{i,t+1}\mathbf{x}^{(i)}_{t+1|t}\right]\\
\widetilde{\mathbf{C}}_{1,t+1} &= \mathbf{C}_1 + \mathbf{C}_2\boldsymbol{\Delta}_{i,t+1}, \mathbf{x}^{(i)}_{t+1|t} = E(\mathbf{x}^{(i)}_{t+1}|\underline{\mathbf{y}}^o_t, s_t = i, \boldsymbol{\theta})\\
\boldsymbol{\Delta}_{i,t+1} &= \left[\frac{\partial vech\left(\mathbf{x}_{t+1}\mathbf{x}'_{t+1}\right)}{\partial \mathbf{x}_{t+1}}\right]_{\mathbf{x}_{t+1}=\mathbf{x}^{(i)}_{t+1|t}}\\
\widetilde{\mathbf{a}}_{it} &= \mathbf{a}_i + \mathbf{A}_2\left[vech(\mathbf{x}^{(i)}_{t|t}\mathbf{x}^{(i)\prime}_{t|t}) - \boldsymbol{\Delta}_{i,t}\mathbf{x}^{(i)}_{t|t}\right]\\
\widetilde{\mathbf{A}}_{1,t} &= \mathbf{A}_1 + \mathbf{A}_2\boldsymbol{\Delta}_{i,t}, \mathbf{x}^{(i)}_{t|t} = E(\mathbf{x}^{(i)}_t|\underline{\mathbf{y}}^o_t, s_t = i, \boldsymbol{\theta})\\
\boldsymbol{\Delta}_{i,t} &= \left[\frac{\partial vech\left(\mathbf{x}_t\mathbf{x}'_t\right)}{\partial \mathbf{x}_t}\right]_{\mathbf{x}_t=\mathbf{x}^{(i)}_{t|t}}
\end{aligned}
$$

Using this locally linearised state space representation, it is possible to use the algorithm by Kim (1994) and fully detailed in Kim and Nelson (1999, page 105) to compute an approximated likelihood function which we call $\widetilde{p}\left(\boldsymbol{\theta}|\underline{\mathbf{y}}^o_T\right)$. The likelihood is then combined with the prior and sampled using a tuned Metropolis-Hastings algorithm.

It is important to keep in mind that this approach introduces two entwined sources of approximation error in the computation of the likelihood; first, the state space is locally linearised at each point in time; secondly Kim (1994)'s algorithm relies on the use of a mixture of a finite number of components to approximate conditional distributions of continuous state variables. We believe that these two sources of approximation have negligible effect on our results for two reasons. First of all, the quadratic terms $\mathbf{A}_2$ and $\mathbf{C}_2$ are quite small for reasonable values of the parameters, while intercept terms of the state and measurement equations are relevantly different across the domain of the discrete states. This is going to make the approximation error deriving from the linearisation (25) and (26) very small.

In addition, it is possible to fully correct for the approximation error by aptly using importance sampling. Let us in fact define a MCMC sample of parameters drawn from their approximated joint posterior distribution as

$$\boldsymbol{\theta}^{(m)}, i = 1, 2, ...M \tag{27}$$

each of these points is associated with an approximate posterior density which is proportional to

$$\widetilde{p}\left(\boldsymbol{\theta}^{(m)}|\underline{\mathbf{y}}_T^o\right) \propto p(\boldsymbol{\theta}^{(m)}) \times \widetilde{p}\left(\underline{\mathbf{y}}_T^o|\boldsymbol{\theta}^{(m)}\right)$$

where $\widetilde{p}$. Typically M, the number of recorded draws will be much smaller than those used to run the MCMC chain, since the correlation of the resulting draws is quite high and "thinning "(i.e. recording only for instance every 10th draw) is often used.

Now let us now call the likelihood computation as obtained by using an efficient SMC procedure as $p\left(\boldsymbol{\theta}^{(m)}|\underline{\mathbf{y}}_T^o\right)$.

It is evident that we can correct for the approximation, induced by using $\widetilde{p}\left(\underline{\mathbf{y}}_T^o|\boldsymbol{\theta}^{(m)}\right)$ instead of $p\left(\boldsymbol{\theta}^{(m)}|\underline{\mathbf{y}}_T^o\right)$, by computing importance weights as

$$w\left(\boldsymbol{\theta}^{(m)}\right) \propto \frac{p(\boldsymbol{\theta}^{(m)}) \times p\left(\underline{\mathbf{y}}_T^o|\boldsymbol{\theta}^{(m)}\right)}{p(\boldsymbol{\theta}^{(m)}) \times \widetilde{p}\left(\underline{\mathbf{y}}_T^o|\boldsymbol{\theta}^{(m)}\right)} = \frac{p\left(\underline{\mathbf{y}}_T^o|\boldsymbol{\theta}^{(m)}\right)}{\widetilde{p}\left(\underline{\mathbf{y}}_T^o|\boldsymbol{\theta}^{(m)}\right)} \tag{28}$$

and reweighting the MCMC output accordingly.

We do not have yet performed this reweighting on our results, which are reported as they are generated using the approximate likelihood $\widetilde{p}\left(\underline{\mathbf{y}}_T^o|\boldsymbol{\theta}^{(m)}\right)$. Preliminary experiments allow us to cautiously anticipate that the reweighting should not have relevant consequences on our results.

Note that this approach based on the extended Kalman Filter linearisation is computationally much faster and more reliable than using sequential Monte Carlo. In order to use SMC effectively to explore the posterior distribution, these algorithms should be used with a very high number of particles and for all the draws of the posterior distribution, even those far on the tails of the posterior distribution where the reliability of likelihood computations based on sequential Monte Carlo is very low. With our algorithm we instead use sequential Monte Carlo only for the subset of accepted draws remaining after "thinning " the chain.

4.2 Data and prior distributions

We estimate the model on quarterly US data over the sample period from 1966Q1 to 2009Q1. Our estimation sample starts in 1966, because this is often argued to be the date after which a Taylor rule provides a reasonable characterisation of Federal Reserve policy.[8]

[8] According to Fuhrer (1996), "since 1966, understanding the behaviour of the short rate has been equivalent to understanding the behaviour of the Fed, which has since that time essentially set the federal

Concerning the macro data, we use measures of consumption, output and inflation. We use both GDP per capita and consumption to impose some discipline on our estimates of the government spending shock. Given that we abstract from investment, consumption in our model captures all interest-sensitive components of private expenditure. As argued by Giannoni and Woodford (2005), assuming habit persistence for the whole level of private expenditure is a reasonable assumption, given that models with capital typically need adjustment costs that imply inertia in the rate of investment spending. We therefore use total real personal consumption per-capita in the information set. Finally, inflation is measured by the consumption deflator (all macro variables are from the FRED database of the St. Louis Fed).

In addition, we use the 3-month nominal interest rate and yields on 3-year and 10-year zero-coupon bonds (from the Federal Reserve Board).

Prior and posterior distributions for our model are presented in Table 1.

Concerning regime switching processes, we assume beta priors for transition probabilities. The distributions imply that persistences in each state are symmetric and have high means. In the prior, we assume that the standard deviations of the structural shocks are identical in the various states.

For the price adjustment cost, inflation indexation and the utility parameters we use priors broadly in line with other macro studies. The main exception is γ, where we are much more agnostic and use a flat prior between 1 and 21. For the policy rule, we use relatively loose priors centred around parameter values estimated from quarterly data over a pre-sample period running from 1953 to 1965.

The priors for the standard deviation and persistence of shocks, as well as for the long run growth rate of technology and for the long run inflation target, are centred on values which allow us to roughly match unconditional data moments in the pre-sample period, given the other parameter.

Funds rate at a target level, in response to movements in inflation and real activity". Goodfriend (1991) argues that even under the period of official reserves targeting, the Federal Reserve had in mind an implicit target for the Funds rate.

4.3 Posterior distributions and goodness of fit

The posterior distributions of macro parameters in Table 1 are roughly consistent with other estimates based solely on macro data. The γ parameter is larger than ψ, but it does not reach very high levels. Our model generates a slope in the term structure of interest rates through a combination of moderate risk-aversion and occasional bursts in the variance of of shocks, especially technology shocks

The standard deviation of measurement errors on 10-year yields is around 25 basis points, thus not exceedingly higher than in more flexible term structure models estimated using only yields data.

Concerning regime-switching parameters, the posterior mode of the transition probabilities suggests that all states are very persistent. The difference between estimated variances in the two regimes is marked for both technology and policy shocks.

Figures 1 and 2 show the implications of our parameter estimates for the population means, variances and cross-correlations of our observable variables. By and large, sample moments are within the posterior distribution of their population counterparts. In a few cases, however, sample moments are very much at the tail of the distribution of theoretical moments.

4.4 Regime switches and risk premia

Figure 3 displays 1-step-ahead forecasts and realised variables. This figure shows that the model can track yields data relatively well. In comparison, the fit is less good for inflation. One-step-ahead forecast errors are larger for the rates of growth of consumption and output.

Figure 3 also illustrates the clear heteroskedasticity in the residuals. In the early eighties, for example, there are clear increases in the variance of yields forecast errors. Our assumption of regime-switching in the variance of shocks helps the model to capture these patterns in the data.

Figures 4 and 5 display filtered and smoothed estimates of the discrete states together with the official NBER recession dates. In the figure, 1 denotes the low-variance state, 0 the high-variance state.

The regime associated with the policy shock hovers between the high and low variance regimes over the seventies; stays in the high variance regime in 1980 and remains there until 1983; it then drifts back to the low-variance regime until the beginning of the Great recession.

The variance of productivity growth shocks moves between high and low-variance regimes for most of the sample. On average, it tends to stay closer tothe high-variance regime over the eighties and nineties.

The technology and monetary policy states can be composed to define 4 possible combinations of regimes. This is done to construct Figures 6, which displays expected excess holding period returns as defined in equation (20), with confidence sets.

Excess returns are increasing in the maturity of bonds and fluctuate around 120 basis points at the 10-year maturity. The notable features emerging from Figure 6 is that regime-switching can induce a non-negligible amount of variability in expected excess holding period returns over time. This is a desirable feature to explain observed deviations of the data from features consistent with the expectations hypothesis (see e.g. Dai and Singleton, 2002). Variability is especially high at the time of the monetarist experiment in the early 1980s. This is also encouraging, because deviations of yields from values consistent with the expectations hypothesis are known to be particularly marked during the Volcker tenure. For example, Rudebusch and Wu (2006) note that the performance of the expectations hypothesis improves after 1988 and until 2002.

5 Conclusions

We have estimated the second order approximation of a macro-yield curve model with Epstein-Zin-Weil preferences, in which the variance of structural shocks is subject to changes of regime.

Our empirical results support the regime switching specification. Different regimes can fit the heteroskedasticity of economic variables. Estimated regimes also bear an intuitively appealing structural interpretation. Finally, changes in regimes generate nonnegligible changes in risk premia.

The inclusion of yields data in the estimation set does not alter the basic functioning of the macro-model. Most parameters are estimated to be close to the values obtained in other studies, in which solely macro data enter the econometrician's information set. The main exception is the interest rate smoothing coefficient in the monetary policy rule, which is found to be much higher than in studies which do not look at yields data. A high value of this coefficient helps to generate persistence in the short rate, hence to transmit movements in the policy rate to long-term yields.

