working paper

11 33

The Term Structure of Inflation Compensation in the Nominal Yield Curve

Mehmet Pasaogullari and Simeon Tsonev

FEDERAL RESERVE BANK OF CLEVELAND

Working Paper 11-33 December 2011

The Term Structure of Inflation Compensation in the Nominal Yield Curve

Mehmet Pasaogullari and Simeon Tsonev

We propose a DSGE model with regime switching in the central bank's inflation target to explain inflation compensation in the UK. Taking advantage of the well-documented change in UK monetary policy to adopt inflation targeting, we estimate our model using nominal and inflation-linked Treasury bond data from the UK from 1985 to 2007. We find that this model can account for the term structure of inflation compensation in the nominal yield curve by generating regime-dependent conditional expectations of future inflation.

JEL Codes: E31, E43, E52, G12.

Keywords: Term structure of interest rates, real yield curve, inflation risk premia, regime-switching DSGE model, Bayesian estimation, inflation target.

Mehmet Pasaogullari is at the Federal Reserve Bank of Cleveland; he can be reached at mehmetpasa@clev.frb.org. Simeon Tsonev is at Credit Suisse. The authors are greatly indebted to Michael Woodford, Alexei Onatski, Bruce Preston and Marc Giannoni for their support, advice and encouragement. They also thank Jon Steinsson and Ricardo Reis for helpful discussions. They greatly benefited from the suggestions of seminars participants at the Bank of England, Columbia University, the Federal Reserve Bank of Cleveland, the Federal Reserve Board of Governors, and Sabanci University.

1 Introduction

Understanding the dynamics of the term structure of interest rates is of great significance in macroeconomics. Bond prices contain important information about expectations of future monetary policy and inflation. With regards to the latter, the development of inflation-indexed bond markets in several countries in recent years has provided a useful observation of real rates and risk-adjusted inflation expectations. We make use of these data to better evaluate the ability of theoretical models to explain the link between the economy and the yield curve. In addition to its significance for the conduct of monetary policy, studying the term structure can provide insights into the asset pricing implications of macroeconomic models.

The question our paper is concerned with is what can explain the slope of the nominal yield curve. Is it the term structure of real interest rates or the term structure of inflation compensation (which itself reflects both inflation expectations and inflation risk premia)? So far there has been considerable disagreement about this issue in the literature, with most structural macro models pointing to the former and many atheoretical empirical finance models pointing to the latter explanation. Our goal is to decompose the nominal yield curve into its three constituent parts. We define inflation risk premia (henceforth, IRP's) as the difference between nominal yields and the sum of real yields and inflation expectations. In other words, IRP's measure the departure from the Fisher hypothesis of interest rate parity.

We are interested in explaining the recently observed upward slope of inflation compensation in UK data as well as determining the relative roles of real rates and inflation compensation in the dynamics of the nominal term structure. In particular, even after the adoption of inflation targeting in the UK in 1992, long term inflation compensation has been significantly higher than short term inflation compensation. To the extent that unconditional inflation expectations are constant (as one might expect in an inflation targeting setting), an upward (or downward) average slope of inflation compensation should reflect inflation risk premia rising (or falling) with maturity. Producing IRP in that magnitude within the structural models has been unsuccessful in the structural macro models.

Alternatively, we propose a different explanation of the observed term structure of inflation compensation, namely, regime switching in the conditional expectations of long-run inflation. We take advantage of the relatively long history of observed inflation-linked treasury yields in the UK, spanning a period over which there was at least one well-documented discrete change in the conduct of monetary policy in that country. The results from an estimated DSGE model with Markov switching in the central bank's inflation target indicate that the observed average slope of inflation compensation reflects long-run expectations of rather infrequent regime changes rather than risk premia. We find that the change in monetary policy can explain the observed change in the slope of inflation compensation in the UK. We also find that before the introduction of inflation targeting in 1992, the variation in the nominal slope primarily reflected the variation of the term structure of inflation compensation, whereas after that time it was substantially affected by changes in the real slope. The paper proceeds as follows. In section 2 we discuss the recent literature. In section 3 we document the empirical facts that we focus on in the paper. In section 4 we present our structural model and estimation results. Section 5 concludes.

2 Explaining the nominal slope

Our paper attempts to throw light on the decomposition of inflation compensation into an expectations part and a risk premium part. With regards to the implications for inflation risk premia, there seems to be a disagreement between the literature on New Keynesian (henceforth, NK) models on the one hand and the no-arbitrage macro-finance VAR literature on the other. Because the latter uses a very flexible specification for the stochastic discount factor that prices all assets, it has estimated significant inflation risk premia in the nominal term structure. Models with micro-founded stochastic discount factors however, either have difficulty matching this empirical finding or face serious trade-offs in their fit to other variables when they do.

Before we continue, we need to introduce precise definitions of the terms we use throughout the paper. If y_t^n is the continuously compounded yield on a n-period nominal bond and r_t^n is the yield on the corresponding real bond, then inflation compensation is simply the difference between the two: $y_t^n - r_t^n$. The IRP is that quantity less inflation expectations: $IRP_t^n = y_t^n - r_t^n - \frac{1}{n}\sum_{i=1}^{n} E_t\pi_{t+i}$, where π_{t+i} is the inflation rate in period t.

What we refer to as the "term spread" is the simple difference between long-term and short-term yields: $y_t^n - y_t^1$. Conditional on information at a certain point in time the term spread could in general depend on expectations of the future path of real short rates and inflation. Unconditionally, however, real short rates and inflation should be constant in a stationary setting and therefore the average term spread reflects only investors' preferences for holding a long-term bond over investing in short-term bonds and rolling those over a period equal to the maturity of the long-term bond. We call this quantity the "term premium". It is simply the difference between the current long-term yield and the average of expected short term rates: $y_t^n - \frac{1}{n}\sum_{i=0}^{n-1} E_t y_{t+i}^1$.

Therefore, the nominal term spread or slope can be decomposed into three components: the real slope $(r_t^n - r_t^1)$, the slope of inflation expectations $(\frac{1}{n}\sum_{i=1}^{n} E_t \pi_{t+i} - E_t \pi_{t+1})$ and the slope of inflation risk premia $(IRP_t^n - IRP_t^1)$. We now present a summary of alternative views on this decomposition.

Hordahl et al. (2007) (henceforth, HTV) construct a theoretical NK model which they try to calibrate to data in terms of its implications for the term structure and moments of macro variables. They find that the average term premia in the nominal yield curve have little to do with inflation risk and are due to real risk premia. Ravenna and Seppala (2007a, 2007b) (henceforth, RS) perform a similar calibration exercise, working with a NK model in the same class with habits, no inflation indexation, persistent technology and preference shocks and a transitory monetary policy shock. Like HTV, they find that real term premia account almost entirely for the upward average slope of the yield curve and that inflation risk premia are negligible and even negative.

Piazzesi and Schneider (2006) use an endowment-economy model with exogenous consumption and inflation processes. Their model generates significant inflation risk premia that rise with maturity. This is achieved through the combined effect of two features. First, they estimate the joint process for consumption growth and inflation using US postwar data and find that the correlation between consumption growth and lagged inflation is negative. Then, by introducing recursive utility for the representative agent, they obtain a real pricing kernel that depends negatively on revisions in expectations of future consumption growth. This way, higher inflation can have opposite effects on the yields of real and nominal bonds respectively. Feeding the estimated process for consumption and inflation through their preference structure, Piazzesi and Schneider find that the average nominal curve is upward sloping and the average real curve is downward sloping.

Rudebusch and Swanson (2008a, 2008b) have recently claimed that the relatively better success of endowment-economy models is mainly due to the fact that unlike many production-based models they use the "right" covariance between consumption and inflation by construction. On the other hand, models with endogenous production and labor supply fail to produce enough volatility in the consumption process and specifically in conditional expectations of long-run consumption growth. Therefore the long-term marginal rate of substitution of the representative agent is not volatile enough and those models do not generate substantial risk premia. Models such as Piazzesi and Schneider (2006) also employ more exotic preference specifications to translate the "right" process for consumption into the "right" process for marginal utility.

In contrast to NK models, the empirical no-arbitrage macro-finance VAR literature often finds significant inflation risk premia. In particular, Ang, Bekaert and Wei (2008) find that the US real term structure has been flat on average while the upward sloping nominal term structure has reflected high inflation risk premia that increase with maturity. They use a no-arbitrage VAR with a Markov regime switching specification for the joint dynamics of the state variables, without imposing any economic structure on the correlation of inflation and the unobservable real factors.

Their results are consistent with what one finds in studies from countries such as the UK. In particular, Evans (1998) estimates the real term structure for the UK and, using EH regressions, finds evidence for inflation risk premia. Risa (2001) also uses UK nominal and index-linked bond data and estimates a no-arbitrage VAR. He finds a variable and high on average inflation risk premium for the UK, which has however fallen over time.

Empirical evidence for sizeable and time-varying inflation risk premia was also found by Hordahl and Tristani (2007) for European data as well as by D'Amico, Kim and Wei (2008) for US data. Both papers rely on no-arbitrage macro-finance VAR's for their results and the latter also utilizes inflation indexed bond data.

Our paper also draws on previous empirical studies estimating term structure models with regime shifts. We already mentioned the empirical findings of Ang et al. (2008). From Bansal and Zhou (2002) we borrow the bond pricing methodology for our regime-switching model. Our paper is also close in spirit to Bikbov (2005), who estimates a linearized NK model with regime switching. Where we depart from his study is in our use of a micro-founded rather than an exogenous pricing kernel. In addition, we focus our attention on the model's ability to produce an upward sloping term structure on inflation compensation. To the extent that we estimate a production-based model with a Markov-switching inflation target for the central bank, our exercise resembles that of Liu et al. (2008), the main difference being that we utilize term structure data in our estimation.

3 A first look at the data

Our goal in this section is to document the salient features of the data which a good model should be expected to reproduce. We demonstrate that inflation compensation has been upward sloping on average in recent UK data. We also show that inflation compensation appears to have been relatively stable in recent years with most of the variation in the nominal slope coming from the real slope. On the other hand, it appears that variations in inflation compensation played a much larger role in determining variations in nominal bond yields in previous periods in the UK.

Establishing the "stylized facts" has been relatively straightforward for nominal yields, where previous studies have focused on replicating the average slope of the yield curve, the term structure of yield volatilities and rejections of the EH as demonstrated by OLS regressions. Data on these variables is readily available and the statistics are easily computable.

With real yields and inflation compensation, defined as the difference between nominal and real yields, agreeing on those "stylized facts" is a bit more difficult. The reason for that is that inflation compensation as defined above (also referred to by bond market participants as "breakeven inflation") contains expectations as well as risk premia. Separating the two is very difficult in the data (e.g. using inflation forecasts from surveys) and previous studies has often relied on specific modeling assumptions.

3.1 Data description

We focus on the UK data in this study, because of the availability of longer time series for real yields. We use data on UK zero-coupon yields compiled and computed by the Bank of England and available on the Bank's web site. These go back to 1985 for real yields and 1978 for nominal yields. We end the sample period at December 2007 so as to exclude the financial crisis period. The yields have been estimated from the prices of traded securities using a spline methodology and inflation-linked yields have been adjusted for the indexation lag and for seasonality.[1]

For the estimation of our DSGE model for the UK we use quarterly data on Retail Price Index (henceforth, RPI) inflation and household consumption take from the Office

[1] For more details on the methodology, please refer to Anderson and Sleath (2001). They report that their method produces very stable yield curves in the sense that they show little sensitivity to measurement error in the prices of the underlying bonds.

of National Statistics (ONS). We motivate the choice of the RPI for measuring the price level by the fact that this is the index which UK inflation protected bonds are linked to. Because the original series is not seasonally adjusted we follow Risa (2001) and adjust it using the X-12 method.

3.2 Yield curve slope

Table 1 shows the means of the yield levels in the UK for different periods. Clearly, the levels of real and nominal yields and breakeven inflation all decreased in the recent subsample. However, we are less interested in the levels, which might reflect secular trends in inflation, than in the differences between yields at long and short maturities. The slopes of the different yield curves contain a lot more information about expectations and risk premia.

We focus on the slope between 10-year and 2-year yields because inflation-indexed yields with shorter maturities suffer from measurement error issues and are not reliably estimated. Securities with less than 2 years to maturity have been excluded from the fitting procedure due to their erratic behavior resulting from the effects of seasonality and indexation lag. In particular, prices for bonds with little time remaining to maturity become more sensitive to inflation accretion and short term inflation expectations can be extremely volatile.

Figure 1 presents the data for our two subsamples - before and after October 1992. We can see that in different periods, characterized by different types of monetary policy, the nominal curve slope has been driven by different components. In particular, the slope of inflation compensation has been a lot more stable in recent times. A comparison between the first two panels shows that the movements of the nominal slope have been more correlated with the movements in the breakeven inflation slope in the first subsample and more correlated with movements in the real curve slope in the second. The correlations in the first subsample were 0.63 between the nominal and breakeven slope and 0.26 between the nominal and real slope. In the second subsample these figures stood at 0.39 each. If we restrict the second subsample to only the period after the Bank of England acquired operational independence in 1997, the correlations become 0.27 and 0.49 respectively, thus reversing their relative magnitude.

That being said, the last panel of Figure 1 indicates that while more stable, inflation compensation has been upward sloping in recent times. The difference between 2-year and 10-year breakeven inflation has averaged about 40 basis points. Our main goal in this paper is to explain this fact given the Bank of England's commitment to a low inflation target.

One can see in the third panel of Figure 1 that before October 1992 the breakeven inflation slope (as well as the nominal curve slope) has been negative on average. This has partly preceded and partly coincided with a period of steady decline in UK inflation rates and nominal interest rates. As the figure shows, UK long-term interest rates decreased gradually throughout the 1990's.

Therefore it seems reasonable to hypothesize that the slope of breakeven inflation in our first subsample has reflected expectations of future inflation rates rather than risk premia. By the same token it could be the case that the slope of breakeven inflation in the second subsample also has reflected certain expectations about inflation. The UK evidence may suggest a shift from a period with relatively high inflation and expectations of decreasing inflation to a period characterized by low inflation and expectations of increasing inflation. This would have dramatically steepened the slope of the inflation compensation term structure from possibly very negative to positive.

4 DSGE Model Estimation

4.1 Markov regime switching in the inflation target

In this paper, we propose a mechanism whereby the average term structure of inflation expectations can have a positive slope for a prolonged period of time. This can be achieved through a level of inflation that alternates between different regimes with relatively low frequency. This way, in each regime the current level of inflation is below or above the long-term expectations of inflation, which eventually converge to the ergodic mean of the regime-switching process. Hence, the observed slope of inflation compensation depends on the level of inflation in each regime and on the probabilities of moving from one regime to another.

We have already seen that the behavior of bond yields has been quite different over different periods in time. It could be argued these periods coincide with changes in the behavior of macroeconomic variables. For example, many authors document the so called "Great Moderation" in the US, while still disagreeing on its ultimate causes – improved monetary policy or reduced exogenous volatility. Other papers, such as Schorfheide (2005), make a case that there have been different regimes for the Fed's inflation target. The work of Bekaert et al. (2001) and Bansal and Zhou (2002) also suggest that regime switching can be useful in explaining the historical evolution of yields and bond returns.

All of this serves as motivation to explore the role of Markov regime switching in accounting for the term structure. We explore this issue in the context of a DSGE model, because we wish to take account of the way in which the dynamics of real variables as well as inflation should be different as a consequence of the switches between policy regimes (and awareness of the possibility of switches), and we need a structural model in order to analyze this. We estimate our model using UK data because it provides us with a good case study. Besides the availability of data on real yields, it is characterized by a well defined break in the way monetary policy was conducted - the adoption of inflation targeting.

After the "Black Wednesday" of 16 September 1992, on 8 October that year the UK government announced it was leaving the European ERM mechanism and that it would adopt an inflation target in a wide band between 1% and 4% annually, based on the retail price index excluding mortgage interest payments (RPIX). The stated objective was to have inflation below the mid-point of this range (2.5%) by the end of the 1992-1997 parliament. On 14 June 1995 the Chancellor of the Exchequer announced an inflation target of 2.5% or less. In May 1997 the Bank of England was given operational independence in achieving the government's inflation objective, which was itself amended to a symmetrical target of 2.5% annual RPIX inflation. Finally, in April 2003 the targeted price level measure was changed to the harmonized index of consumer prices (HICP or CPI) and on 12 December that year the target was lowered to 2%. One might conjecture that more than of these events could correspond to a change in policy regime. In our empirical work, we do not pre-judge the timing of any regime changes, and instead estimate the likelihood that a regime change has occurred at any date.

4.2 Model Description

The model we consider consists of three building blocks - a representative consumer who maximizes utility and supplies labor, a private sector with monopolistically competitive firms and a central bank that sets monetary policy. We focus on the most basic version of the sticky price model discussed in Woodford (2003), abstracting from investment, fiscal policy and a number of other features such as certain types of nominal and real frictions.

4.2.1 Consumers

The representative household has a time-separable utility function with internal habits and additively separable disutility of labor.

$$U = \max E_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{\left(C_t - \delta C_{t-1}\right)^{1-\sigma}}{1-\sigma} \xi_t - \chi L_t^{\omega} \right] \tag{1}$$

Its preferences over the consumption stream are characterized by the subjective discount factor β, the risk aversion parameter σ and the internal habit persistence parameter δ. We also tried using external habits and found that this does not materially change the results. The curvature of the disutility of labor is determined by the ω parameter. Here E_t denotes the expectation operator conditional on information up to time t. There is also a preference shock to the consumer's marginal utility. It follows an autoregressive process in logs with a zero mean.

$$\log \xi_{t+1} = \rho_\xi \log \xi_t + \eta^\xi \varepsilon_{t+1}^\xi \tag{2}$$

Here and afterwards ρ_i and η^i denote the first-order autocorrelation and conditional volatility of the exogenous process i respectively. ε_t^i is an i.i.d. zero-mean shock.

Consumers maximize utility subject to a budget constraint whereby the nominal value of their consumption and investments in assets must be less than or equal to their current labor income and asset returns from previous periods. We assume households have access to a set of securities spanning all possible contingencies and including shares of all firms in the economy.

4.2.2 Firms

There is a continuum of producers indexed by i, each producing a differentiated good. All producers have access to the same linear technology where output is given by

$$Y_t\left(i\right) = A_t L_t\left(i\right) \tag{3}$$

where $L_t\left(i\right)$ is the labor demand of firm i. Since the equilibrium wage is the same for all sectors and firms, we implicitly assume that the labor supply of each type i is determined by the labor demand so that the market is cleared. The level of technology is stationary and follows an AR(1) process.

$$\log A_{t+1} = \rho_A \log A_t + \eta^a \varepsilon_{t+1}^a \tag{4}$$

In order to have a non-trivial role for monetary policy, we assume that firms follow Calvo (1983) pricing with a probability of not optimizing prices in a given period equal to α. In addition, we allow firms who do not optimize prices to index them to lagged inflation

$$P_t(i) = P_{t-1}(i)\,\Pi_{t-1}^{\gamma} \tag{5}$$

γ is the parameter governing the degree of indexation and can take values from 0 to 1. Firms maximize the discounted stream of expected nominal profits, valuing future cash flows with the representative consumer's stochastic discount factor. The resulting first order conditions of the firms' and consumers' maximization problem are used to determine the equilibrium in the economy and solve for the dynamics.

4.2.3 Monetary policy

Finally, to close the model, we specify monetary policy as follows. The nominal short interest rate is set by a central bank that follows a Taylor rule with monetary policy inertia where ρ_I is the weight on the lagged interest rate.

$$I_t = \left[\frac{\bar{\Pi}}{\beta}\left(\frac{\Pi_t}{\Pi_t^*}\right)^{\phi_\pi}\left(\kappa\frac{Y_t}{Y_t^n}\right)^{\phi_y}\right]^{1-\rho_I} I_{t-1}^{\rho_I}\upsilon_t \tag{6}$$

ϕ_π and ϕ_y are parameters governing the systematic response of the bank to deviations of inflation and output from the target Π_t^* and the natural rate of output Y_t^n respectively. Y_t^n is defined as equilibrium output in an economy where $\alpha = 0$. κ is the inverse of the steady-state output gap, which depends on the steady-state level of inflation. υ_t is a transitory monetary policy shock:

$$\log \upsilon_t = \eta^m \varepsilon_t^m \tag{7}$$

The value of the target $\Pi_t^* = \Pi^*(s_t)$ depends on the current state of an unobserved discrete-valued S-state Markov switching variable s_t ($s_t = 1, 2, ..S$). We assume that regime changes are governed by a Markov chain with a transition matrix P, whose element $p_{ij} = \mathbb{P}(s_t = i|\, s_{t-1} = j)$ is the probability of moving to regime i given that the current state is j such that $\sum_{i=1}^{S} p_{ij} = 1$ for all j. In our estimation we allow for two regimes in the inflation target of the central bank and set $S = 2$.

4.3 Model solution and bond price computation

Detailed derivations of the first order conditions for the model can be found in Appendix A. We log-linearize them around the non-stochastic steady state of the model, where we set inflation equal to its ergodic mean. In its log-linear form, the model can be written as

$$BX_t = M(s_t) + AE_t X_{t+1} + CX_{t-1} + D\varepsilon_t \tag{8}$$

where X_t is a vector containing the model's endogenous variables. Besides the Markov chain shock that determines the regime every period we have three additional shocks, $\varepsilon_t^a, \varepsilon_t^m$ and ε_t^ξ, which are collected in the vector ε_t. As shown in Appendix B.1, the solution to our model is a regime-switching VAR of the form

$$X_{t+1} = \mu(s_{t+1}) + \Phi X_t + \Sigma\varepsilon_{t+1} \tag{9}$$

where the intercept term depends on the regime probabilities p_{ij} as well as the regime-dependent inflation target $\pi^*(s_t)$.

Because there is no regime switching in the A, B and C coefficient matrices we can first apply standard linear rational expectations methods to solve for the transition matrix Φ and then use the set of equations

$$[B - A\Phi]\,\mu(i) = M(i) + A{\textstyle\sum_j} p_{ij}\mu(j)$$

to obtain $\mu(s_{t+1})$.

Appendix B.2 provides details on the bond price computation algorithm under regime switching. It is shown by Ang et al. (2008) that when Φ is not regime-dependent, bond prices are computed exactly. In the case when Φ does depend on regimes, bond prices can be computed approximately (with a very small approximation error) as shown by Bansal and Zhou (2002). Finally, using the model solution and bond pricing method we can compute the expectations and variances of yields and inflation conditional on different regimes. Details of the conditional moment computations can be found in Appendix B.3.

The resulting equations for the real and nominal bond yields with n quarters to maturity are affine in X_t and with regime dependent intercept terms:

$$\begin{aligned} y_t^n &= a_n^N(s_t) + b_n^N X_t \\ r_t^n &= a_n^R(s_t) + b_n^R X_t \end{aligned}$$

4.4 Estimation Methodology

We estimate our model using Bayesian techniques, which allow us to specify priors for the distribution of certain parameters, based on previous studies and on certain data moments. We use the Hamilton-Kim filter for computing the likelihood as well for finding the probabilities of the two regimes.[2] This allows us to compute the posterior distribution of the model parameters up to an integrating constant. Here, we show the posterior mode estimates that maximizes the posterior likelihood of the model.

4.4.1 Likelihood computation

In our estimation we use eight UK time series. These are per capita consumption growth, RPI inflation, nominal bond yields with maturities of 1, 2, 5 and 10 years and real bond yields with maturities of 5 and 10 years. The data, represented by the vector Y_t, is related to the state variables X_t by the observation equation

$$Y_t = A\left(s_t\right) + BX_t + \sigma_e e_t \tag{10}$$

The intercept term $A(s_t)$ contains the sample means of inflation and consumption growth as well as regime-dependent intercept terms for bond yields. The coefficients in B contain the bond pricing coefficients b_n^N and b_n^R as well as the identity mapping between the model variables Δy_t and π_t and their data counterparts: demeaned per capita consumption growth and RPI inflation.

We allow for measurement errors in the observation equation represented by the i.i.d. vector e_t. We assume that the random vectors e_t and ε_t are jointly normal

$$\begin{pmatrix} e_t \\ \varepsilon_t \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}\right)$$

Thus we have a state space system of the form

$$\begin{aligned} Y_t &= A\left(s_t\right) + BX_t + \sigma_e e_t \\ X_{t+1} &= \mu\left(s_{t+1}\right) + \Phi X_t + \Sigma\varepsilon_{t+1} \end{aligned}$$

The matrices $A\left(s_t\right)$, B, $\mu\left(s_t\right)$, Φ and Σ all depend on the model's structural parameters, which we collect in the vector $\tilde{\Theta}$. In our case

$$\tilde{\Theta} = \left[\beta, \sigma, \delta, \chi, \omega, \alpha, \theta, \gamma, \phi_\pi, \phi_y, \rho_i, \rho_a, \rho_\xi, \eta^a, \eta^m, \eta^\xi, \pi\left(1\right), \pi\left(2\right), p_{11}, p_{22}, \sigma_e\right]'$$

Let Y^t denote the data up to and including period t: $Y^t = (Y_0, Y_1, ..., Y_t)$. Our goal is to compute the likelihood of the data given the parameters

$$\mathcal{L}\left(Y^T \middle| \tilde{\Theta}\right) = \prod_{t=2}^{T} \mathcal{L}\left(Y_t \middle| Y_{t-1}, \tilde{\Theta}\right) \mathcal{L}\left(Y_1 \middle| \tilde{\Theta}\right)$$

[2] Please refer to Kim(1994) and Hamilton (1989) for details. Our exposition of the methodology follows Kim and Nelson (1999).

where Y_1 is a function of X_1 and s_1, which are assumed drawn from their ergodic distributions.

To compute $\mathcal{L}\left(Y_t|\, Y_{t-1}, \tilde{\Theta}\right)$ we need to make an inference about the unobserved states X_t. We do that by using the algorithms proposed by Hamilton (1989) and Kim (1994). Appendix B.4 provides the details.

4.4.2 Estimating the posterior mode

Given a prior distribution $p(\Theta)$ and the likelihood $\mathcal{L}\left(Y^T\middle|\, \tilde{\Theta}\right)$, we can compute the value of the posterior at a given set of parameters up to an integrating constant

$$p\left(\Theta|\, Y_t\right) \propto \mathcal{L}\left(Y^T\middle|\, \tilde{\Theta}\right) p\left(\Theta\right) \tag{11}$$

As discussed in Liu et al. (2008), since the posterior density function is in general non-Gaussian and with a complicated shape, it is important to find the its mode by maximizing (11). This we do as follows.