Appendix

A The household problem

Using the definitions $U_{1,t} \equiv \partial U \left[u_t, \left(\mathrm{E}_t V_{t+1}^{1-\gamma} \right)^{\frac{1}{1-\gamma}} \right] / \partial u_t$ and $U_{2,t} \equiv \partial U \left[u_t, \left(\mathrm{E}_t V_{t+1}^{1-\gamma} \right)^{\frac{1}{1-\gamma}} \right] / \partial \left(\mathrm{E}_t V_{t+1}^{1-\gamma} \right)^{\frac{1}{1-\gamma}}$, we can write the first order conditions for the optimum as

$$\Lambda_t P_t = U_{1,t} v\left(N_t\right) - h \mathrm{E}_t U_{1,t+1} v\left(N_{t+1}\right) U_{2,t} \left(\frac{\left[\mathrm{E}_t J_{t+1}^{1-\gamma} \right]^{\frac{1}{1-\gamma}}}{J_{t+1}} \right)^{\gamma}$$

$$U_{1,t}\left(C_t - hC_{t-1}\right) v'\left(N_t\right) = \Lambda_t \frac{1-\theta_w}{\theta_w} w_t$$

$$Q_{t,t+1} = \frac{\Lambda_{t+1}}{\Lambda_t} U_{2,t} \left(\frac{\left[\mathrm{E}_t J_{t+1}^{1-\gamma} \right]^{\frac{1}{1-\gamma}}}{J_{t+1}} \right)^{\gamma} \tag{29}$$

plus the envelope conditions

$$J_{W,t} = \Lambda_t$$

$$J_{C,t} = -h U_{1,t} v\left(N_t\right)$$

where we also defined $J_t \equiv J\left(W_t, C_{t-1}\right)$, $J_{C,t} \equiv \partial J\left(W_t, C_{t-1}\right) / \partial C_{t-1}$.

Note that the two derivatives $U_{1,t}$ and $U_{2,t}$ can be rewritten as

$$U_{1,t} = (1-\beta) \left\{ (1-\beta)\left(C_t - hC_{t-1}\right)^{1-\psi} \left[v\left(N_t\right)\right]^{1-\psi} + \beta \left[\mathrm{E}_t J_{t+1}^{1-\gamma} \right]^{\frac{1-\psi}{1-\gamma}} \right\}^{\frac{\psi}{1-\psi}} \left(C_t - hC_{t-1}\right)^{-\psi} \left[v\left(N_t\right)\right]^{-\psi}$$

and

$$U_{2,t} \left[\mathrm{E}_t J_{t+1}^{1-\gamma} \right]^{\frac{\psi}{1-\gamma}} = \beta \left\{ (1-\beta)\left(C_t - hC_{t-1}\right)^{1-\psi} \left[v\left(N_t\right)\right]^{1-\psi} + \beta \left[\mathrm{E}_t J_{t+1}^{1-\gamma} \right]^{\frac{1-\psi}{1-\gamma}} \right\}^{\frac{\psi}{1-\psi}}$$

Moreover, at the optimum, the maximum value function will obey the recursion

$$J_t = \left\{ (1-\beta) \left[\left(C_t - hC_{t-1}\right) v\left(N_t\right) \right]^{1-\psi} + \beta \left[\mathrm{E}_t J_{t+1}^{1-\gamma} \right]^{\frac{1-\psi}{1-\gamma}} \right\}^{\frac{1}{1-\psi}}$$

Using these equations, we can rewrite the FOCs in the text as

$$\widetilde{w}_t = -\frac{v'\left(N_t\right)}{v\left(N_t\right)} \frac{\theta_w}{\theta_w - 1} \frac{\left[\left(C_t - hC_{t-1}\right) v\left(N_t\right) \right]^{1-\psi}}{\widetilde{\Lambda}_t}$$

and

$$Q_{t,t+1} = \beta \frac{\widetilde{\Lambda}_{t+1}}{\widetilde{\Lambda}_t} \frac{1}{\pi_{t+1}} \left(\frac{\left[\mathrm{E}_t J_{t+1}^{1-\gamma} \right]^{\frac{1}{1-\gamma}}}{J_{t+1}} \right)^{\gamma - \psi}$$

where

$$\widetilde{\Lambda}_t \equiv [v(N_t)]^{1-\psi}(C_t - hC_{t-1})^{-\psi} - \beta h \mathrm{E}_t [v(N_{t+1})]^{1-\psi}(C_{t+1} - hC_t)^{-\psi}\left(\frac{\left[\mathrm{E}_t J_{t+1}^{1-\gamma}\right]^{\frac{1}{1-\gamma}}}{J_{t+1}}\right)^{\gamma-\psi}$$

and $\widetilde{w}_t \equiv w_t/P_t$.

Note that in the absence of labour-leisure choice ($v(N_t) = 1$ for all t), we would obtain

$$Q_{t,t+1} = \frac{\beta}{\pi_{t+1}}\frac{\widetilde{\Lambda}_{t+1}}{\widetilde{\Lambda}_t}\left(\frac{\left[\mathrm{E}_t J_{t+1}^{1-\gamma}\right]^{\frac{1}{1-\gamma}}}{J_{t+1}}\right)^{\gamma-\psi}$$

$$\widetilde{\Lambda}_t \equiv (C_t - hC_{t-1})^{-\psi} - \beta h \mathrm{E}_t (C_{t+1} - hC_t)^{-\psi}\left(\frac{\left[\mathrm{E}_t J_{t+1}^{1-\gamma}\right]^{\frac{1}{1-\gamma}}}{J_{t+1}}\right)^{\gamma-\psi}$$

If habits were also set to zero, we would go back to the standard Epstein-Zin-Weil case

$$Q_{t,t+1} = \frac{\beta}{\pi_{t+1}}\left(\frac{C_{t+1}}{C_t}\right)^{-\psi}\left(\frac{\left[\mathrm{E}_t J_{t+1}^{1-\gamma}\right]^{\frac{1}{1-\gamma}}}{J_{t+1}}\right)^{\gamma-\psi}$$

$$\widetilde{\Lambda}_t = C_t^{-\psi}$$

Finally, when $\gamma = \psi$

$$\widetilde{w}_t = -\frac{v'(N_t)}{v(N_t)}\frac{\theta_{w,t}}{\theta_{w,t}-1}\frac{[(C_t - hC_{t-1})\,v(N_t)]^{1-\psi}}{\widetilde{\Lambda}_t}$$

$$Q_{t,t+1} = \frac{\beta}{\pi_{t+1}}\frac{\widetilde{\Lambda}_{t+1}}{\widetilde{\Lambda}_t}$$

$$\widetilde{\Lambda}_t = [v(N_t)]^{1-\psi}(C_t - hC_{t-1})^{-\psi} - \beta h \mathrm{E}_t [v(N_{t+1})]^{1-\psi}(C_{t+1} - hC_t)^{-\psi}$$

B The approximate SDF and short rate

Equation (6) in Piazzesi and Schneider (2006) derives the stochastic discount factor for a model with non-expected utility, exogenous labour supply, exogenous consumption process and $\psi = 1$. A similar expression can be derived in our model. In this appendix, we also allow for a stochastic trend in technology growth, so that

$$\begin{aligned} A_t &= Z_t B_t \\ B_t &= B_{t-1}\Xi_t \\ \Xi_t &= \overline{\Xi}^{1-\rho_\xi}\Xi_{t-1}^{\rho_\xi}e^{\varepsilon_t^\xi} \\ Z_t &= Z_{t-1}^{\rho_z}e^{\varepsilon_t^z} \end{aligned}$$

In order to approximate the model around a deterministic steady state, we first detrend all variables by the growing level of technology. More specifically, for the stochcastic discount factor we obtain

$$Q_{t,t+1} = \beta \frac{1}{\Xi_{t+1}^{\gamma}} \frac{\widetilde{\widetilde{\Lambda}}_{t+1}}{\widetilde{\widetilde{\Lambda}}_{t}} \frac{1}{\Pi_{t+1}} \left(\frac{\left[\mathrm{E}_t \Xi_{t+1}^{1-\gamma} \widetilde{J}_{t+1}^{1-\gamma} \right]^{\frac{1}{1-\gamma}}}{\widetilde{J}_{t+1}} \right)^{\gamma-\psi}$$

where Ξ_t is the rate of growth of technology, $\widetilde{\widetilde{\Lambda}}_t = \widetilde{\Lambda}_t / B_t^{-\psi}$ and $\widetilde{J}_t = J_t / B_t^{-\psi}$ where

$$\widetilde{\widetilde{\Lambda}}_t = \left[v\left(N_t\right) \right]^{1-\psi} \left(\widetilde{C}_t - h \frac{\widetilde{C}_{t-1}}{\Xi_t} \right)^{-\psi} - \beta h \mathrm{E}_t \left[v\left(N_{t+1}\right) \right]^{1-\psi} \Xi_{t+1}^{-\gamma} \left(\widetilde{C}_{t+1} - h \frac{\widetilde{C}_t}{\Xi_{t+1}} \right)^{-\psi} \frac{\left[\mathrm{E}_t \left(\widetilde{J}_{t+1} \Xi_{t+1} \right)^{1-\gamma} \right]^{\frac{\gamma-\psi}{1-\gamma}}}{\widetilde{J}_{t+1}^{\gamma-\psi}}$$

Note that using the definition $D_t \equiv \mathrm{E}_t \Xi_{t+1}^{1-\gamma} \widetilde{J}_{t+1}^{1-\gamma}$ the stochastic discount factor can be rewritten as

$$Q_{t,t+1} = \beta \frac{1}{\Xi_{t+1}^{\gamma}} \frac{\widetilde{\widetilde{\Lambda}}_{t+1}}{\widetilde{\widetilde{\Lambda}}_{t}} \frac{1}{\Pi_{t+1}} \left(\frac{D_t^{\frac{1}{1-\gamma}}}{\widetilde{J}_{t+1}} \right)^{\gamma-\psi}$$

or in deviation from the steady state

$$\begin{aligned} \widehat{q}_{t,t+1} &= \Delta \widehat{\widetilde{\widetilde{\lambda}}}_{t+1} - \psi \mathrm{E}_t \xi_{t+1} - \gamma \left(\widehat{\xi}_{t+1} - \mathrm{E}_t \widehat{\xi}_{t+1} \right) - \widehat{\pi}_{t+1} - (\gamma - \psi) \left(\widehat{\widetilde{j}}_{t+1} - \mathrm{E}_t \widehat{\widetilde{j}}_{t+1} \right) \\ & -\frac{1}{2} (\gamma - \psi)(\gamma - 1) \left(\mathrm{Var}_t \left[\widehat{\widetilde{j}}_{t+1} \right] + \mathrm{Var}_t \left[\widehat{\xi}_{t+1}^2 \right] + 2 \mathrm{Cov}_t \left[\widehat{\xi}_{t+1} \widehat{\widetilde{j}}_{t+1} \right] \right) \end{aligned} \tag{30}$$

When $\psi = 1$ this expression boils down to

$$\begin{aligned} \widehat{q}_{t,t+1} &= \Delta \widehat{\widetilde{\widetilde{\lambda}}}_{t+1} - \mathrm{E}_t \xi_{t+1} - \gamma \left(\widehat{\xi}_{t+1} - \mathrm{E}_t \widehat{\xi}_{t+1} \right) - \widehat{\pi}_{t+1} \\ & - (\gamma - 1) \left(\widehat{\widetilde{j}}_{t+1} - \mathrm{E}_t \widehat{\widetilde{j}}_{t+1} \right) - \frac{1}{2} (\gamma - 1)^2 \mathrm{Var}_t \left[\widehat{\widetilde{j}}_{t+1} \right] \\ & -\frac{1}{2} (\gamma - 1)^2 \mathrm{Var}_t \left[\widehat{\xi}_{t+1}^2 \right] - (\gamma - 1)^2 \mathrm{Cov}_t \left[\widehat{\xi}_{t+1} \widehat{\widetilde{j}}_{t+1} \right] \end{aligned}$$

which corresponds to the case considered by Piazzesi and Schneider (2006) when there is no growth and temporary utility depends on consumption only so that $\widetilde{\widetilde{\Lambda}}_t = \widetilde{C}_t^{-1}$ and $\Delta \widehat{\widetilde{\widetilde{\lambda}}}_{t+1} = -\Delta \widehat{c}_{t+1}$.