We first make 1,000,000 draws from the prior distribution and estimate the value of posterior for each one of them. Then we use the fifty with the highest posterior values as initial parameter vectors for a sequence of non-linear optimizations.[3] Our results are based on the highest local maximum of the posterior that we find.

4.4.3 Priors

The dimensionality of $\tilde{\Theta}$ is large, with 21 different parameters. Therefore we split it into two parts: Θ, which we estimate, and $\bar{\Theta}$, which we fix a priori. The elements of $\bar{\Theta}$ were determined as follows. We set the discount factor $\beta = 0.995$, which corresponds to a 2% steady-state real short rate. We set the CES parameter $\theta = 10$ corresponding to a mark-up of 11%, which is in the neighborhood of the one estimated by previous DSGE models. We set the Calvo parameter α to 0.75 and the Taylor rule's output gap response coefficient ϕ_y to 0.1 (corresponding to a 0.4 coefficient for the annualized output gap).

For the remaining sixteen parameters, we give priors that are in line with earlier studies employing Bayesian methods for estimating DSGE models. The prior distributions of those parameters are summarized in the first three columns of Table 2.

[3]We used the optimisation routines available on Chris Sims's web site as well as the fminsearch function from MATLAB's Optimization Toolbox, which is based on the Nelder-Mead search method.

The prior distribution of the utility curvature parameter σ is Gamma with a mean of 3 and a standard deviation of 1. We allow for a higher value than usual for the utility function curvature parameter in order to improve the fit to bond yields.

The inverse Frisch elasticity ω has a mean of 3 and a standard deviation of 0.75 as in Rabanal and Rubio-Ramirez (2005). The $5\% - 95\%$ bounds for the prior distribution of ω correspond to a Frisch elasticity between 0.123 and 0.53.

We allow for a rather loose prior for the monetary policy inertia parameter in the Taylor rule ρ_I with mean of 0.7 and standard deviation 0.2. The $5\% - 95\%$ bounds for the prior distribution are 0.32 and 0.97 respectively.

The prior for the AR(1) coefficient for the technology process mean value of 0.9 and standard deviation of 0.05. We also tried looser priors for this parameter but the estimates did not change significantly. The prior for the preference shock persistence has a mean of 0.6 and standard deviation of 0.2 as in Justiniano and Primicieri (2007).

We use an Inverse Gamma (IG) distribution for all volatility parameters to guarantee positivity. Our prior for the volatility of the technology shock has a mean value of 2.5% and a standard error of 1%. The mean level is on the higher end of previous studies but we have to compensate for the fact that our model lacks other important sources of volatility such as capital adjustment or variable capacity utilization, which other studies find to add to the volatility of the output and consumption.

For the monetary policy shock we chose a mean of 0.05% implying that the central bank cannot deviate by more than 0.4% from its target interest rate 95% of the time.

The mean and standard deviation of the volatility of the measurement error were set to 10 basis points each.

We set the mean values for the inflation target in each regime close to the mean values for RPI inflation in the pre and post ERM crisis subsamples. The standard deviation for those parameters were set to 0.2% and 0.1% respectively. The low standard deviation for the high inflation target is to let the estimation easily distinguish between the low and high inflation regimes for a given period.

Finally, we set the mean of the probability of a transition from a low to a low inflation regime to 0.9, corresponding to an average duration of 10 quarters. We set a very loose prior for the transition probability from a high to a high inflation regime in order to give the data enough flexibility to distinguish between the low and high regimes. Our prior mean for p_{22} is 0.6 corresponding to an average duration of 2.5 quarters.

4.5 Estimation results

4.5.1 Parameter estimates

We present the medians and standard deviations of the estimated posterior distributions in Table 2. In general the data is quite informative for the estimated parameters. Given the relatively flat UK term structure, the estimated median for σ of 2.58 is much lower than the one used by HTV to fit US term premia. The estimated inflation target in the high regime (regime 2) is 5.37% , which is close to the 5.48% mean of the first subsample. The median of the inflation target in regime 1 is estimated at 2.22% which is lower than the post-1992 sample mean of inflation of 2.71% but in between the actual announced inflation targets of the Bank of England of 2.5% and 2%.

The transition probabilities p_{11} and p_{22} have estimated medians of 0.98 and 0.95 respectively. This implies that the low-inflation regime is more persistent. Despite its very high persistence it is not an absorbing state. Therefore, even under the current regime of inflation targeting there is a non-negligible implied probability of reverting to a high inflation regime. This has a significant impact on the shape of the inflation compensation term structure.

4.5.2 Model fit

The estimated parameter values are of less interest themselves than their implications for the model variables behavior. Table 3 gives the correlations between the observed values of the variables and estimated values of the variables when the measurement errors that are used in estimation are set to zero. Overall, the fit of the estimated variables are quite good. However, it has to be underlined that estimating the real yields and longer maturity nominal yields precisely proved to be harder. This is not surprising given that the model has itself a short-term nominal variable (the policy rate set by the Central Bank) and the high correlation between the shorter term yields. Still, the correlation values suggest that the model is able to explain the movement of the breakeven rates, which are actually not used in the estimation.

In three panels of the Figure 2 (the fourth is on the lower right panel), we see the fit of the selected variables again when the measurement error terms are set to zero. It can be seen that the model has a very good fit to observed consumption growth, inflation and 1-year nominal yield. This is not surprising given the large dimensionality of the parameter vector.

The last panel of Figure 2 shows the estimated regime probability of low-inflation regime. The transition between the high and low regimes is very clear. Our model estimation suggests an inflationary target switch happened at the end of 1991. This seems to be at odds with the fact that the Bank of England adopted inflation targeting regime in 1992:Q4. To shed into light for this result, in Figures 3 and 4, we draw the observed inflation rate, observed breakeven inflation spread respectively (defined as the 10year minus 5 year breakeven rate) along with the model implied values. It has to be noted that our identification for the regime switches for the inflation target basically exploits the inflation rate and the breakeven inflation spread. Given the low (high) inflation target regime, inflation rate will be lower (higher) and because of higher (lower) future expected inflation breakeven inflation rate will be higher (lower).

Although the Bank of England adopted the inflation targeting in 1992:Q4, Figure 3 shows that the inflation dropped significantly between 1990:Q2 and 1991:Q3 over 9.6 percent. Hence, the model gives a higher probability of being in the lower inflation regime starting from the end of 1991 given this significant drop in inflation. As we can see form Figures 2 to 4 that there has been a previous short-lived regime switch from high inflation regime to low inflation regime 1986:Q1. In addition, the estimation suggests another short-lived switch, this time from low-inflation regime to high, in 1994:Q4. A look at Figure4 shows that the break-even inflation rates are the major culprit for these short-lived switches as the observed breakeven rate spread nearly coincides with the model-implied one.

Figure 5 shows the fit to variables that are functions of the variables used in the observation equation. The fit to the breakeven inflation rates is quite good whereas the fit to the nominal and real slope terms is somewhat worse but they still capture the trends in the data.

4.5.3 Implications for breakeven inflation

We now compare our results with the observed data moments for the nominal, real and breakeven inflation term structures. Figure 6 depicts the subsample averages predicted values of the term structure for the pre- and post 1992:Q4, where the BOE adopted inflation targeting. Although, we found that the model shows an earlier date for the regime switch, we still want to use 1992:Q4 to compare the predicted values from estimation to the observed values drawn in Figure 1. As earlier, we use the 2 year maturity as the level term of the term structure and the spreads are computed with this maturity. The only bit where we differ from the data is the slightly positive slope of the estimated real term structure in the second sample. This could possibly be a result of the institutional features such as the regulations that require UK pension funds to invest in inflation indexed assets.

The bottom two panels of Figure 6 show the model implied unconditional average shape of the term structure in both regimes. While they are slightly different in magnitude (especially for the breakeven inflation slope in the high inflation regime) from the sample means, they do have the same general shape. Hence we can conclude that our model explains the observed term structure of breakeven inflation fairly well. It produces a breakeven slope in the second subsample that is very close to the one observed in data and generates and even higher slope for the low inflation regime in the population.

As noted earlier, pre- and post 1992:Q4 contain regime switches. To show that the predicted values of term structures are not driven by these switches but rather they conform to model implied unconditional term structures, we also subdivide the top panel according to the predicted regime probabilities in Figure 7. As the figure conforms, the similar patterns are mainly dictated by the fact that a high (low) inflation regime has a downward (upward) sloping breakeven inflation curve.

If we look at Figure 8 we can see the implied IRP's (which are the same in both regimes). Note that, given our model structure, unconditional IRP's do not differ across regimes. They can account for only a tiny fraction of the inflation compensation slope. The highest value for the IRP is slightly higher than 5 basis points. This is not surprising given other structural macro model results. The effect of regime-dependent conditional expectations clearly dominates.

Finally, Figure 9 shows the comovement of the nominal, real and breakeven 10y-2y slope in the two subsamples. We can see that in the first subsample the fitted nominal slope is much more correlated with breakeven inflation. The correlation between the two is 0.82 whereas the correlation between the nominal and real slope is only 0.29. In the second subsample the nominal-real correlation increases to 0.72 just like it does in the data, whereas the nominal-breakeven correlation decreases to 0.66. These correlations are higher as the ones in the data but the direction of the changes is very similar. We conclude that the relative role of the real slope in the dynamics of the nominal curve increased whereas that of inflation compensation decreased. Thus our model agrees with our preliminary observations of the data.

5 Conclusions

In this paper we investigate the term structure of inflation compensation and its contribution to the nominal curve slope. We attempt to answer the question of why inflation compensation in the UK has been upward sloping during the time the Bank of England has been committed to a low and stable inflation target and propose a mechanism. In particular, we estimate a DSGE model with Markov regime-switching in the inflation target of the central bank and find that the period before the adoption of inflation targeting is basically characterized by a high inflation target regime whereas the period after is mostly a low-inflation target regime. However, the model assign the regime switch date before 1992:Q4 mostly because of a rapid disinflation in the previous year. The model also matches the breakeven inflation slope observed in different subsamples of UK data. Thus our results attribute the slope of the UK breakeven inflation to the conditional expectations of long-run inflation which changed substantially with the transition to a new monetary policy regime.

Our model is a rational expectations model, where economic agents know the structure of the economy as well as the statistical distributions of the exogenous shock including the process for the regime-switching inflation target. Therefore the term structure of inflation compensation is primarily determined by expectations. However, we have not ruled out the existence of larger risk premia than our simple model suggests. For example, one could incorporate risk premia via different mechanisms involving the introduction of a wedge between the beliefs of agents and the actual distribution of the underlying process for inflation.

Our regime-switching model can also be extended to allow for subjective transition probabilities, different from the ones governing the actual inflation target process. In this extended model breakeven inflation could contain a significant component that is not related to inflation expectations under the objective physical measure without affecting the real term structure to a large extent.

References

[1] Anderson, N., and J. Sleath (2001): "New estimates of the UK real and nominal yield curves," Manuscript, Bank of England.

[2] Ang, A., G. Bekaert, and M. Wei (2008): "The Term Structure of Real Rates and Expected Inflation," *Journal of Finance*, 63, 797-849.

[3] Backus, D., and J. Wright (2007): "Cracking the Conundrum," Manuscript, New York University.

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

[5] Bansal and Zhou (2004): "Regime Shifts, Risk Premiums in the Term Structure, and the Business Cycle," *Journal of Business and Economic Statistics*, 22, 396-409.

[6] Beechey, M., and J. Wright (2008): "The High-Frequency Impact of News on Long-Term Yields and Forward Rates: Is It Real?" Manuscript, Federal Reserve Board of Governors.

[7] Bekaert, G., R. Hodrick, and D. Marshall (2001). "Peso problem explanations for term structure anomalies," *Journal of Monetary Economics*, 48, 241-270.

[8] Bikbov, R. (2005): "Monetary Policy Regimes and The Term Structure of Interest Rates," Manuscript, Columbia University.

[9] Calvo, G. A. (1983): "Staggered Prices in a Utility maximizing Framework," *Journal of Monetary Economics*, 12, 983-98.

[10] D'Amico, S., D. Kim, and M. Wei (2008): "Tips from TIPS: the informational content of Treasury Inflation-Protected Security prices," Manuscript, Board of Governors of The Federal Reserve.

[11] Evans, M.D. (1998): "Real Rates, Expected Inflation, and Inflation Risk Premia," *Journal of Finance*, 53, 187-218.

[12] Gurkaynak, R., A. Levin and E. T. Swanson (2006): "Does inflation targeting anchor long-run inflation expectations? Evidence from long-term bond yields in the U.S., U.K., and Sweden," Manuscript, Federal Reserve Bank of San Francisco.

[13] Gurkaynak, R., B. Sack and J. Wright (2007): "The U.S. Treasury yield curve: 1961 to the present," *Journal of Monetary Economics*, 54, 2291-2304.

[14] Gurkaynak, R., B. Sack and J. Wright (2008): "The TIPS yield curve and inflation compensation," Manuscript, Federal Reserve Board of Governors.

[15] Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357-84.

[16] Hordahl, P., O. Tristani, and D. Vestin (2008): "The Yield Curve and Macroeconomic Dynamics," Economic Journal, 118, 1937-1970.

[17] Hordahl, P., and O. Tristani (2007): "Inflation Risk Premia in the Term Structure of Interest Rates," Manuscript, European Central Bank.

[18] Justiniano, A., and G. Primicieri (2007): "The Time Varying Volatility of Macroeconomic Fluctuations," Manuscript, Northwestern University.

[19] Kim, C. (1994): "Dynamic Linear Models with Markov Switching," *Journal of Econometrics*, 60, 1-22.

[20] Kim C., and C. R. Nelson (1999): *State-Space Models with Regime Switching*, Cambridge: MIT Press.

[21] Liu, Z., D. F. Waggoner and T. Zha (2008): "Has the Federal Reserve's Inflation Target Changed?,".Manuscript, Federal Reserve Bank of Atlanta.

[22] McCulloch, J. H., and H. Qwon (1993): "U.S. Term Structure Data, 1947-1991," Manuscript, Ohio State University.

[23] Nelson, C. R., and A. F. Siegel (1987): "Parsimonious Modeling Of Yield Curves," *Journal of Business*, 60, 473–489.

[24] Piazzesi, M., and M. Schneider (2007): "Equilibrium Yield Curves," NBER Macroannual 2006, 389-442.

[25] Rabanal, P. and J. F. Rubio-Ramirez (2005): "Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach," *Journal of Monetary Economics*, 52, 1151-1166.

[26] Ravenna, F., and J. I. Seppala (2007a): "Monetary Policy and Rejections of Expectations Hypothesis," Manuscript, University of California Santa Cruz.

[27] Ravenna, F., and J. I. Seppala (2007b): "Monetary Policy, Expected Inflation and Inflation Risk Premia," Manuscript, University of California Santa Cruz.

[28] Risa, S. (2001): "Nominal and Inflation Indexed Yields: Separating Expected Inflation and Inflation Risk Premia," Manuscript, Columbia University.

[29] Rudebusch, G. D., E. T. Swanson, and T. Wu (2006): "The bond yield "conundrum" from a macro-finance perspective," Manuscript, Federal Reserve Bank of San Francisco.

[30] Rudebusch, G. D., and E. T. Swanson (2008a): "Examining the Bond Premium Puzzle with a DSGE Model," Manuscript, Federal Reserve Bank of San Francisco.

[31] Rudebusch, G.D., and E.T. Swanson (2008b): "Long-Run Inflation Risk and the Post-War Term Premium," Manuscript, Federal Reserve Bank of San Francisco.

[32] Schorfheide, F. (2005): "Learning and Monetary Policy Shifts," *Review of Economic Dynamics*, 8, 392-419.

[33] Svensson, L. E. O. (1995): "Estimating Forward Interest Rates with the Extended Nelson & Siegel Method," *Quarterly Review*, Sveriges Riksbank, 3, 13-26.

[34] Woodford, M. (2003): *Interest and Prices*, Princeton: Princeton University Press.

TABLE 1: Sample means of nominal and real yields

	1y	2y	5y	10y
UK (1985:01-1992:10)				
Nominal yield	10.58	10.29	10.09	9.97
Real yield	2.96	3.19	3.61	3.90
Breakeven inflation	7.62	7.10	6.48	6.07
UK (1985:01-1992:10)				
Nominal yield	5.35	5.52	5.76	5.81
Real yield	2.72	2.64	2.58	2.54
Breakeven inflation	2.63	2.88	2.54	3.27

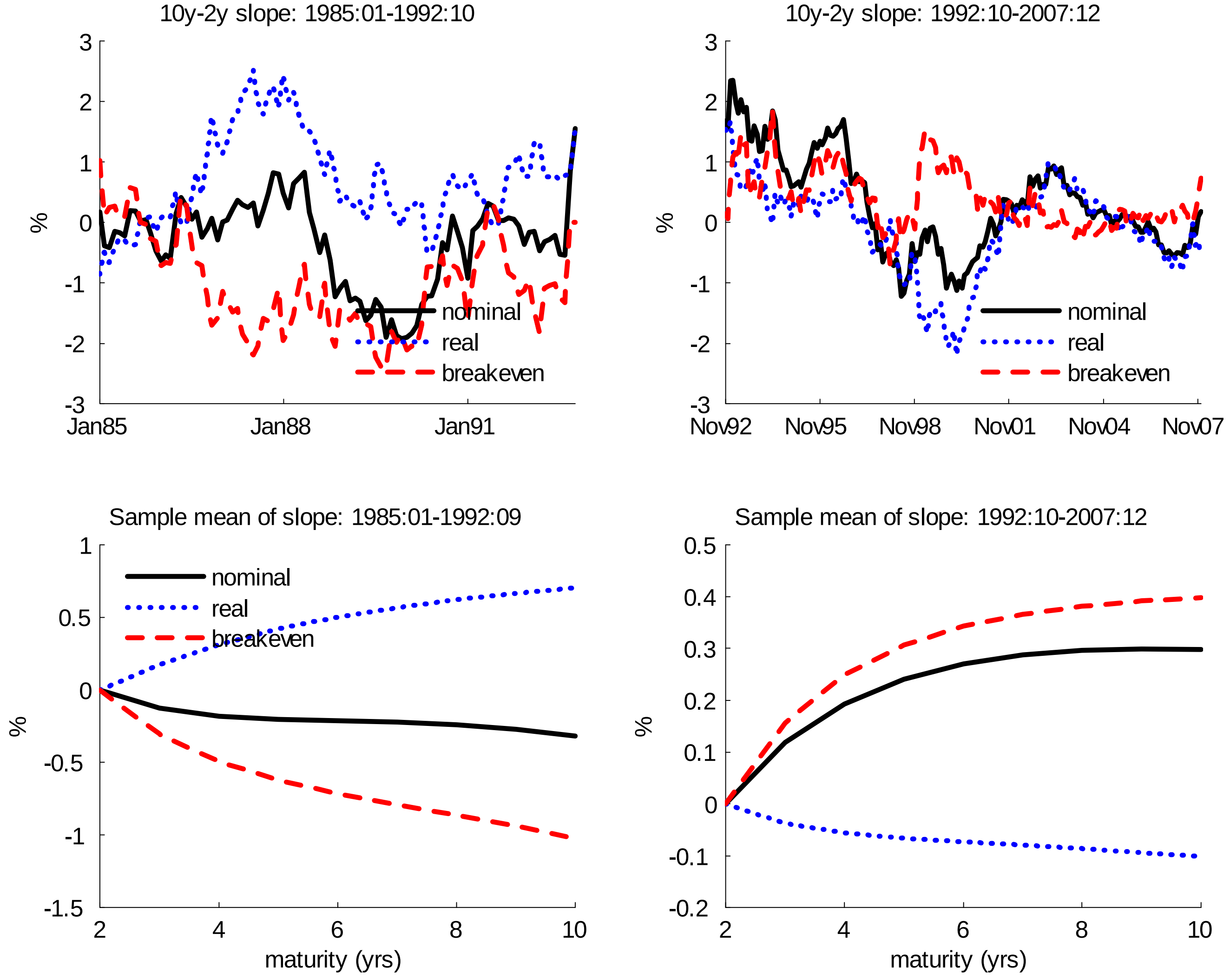
TABLE 2: Model Parameters: Priors and Estimates

Parameter	Distribution	Prior Mean	Prior Std. Dev.	Posterior Mode
σ , Risk aversion	Gamma	3.00	1.00	2.58
δ , Habit	Beta	0.40	0.15	0.36
ω , Curvature for labor disutility	Gamma	3.00	0.75	7.03
γ , Indexation	Beta	0.50	0.20	0.27
ϕ_{π}, Taylor rule, inflation	Gamma	2.00	0.30	2.01
ρ_{I}, Taylor rule, lagged interest rate	Beta	0.70	0.20	0.85
ρ_{a}, Persistence, Technology shock	Beta	0.90	0.05	0.99
ρ_{ξ}, Persistence, Preference shock	Beta	0.60	0.20	0.93
$100*\eta_{a}$, Standard deviation, technology shock	IG	2.50	1.00	0.01
$100*\eta_{m}$, Standard deviation, Taylor rule	IG	0.05	0.10	0.00
$100*\eta_{\xi}$, Standard deviation, preference shock	IG	1.00	0.50	0.03
π (1), Inflation target, low inflation regime	Normal	2.40	0.80	2.22
π (2), Inflation target, high inflation regime	Normal	6.00	0.40	5.37
p_{11}, Transition probability from low regime to low regime	Beta	0.90	0.05	0.98
p_{22}, Transition probability from high regime to high regime	Beta	0.60	0.13	0.95
$100*\eta_{m}$, Standard deviation, measurement error	IG	0.10	0.10	0.00
β , Discount rate	Fixed	0.995		
χ , Coefficient for labor disutility	Fixed	1.00		
α , Calvo parameter	Fixed	0.75		
θ , Dixit-Stiglitz CES parameter	Fixed	10		
ϕ_{y}, Taylor rule, output gap	Fixed	0.10		

TABLE 3: Correlations for Observed and Fitted Variables

	Level	Change
Consumption growth	1.0000	1.0000
Inflation	0.9985	0.9983
1 year nominal yield	0.9912	0.9308
2 year nominal yield	0.9979	0.9692
5 year nominal yield	0.9898	0.8958
10 year nominal yield	0.9678	0.8182
5 year real yield	0.8971	0.5894
10 year real yield	0.9488	0.6451
5 year breakeven inflation rate	0.9433	0.4842
10 year breakeven inflation rate	0.8643	0.2897
Spread between 10 year and 1 year nominal yields	0.6968	0.2815
Spread between 10 year and 5 year real yields	0.1631	0.1951
Spread between 10 year and 1 year breakeven inflation rates	0.5812	0.7034

FIGURE 1: UK term structure



Technical Appendix

December 29, 2011

Contents

A Model

A.1 Consumers

There is a continuum of goods in the economy indexed by i in the unit interval. Each good is produced by a separate firm. The representative consumer's preferences over the different goods are described by the usual CES Dixit-Stiglitz aggregator:

$$C_t = \left[\int_0^1 C_t(i)^{\frac{\theta-1}{\theta}} \, di \right]^{\frac{\theta}{\theta-1}} \tag{1}$$

where $C(i)$ is the consumption of good i. θ is a preference parameter describing the elasticity of substitution between good i and good j. Consumer utility depends on the aggregate consumption good C_t and labor supply $L_t(i)$ to each firm i in an additively separable manner. We assume time-separable preferences with constant relative risk aversion over the consumption aggregate:

$$U = \max E_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{(C_t - \delta C_{t-1})^{1-\sigma}}{1-\sigma} \xi_t - \chi L_t^{\omega} \right] \tag{2}$$

Here, $\beta < 1$ is a subjective discount factor, σ is the inverse of the intertemporal elasticity of substitution and ω and χ are parameters describing the disutility from labor.

The flow budget constraint of the representative agent is:

$$\int_0^1 P_t(i) C_t(i) \, di + B_t \leq w_t L_t + \int_0^1 \Xi_t(i) \, di + W_t \tag{3}$$

where $P_t(i)$ is the nominal price of good i. From the consumer's intratemporal cost-minimization problem it can be shown that

$$\int_0^1 P_t(i) C_t(i) \, di = P_t C_t \tag{4}$$

where the aggregate price index P_t is given by

$$P_t = \left[\int_0^1 P_t(i)^{1-\theta} \, di \right]^{\frac{1}{1-\theta}} \tag{5}$$

$\Xi_t(i)$ are the profits of firm i, w_t is the nominal wage paid by firms, W_t is beginning of period wealth, consisting of the nominal value of all asset holdings and B_t represents the value end of period asset holdings. Complete markets imply the existence of a stochastic discount factor $Q_{t,t+1}$ such that

$$B_t = E_t \left[Q_{t,t+1} W_{t+1} \right] \tag{6}$$

The Lagrangian for the consumer's utility maximization problem is:

$$\mathcal{L} = E_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{(C_t - \delta C_{t-1})^{1-\sigma}}{1-\sigma} \xi_t - \chi L_t^{\omega} di \right]$$
$$- E_0 \sum_{t=0}^{\infty} \beta^t \mu_t \left[P_t C_t + Q_{t,t+1} W_{t+1} - \left(w_t L_t + \int_0^1 \Xi(i)\, di \right) - W_t \right] \tag{7}$$

The first-order conditions with respect to the control variables C_t, L_t and W_{t+1} imply:

$$Q_{t,t+1} = \beta \frac{\mu_{t+1}}{\mu_t} = \beta \frac{P_t}{P_{t+1}} \frac{\Lambda_{t+1}}{\Lambda_t} \tag{8}$$

where $\Lambda_t \equiv \mu_t P_t = (C_t - \delta C_{t-1})^{-\sigma} \xi_t - \beta\delta E_t \left[(C_{t+1} - \delta C_t)^{-\sigma} \xi_{t+1} \right]$, and

$$\frac{w_t}{P_t} = \frac{\chi \omega L_t^{\omega - 1}}{\Lambda_t} \tag{9}$$

A.2 Firms

There is a continuum of monopolistically competitive firms in the economy, each with some price-setting power. The degree of market power of each firm depends on the CES parameter of the representative consumer's utility function. The demand curve for each firm's product is derived from the consumer's intratemporal optimization problem and market clearing. Together, they imply:

$$Y_t(i) = \left[\frac{P_t(i)}{P_t} \right]^{-\theta} Y_t \tag{10}$$

Firms are price setters but not all of them get to optimize their price in each period. Following Calvo (1983) and Christiano et al. (2001), we introduce both staggered price-setting and inflation indexation. In particular, in each period every firm has a probability $1-\alpha$ of re-optimising its price. Firms who do not get to optimize their price in period t, set it according to the formula:

$$P_t(i) = P_{t-1}(i)\, \Pi_{t-1}^{\gamma} \tag{11}$$

where

$$\Pi_t \equiv \frac{P_t}{P_{t-1}}$$

Firms maximize profits which are given by

$$\Xi_t(i) = P_t(i)\, Y_t(i) - TC_t(i)$$

where $TC_t = w_t(i) L_t(i)$, $Y_t(i) = A_t L_t(i)$, A_t is an exogenous aggregate productivity shock. Assuming competitive factor markets, firms take the common wage $w_t(i) = w_t$ as given. Firm i's nominal marginal cost of producing an additional unit of output is:

$$MC_t(i) = MC_t = \frac{w_t}{A_t} \tag{12}$$

Substituting the labor supply condition (9) into the above equation we get that the real marginal cost is:

$$mc_t \equiv \frac{MC_t}{P_t} = \frac{\chi\omega L_t^{\omega-1}}{A_t\Lambda_t} = \frac{\chi\omega}{\Lambda_t}\frac{\left(\frac{Y_t s_t}{A_t}\right)^{\omega-1}}{A_t}$$
$$= \frac{\chi\omega}{\Lambda_t}\frac{(Y_t s_t)^{\omega-1}}{A_t^{\omega}}$$

$$mc_t \equiv \frac{MC_t}{P_t} = \frac{\chi\omega L_t^{\omega-1}}{A_t\Lambda_t} = \frac{\chi\omega}{\Lambda_t}\frac{(Y_t s_t)^{\omega-1}}{A_t^{\omega}} \quad (13)$$