We now wish to rewrite $\widetilde{J}_t$ in terms of its consumption and labour determinants. Using the definitions $D_t \equiv \mathrm{E}_t \Xi_{t+1}^{1-\gamma} \widetilde{J}_{t+1}^{1-\gamma}$, $\overleftrightarrow{c}_t \equiv \widetilde{C}_t - h \widetilde{C}_{t-1} / \Xi_t$ (for the consumpion surplus) and $\widetilde{u}_t \equiv \overleftrightarrow{c}_t \left[v\left(N_t\right) \right]$, we can simplify the notation and write

$$\widetilde{J}_t^{1-\psi} = (1 - \beta) \widetilde{u}_t^{1-\psi} + \beta D_t^{\frac{1-\psi}{1-\gamma}}$$

and

$$\widetilde{\widetilde{\Lambda}}_t = v\left(N_t\right)\widetilde{u}_t^{-\psi} - \beta h \mathrm{E}_t v\left(N_{t+1}\right)\widetilde{u}_{t+1}^{-\psi}\Xi_{t+1}^{-\gamma}\frac{D_t^{\frac{\gamma-\psi}{1-\gamma}}}{\widetilde{J}_{t+1}^{\gamma-\psi}}$$

Note that $\widetilde{u}_t$ can be approximated as

$$\widehat{\widetilde{u}}_t = \widehat{\overleftrightarrow{c}}_t + \frac{v'N}{v}\widehat{l}_t + \frac{1}{2}\left(1+\frac{v''N}{v'}-\frac{v'N}{v}\right)\frac{v'N}{v}\widehat{l}_t^2$$

Using these definitions, the value function can be expanded to second order as

$$\begin{aligned}\widehat{\widetilde{j}}_t - \widehat{\widetilde{u}}_t &= \beta\Xi^{1-\psi}\mathrm{E}_t\left[\widehat{\widetilde{j}}_{t+1}+\widehat{\xi}_{t+1}-\widehat{\widetilde{u}}_{t+1}+\Delta\widehat{\widetilde{u}}_{t+1}\right]+\frac{1-\gamma}{2}\beta\Xi^{1-\psi}\mathrm{Var}_t\left[\widehat{\widetilde{j}}_{t+1}+\widehat{\xi}_{t+1}\right] \\ &\quad +\frac{1}{2}\left(1-\psi\right)\left(1-\beta\Xi^{1-\psi}\right)\beta\Xi^{1-\psi}\left(\mathrm{E}_t\left[\widehat{\widetilde{j}}_{t+1}+\widehat{\xi}_{t+1}-\widehat{\widetilde{u}}_{t+1}+\Delta\widehat{\widetilde{u}}_{t+1}\right]\right)^2 \qquad (31)\end{aligned}$$

To solve equation (31) forward, note that second order terms can be evaluated using the first order approximation

$$\widehat{\widetilde{j}}_t - \widehat{\widetilde{u}}_t = \beta\Xi^{1-\psi}\mathrm{E}_t\left[\widehat{\widetilde{j}}_{t+1}+\widehat{\xi}_{t+1}-\widehat{\widetilde{u}}_{t+1}+\Delta\widehat{\widetilde{u}}_{t+1}\right]$$

Hence

$$\mathrm{Var}_t\left[\widehat{\widetilde{j}}_{t+1}+\widehat{\xi}_{t+1}\right] = \mathrm{Var}_t\left[\sum_{i=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^i\mathrm{E}_{t+1}\left[\widehat{\xi}_{t+1+i}+\Delta\widehat{\widetilde{u}}_{t+1+i}\right]\right]$$

and

$$\left(\mathrm{E}_t\left[\widehat{\widetilde{j}}_{t+1}+\widehat{\xi}_{t+1}-\widehat{\widetilde{u}}_{t+1}+\Delta\widehat{\widetilde{u}}_{t+1}\right]\right)^2 = \left(\sum_{i=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^i\mathrm{E}_t\left[\widehat{\xi}_{t+1+i}+\Delta\widehat{\widetilde{u}}_{t+1+i}\right]\right)^2$$

First order terms imply

$$\begin{aligned}&\mathrm{E}_t\left[\widehat{\widetilde{j}}_{t+1}+\widehat{\xi}_{t+1}-\widehat{\widetilde{u}}_{t+1}+\Delta\widehat{\widetilde{u}}_{t+1}\right] \\ = \; &\frac{1-\gamma}{2}\mathrm{E}_t\sum_{j=1}^{\infty}\left(\beta\Xi^{1-\psi}\right)^j\mathrm{Var}_{t+j}\mathrm{E}_{t+1+j}\left[\sum_{i=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^i\left(\widehat{\xi}_{t+1+j+i}+\Delta\widehat{\widetilde{u}}_{t+1+j+i}\right)\right] \\ &+\frac{1-\psi}{2}\left(1-\beta\Xi^{1-\psi}\right)\sum_{j=1}^{\infty}\left(\beta\Xi^{1-\psi}\right)^j\mathrm{E}_t\left[\left(\sum_{i=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^i\mathrm{E}_{t+j}\left[\widehat{\xi}_{t+1+j+i}+\Delta\widehat{\widetilde{u}}_{t+1+j+i}\right]\right)^2\right] \\ &+\sum_{j=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^j\mathrm{E}_t\left[\widehat{\xi}_{t+1+j}+\Delta\widehat{\widetilde{u}}_{t+1+j}\right]\end{aligned}$$

Putting everything together

$$\begin{aligned}\widehat{\widetilde{j}}_t - \widehat{\widetilde{u}}_t &= \sum_{j=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{j+1}\mathrm{E}_t\left[\widehat{\xi}_{t+1+j}+\Delta\widehat{\widetilde{u}}_{t+1+j}\right] \\ &\quad +\frac{1-\gamma}{2}\sum_{j=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{j+1}\mathrm{E}_t\mathrm{Var}_{t+j}\mathrm{E}_{t+1+j}\left[\sum_{i=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^i\left(\widehat{\xi}_{t+1+j+i}+\Delta\widehat{\widetilde{u}}_{t+1+j+i}\right)\right] \\ &\quad +\frac{1-\psi}{2}\left(1-\beta\Xi^{1-\psi}\right)\sum_{j=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{j+1}\mathrm{E}_t\left[\left(\sum_{i=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^i\mathrm{E}_{t+j}\left[\widehat{\xi}_{t+1+j+i}+\Delta\widehat{\widetilde{u}}_{t+1+j+i}\right]\right)^2\right]\end{aligned}$$

where

$$\Delta\widehat{\widehat{u}}_{t+1} = \Delta\widehat{\overleftrightarrow{c}}_{t+1} + \frac{v'N}{v}\Delta\widehat{l}_{t+1} + \frac{1}{2}\left(1 + \frac{v''N}{v'} - \frac{v'N}{v}\right)\frac{v'N}{v}\left(\widehat{l}^2_{t+1} - \widehat{l}^2_t\right)$$

and

$$\Delta\widehat{\overleftrightarrow{c}}_{t+1} \equiv \frac{1}{\Xi - h}\left(\Xi\Delta\widehat{\widehat{c}}_{t+1} - h\Delta\widehat{\widehat{c}}_t + h\Delta\widehat{\xi}_{t+1}\right) - \frac{1}{2}\frac{h\Xi}{(\Xi - h)^2}\left[\left(\Delta\widehat{\widehat{c}}_{t+1} + \widehat{\xi}_{t+1}\right)^2 - \left(\Delta\widehat{\widehat{c}}_t + \widehat{\xi}_t\right)^2\right]$$

To use this expressions in the stochastic discount factor of equation (30), note that variances and convariances can be evaluated using the first order approximation

$$\widehat{\widehat{j}}_t = \widehat{\widehat{u}}_t + \sum_{j=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{j+1}\mathrm{E}_t\left[\widehat{\xi}_{t+1+j} + \Delta\widehat{\widehat{u}}_{t+1+j}\right]$$

which yields

$$\begin{aligned}\mathrm{Var}_t\left[\widehat{\widehat{j}}_{t+1}\right] &= \mathrm{Var}_t\left[\widehat{\xi}_{t+1}\right] + \mathrm{Var}_t\left[\sum_{j=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{j}\mathrm{E}_{t+1}\left[\widehat{\xi}_{t+1+j} + \Delta\widehat{\widehat{u}}_{t+1+j}\right]\right] \\ &\quad -2\mathrm{Cov}_t\left[\widehat{\xi}_{t+1}, \sum_{j=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{j}\mathrm{E}_{t+1}\left[\widehat{\xi}_{t+1+j} + \Delta\widehat{\widehat{u}}_{t+1+j}\right]\right]\end{aligned}$$

and

$$\mathrm{Cov}_t\left[\widehat{\xi}_{t+1}\widehat{\widehat{j}}_{t+1}\right] = -\mathrm{Var}_t\left[\widehat{\xi}_{t+1}\right] + \mathrm{Cov}_t\left[\widehat{\xi}_{t+1}, \sum_{j=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{j}\mathrm{E}_{t+1}\left[\widehat{\xi}_{t+1+j} + \Delta\widehat{\widehat{u}}_{t+1+j}\right]\right]$$

Finally

$$\begin{aligned}\widehat{\widehat{j}}_{t+1} - \mathrm{E}_t\widehat{\widehat{j}}_{t+1} &= -\left(\mathrm{E}_{t+1} - \mathrm{E}_t\right)\widehat{\xi}_{t+1} + \sum_{j=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{j}\left(\mathrm{E}_{t+1} - \mathrm{E}_t\right)\left[\widehat{\xi}_{t+1+j} + \Delta\widehat{\widehat{u}}_{t+1+j}\right] + \frac{1-\gamma}{2}\times \\ &\quad \times\sum_{j=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{j+1}\left(\mathrm{E}_{t+1} - \mathrm{E}_t\right)\mathrm{Var}_{t+1+j}\mathrm{E}_{t+2+j}\left[\sum_{i=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{i}\left(\widehat{\xi}_{t+2+j+i} + \Delta\widehat{\widehat{u}}_{t+2+j+i}\right)\right] \\ &\quad +\frac{1-\psi}{2}\left(1 - \beta\Xi^{1-\psi}\right)\times \\ &\quad \times\sum_{j=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{j+1}\left(\mathrm{E}_{t+1} - \mathrm{E}_t\right)\left[\left(\sum_{i=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{i}\mathrm{E}_{t+1+j}\left[\widehat{\xi}_{t+2+j+i} + \Delta\widehat{\widehat{u}}_{t+2+j+i}\right]\right)^2\right]\end{aligned}$$

The second order approximation of the stochastic discount factor is therefore

$$\begin{aligned}
\widehat{q}_{t,t+1} &= \Delta\widehat{\widetilde{\lambda}}_{t+1} - \psi\widehat{\xi}_{t+1} - \widehat{\pi}_{t+1} - (\gamma-\psi)\sum_{j=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{j}\left(\mathrm{E}_{t+1}-\mathrm{E}_{t}\right)\left(\widehat{\xi}_{t+1+j} + \Delta\widehat{\widetilde{u}}_{t+1+j}\right) + \\
&\quad -\frac{1}{2}(\gamma-\psi)(\gamma-1)\,\mathrm{Var}_{t}\left[\sum_{i=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{i}\mathrm{E}_{t+1}\left(\widehat{\xi}_{t+1+i} + \Delta\widehat{\widetilde{u}}_{t+1+i}\right)\right] + \frac{1}{2}(\gamma-\psi)(\gamma-1)\times \\
&\quad \times\left(\mathrm{E}_{t+1}-\mathrm{E}_{t}\right)\sum_{j=1}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{j}\mathrm{Var}_{t+j}\mathrm{E}_{t+1+j}\left[\sum_{i=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{i}\left(\widehat{\xi}_{t+1+j+i} + \Delta\widehat{\widetilde{u}}_{t+1+j+i}\right)\right] + \\
&\quad + \frac{(\gamma-\psi)(\psi-1)\left(1-\beta\Xi^{1-\psi}\right)}{2}\times \\
&\quad \times\left(\mathrm{E}_{t+1}-\mathrm{E}_{t}\right)\sum_{j=1}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{j}\left[\left(\sum_{i=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{i}\mathrm{E}_{t+j}\left[\widehat{\xi}_{t+1+j+i} + \Delta\widehat{\widetilde{u}}_{t+1+j+i}\right]\right)^{2}\right]
\end{aligned}$$

Note that we could define

$$\overleftrightarrow{\Lambda}_{t} \equiv v(N_t) - \beta h\mathrm{E}_{t}v(N_{t+1})\frac{\widetilde{u}_{t+1}^{-\psi}}{\widetilde{u}_{t}^{-\psi}}\Xi_{t+1}^{-\gamma}\frac{D_{t}^{\frac{\gamma-\psi}{1-\gamma}}}{\widetilde{J}_{t+1}^{\gamma-\psi}}$$

to write

$$\widetilde{\widetilde{\Lambda}}_{t} \equiv \widetilde{u}_{t}^{-\psi}\overleftrightarrow{\Lambda}_{t}$$

such that

$$\widehat{\overleftrightarrow{\lambda}}_{t} = \frac{v'N}{v}\widehat{l}_{t} + \frac{1}{2}\left[1 + \frac{v''N}{v'} - \frac{v'N}{v}\right]\frac{v'N}{v}\widehat{l}_{t}^{2}$$

We could also define a "growth adjusted utility" $\Delta\overleftrightarrow{u}_{t} \equiv \Delta\widehat{\widetilde{u}}_{t} + \widehat{\xi}_{t}$ to finally write

$$\begin{aligned}
\widehat{q}_{t,t+1} &= \Delta\widehat{\overleftrightarrow{\lambda}}_{t+1} - \psi\Delta\overleftrightarrow{u}_{t+1} - \widehat{\pi}_{t+1} - (\gamma-\psi)\sum_{j=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{j}\left(\mathrm{E}_{t+1}-\mathrm{E}_{t}\right)\Delta\overleftrightarrow{u}_{t+1+j} \\
&\quad -\frac{1}{2}(\gamma-\psi)(\gamma-1)\,\mathrm{Var}_{t}\left[\sum_{i=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{i}\mathrm{E}_{t+1}\Delta\overleftrightarrow{u}_{t+1+i}\right] \\
&\quad +\frac{1}{2}(\gamma-\psi)(\gamma-1)\left(\mathrm{E}_{t+1}-\mathrm{E}_{t}\right)\sum_{j=1}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{j}\mathrm{Var}_{t+j}\mathrm{E}_{t+1+j}\left[\sum_{i=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{i}\Delta\overleftrightarrow{u}_{t+1+j+i}\right] \\
&\quad +\frac{(\gamma-\psi)(\psi-1)\left(1-\beta\Xi^{1-\psi}\right)}{2}\left(\mathrm{E}_{t+1}-\mathrm{E}_{t}\right)\sum_{j=1}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{j}\left[\left(\sum_{i=0}^{\infty}\left(\beta\Xi^{1-\psi}\right)^{i}\mathrm{E}_{t+j}\Delta\overleftrightarrow{u}_{t+1+j+i}\right)^{2}\right]
\end{aligned}$$

To a second order approximation, the short rate is given by

$$\widehat{i}_{t} = -\mathrm{E}_{t}\widehat{q}_{t,t+1} - \frac{1}{2}\mathrm{Var}_{t}\widehat{q}_{t,t+1}$$

From the SDF we obtain

$$
\begin{aligned}
\widehat{i}_t &= -\mathrm{E}_t \Delta \widehat{\overleftrightarrow{\lambda}}_{t+1} + \mathrm{E}_t \psi \Delta \overleftrightarrow{u}_{t+1} + \mathrm{E}_t \widehat{\pi}_{t+1} - \frac{1}{2} \mathrm{Var}_t \left(\Delta \widehat{\overleftrightarrow{\lambda}}_{t+1} - \psi \Delta \overleftrightarrow{u}_{t+1} - \widehat{\pi}_{t+1} \right) \\
&+ \frac{1}{2} (\gamma - \psi)(\psi - 1) \mathrm{Var}_t \left[\mathrm{E}_{t+1} \sum_{i=0}^{\infty} \left(\beta \Xi^{1-\psi} \right)^i \Delta \overleftrightarrow{u}_{t+1+i} \right] \\
&+ (\gamma - \psi) \mathrm{Cov}_t \left[\left(\Delta \widehat{\overleftrightarrow{\lambda}}_{t+1} - \psi \Delta \overleftrightarrow{u}_{t+1} - \widehat{\pi}_{t+1} \right), \mathrm{E}_{t+1} \sum_{i=0}^{\infty} \left(\beta \Xi^{1-\psi} \right)^i \Delta \overleftrightarrow{u}_{t+1+i} \right]
\end{aligned}
$$

C Solution for bond prices

Recall that

$$
\begin{aligned}
\mathbf{x}_{t+1} &= P \widehat{\mathbf{x}}_t + \frac{1}{2} \left(I_{n_x} \otimes \widehat{\mathbf{x}}_t' \right) G \widehat{\mathbf{x}}_t + k_{x,s_t} \widetilde{\sigma}^2 + \widetilde{\sigma} \Sigma_t \mathbf{u}_{t+1} \\
\mathbf{s}_{t+1} &= \boldsymbol{\kappa}_0 + \boldsymbol{\kappa}_1 \mathbf{s}_t + \boldsymbol{\nu}_{t+1} \\
\widehat{\overleftrightarrow{\lambda}}_t &= F_\lambda \widehat{\mathbf{x}}_t + \frac{1}{2} \left(I_{n_y} \otimes \widehat{\mathbf{x}}_t' \right) E_\lambda \widehat{\mathbf{x}}_t + k_{\lambda,s_t} \widetilde{\sigma}^2 \\
\widehat{\overleftrightarrow{\pi}}_t &= F_\pi \widehat{\mathbf{x}}_t + \frac{1}{2} \left(I_{n_y} \otimes \widehat{\mathbf{x}}_t' \right) E_\pi \widehat{\mathbf{x}}_t + k_{\pi,s_t} \widetilde{\sigma}^2
\end{aligned}
$$

where P is a $n_x \times n_x$ matrix, G is a $n_x^2 \times n_x$ matrix, $k_{x,s}$ is an $n_x \times 1$ vector (whose elements are state dependent), $\Sigma_t \equiv \Sigma(\mathbf{s}_t)$ is a $n_x \times n_u$ matrix, F_λ and F_π are $1 \times n_x$ vectors, E_λ and E_π are $n_x \times n_x$ matrices, and finally $k_{\lambda,s}$ and $k_{\pi,s}$ are (state dependent) scalars.

C.1 1-period bonds

To derive the price of 1-period bonds, note first that a second order approximation to the stochastic discount factor is

$$
\widehat{q}_{t,t+1} = (F_\lambda - F_\pi) \widehat{\mathbf{x}}_{t+1} + \frac{1}{2} \widehat{\mathbf{x}}_{t+1}' (E_\lambda - E_\pi) \widehat{\mathbf{x}}_{t+1} - F_\lambda \widehat{\mathbf{x}}_t - \frac{1}{2} \widehat{\mathbf{x}}_t' E_\lambda \widehat{\mathbf{x}}_t - k_{\pi,s} \widetilde{\sigma}^2
$$

or

$$
\begin{aligned}
\widehat{q}_{t,t+1} &= ((F_\lambda - F_\pi) P - F_\lambda) \widehat{\mathbf{x}}_t + \frac{1}{2} (F_\lambda - F_\pi) \left(I_{n_x} \otimes \widehat{\mathbf{x}}_t' \right) G \widehat{\mathbf{x}}_t \\
&+ \frac{1}{2} \widehat{\mathbf{x}}_t' P' (E_\lambda - E_\pi) P \widehat{\mathbf{x}}_t - \frac{1}{2} \widehat{\mathbf{x}}_t' E_\lambda \widehat{\mathbf{x}}_t \\
&+ \widetilde{\sigma}^2 (F_\lambda - F_\pi) k_{x,s} - k_{\pi,s} \widetilde{\sigma}^2 \\
&+ \widetilde{\sigma} (F_\lambda - F_\pi) \Sigma_t \mathbf{u}_{t+1} + \frac{1}{2} \widetilde{\sigma} \widehat{\mathbf{x}}_t' P' (E_\lambda - E_\pi) \Sigma_t \mathbf{u}_{t+1} \\
&+ \frac{1}{2} \widetilde{\sigma} \mathbf{u}_{t+1}' \Sigma_t' (E_\lambda - E_\pi) P \widehat{\mathbf{x}}_t + \frac{1}{2} \widetilde{\sigma}^2 \mathbf{u}_{t+1}' \Sigma_t' (E_\lambda - E_\pi) \Sigma_t \mathbf{u}_{t+1}
\end{aligned}
$$