From here, we get that in every period $t+s>t$, for a firm that has not re-optimized its price since period t, the real marginal cost mc_{t+s} is:

$$mc_{t+s} \equiv \frac{MC_{t+s}}{P_{t+s}} = \frac{\chi\omega}{\Lambda_{t+s}}\frac{(Y_{t+s}s_{t+s})^{\omega-1}}{A_{t+s}^{\omega}} \quad (14)$$

Note that

$$TC_{t+s}(i) = MC_{t+s}Y_{t+s}(i) = mc_{t+s}P_{t+s}Y_{t+s}(i) \quad (15)$$

The intertemporal profit maximization problem of the firms optimising their price in period t is:

$$\max_{P_t(i)} E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[P_{t+s}(i)\,Y_{t+s}(i) - mc_{t+s}P_{t+s}Y_{t+s}(i)\right]$$

$$Y_{t+s}(i) = \left(\frac{P_{t+s}(i)}{P_{t+s}}\right)^{-\theta}Y_{t+s}$$

$$\max_{P_t(i)} E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[P_{t+s}(i)\left(\frac{P_{t+s}(i)}{P_{t+s}}\right)^{-\theta}Y_{t+s} - mc_{t+s}P_{t+s}\left(\frac{P_{t+s}(i)}{P_{t+s}}\right)^{-\theta}Y_{t+s}\right]$$

$$\max_{P_t(i)} E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[P_{t+s}(i)\left(\frac{P_{t+s}(i)}{P_{t+s}}\right)^{-\theta}Y_{t+s} - mc_{t+s}P_{t+s}\left(\frac{P_{t+s}(i)}{P_{t+s}}\right)^{-\theta}Y_{t+s}\right]$$

$$\max_{P_t(i)} E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[P_t(i)^{1-\theta}\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{\gamma(1-\theta)}P_{t+s}^{\theta}Y_{t+s} - P_t(i)^{-\theta}mc_{t+s}P_{t+s}^{1+\theta}\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{-\gamma\theta}Y_{t+s}\right]$$

The FOC is:

$$E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[(\theta-1)\,P_t^*(i)^{-\theta}\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{\gamma(1-\theta)}P_{t+s}^{\theta}Y_{t+s}\right]$$
$$= E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[\theta P_t^*(i)^{-\theta-1}mc_{t+s}P_{t+s}^{1+\theta}\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{-\gamma\theta}Y_{t+s}\right]$$

$$(\theta-1)\,P_t^*(i)^{-\theta}\,E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{\gamma(1-\theta)}P_{t+s}^{\theta}Y_{t+s}\right]$$

$$=\theta P_t^*(i)^{-\theta-1}\,E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[mc_{t+s}P_{t+s}^{1+\theta}\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{-\gamma\theta}Y_{t+s}\right]$$

$$\frac{\theta P_t^*(i)^{-\theta-1}}{(\theta-1)\,P_t^*(i)^{-\theta}}=\frac{E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{\gamma(1-\theta)}P_{t+s}^{\theta}Y_{t+s}\right]}{E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[mc_{t+s}P_{t+s}^{1+\theta}\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{-\gamma\theta}Y_{t+s}\right]}$$

So the optimal price is

$$P_t^*(i)=\frac{\theta}{\theta-1}\frac{E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[mc_{t+s}P_{t+s}^{1+\theta}\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{-\gamma\theta}Y_{t+s}\right]}{E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{\gamma(1-\theta)}P_{t+s}^{\theta}Y_{t+s}\right]}$$

$$P_t^*(i)=\frac{\theta}{\theta-1}\frac{E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[mc_{t+s}P_{t+s}^{1+\theta}\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{-\gamma\theta}Y_{t+s}\right]P_t^{-1-\theta}}{E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{\gamma(1-\theta)}P_{t+s}^{\theta}Y_{t+s}\right]P_t^{-1-\theta}}$$

$$\frac{P_t^*(i)}{P_t}=\frac{\theta}{\theta-1}\frac{E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[mc_{t+s}\left(\frac{P_{t+s}}{P_t}\right)^{1+\theta}\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{-\gamma\theta}Y_{t+s}\right]}{E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{\gamma(1-\theta)}\left(\frac{P_{t+s}}{P_t}\right)^{\theta}Y_{t+s}\right]}=\frac{\theta}{\theta-1}\frac{DC_t}{DR_t}$$

Numerator

$$DC_t\equiv E_t\sum_{s=0}^{\infty}\alpha^s Q_{t,t+s}\left[mc_{t+s}\left(\frac{P_{t+s}}{P_t}\right)^{1+\theta}\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{-\gamma\theta}Y_{t+s}\right]$$

$$DC_{t+1}=E_t\sum_{s=0}^{\infty}\alpha^s Q_{t+1,t+1+s}\left[mc_{t+1+s}\left(\frac{P_{t+1+s}}{P_{t+1}}\right)^{1+\theta}\left(\frac{P_{t+s}}{P_t}\right)^{-\gamma\theta}Y_{t+1+s}\right]$$

$$DC_t=mc_t\left(\frac{P_t}{P_t}\right)^{1+\theta}\left(\frac{P_{t-1}}{P_{t-1}}\right)^{-\gamma\theta}Y_t+E_t\sum_{s=1}^{\infty}\alpha^s Q_{t,t+s}\left[mc_{t+s}\left(\frac{P_{t+s}}{P_t}\right)^{1+\theta}\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{-\gamma\theta}Y_{t+s}\right]$$

$$DC_t=mc_tY_t+E_t\sum_{s=0}^{\infty}\alpha^{s+1}Q_{t,t+1+s}\left[mc_{t+1+s}\left(\frac{P_{t+1+s}}{P_t}\right)^{1+\theta}\left(\frac{P_{t+s}}{P_{t-1}}\right)^{-\gamma\theta}Y_{t+1+s}\right]$$

$$DC_t=mc_tY_t+\alpha E_tQ_{t,t+1}\sum_{s=0}^{\infty}\alpha^s Q_{t+1,t+1+s}\left[mc_{t+1+s}\left(\frac{P_{t+1+s}}{P_t}\frac{P_{t+1}}{P_{t+1}}\right)^{1+\theta}\left(\frac{P_{t+s}}{P_{t-1}}\frac{P_t}{P_t}\right)^{-\gamma\theta}Y_{t+1+s}\right]$$

$$DC_t = mc_t Y_t$$

$$+ \alpha \left(\frac{P_t}{P_{t-1}}\right)^{-\gamma\theta} E_t Q_{t,t+1} \left(\frac{P_{t+1}}{P_t}\right)^{1+\theta} \sum_{s=0}^{\infty} \alpha^s Q_{t+1,t+1+s} \left[mc_{t+1+s} \left(\frac{P_{t+1+s}}{P_{t+1}}\right)^{1+\theta} \left(\frac{P_{t+s}}{P_t}\right)^{-\gamma\theta} Y_{t+1+s} \right]$$

$$DC_t = \frac{\chi\omega}{\Lambda_t} \frac{(Y_t s_t)^{\omega-1}}{A_t^{\omega}} Y_t + \alpha \Pi_t^{-\gamma\theta} E_t Q_{t,t+1} \Pi_{t+1}^{1+\theta} DC_{t+1}$$

$$DC_t = \chi\omega \frac{s_t^{\omega-1}}{\Lambda_t} \frac{Y_t^{\omega}}{A_t^{\omega}} + \alpha \Pi_t^{-\gamma\theta} E_t Q_{t,t+1} \Pi_{t+1}^{1+\theta} DC_{t+1}$$

$$DC_t = \chi\omega \frac{s_t^{\omega-1}}{\Lambda_t} \frac{Y_t^{\omega}}{A_t^{\omega}} + \alpha\beta \Pi_t^{-\gamma\theta} E_t \left[\frac{\Lambda_{t+1}}{\Lambda_t} \Pi_{t+1}^{\theta} DC_{t+1} \right]$$

Denominator

$$DR_t \equiv E_t \sum_{s=0}^{\infty} \alpha^s Q_{t,t+s} \left[\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{\gamma(1-\theta)} \left(\frac{P_{t+s}}{P_t}\right)^{\theta} Y_{t+s} \right]$$

$$DR_{t+1} = E_{t+1} \sum_{s=0}^{\infty} \alpha^s Q_{t+1,t+s} \left[\left(\frac{P_{t+s}}{P_t}\right)^{\gamma(1-\theta)} \left(\frac{P_{t+1+s}}{P_{t+1}}\right)^{\theta} Y_{t+1+s} \right]$$

$$DR_t = Y_t + E_t \sum_{s=1}^{\infty} \alpha^s Q_{t,t+s} \left[\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^{\gamma(1-\theta)} \left(\frac{P_{t+s}}{P_t}\right)^{\theta} Y_{t+s} \right]$$

$$DR_t = Y_t + E_t \sum_{s=0}^{\infty} \alpha^{s+1} Q_{t,t+1+s} \left[\left(\frac{P_{t+s}}{P_{t-1}}\right)^{\gamma(1-\theta)} \left(\frac{P_{t+1+s}}{P_t}\right)^{\theta} Y_{t+1+s} \right]$$

$$DR_t = Y_t + \alpha E_t Q_{t,t+1} \sum_{s=0}^{\infty} \alpha^s Q_{t+1,t+1+s} \left[\left(\frac{P_{t+s}}{P_{t-1}} \frac{P_t}{P_t}\right)^{\gamma(1-\theta)} \left(\frac{P_{t+1+s}}{P_t} \frac{P_{t+1}}{P_{t+1}}\right)^{\theta} Y_{t+1+s} \right]$$

$$DR_t = Y_t + \alpha \left(\frac{P_t}{P_{t-1}}\right)^{\gamma(1-\theta)} E_t Q_{t,t+1} \left(\frac{P_{t+1}}{P_t}\right)^{\theta} \sum_{s=0}^{\infty} \alpha^s Q_{t+1,t+1+s} \left[\left(\frac{P_{t+s}}{P_t}\right)^{\gamma(1-\theta)} \left(\frac{P_{t+1+s}}{P_{t+1}}\right)^{\theta} Y_{t+1+s} \right]$$

$$DR_t = Y_t + \alpha \Pi_t^{\gamma(1-\theta)} E_t Q_{t,t+1} \Pi_{t+1}^{\theta} DR_{t+1}$$

$$DR_t = Y_t + \alpha\beta \Pi_t^{\gamma(1-\theta)} E_t \left[\frac{\Lambda_{t+1}}{\Lambda_t} \Pi_{t+1}^{\theta-1} DR_{t+1} \right]$$

Hence the relative price of the optimizing firm can be expressed in recursive form with the following equations:

$$\frac{P_t^*(i)}{P_t} = \frac{\theta}{\theta-1} \frac{DC_t}{DR_t} \tag{16}$$

$$DR_t = Y_t + \alpha\beta \Pi_t^{\gamma(1-\theta)} E_t \left[\frac{\Lambda_{t+1}}{\Lambda_t} \Pi_{t+1}^{\theta-1} DR_{t+1} \right] \tag{17}$$

$$DC_t = \chi\omega\frac{s_t^{\omega-1}}{\Lambda_t}\left(\frac{Y_t}{A_t}\right)^{\omega} + \alpha\beta\Pi_t^{-\gamma\theta}E_t\left[\frac{\Lambda_{t+1}}{\Lambda_t}\Pi_{t+1}^{\theta}DC_{t+1}\right] \tag{18}$$