To second order, the price of a 1-period bond is

$$\widehat{b}_{t,1} = -i_t = \mathrm{E}_t\left[\widehat{q}_{t+1}\right] + \frac{1}{2}\left(\mathrm{E}_t\left[\widehat{q}_{t+1}^2\right] - \left(\mathrm{E}_t\left[\widehat{q}_{t+1}\right]\right)^2\right)$$

for which we need

$$\mathrm{E}_t\left[\widehat{q}_{t+1}^2\right] - \left(\mathrm{E}_t\left[\widehat{q}_{t+1}\right]\right)^2 = \widetilde{\sigma}^2\left(F_\lambda - F_\pi\right)\Sigma_t\Sigma_t'\left(F_\lambda - F_\pi\right)'$$

and

$$\begin{aligned}
\mathrm{E}_t\left[\widehat{q}_{t,t+1}\right] &= \left(F_\lambda - F_\pi\right)P\widehat{\mathbf{x}}_t + \frac{1}{2}\left(F_\lambda - F_\pi\right)\left(I_{n_x}\otimes\widehat{\mathbf{x}}_t'\right)G\widehat{\mathbf{x}}_t + \widetilde{\sigma}^2\left(F_\lambda - F_\pi\right)k_{x,s} \\
&\quad + \frac{1}{2}\widehat{\mathbf{x}}_t'P'\left(E_\lambda - E_\pi\right)P\widehat{\mathbf{x}}_t + \frac{1}{2}\widetilde{\sigma}^2\mathrm{E}_t\left[\mathbf{u}_{t+1}'\Sigma_t'\left(E_\lambda - E_\pi\right)\Sigma_t\mathbf{u}_{t+1}\right] \\
&\quad - F_\lambda\widehat{\mathbf{x}}_t - \frac{1}{2}\widehat{\mathbf{x}}_t'E_\lambda\widehat{\mathbf{x}}_t - k_{\pi,s}\widetilde{\sigma}^2
\end{aligned}$$

Now note that, for any matrix A and vector x,

$$\begin{aligned}
\mathrm{E}\left[x'Ax\right] &= \mathrm{E}\left[\mathrm{vec}\left(x'Ax\right)\right] \\
&= \mathrm{E}\left[x'\otimes x'\right]\mathrm{vec}\left(A\right) \\
&= \left(\mathrm{vec}\left(\mathrm{E}\left[xx'\right]\right)\right)'\mathrm{vec}\left(A\right)
\end{aligned}$$

where the vec operator transforms a matrix into a vector by stacking its columns. It follows that

$$\begin{aligned}
\mathrm{E}_t\left[\mathbf{u}_{t+1}'\Sigma_t{}'\left(E_\lambda - E_\pi\right)\Sigma_t\mathbf{u}_{t+1}\right] &= \left(\mathrm{vec}\left(\mathrm{I}\right)\right)\mathrm{vec}\left(\Sigma_t'\left(E_\lambda - E_\pi\right)\Sigma_t\right) \\
&= \mathrm{tr}\left(\Sigma_t{}'\left(E_\lambda - E_\pi\right)\Sigma_t\right)
\end{aligned}$$

where tr represents the trace, i.e. the sum of the diagonal elements of a matrix.

Hence,

$$\begin{aligned}
\widehat{b}_{t,1} &= \left(\left(F_\lambda - F_\pi\right)P - F_\lambda\right)\widehat{\mathbf{x}}_t + \widetilde{\sigma}^2\left(\left(F_\lambda - F_\pi\right)k_{x,s} - k_{\pi,s}\right) \\
&\quad + \frac{1}{2}\widetilde{\sigma}^2\mathrm{tr}\left(\Sigma_t'\left(E_\lambda - E_\pi\right)_t\right) + \frac{1}{2}\widetilde{\sigma}^2\left(F_\lambda - F_\pi\right)\Sigma_t\Sigma_t'\left(F_\lambda - F_\pi\right)' \\
&\quad + \frac{1}{2}\widehat{\mathbf{x}}_t'\left(P'\left(E_\lambda - E_\pi\right)P - E_\lambda\right)\widehat{\mathbf{x}}_t + \frac{1}{2}\left(F_\lambda - F_\pi\right)\left(I_{n_x}\otimes\widehat{\mathbf{x}}_t'\right)G\widehat{\mathbf{x}}_t
\end{aligned}$$

Finally, note that

$$
\begin{aligned}
\left(F_\lambda - F_\pi\right)\left(I_{n_x} \otimes \widehat{\mathbf{x}}_t'\right) G\widehat{\mathbf{x}}_t &= \left(F_\lambda - F_\pi\right) \begin{pmatrix} \widehat{\mathbf{x}}_t' & 0 & \cdots & 0 \\ 0 & \widehat{\mathbf{x}}_t' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \widehat{\mathbf{x}}_t' \end{pmatrix} G\widehat{\mathbf{x}}_t \\
&= \left[\ \left(F_{\lambda,1} - F_{\pi,1}\right)\widehat{\mathbf{x}}_t' \quad \left(F_{\lambda,2} - F_{\pi,2}\right)\widehat{\mathbf{x}}_t' \quad \cdots \quad \left(F_{\lambda,n_x} - F_{\pi,n_x}\right)\widehat{\mathbf{x}}_t'\ \right] G\widehat{\mathbf{x}}_t \\
&= \left[\widehat{\mathbf{x}}_t'\left(F_{\lambda,1} - F_{\pi,1}\right)\mathbf{G}_1 + \widehat{\mathbf{x}}_t'\left(F_{\lambda,2} - F_{\pi,2}\right)\mathbf{G}_2 + \cdots + \widehat{\mathbf{x}}_t'\left(F_{\lambda,n_x} - F_{\pi,n_x}\right)\mathbf{G}_{n_x}\right]\widehat{\mathbf{x}}_t \\
&= \widehat{\mathbf{x}}_t' \left[\sum_{j=1}^{n_x}\left(F_{\lambda,j} - F_{\pi,j}\right)\mathbf{G}_j\right]\widehat{\mathbf{x}}_t
\end{aligned}
$$

where $F_{\lambda,i}$ and $F_{\pi,i}$ denote the i-th elements of vectors F_λ and F_π, respectively, and $\mathbf{G}_i$ denotes the i-th $n_x \times n_x$ matrix which are vertically stacked to make up G. We can therefore rewrite the 1-period bond as

$$
\widehat{b}_{t,1} = F_{B,1}\widehat{\mathbf{x}}_t + \frac{1}{2}\widehat{\mathbf{x}}_t' E_{B,1}\widehat{\mathbf{x}}_t + k_{B,1,s}\widetilde{\sigma}^2
$$

where

$$
\begin{aligned}
F_{B_1} &\equiv \left(F_\lambda - F_\pi\right) P - F_\lambda \\
k_{B_1,s} &\equiv \left(F_\lambda - F_\pi\right) k_{x,s} - k_{\pi,s} + \operatorname{tr}\left(\Sigma_t'\left(E_\lambda - E_\pi\right)\Sigma_t\right) + \left(F_\lambda - F_\pi\right)\Sigma_t\Sigma_t'\left(F_\lambda - F_\pi\right)' \\
E_{B_1} &\equiv P'\left(E_\lambda - E_\pi\right) P - E_\lambda + \sum_{j=1}^{n_x}\left(F_{\lambda,j} - F_{\pi,j}\right)\mathbf{G}_j
\end{aligned}
$$

Note also that, by construction, $\widehat{b}_{t,1} = -\widehat{i}_t$, so $F_{B_1} = -F_i$, $E_{B_1} = -E_i$ and $k_{B_1,s} = -k_{i,s}$. Note that this definition also allows us to rewrite $\widehat{q}_{t,t+1}$ as

$$
\begin{aligned}
\widehat{q}_{t,t+1} &= F_{B,1}\widehat{\mathbf{x}}_t + \frac{1}{2}\widehat{\mathbf{x}}_t' E_{B_1}\widehat{\mathbf{x}}_t + \widetilde{\sigma}^2\left(\left(F_\lambda - F_\pi\right) k_{x,s} - k_{\pi,s}\right) \\
&\quad + \widetilde{\sigma}\left(F_\lambda - F_\pi\right)\Sigma_t\mathbf{u}_{t+1} + \widetilde{\sigma}\widehat{\mathbf{x}}_t' P'\left(E_\lambda - E_\pi\right)\Sigma_t\mathbf{u}_{t+1} + \frac{1}{2}\widetilde{\sigma}^2\mathbf{u}_{t+1}'\Sigma_t'\left(E_\lambda - E_\pi\right)\Sigma_t\mathbf{u}_{t+1}
\end{aligned}
$$

C.2 2-period bonds

2-period bond prices can be written as (up to a second order approximation)

$$
\widehat{b}_{t,2} = \widehat{b}_{t,1} + \mathrm{E}_t\left[\widehat{b}_{t+1,1}\right] + \frac{1}{2}\mathrm{Var}_t\left[\widehat{b}_{t+1,1}\right] + \mathrm{Cov}_t\left[\widehat{q}_{t+1}, \widehat{b}_{t+1,1}\right]
$$

Based on 1-period prices, we can derive

$$
\begin{aligned}
\mathrm{E}_t\left[\widehat{b}_{t+1,1}\right] &= F_{B_1} P\widehat{\mathbf{x}}_t + \frac{1}{2} F_{B_1}\left(I_{n_x} \otimes \widehat{\mathbf{x}}_t'\right) G\widehat{\mathbf{x}}_t + \frac{1}{2}\widehat{\mathbf{x}}_t' P' E_{B_1} P\widehat{\mathbf{x}}_t \\
&\quad + F_{B_1} k_{x,s}\widetilde{\sigma}^2 + \widetilde{\sigma}^2 k_{B_1,s} + \frac{1}{2}\widetilde{\sigma}^2 \operatorname{tr}\left[\Sigma_t' E_{B_1}\Sigma_t\right]
\end{aligned}
$$

and

$$\begin{aligned}
\mathrm{E}_t\left[\widehat{b}_{t+1,1}\widehat{b}'_{t+1,1}\right]-\mathrm{E}_t\left[\widehat{b}'_{t+1,1}\right]\mathrm{E}_t\left[\widehat{b}_{t+1,1}\right] &= \widetilde{\sigma}^2 F_{B_1}\Sigma'_t\Sigma'_t F'_{B_1} \\
\mathrm{E}_t\left[\widehat{b}_{t+1,1}\widehat{q}'_{t+1}\right]-\mathrm{E}_t\left[\widehat{b}_{t+1,1}\right]\mathrm{E}_t\left[\widehat{q}'_{t+1}\right] &= F_{B_1}\Sigma_t\Sigma'_t\left(F_\lambda - F_\pi\right)'
\end{aligned}$$

It follows that

$$\widehat{b}_{t,2} = F_{B_2}\widehat{\mathbf{x}}_t + \frac{1}{2}\widehat{\mathbf{x}}'_t E_{B_2}\widehat{\mathbf{x}}_t + k_{B_2,s}\widetilde{\sigma}^2$$

where

$$\begin{aligned}
F_{B_2} &= F_{B_1}\left(I+P\right) \\
E_{B_2} &= E_{B_1} + P'E_{B_1}P + \sum_{j=1}^{n_x} F_{B_1,j}\mathbf{G}_j \\
k_{B_2,s} &= k_{B_1,s} + F_{B_1}k_{x,s} + \mathrm{tr}\left(\Sigma'_t E_{B_1}\Sigma_t\right) + F_{B_1}\Sigma_t\Sigma'_t F'_{B_1} + 2F_{B_1}\Sigma_t\Sigma'_t\left(F_\lambda - F_\pi\right)'
\end{aligned}$$

C.3 n-period bonds

Using the same procedure, we find that n-period bond prices can be written as

$$\widehat{b}_{t,n} = F_{B_n}\widehat{\mathbf{x}}_t + \frac{1}{2}\widehat{\mathbf{x}}'_t E_{B_n}\widehat{\mathbf{x}}_t + k_{B_n,s}\widetilde{\sigma}^2$$

where for $n > 1$

$$\begin{aligned}
F_{B_n} &= F_{B_1} + F_{B_{n-1}}P \\
E_{B_n} &= E_{B_1} + P'E_{B_{n-1}}P + \sum_{j=1}^{n_x} F_{B_{n-1},j}\mathbf{G}_j \\
k_{B_n,s} &= k_{B_1,s} + k_{B_{n-1},s} + F_{B_{n-1}}k_{x,s} + \mathrm{tr}\left(\Sigma'_t E_{B_{n-1}}\Sigma_t\right) \\
&\quad + F_{B_{n-1}}\Sigma_t\Sigma'_t F'_{B_{n-1}} + 2F_{B_{n-1}}\Sigma_t\Sigma'_t\left(F_\lambda - F_\pi\right)'
\end{aligned}$$

References

[1] Amisano, G. and O. Tristani (2010a), “Euro area inflation persistence in an estimated nonlinear DSGE model,” Journal of Economic Dynamics and Control, 34, 1837-1858.

[2] Amisano, G. and O. Tristani (2010b), “Exact likelihood computation for nonlinear DSGE models with heteroskedastic innovations,” mimeo, European Central Bank.

[3] Andreasen, M.M. (2008), "Explaining macroeconomic and term-structure dynamics jointly in a non-linear DSGE model," mimeo, University of Aarhus.

[4] Ang, Andrew and G. Bekaert (2002a), "Regime Switches in Interest Rates", Journal of Business and Economic Statistics 20, 163-182.

[5] Ang, Andrew and G. Bekaert (2002b), "Short Rate Nonlinearities and Regime Switches", Journal of Economic Dynamics and Control 26, 1243-1274.

[6] Ang, Andrew, G. Bekaert and Min Wei (2008), "The Term Structure of Real Rates and Expected Inflation," Journal of Finance 63, 797-849.

[7] Atkeson, A. and P. Kehoe (2008), "On the need for a new approach to analyzing monetary policy", Federal Reserve Bank of Minneapolis Working Paper No. 662.

[8] Backus, D.K., B.R. Routledge and S.E. Zin (2004), "Exotic Preferences for Macroeconomists", NBER Macroeconomics Annual.

[9] Backus, D.K., B.R. Routledge and S.E. Zin (2005), "Recursive Preferences", New York University, Leonard N. Stern School of Business Working Paper No. 05-19.

[10] Backus, D.K., B.R. Routledge and S.E. Zin (2007), "Asset prices in business cycle analysis", mimeo, New York University

[11] Backus, D.K. and J.H. Wright (2007), "Cracking the conundrum", Brookings Papers on Economic Activity 38 (2007-1), pages 293-329

[12] Bansal, R. and Hao Zhou (2002), "Term Structure of Interest Rates with Regime Shifts", Journal of Finance 57, 1997-2043.

[13] Bansal, R., G. Tauchen and Hao Zhou (2004), "Regime-Shifts, Risk Premiums in the Term Structure, and the Business Cycle", Journal of Business and Economic Statistics.

[14] Bansal, R. and I. Shaliastovich (2008), "A Long-Run Risks Explanation of Predictability Puzzles in Bond and Currency Markets", mimeo, Duke University.

[15] Bekaert, G., Seonghoon Cho and Antonio Moreno (2006), "New-Keynesian Macroeconomics and the Term Structure," mimeo.

[16] Bikbov, R. and M. Chernov (2008), "Monetary Policy Regimes and The Term Structure of Interest Rates", mimeo, London Business School.

[17] Binsbergen, J.H, J. Fernandez-Villaverde, R.S.J. Koijen and J. Rubio-Ramirez (2008), “Likelihood estimation of DSGE models with Epstein-Zin preferences”, mimeo.

[18] Campbell, J.Y. (1999), "Asset prices, consumption, and the business cycle," in M. Woodford, and J.B. Taylor, eds.: Handbook of Macroeconomics, pp. 1231–1303, Amsterdam: North-Holland.

[19] Canova, Fabio and Luca Sala (2006), "Back to Square One: Identification Issues in DSGE Models", ECB Working Paper No. 583.

[20] Cochrane, John H. (2008), "Risk and regimes in the bond market: Comments on Atkeson and Kehoe," [...].

[21] Dai, G. and K. Singleton (2002), "Expectations puzzles, time-varying risk premia, and affine models of the term structure," *Journal of Financial Economics* 63, 415-441.

[22] Dai, Qiang, K. Singleton and Wei Yang (2008), "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields," Review of Financial Studies, forthcoming.

[23] De Graeve, Ferre, Marina Emiris and Raf Wouters (2007), A Structural Decomposition of the US Yield Curve, mimeo.

[24] Doh, Taeyoung (2006), "What Moves the Yield Curve? Lessons from an Estimated Nonlinear Macro Model," manuscript, University of Pennsylvania.

[25] Epstein, L.G. and S.E. Zin (1989), "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework", Econometrica 57, 937–969

[26] Fuhrer, J. (1996), "Monetary policy shifts and long-term interest rates," *Quarterly Journal of Economics* 111, pp. 1183-1209.

[27] Gallant, A.R., P.E. Rossi and G. Tauchen (1993), "Nonlinear dynamic structures", Econometrica 57, 357-85.

[28] Gallmeyer, M.F., B. Hollifield, F. Palomino and S.E. Zin (2007), "Arbitrage-free bond pricing with dynamic macroeconomic models", mimeo, Carnegie Mellon University.

[29] Giannoni, Marc P. and Michael Woodford (2005), "Optimal Inflation Targeting Rules," in: B.S. Bernanke and M. Woodford, eds., Inflation Targeting, University of Chicago Press.

[30] Goodfriend, M. (1991), "Interest Rates and the Conduct of Monetary Policy," *Carnegie-Rochester Conference Series on Public Policy* 34, 7–30.

[31] Guan, Yun and Jan Verschelde (2008), "PHClab: A MATLAB/Octave interface to PHCpack," in Michael E. Stillman, Nobuki Takayama, and Jan Verschelde, eds., IMA Volume 148: Software for Algebraic Geometry, pp. 15-32, Springer-Verlag.

[32] Guariglia, A. and M. Rossi (2002), "Consumption, habit formation, and precautionary saving: evidence from the British Household Panel Survey", Oxford Economic Papers 54, 1-19

[33] Gürkaynak, Refet S., Brian Sack and Eric Swanson (2005), "The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models," American Economic Review 95, 425-436

[34] Hamilton, James D. (1988), "Rational-expectations econometric analysis of changes in regime : An investigation of the term structure of interest rates", Journal of Economic Dynamics and Control 12, 385-423

[35] Hamilton, James D. (1994), Time series analysis, (Princeton University Press, Princeton).

[36] Hamilton, James D. (2008), "Macroeconomics and ARCH", NBER Working Paper No. 14151.

[37] Hördahl, P., O. Tristani and D. Vestin (2008), "The yield curve and macroeconomic dynamics", Economic Journal 118, 1937-70.

[38] Ireland, P. (1997), "A small, structural, quarterly model for monetary policy evaluation", Carnegie-Rochester Conference Series on Public Policy 47, 83-108.

[39] Judd, K. (1998), Numerical Methods in Economics, Cambridge (MA): MIT Press.

[40] Koskievic, J.-M. (1999), "An intertemporal consumption–leisure model with non-expected utility", Economics Letters 64, 285–289

[41] Kreps, D.M. and E.L. Porteus (1978), "Temporal resolution of uncertainty and dynamic choice theory", Econometrica 46, 185–200.

[42] Lombardo, G. and D. Vestin (2008), "Welfare implications of Calvo vs. Rotemberg-pricing assumptions", Economics Letters, forthcoming.

[43] Lubik, Thomas and Frank Schorfheide (2004), "Testing for Indeterminacy: An Application to U.S. Monetary Policy," American Economic Review 94, 190-217

[44] McDonnell, M. M. and Perez-Quiros, G. (2000) "Output Fluctuations in the United States: What has Changed Since the Early 1980s?", American Economic Review 90, 1464-1467

[45] A. Morgan (1987), Solving polynomial systems using continuation for engineering and scientific problems. Englewood Cliffs, N.J.: Prentice-Hall.

[46] Naik, Vasant and Moon Hoe Lee (1997), "Yield curve dynamics with discrete shifts in economic regimes: theory and estimation", Working Paper Faculty of Commerce, University of British Columbia

[47] Piazzesi, M. and M. Schneider (2006), "Equilibrium yield curves", NBER Working Paper No. 12609.

[48] Ravenna, F. and J. Seppala (2007a), "Monetary Policy, Expected Inflation and Inflation Risk Premia", mimeo, University of California Santa Cruz

[49] Ravenna, F. and J. Seppala (2007b), "Monetary Policy and Rejections of the Expectations Hypothesis", mimeo, University of California Santa Cruz

[50] Rotemberg, J.J. (1982), "Sticky Prices in the United States", Journal of Political Economy 90, 1187-1211.

[51] Rudebusch, Glenn D. (2002), "Term structure evidence on interest rate smoothing and monetary policy inertia", Journal of Monetary Economics 49, 1161-1187.

[52] Rudebusch, Glenn D., Brian Sack, and Eric Swanson (2007), “Macroeconomic Implications of Changes in the Term Premium,” Federal Reserve Bank of St. Louis, Review 89, 241–69.

[53] Rudebusch, Glenn D. and Eric Swanson (2008), "Examining the bond premium puzzle with a DSGE model", Journal of Monetary Economics 55, 111-26.

[54] Rudebusch, G.D. and E.T. Swanson (2009), "The Bond Premium in a DSGE Model with Long-Run Real and Nominal Risks", mimeo.

[55] Rudebusch, G. and T. Wu (2006), “Accounting for a Shift in Term Structure Behavior with No-Arbitrage and Macro-Finance Models,” *Journal of Money, Credit and Banking*, forthcoming.

[56] Schmitt-Grohé, S. and M. Uribe (2004), "Optimal fiscal and monetary policy under sticky prices", Journal of Economic Theory 114, 198–230

[57] Schorfeide, F. (2005), “Learning and Monetary Policy Shifts,”Review of Economic Dynamics 8, 392-419.

[58] Shiller, Robert J. (1990), "The term structure of interest rates," in B.M. Friedman and F.H. Hahn, eds., Handbook of Monetary Economics, Vol. I, Chapter 13, 627-722.

[59] Sims, Christopher A. and Tao Zha (2007), Were There Regime Switches in U.S. Monetary Policy?, American Economic Review 96, 54 - 81.

[60] Uhlig, H, (2007), “Leisure, growth, and long run risk,” mimeo.

[61] Verschelde, Jan (1999), "Algorithm 795: PHCpack: A general-purpose solver for polynomial systems by homotopy continuation," ACM Transactions on Mathematical Software 25(2): 251-276.

[62] Wachter, J.A (2006), "A consumption-based model of the term structure of interest rates," Journal of Financial Economics 79, 365-99.

[63] Weil, P. (1990), "Nonexpected Utility in Macroeconomics", Quarterly Journal of Economics 105, pp. 29-42.