By symmetry, the optimal price $P_t^*(i)$ for each firm that gets to optimize in period t should be the same and denote it by P_t^*. Finally, note that the aggregate price level can be written as:

$$P_t = \left[(1-\alpha)P_t^{*(1-\theta)} + \alpha\left(P_{t-1}\left(\frac{P_{t-1}}{P_{t-2}}\right)^{\gamma}\right)^{1-\theta}\right]^{\frac{1}{1-\theta}}$$

$$\begin{aligned}
P_t^{1-\theta} &= \left[(1-\alpha)P_t^{*(1-\theta)} + \alpha\left(P_{t-1}\left(\frac{P_{t-1}}{P_{t-2}}\right)^{\gamma}\right)^{1-\theta}\right] \\
1 &= (1-\alpha)\frac{P_t^{*(1-\theta)}}{P_t^{1-\theta}} + \alpha\left(\frac{P_{t-1}}{P_t}\left(\frac{P_{t-1}}{P_{t-2}}\right)^{\gamma}\right)^{1-\theta} \\
(1-\alpha)\frac{P_t^{*(1-\theta)}}{P_t^{1-\theta}} &= 1-\alpha\left(\frac{P_{t-1}}{P_t}\left(\frac{P_{t-1}}{P_{t-2}}\right)^{\gamma}\right)^{1-\theta} \\
(1-\alpha)\left(\frac{P_t^*}{P_t}\right)^{1-\theta} &= 1-\alpha\left(\Pi_t^{-1}\Pi_{t-1}^{\gamma}\right)^{1-\theta} \\
\left(\frac{P_t^*}{P_t}\right)^{1-\theta} &= \frac{1-\alpha\left(\Pi_t^{-1}\Pi_{t-1}^{\gamma}\right)^{1-\theta}}{1-\alpha}
\end{aligned}$$

and therefore:

$$\frac{P_t^*}{P_t} = \left[\frac{1-\alpha\left(\frac{\Pi_t}{\Pi_{t-1}^{\gamma}}\right)^{\theta-1}}{(1-\alpha)}\right]^{\frac{1}{1-\theta}} \tag{19}$$

Finally, we get

$$\left[\frac{1-\alpha\left(\frac{\Pi_t}{\Pi_{t-1}^{\gamma}}\right)^{\theta-1}}{(1-\alpha)}\right]^{\frac{1}{1-\theta}} = \frac{\theta}{\theta-1}\frac{DC_t}{DR_t} \tag{20}$$

$$DR_t = Y_t + \alpha\beta\Pi_t^{\gamma(1-\theta)}E_t\left[\frac{\Lambda_{t+1}}{\Lambda_t}\Pi_{t+1}^{\theta-1}DR_{t+1}\right] \tag{21}$$

$$DC_t = \chi\omega\frac{s_t^{\omega-1}}{\Lambda_t}\left(\frac{Y_t}{A_t}\right)^{\omega} + \alpha\beta\Pi_t^{-\gamma\theta}E_t\left[\frac{\Lambda_{t+1}}{\Lambda_t}\Pi_{t+1}^{\theta}DC_{t+1}\right] \tag{22}$$

A.3 Monetary policy

The central bank's policy is described by an interest rate rule of the type:

$$I_t = \bar{I}_t^{1-\rho_I} I_{t-1}^{\rho_I} \upsilon_t$$

where $\bar{I}_t$ is defined by

$$\bar{I}_t = \frac{\bar{\Pi}}{\beta} \left(\frac{\Pi_t}{\Pi_t^*} \right)^{\phi_\pi} \left(\kappa \frac{Y_t}{Y_t^n} \right)^{\phi_y}$$

and υ_t is a monetary policy shock. ρ_I is a parameter governing the degree of monetary policy inertia. κ is defined as

$$\kappa^{-1} \equiv \left[\frac{1 - \alpha \bar{\Pi}^{(\theta-1)(1-\gamma)}}{(1-\alpha)} \right]^{\frac{\theta(\omega-1)+1}{\theta-1} \frac{1}{1-\omega-\sigma}} \left(\frac{1 - \alpha\beta\bar{\Pi}^{(\gamma-1)(1-\theta)}}{1 - \alpha\beta\bar{\Pi}^{(1-\gamma)\theta}} \right)^{\frac{1}{1-\omega-\sigma}} \bar{s}^{\frac{\omega-1}{1-\omega-\sigma}}$$

so that in steady state

$$\kappa \frac{Y_t}{Y_t^n} = \kappa \frac{\bar{Y}}{\bar{Y}^n} = 1$$

Clearly, when $\bar{\Pi} = 1$ or $\gamma = 1$, we have $\kappa = 1$.

A.4 Exogenous disturbances

There are several sources of uncertainty in our baseline model. These are the technology shock A_t, the monetary policy shock υ_t, the preference shock ξ_t and the inflation target Π_t^*. We assume that the processes for those are given by

$$\log A_{t+1} = \rho_A \log A_t + \eta^a \varepsilon_{t+1}^a \tag{23}$$

$$\log \upsilon_{t+1} = \eta^m \varepsilon_{t+1}^m \tag{24}$$

$$\log \xi_{t+1} = \rho_\xi \log \xi_t + \eta^\xi \varepsilon_{t+1}^\xi \tag{25}$$

Here, ρ_A, $\rho_\xi \in [0,1]$ are autoregressive parameters and η^a, η^m, η^ξ are volatility scaling parameters. $\{\varepsilon_t^a\}$, $\{\varepsilon_t^m\}$, $\left\{\varepsilon_t^\xi\right\}$ and are independent standard normally distributed shock processes uncorrelated with each other.

In our main model Π_t^* follows a two-state Markov-switching process with transition probability matrix P. In our alternative model, it follows an autoregressive process given by

$$\log\left(\Pi_t^*\right) = (1-\rho_\pi) \log\left(\bar{\Pi}\right) + \rho_\pi \log\left(\Pi_{t-1}^*\right) + \eta_\pi \varepsilon_t^\pi \tag{26}$$

A.4.1 Aggregation

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

$$A_t L_t(i) = \left(\frac{P_t\left(i\right)}{P_t}\right)^{-\theta} Y_t$$

$$L_t(i) = \left(\frac{P_t\left(i\right)}{P_t}\right)^{-\theta} \frac{Y_t}{A_t}$$

$$L_t(i) = \left(\frac{P_t\left(i\right)}{P_t}\right)^{-\theta} \left(\frac{Y_t}{A_t}\right)$$

$$L_t = \int_0^1 L_t(i)di = \left(\frac{Y_t}{A_t}\right) \int_0^1 \left(\frac{P_t\left(i\right)}{P_t}\right)^{-\theta} di = \left(\frac{Y_t}{A_t}\right) s_t$$

$$s_t \equiv \int_0^1 \left(\frac{P_t\left(i\right)}{P_t}\right)^{-\theta} di$$

$$s_t = \int_0^{1-\alpha} \left(\frac{P_t^*}{P_t}\right)^{-\theta} di + \int_{1-\alpha}^1 \left(\frac{P_t\left(i\right)}{P_t}\right)^{-\theta} di$$

$$s_t = \int_0^{1-\alpha} \left(\frac{P_t^*}{P_t}\right)^{-\theta} di + \int_{1-\alpha}^1 \left(\frac{P_t\left(i\right)P_{t-1}}{P_{t-1}P_t}\frac{\Pi_{t-1}^{\gamma}}{\Pi_{t-1}^{\gamma}}\right)^{-\theta} di$$

$$\int_{1-\alpha}^1 \left(\frac{P_t\left(i\right)P_{t-1}}{P_{t-1}P_t}\frac{\Pi_{t-1}^{\gamma}}{\Pi_{t-1}^{\gamma}}\right)^{-\theta} di = \Pi_t^{\theta} \int_{1-\alpha}^1 \left(\frac{P_t\left(i\right)}{P_{t-1}}\frac{\Pi_{t-1}^{\gamma}}{\Pi_{t-1}^{\gamma}}\right)^{-\theta} di$$

$$\int_{1-\alpha}^1 \left(\frac{P_t\left(i\right)P_{t-1}}{P_{t-1}P_t}\right)^{-\theta} di = \Pi_t^{\theta}\Pi_{t-1}^{-\gamma\theta} \int_{1-\alpha}^1 \left(\frac{P_t\left(i\right)}{P_{t-1}\Pi_{t-1}^{\gamma}}\right)^{-\theta} di = \alpha\Pi_t^{\theta}\Pi_{t-1}^{-\gamma\theta} s_{t-1}$$

$$P_t\left(i\right) = P_{t-1}\left(i\right)\Pi_{t-1}^{\gamma}$$

$$\int_{1-\alpha}^1 \left(\frac{P_t\left(i\right)P_{t-1}}{P_{t-1}P_t}\right)^{-\theta} di = \Pi_t^{\theta}\Pi_{t-1}^{-\gamma\theta} \int_{1-\alpha}^1 \left(\frac{P_{t-1}\left(i\right)}{P_{t-1}}\right)^{-\theta} di = \alpha\Pi_t^{\theta}\Pi_{t-1}^{-\gamma\theta} s_{t-1}$$

$$\int_0^{1-\alpha} \left(\frac{P_t^*}{P_t}\right)^{-\theta} di = \left[\frac{1-\alpha\left(\frac{\Pi_t}{\Pi_{t-1}^{\gamma}}\right)^{\theta-1}}{(1-\alpha)}\right]^{\frac{\theta}{\theta-1}}$$

$$s_t = (1-\alpha)\left[\frac{1-\alpha\left(\frac{\Pi_t}{\Pi_{t-1}^{\gamma}}\right)^{\theta-1}}{(1-\alpha)}\right]^{\frac{\theta}{\theta-1}} + \alpha\Pi_t^{\theta}\Pi_{t-1}^{-\gamma\theta} s_{t-1}$$

$$Y_t = A_t L_t s_t^{-1}$$

A.4.2 Natural rate of output

If prices were flexible, the representative agent would optimize $P_t(i)$ in every period in order to maximize

$$\Xi_t(i) = P_t(i) Y_t(i) - mc_t P_t Y_t(i)$$
$$\Xi_t(i) = P_t(i) \left(\frac{P_t(i)}{P_t}\right)^{-\theta} Y_t - mc_t P_t \left(\frac{P_t(i)}{P_t}\right)^{-\theta} Y_t$$

The first order condition for this maximization problem can be shown to be:

$$(\theta - 1) P_t(i)^{-\theta} P_t^{\theta} Y_t^n = \theta \left(\frac{P_t(i)}{P_t}\right)^{-(1+\theta)} mc_t Y_t^n$$
$$(\theta - 1) P_t(i)^{-\theta} P_t^{\theta} Y_t^n = \chi\omega\theta \frac{1}{\Lambda_t} \left(\frac{P_t(i)}{P_t}\right)^{-(1+\theta)} \left(\frac{Y_t^n}{A_t}\right)^{\omega}$$

Using the symmety of the individual price setters' decisions, we can deduce that for each i we have $P_t(i) = P_t$ and the above expression simplifies to:

$$(\theta - 1) Y_t^n = \chi\omega \frac{\theta}{\Lambda_t} \left(\frac{Y_t^n}{A_t}\right)^{\omega}$$
$$Y_t^n = \frac{\chi\omega}{\Lambda_t} \frac{\theta}{\theta - 1} \left(\frac{Y_t^n}{A_t}\right)^{\omega}$$
$$(Y_t^n)^{1-\omega} = \frac{\chi\omega}{\Lambda_t} \frac{\theta}{\theta - 1} A_t^{-\omega}$$

When there is no habit formation in the utility function $\Lambda_t = (Y_t^n)^{-\sigma}$ and the above expression simplifies to:

$$(Y_t^n)^{1-\omega-\sigma} = \chi\omega \frac{\theta}{\theta - 1} A_t^{-\omega}$$

Hence, the natural rate of output Y_t^n is defined as:

$$Y_t^n = \left[\chi\omega \frac{\theta}{\theta - 1} A_t^{-\omega}\right]^{\frac{1}{1-\omega-\sigma}}$$

With habit formation $\Lambda_t = \left(Y_t^n - \delta Y_{t-1}^n\right)^{-\sigma} \xi_t - \beta\delta E_t \left[\left(Y_{t+1}^n - \delta Y_t^n\right)^{-\sigma} \xi_{t+1}\right]$ and

$$\left\{\left(Y_t^n - \delta Y_{t-1}^n\right)^{-\sigma} \xi_t - \beta\delta E_t \left[\left(Y_{t+1}^n - \delta Y_t^n\right)^{-\sigma} \xi_{t+1}\right]\right\} (Y_t^n)^{1-\omega} = \chi\omega \frac{\theta}{\theta - 1} A_t^{-\omega}$$