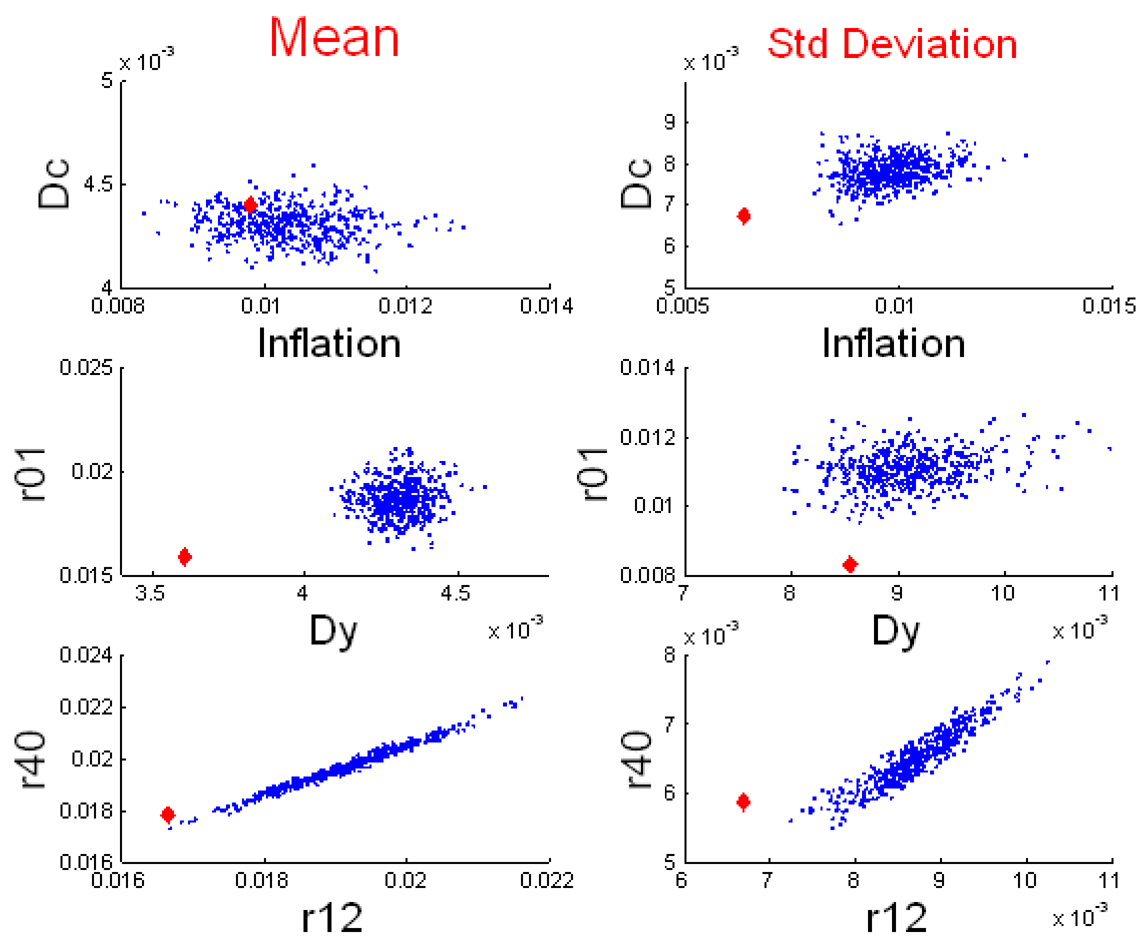
[64] Woodford, M. (2003), Interest and Prices, Princeton University Press.

Table 1: Parameter values

	post mean	**post sd**	**post low q**	**post up q**	**prior mean**	**prior sd**
$p_{\eta,11}$	0.9598	0.0159	0.9196	0.9829	0.9000	0.0650
$p_{\eta,00}$	0.9259	0.0237	0.8694	0.9612	0.9000	0.0650
$p_{z,11}$	0.8746	0.0117	0.8492	0.8918	0.9000	0.0650
$p_{z,00}$	0.9081	0.0144	0.8777	0.9324	0.9000	0.0650
$\sigma_{\eta,1}$	0.0017	0.0001	0.0015	0.0019	0.0028	0.0006
$\sigma_{\eta,0}$	0.0043	0.0005	0.0035	0.0056	0.0038	0.0010
$\sigma_{z,1}$	0.0173	0.0006	0.0162	0.0185	0.0038	0.0012
$\sigma_{z,0}$	0.0271	0.0022	0.0229	0.0313	0.0062	0.0030
σ_G	0.0347	0.0025	0.0303	0.0404	0.0305	0.0148
σ_μ	0.0791	0.0067	0.0668	0.0930	0.0119	0.0051
Π	1.0120	0.0009	1.0106	1.0138	1.0160	0.0033
Ξ	1.0043	0.0001	1.0042	1.0045	1.0030	0.0017
ρ_G	0.8558	0.0110	0.8346	0.8773	0.9000	0.0299
ρ_z	0.9812	0.0019	0.9774	0.9847	0.9000	0.0299
ρ_μ	0.8613	0.0205	0.8157	0.8970	0.7000	0.0454
ψ_π	0.3952	0.0254	0.3495	0.4444	0.5000	0.0998
ψ_y	0.0012	0.0010	−0.0008	0.0029	0.0010	0.0010
ρ_i	0.7862	0.0202	0.7480	0.8268	0.8000	0.1996
ι	0.8497	0.0488	0.7433	0.9231	0.9000	0.0901
ϕ	1.3561	0.0883	1.2072	1.5579	2.0000	0.4014
γ	3.2083	0.4523	2.3585	4.1202	11.0000	7.1268
ψ	2.0526	0.0811	1.9010	2.2120	1.5000	0.3544
ζ	7.7264	1.0208	5.6782	9.4305	10.0000	4.9945
h	0.5476	0.0071	0.5326	0.5610	0.4000	0.1480
θ	5.7413	0.5671	4.6872	6.7827	7.0000	1.9998
β	0.9991	0.0005	0.9980	0.9998	0.9975	0.0025
$\sigma_{me,20}$	0.0006	0.0001	0.0005	0.0008	0.0014	0.0010

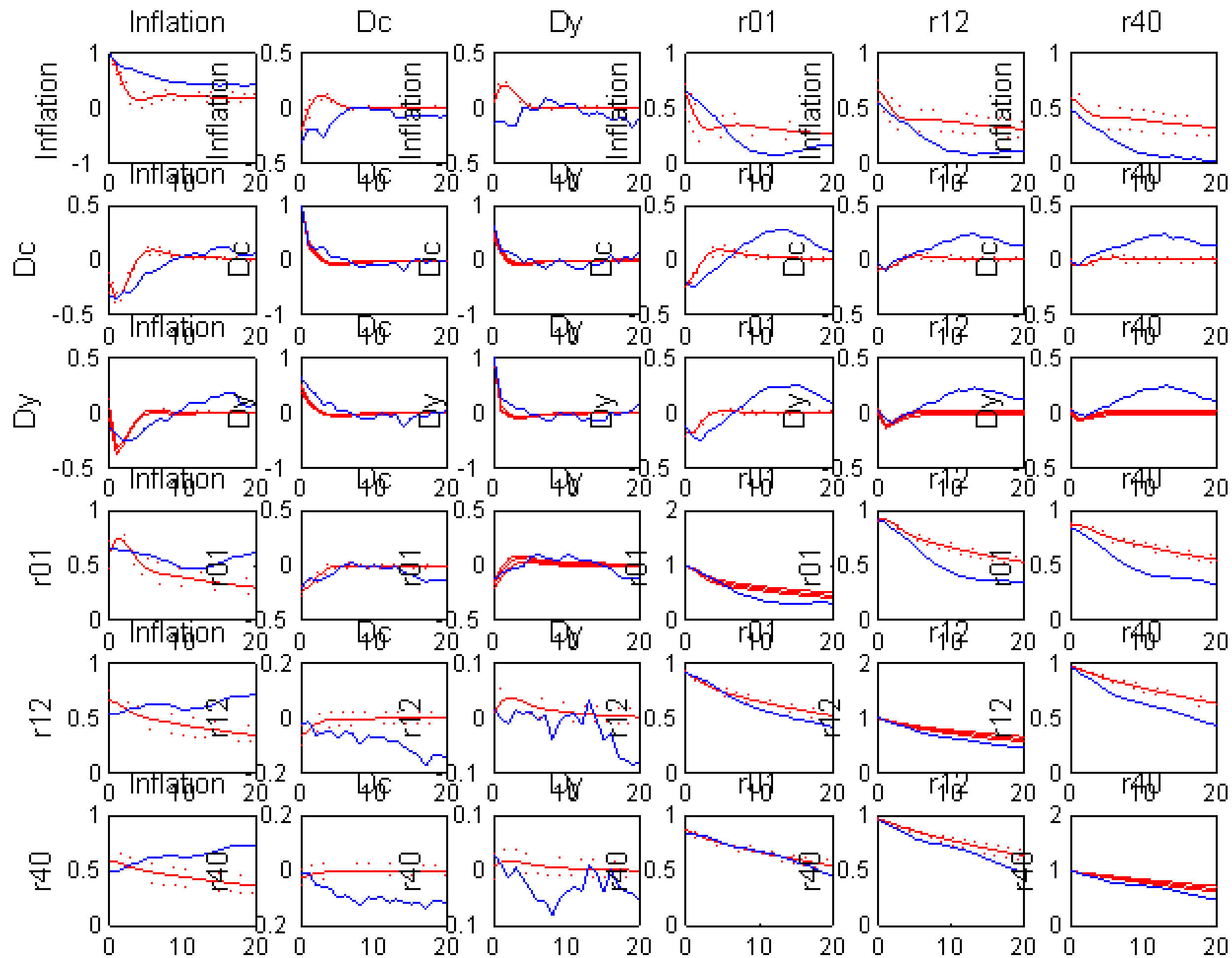
Legend: "sd" denotes the standard deviation; "low q" and "up q" denote the 5th and 95th percentiles of the distribution. Priors: beta distribution for β, h, ι, ζ, ρ_ξ, ρ_g, ρ_π; gamma distribution for ψ_π, ψ_y and all standard deviations; shifted gamma distribution (domain from 1 to ∞) for γ, ϕ, Ξ, Π^*; normal distribution for ρ_i. Posterior distributions are based on 50,000 draws.