A.5 Model equations

$$\Lambda_t = (C_t - \delta C_{t-1})^{-\sigma} \xi_t - \beta\delta E_t \left[(C_{t+1} - \delta C_t)^{-\sigma} \xi_{t+1}\right] \tag{27}$$

$$\frac{1}{I_t} = E_t\left[\frac{\beta}{\Pi_{t+1}}\frac{\Lambda_{t+1}}{\Lambda_t}\right] \tag{28}$$

$$Y_t = A_t L_t s_t^{-1} \tag{29}$$

$$\left[\frac{1-\alpha\left(\frac{\Pi_t}{\Pi_{t-1}^{\gamma}}\right)^{\theta-1}}{(1-\alpha)}\right]^{\frac{1}{1-\theta}} = \frac{\theta}{\theta-1}\frac{DC_t}{DR_t} \tag{30}$$

$$DR_t = Y_t + \alpha\beta\Pi_t^{\gamma(1-\theta)}E_t\left[\frac{\Lambda_{t+1}}{\Lambda_t}\Pi_{t+1}^{\theta-1}DR_{t+1}\right] \tag{31}$$

$$DC_t = \chi\omega\frac{s_t^{\omega-1}}{\Lambda_t}\left(\frac{Y_t}{A_t}\right)^{\omega} + \alpha\beta\Pi_t^{-\gamma\theta}E_t\left[\frac{\Lambda_{t+1}}{\Lambda_t}\Pi_{t+1}^{\theta}DC_{t+1}\right] \tag{32}$$

$$s_t = (1-\alpha)\left[\frac{1-\alpha\left(\frac{\Pi_t}{\Pi_{t-1}^{\gamma}}\right)^{\theta-1}}{(1-\alpha)}\right]^{\frac{\theta}{\theta-1}} + \alpha\Pi_t^{\theta}\Pi_{t-1}^{-\gamma\theta}s_{t-1} \tag{33}$$

$$I_t = \left[\frac{\bar{\Pi}}{\beta}\left(\frac{\Pi_t}{\Pi_t^*}\right)^{\phi_\pi}\left(\kappa\frac{Y_t}{Y_t^n}\right)^{\phi_y}\right]^{1-\rho_I} I_{t-1}^{\rho_I}\upsilon_t \tag{34}$$

$$\left\{\left(Y_t^n-\delta Y_{t-1}^n\right)^{-\sigma}\xi_t - \beta\delta E_t\left[\left(Y_{t+1}^n-\delta Y_t^n\right)^{-\sigma}\xi_{t+1}\right]\right\}\left(Y_t^n\right)^{1-\omega} = \chi\omega\frac{\theta}{\theta-1}A_t^{-\omega} \tag{35}$$

$$\log A_{t+1} = (1-\rho_A)\log\bar{A} + \rho_A\log A_t + \eta^a\varepsilon_{t+1}^a \tag{36}$$

$$\log\upsilon_{t+1} = \eta^m\varepsilon_{t+1}^m \tag{37}$$

$$\log\xi_{t+1} = \rho_\xi\log\xi_t + \eta^\xi\varepsilon_{t+1}^\xi \tag{38}$$

A.6 Steady state

We are interested in model solutions around the non-stochastic steady state value of inflation $\bar{\Pi}$ which we allow to differ from 1. Using the Euler equation (28) we obtain a relationship between the steady-state interest rate and the representative agent's discount factor.

$$\bar{I}_0 = \bar{I}_{-1} = \bar{\Pi}\beta^{-1}$$

From the equation for Y^n we get

$$\left\{\left(Y_t^n-\delta Y_{t-1}^n\right)^{-\sigma}\xi_t - \beta\delta E_t\left[\left(Y_{t+1}^n-\delta Y_t^n\right)^{-\sigma}\xi_{t+1}\right]\right\}\left(Y_t^n\right)^{1-\omega} = \chi\omega\frac{\theta}{\theta-1}A_t^{-\omega}$$

$$\left[\left(\bar{Y}^n-\delta\bar{Y}^n\right)^{-\sigma} - \beta\delta\left(\bar{Y}^n-\delta\bar{Y}^n\right)^{-\sigma}\right]\left(\bar{Y}^n\right)^{1-\omega} = \chi\omega\frac{\theta}{\theta-1}$$

$$\left[(1-\beta\delta)(1-\delta)^{-\sigma}\right]\left(\bar{Y}^n\right)^{1-\omega-\sigma} = \chi\omega\frac{\theta}{\theta-1}$$

$$\bar{Y}^n = \left[\chi\omega\frac{\theta}{\theta-1}\frac{(1-\delta)^\sigma}{1-\beta\delta}\right]^{\frac{1}{1-\omega-\sigma}}$$

Using the definition for the marginal utility of consumption, we get:

$$\bar{\Lambda} = (1-\beta\delta)(1-\delta)^{-\sigma}\bar{Y}^{-\sigma}$$

and from the aggregation equation we have:

$$s_t = (1-\alpha)\left[\frac{1-\alpha\left(\frac{\Pi_t}{\Pi_{t-1}^\gamma}\right)^{\theta-1}}{(1-\alpha)}\right]^{\frac{\theta}{\theta-1}} + \alpha\Pi_t^\theta\Pi_{t-1}^{-\gamma\theta}s_{t-1}$$

$$\bar{s} = (1-\alpha)\left[\frac{1-\alpha\bar{\Pi}^{(\theta-1)(1-\gamma)}}{(1-\alpha)}\right]^{\frac{\theta}{\theta-1}} + \alpha\bar{\Pi}^{\theta(1-\gamma)}\bar{s}$$

$$\bar{s} = (1-\alpha)\left[\frac{1-\alpha\bar{\Pi}^{(\theta-1)(1-\gamma)}}{(1-\alpha)}\right]^{\frac{\theta}{\theta-1}}\left(1-\alpha\bar{\Pi}^{\theta(1-\gamma)}\right)^{-1}$$

$$\bar{L} = \bar{Y}\bar{s}$$

From the firm's FOC's for discounted profits maximization we obtain:

$$DR_t = Y_t + \alpha\beta\Pi_t^{\gamma(1-\theta)}E_t\left[\frac{\Lambda_{t+1}}{\Lambda_t}\Pi_{t+1}^{\theta-1}DR_{t+1}\right]$$

$$\overline{DR} = \bar{Y} + \alpha\beta\bar{\Pi}^{\gamma(1-\theta)}\bar{\Pi}^{\theta-1}\overline{DR}$$

$$\overline{DR} = \frac{\bar{Y}}{1-\alpha\beta\bar{\Pi}^{(\gamma-1)(1-\theta)}}$$

$$DC_t = \chi\omega\frac{s_t^{\omega-1}}{\Lambda_t}\left(\frac{Y_t}{A_t}\right)^\omega + \alpha\beta\Pi_t^{-\gamma\theta}E_t\left[\frac{\Lambda_{t+1}}{\Lambda_t}\Pi_{t+1}^\theta DC_{t+1}\right]$$

$$\overline{DC} = \chi\omega\frac{\bar{s}^{\omega-1}}{\bar{\Lambda}}\bar{Y}^\omega + \alpha\bar{\Pi}^{-\gamma\theta}\beta\bar{\Pi}^\theta\overline{DC}$$

$$\overline{DC} = \chi\omega\frac{\bar{s}^{\omega-1}}{\bar{\Lambda}}\frac{\bar{Y}^\omega}{1-\alpha\beta\bar{\Pi}^{(1-\gamma)\theta}}$$

$$\overline{DC} = \chi\omega\frac{\bar{Y}^\omega}{1-\alpha\beta\bar{\Pi}^{(1-\gamma)\theta}}\frac{(1-\delta)^\sigma}{1-\beta\delta}\bar{Y}^\sigma\bar{s}^{\omega-1}$$

Also we have that

$$\left[\frac{1-\alpha\left(\frac{\Pi_t}{\Pi_{t-1}^\gamma}\right)^{\theta-1}}{(1-\alpha)}\right]^{\frac{1}{1-\theta}} = \frac{\theta}{\theta-1}\frac{DC_t}{DR_t}$$

Hence

$$\overline{DR} = \left[\frac{1-\alpha\bar{\Pi}^{(\theta-1)(1-\gamma)}}{(1-\alpha)}\right]^{\frac{1}{\theta-1}} \frac{\theta}{\theta-1}\overline{DC}$$

$$\overline{DR} = \left[\frac{1-\alpha\bar{\Pi}^{(\theta-1)(1-\gamma)}}{(1-\alpha)}\right]^{\frac{1}{\theta-1}} \frac{\theta}{\theta-1}\chi\omega\frac{\bar{s}^{\omega-1}}{\bar{\Lambda}}\frac{\bar{Y}^{\omega}}{1-\alpha\beta\bar{\Pi}^{(1-\gamma)\theta}}$$

$$\frac{\bar{Y}}{1-\alpha\beta\bar{\Pi}^{(\gamma-1)(1-\theta)}} = \left[\frac{1-\alpha\bar{\Pi}^{(\theta-1)(1-\gamma)}}{(1-\alpha)}\right]^{\frac{1}{\theta-1}} \frac{\theta}{\theta-1}\chi\omega\frac{\bar{s}^{\omega-1}}{\bar{\Lambda}}\frac{\bar{Y}^{\omega}}{1-\alpha\beta\bar{\Pi}^{(1-\gamma)\theta}}$$

$$\bar{Y}^{1-\omega-\sigma} = \left[\frac{1-\alpha\bar{\Pi}^{(\theta-1)(1-\gamma)}}{(1-\alpha)}\right]^{\frac{1}{\theta-1}} \frac{1-\alpha\beta\bar{\Pi}^{(\gamma-1)(1-\theta)}}{1-\alpha\beta\bar{\Pi}^{(1-\gamma)\theta}}\chi\omega\frac{\theta}{\theta-1}\frac{(1-\delta)^{\sigma}}{(1-\beta\delta)}\bar{s}^{\omega-1}$$

$$\bar{Y} = \left[\frac{1-\alpha\bar{\Pi}^{(\theta-1)(1-\gamma)}}{(1-\alpha)}\right]^{\frac{1}{\theta-1}\frac{1}{1-\omega-\sigma}} \left(\frac{1-\alpha\beta\bar{\Pi}^{(\gamma-1)(1-\theta)}}{1-\alpha\beta\bar{\Pi}^{(1-\gamma)\theta}}\right)^{\frac{1}{1-\omega-\sigma}} \left[\chi\omega\frac{\theta}{\theta-1}\frac{(1-\delta)^{\sigma}}{(1-\beta\delta)}\bar{s}^{\omega-1}\right]^{\frac{1}{1-\omega-\sigma}}$$

$$\bar{Y} = \left[\frac{1-\alpha\bar{\Pi}^{(\theta-1)(1-\gamma)}}{(1-\alpha)}\right]^{\frac{1}{\theta-1}\frac{1}{1-\omega-\sigma}} \left(\frac{1-\alpha\beta\bar{\Pi}^{(\gamma-1)(1-\theta)}}{1-\alpha\beta\bar{\Pi}^{(1-\gamma)\theta}}\right)^{\frac{1}{1-\omega-\sigma}} \bar{s}^{\frac{\omega-1}{1-\omega-\sigma}}\bar{Y}^{n}$$

$$\bar{Y} = \kappa^{-1}\bar{Y}^{n}$$

$$\kappa^{-1} = \left[\frac{1-\alpha\bar{\Pi}^{(\theta-1)(1-\gamma)}}{(1-\alpha)}\right]^{\frac{1}{\theta-1}\frac{1}{1-\omega-\sigma}} \left(\frac{1-\alpha\beta\bar{\Pi}^{(\gamma-1)(1-\theta)}}{1-\alpha\beta\bar{\Pi}^{(1-\gamma)\theta}}\right)^{\frac{1}{1-\omega-\sigma}} \bar{s}^{\frac{\omega-1}{1-\omega-\sigma}}$$

B Regime switching models

B.1 Model set-up

We log-linearize the model equations defined in Appendix A to obtain the following linear system

$$B(s_t) X_t = M(s_t) + A(s_t) E_t X_{t+1} + C(s_t) X_{t-1} + D(s_t) \epsilon_t$$

If we conjecture the solution

$$\begin{aligned} X_{t+1} &= \mu(s_{t+1}) + \Phi(s_{t+1}) X_t + \Sigma(s_{t+1}) \epsilon_{t+1} \\ E_t X_{t+1} &= E_t \mu(s_{t+1}) + E_t \Phi(s_{t+1}) X_t \end{aligned}$$

and plug into the model equations, we match coefficients and get

$$[B(s_t) - A(s_t) E_t \Phi(s_{t+1})] X_t = M(s_t) + A(s_t) E_t \mu(s_{t+1}) + C(s_t) X_{t-1} + D(s_t) \epsilon_t$$

$$\begin{aligned} \mu(s_t) &= [B(s_t) - A(s_t) E_t \Phi(s_{t+1})]^{-1} [M(s_t) + A(s_t) E_t \mu(s_{t+1})] \\ \Phi(s_t) &= [B(s_t) - A(s_t) E_t \Phi(s_{t+1})]^{-1} C(s_t) \\ \Sigma(s_t) &= [B(s_t) - A(s_t) E_t \Phi(s_{t+1})]^{-1} D(s_t) \end{aligned}$$

This can also be written as

$$\begin{aligned} &\left[B(i) - A(i) \textstyle\sum_j \pi_{ij} \Phi(j)\right] \mu(i) = M(i) + A(i) \textstyle\sum_j \pi_{ij} \mu(j) \\ &\left[B(i) - A(i) \textstyle\sum_j \pi_{ij} \Phi(j)\right] \Phi(i) = C(i) \\ &\left[B(i) - A(i) \textstyle\sum_j \pi_{ij} \Phi(j)\right] \Sigma(i) = D(i) \end{aligned}$$

The key equation to solve is

$$\left[B(i) - A(i) \textstyle\sum_j \pi_{ij} \Phi(j)\right] \Phi(i) = C(i)$$

and the rest follow. If, like in our model, A, B and C do not depend on the regime, then Φ does not depend on regime either. The equation simplifies to a well-known generalized eigenvalues problem. We re-write it as

$$A\Phi\Phi - B\Phi + C = 0$$

and could apply familiar solution algorithms. Then we solve for $\mu(s_t)$ and $\Sigma(s_t)$ using

$$\begin{aligned} [B - A\Phi] \mu(i) &= M(i) + A(i) \textstyle\sum_j \pi_{ij} \mu(j) \\ [B - A\Phi] \Sigma(i) &= D(i) \end{aligned}$$

B.2 Bond prices with regime switching

Here we present the most general case for computing bond prices, adapted from Bansal and Zhou (2002). The state vector follows the process

$$x_{t+1} = \mu(s_{t+1}) + \Phi(s_{t+1}) x_t + \Sigma(s_{t+1}) \sqrt{\Psi_t} \varepsilon_{t+1}$$

$$\Psi_t = \operatorname{diag}\{\mathcal{A} + \mathcal{B} x_t\}$$

$$\mathcal{A} \in \mathbb{R}^{n_\varepsilon}$$
$$\mathcal{B} \in \mathbb{R}^{n_\varepsilon \times n_x}$$

In our model, we have $\Phi(s_{t+1}) = \Phi$, $\Sigma(s_{t+1}) = \Sigma$, $\mathcal{A} = I$, $\mathcal{B} = 0$.