Figure 1: Moments of observable variables: sample vs model-implied



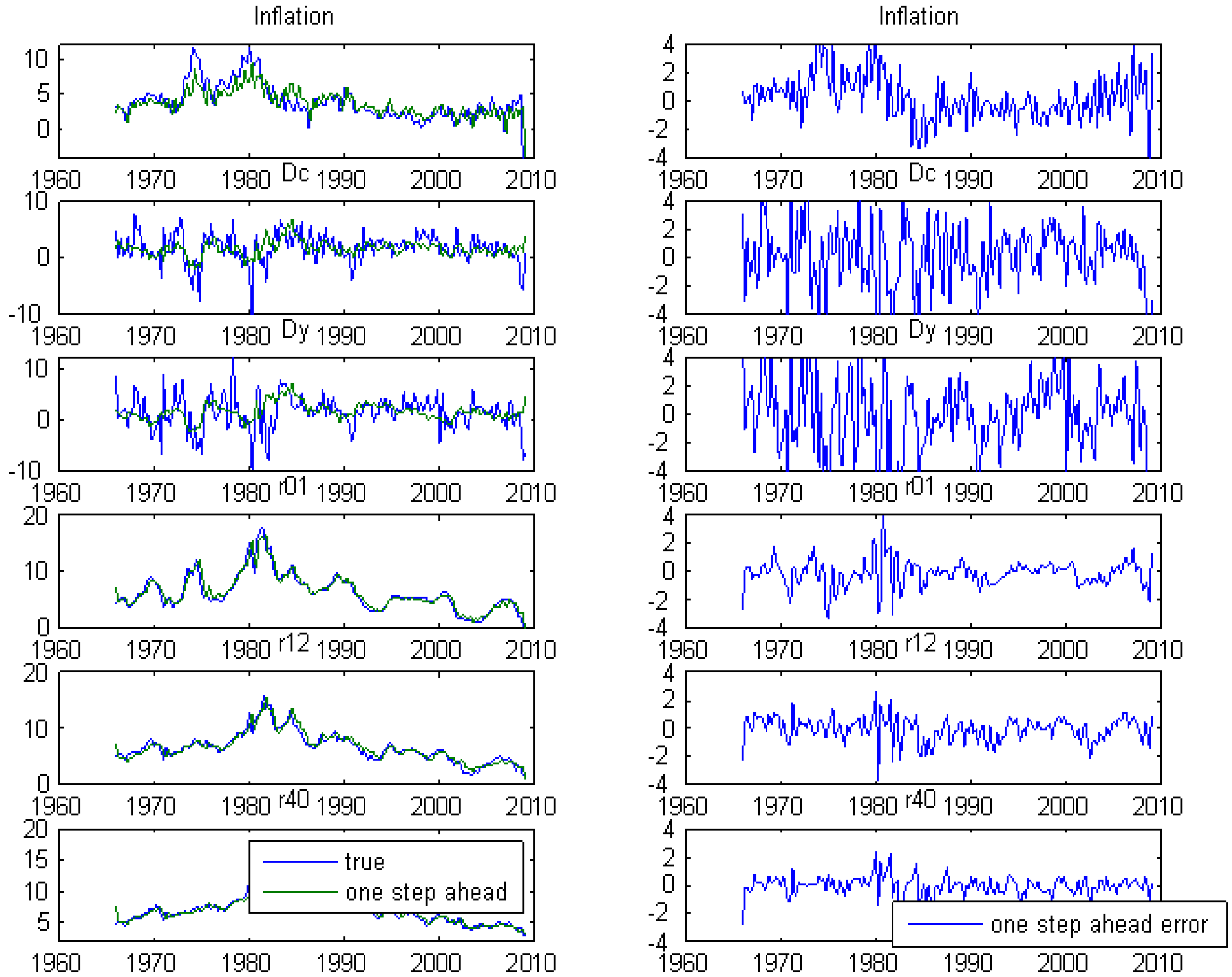
Note: The dots denote combinations of unconditional means and standard deviations implied by the posterior distribution of our parameter estimates. The diamonds indicate the corresponding sample moments.

Figure 2: Cross-correlations of observable variables: sample vs model-implied



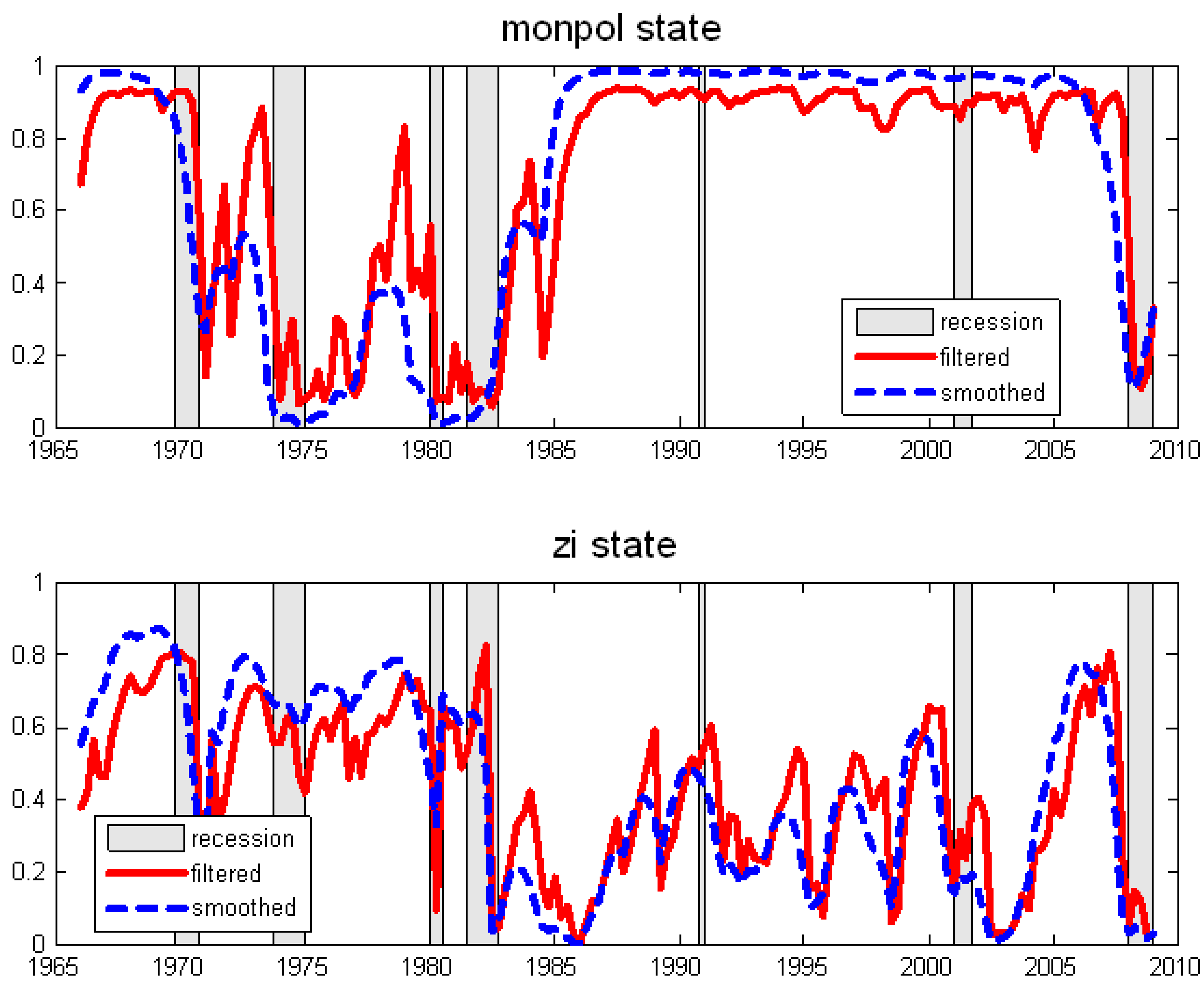
Note: The red solid lines denote sample cross-correlations. The blue lines indicate the distribution of cross- orrelations implied by the posterior distribution of our parameter estimates.

Figure 3: Actual variables and 1-step-ahead predictions



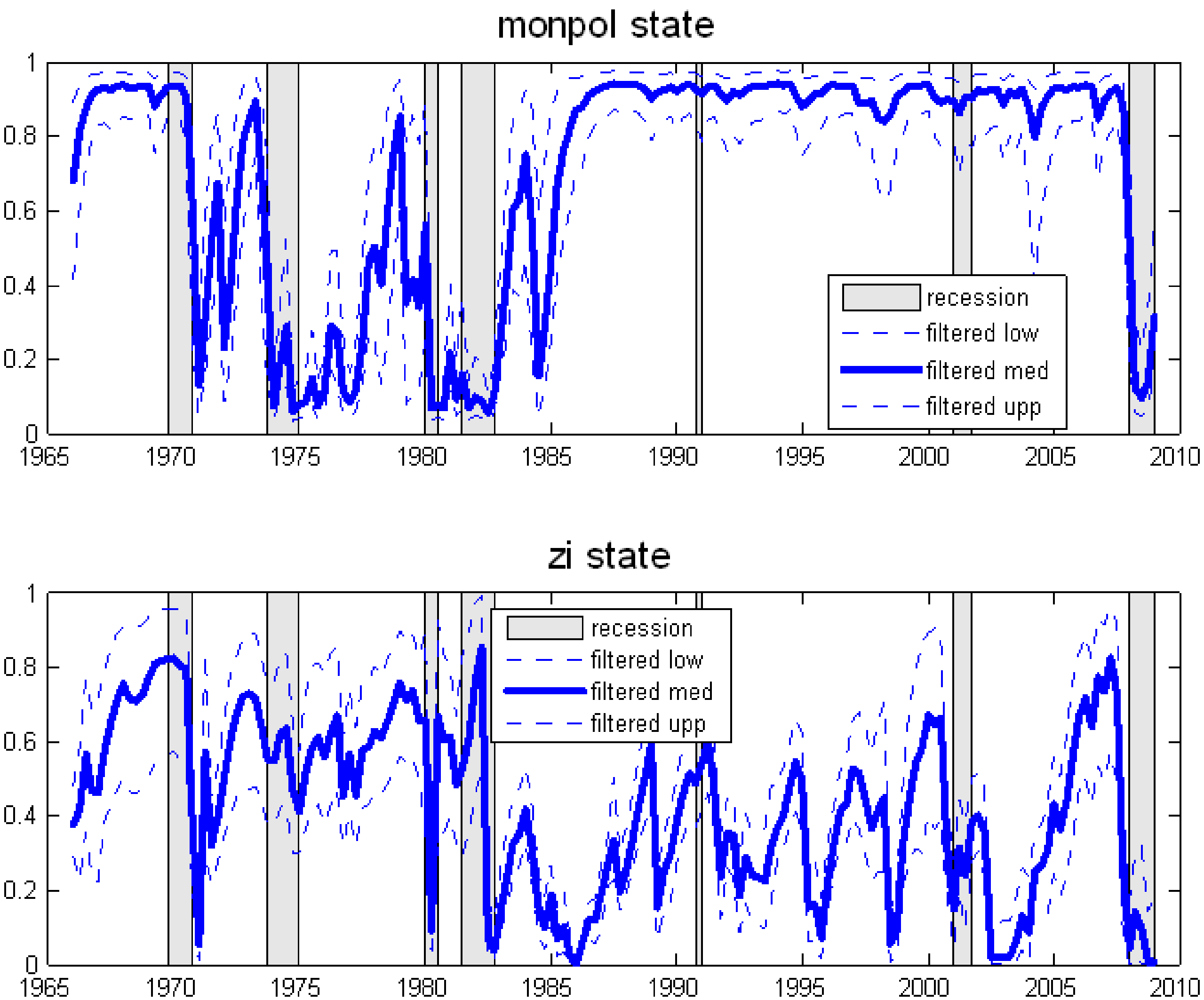
Note: all variables are measured in annual terms.

Figure 4: Filtered and smoothed estimates of the regime-variables



Note: "0" is the high-variance state.

Figure 5: Filtered estimates of the regime-variables with confidence sets



Note: "0" is the high-variance state.

Figure 6: Expected excess holding period returns with confidence sets