The reduced form equation for the pricing kernel is

$$q_{t,t+1} = \lambda_{t+1} - \lambda_t - \pi_{t+1} - \sigma z_{t+1}$$

$$q_{t,t+1} = (e_\lambda - e_\pi - \sigma e_z)\,\mu(s_{t+1}) + [(e_\lambda - e_\pi - \sigma e_z)\,\Phi(s_{t+1}) - e_\lambda]\,x_t + (e_\lambda - e_\pi - \sigma e_z)\,\Sigma(s_{t+1}) \sqrt{\Psi_t} \varepsilon_{t+1}$$

$$q_{t,t+1} = m(s_{t+1}) + f(s_{t+1}) x_t + \sigma_m(s_{t+1}) \sqrt{\Psi_t} \varepsilon_{t+1}$$

$$\exp(b_t^n(s_t)) = \mathbb{E}\left[\exp\left(q_{t,t+1} + b_{t+1}^{n-1}\right) | \mathcal{F}_t\right]$$

$$E_t(\cdot) \equiv \mathbb{E}[\cdot | \mathcal{F}_t]$$

$$n = 1 \Rightarrow b_{t+1}^{n-1} = 0$$

$$\exp\left(b_t^1(s_t)\right) = \mathbb{E}[\exp(q_{t,t+1}) | \mathcal{F}_t] = \mathbb{E}[\mathbb{E}[\exp(q_{t,t+1}) | \mathcal{F}_t, s_{t+1}] | \mathcal{F}_t]$$

$$\exp\left(b_t^1(i)\right) = \textstyle\sum_j \pi_{ij} \mathbb{E}[\exp(q_{t,t+1}) | \mathcal{F}_t, j]$$

$$\exp\left(b_t^1(i)\right) = \textstyle\sum_j \pi_{ij} \exp\left\{\mathbb{E}(q_{t,t+1} | \mathcal{F}_t, j) + \frac{1}{2} Var(q_{t,t+1} | \mathcal{F}_t, j)\right\}$$

$$\exp\left(b_t^1(i)\right) = \textstyle\sum_j \pi_{ij} \exp\left\{m(j) + f(j) x_t + \frac{1}{2} \sigma_m(j) \operatorname{diag}(\sigma_m(j)) [\mathcal{A} + \mathcal{B} x_t]\right\}$$

$$\exp\left(b_t^1(i)\right) = \sum\nolimits_j \pi_{ij} \exp\left\{m(j) + \frac{1}{2}\sigma_m(j)\,\mathrm{diag}\left(\sigma_m(j)\right)\mathcal{A}\right\}$$
$$\times \exp\left\{\left[f(j) + \frac{1}{2}\sigma_m(j)\,\mathrm{diag}\left(\sigma_m(j)\right)\mathcal{B}\right]x_t\right\}$$

$$b_t^1(i) \approx \sum\nolimits_j \pi_{ij}\left\{m(j) + \frac{1}{2}\sigma_m(j)\,\mathrm{diag}\left(\sigma_m(j)\right)\mathcal{A} + \left[f(j) + \frac{1}{2}\sigma_m(j)\,\mathrm{diag}\left(\sigma_m(j)\right)\mathcal{B}\right]x_t\right\}$$

$$A_1(i) = \sum\nolimits_j \pi_{ij}\left[m(j) + \frac{1}{2}\sigma_m(j)\,\mathrm{diag}\left(\sigma_m(j)\right)\mathcal{A}\right]$$
$$B_1(i) = \sum\nolimits_j \pi_{ij}\left[f(j) + \frac{1}{2}\sigma_m(j)\,\mathrm{diag}\left(\sigma_m(j)\right)\mathcal{B}\right]$$

We guess that prices are of the form

$$b_t^n(s_t) = A_n(s_t) + B_n(s_t)\,x_t$$

$$\exp\left(b_t^n(s_t)\right) = \mathbb{E}\left\{\mathbb{E}\left[\exp\left(q_{t,t+1} + b_{t+1}^{n-1}(s_{t+1})\right)\middle|\,\mathcal{F}_t, s_{t+1}\right]\middle|\,\mathcal{F}_t\right\}$$

Conditional on $\{\mathcal{F}_t, s_{t+1}\}$, $q_{t,t+1}$ and b_{t+1}^{n-1} are jointly normal since both depend on the normal shock ε_{t+1}. Then the log of the inner expectation is

$$\log\left(\mathbb{E}\left[\exp\left(q_{t,t+1} + b_{t+1}^{n-1}\right)\middle|\,\mathcal{F}_t, s_{t+1}\right]\right) = \mathbb{E}\left[q_{t,t+1} + b_{t+1}^{n-1}(s_{t+1})\middle|\,\mathcal{F}_t, s_{t+1}\right]$$
$$+\frac{1}{2}Var\left(q_{t,t+1} + b_{t+1}^{n-1}(s_{t+1})\middle|\,\mathcal{F}_t, s_{t+1}\right)$$

$$q_{t,t+1}+b_{t+1}^{n-1}(s_{t+1}) = m(s_{t+1})+f(s_{t+1})\,x_t+\sigma_m(s_{t+1})\sqrt{\Psi_t}\varepsilon_{t+1}+A_{n-1}(s_{t+1})+B_{n-1}(s_{t+1})\,x_{t+1}$$

$$q_{t,t+1} + b_{t+1}^{n-1}(s_{t+1}) = m(s_{t+1}) + f(s_{t+1})\,x_t + \sigma_m(s_{t+1})\sqrt{\Psi_t}\varepsilon_{t+1} + A_{n-1}(s_{t+1})$$
$$+ B_{n-1}(s_{t+1})\left[\mu(s_{t+1}) + \Phi(s_{t+1})\,x_t + \Sigma(s_{t+1})\sqrt{\Psi_t}\varepsilon_{t+1}\right]$$

$$\begin{aligned} q_{t,t+1} + b_{t+1}^{n-1}(s_{t+1}) = {} & m(s_{t+1}) + A_{n-1}(s_{t+1}) + B_{n-1}(s_{t+1})\,\mu(s_{t+1}) \\ & + \left[f(s_{t+1}) + B_{n-1}(s_{t+1})\,\Phi(s_{t+1})\right] x_t \\ & + \left[\sigma_m(s_{t+1}) + B_{n-1}(s_{t+1})\,\Sigma(s_{t+1})\right] \sqrt{\Psi_t}\varepsilon_{t+1} \end{aligned}$$

$$\begin{aligned} \mathbb{E}\left[q_{t,t+1} + b_{t+1}^{n-1}(s_{t+1}) \middle| \mathcal{F}_t, s_{t+1}\right] = {} & m(s_{t+1}) + A_{n-1}(s_{t+1}) + B_{n-1}(s_{t+1})\,\mu(s_{t+1}) \\ & + \left[f(s_{t+1}) + B_{n-1}(s_{t+1})\,\Phi(s_{t+1})\right] x_t \end{aligned}$$

$$\begin{aligned} \frac{1}{2} Var\left(q_{t,t+1} + b_{t+1}^{n-1}(s_{t+1}) \middle| \mathcal{F}_t, s_{t+1}\right) = {} & \frac{1}{2}\left[\sigma_m(s_{t+1}) + B_{n-1}(s_{t+1})\,\Sigma(s_{t+1})\right] \\ & \times \operatorname{diag}\left[\sigma_m(s_{t+1}) + B_{n-1}(s_{t+1})\,\Sigma(s_{t+1})\right]\left[\mathcal{A} + \mathcal{B}x_t\right] \end{aligned}$$

$$\begin{aligned} \frac{1}{2} Var\left(q_{t,t+1} + b_{t+1}^{n-1}(s_{t+1}) \middle| \mathcal{F}_t, s_{t+1}\right) = {} & \frac{1}{2}\left[\sigma_m(s_{t+1}) + B_{n-1}(s_{t+1})\,\Sigma(s_{t+1})\right] \\ & \times \operatorname{diag}\left[\sigma_m(s_{t+1}) + B_{n-1}(s_{t+1})\,\Sigma(s_{t+1})\right]\mathcal{A} \\ & + \frac{1}{2}\left[\sigma_m(s_{t+1}) + B_{n-1}(s_{t+1})\,\Sigma(s_{t+1})\right] \\ & \times \operatorname{diag}\left[\sigma_m(s_{t+1}) + B_{n-1}(s_{t+1})\,\Sigma(s_{t+1})\right]\mathcal{B}x_t \end{aligned}$$

$$\theta(\mathcal{F}_t, s_{t+1}) \equiv \mathbb{E}\left[\exp\left(q_{t,t+1} + b_{t+1}^{n-1}\right) \middle| \mathcal{F}_t, s_{t+1}\right]$$

$$\begin{aligned} \log\left[\theta(\mathcal{F}_t, s_{t+1})\right] = {} & m(s_{t+1}) + A_{n-1}(s_{t+1}) + B_{n-1}(s_{t+1})\,\mu(s_{t+1}) \\ & + \frac{1}{2}\left[\sigma_m(s_{t+1}) + B_{n-1}(s_{t+1})\,\Sigma(s_{t+1})\right] \operatorname{diag}\left[\sigma_m(s_{t+1}) + B_{n-1}(s_{t+1})\,\Sigma(s_{t+1})\right]\mathcal{A} \\ & + \left\{ \begin{array}{c} \left[f(s_{t+1}) + B_{n-1}(s_{t+1})\,\Phi(s_{t+1})\right] \\ +\frac{1}{2}\left[\sigma_m(s_{t+1}) + B_{n-1}(s_{t+1})\,\Sigma(s_{t+1})\right] \operatorname{diag}\left[\sigma_m(s_{t+1}) + B_{n-1}(s_{t+1})\,\Sigma(s_{t+1})\right]\mathcal{B} \end{array} \right\} x_t \end{aligned}$$

$$\exp\left(b_t^n(s_t)\right) = \mathbb{E}\left\{ \theta(\mathcal{F}_t, s_{t+1}) \middle| \mathcal{F}_t \right\}$$

$$\exp\left(b_t^n(i)\right) = \textstyle\sum_j \pi_{ij}\theta(i,j)$$

$$b_t^n(i) \approx \textstyle\sum_j \pi_{ij} \log\left[\theta(i,j)\right]$$

$$A_n(i) + B_n(i)\, x_t \approx \textstyle\sum_j \pi_{ij} \log\left[\theta(i,j)\right]$$

$$A_n(i) = \textstyle\sum_j \pi_{ij} \left\{ \begin{array}{c} m(j) + A_{n-1}(j) + B_{n-1}(j)\, \mu(j) \\ +\frac{1}{2}\left[\sigma_m(j) + B_{n-1}(j)\, \Sigma(j)\right] \operatorname{diag}\left[\sigma_m(j) + B_{n-1}(j)\, \Sigma(j)\right] \mathcal{A} \end{array} \right\}$$

$$B_n(i) = \textstyle\sum_j \pi_{ij} \left\{ \begin{array}{c} \left[f(j) + B_{n-1}(j)\, \Phi(j)\right] \\ +\frac{1}{2}\left[\sigma_m(j) + B_{n-1}(j)\, \Sigma(j)\right] \operatorname{diag}\left[\sigma_m(j) + B_{n-1}(j)\, \Sigma(j)\right] \mathcal{B} \end{array} \right\}$$

B.3 Computing model moments

B.3.1 Conditional expectation $\mathbb{E}\left(X_t | s_t\right)$

$$X_{t+1} = \mu(s_{t+1}) + \Phi X_t + \Sigma(s_{t+1})\, \varepsilon_{t+1}$$

$$\mathbb{E}\left(X_{t+1} | s_{t+1}\right) = \mathbb{E}\left(\mu(s_{t+1}) | s_{t+1}\right) + \Phi \mathbb{E}\left(X_t | s_{t+1}\right)$$

$$\mathbb{E}\left(X_t | s_{t+1} = i\right) = \textstyle\sum_{j=1}^{n_s} \mathbb{E}\left(X_t | s_t\right) \mathbb{P}\left(s_t = j | s_{t+1} = i\right)$$

$$\mathbb{P}\left(s_t = j | s_{t+1} = i\right) \equiv b_{ij} = p_{ji} \frac{\pi_j}{\pi_i}$$

$$\begin{aligned} p_{ji} &= \mathbb{P}\left(s_{t+1} = i | s_t = j\right) \\ \pi_j &= \mathbb{P}\left(s_t = j\right) \end{aligned}$$

$$\mathbb{E}\left(X_{t+1} | s_{t+1} = i\right) = \mathbb{E}\left(\mu(s_{t+1}) | s_{t+1}\right) + \Phi \textstyle\sum_{j=1}^{n_s} \mathbb{E}\left(X_t | s_t = j\right) b_{ij}$$

$$\left[\mathbb{E}\left(X_{t+1} | s_{t+1} = i\right)\right]_{n_x \times 1} = \left[\mathbb{E}\left(\mu(s_{t+1}) | s_{t+1}\right)\right]_{n_x \times 1} + \left[\Phi\right]_{n_x \times n_x} \left[\textstyle\sum_{j=1}^{n_s} \left[\mathbb{E}\left(X_t | s_t = j\right)\right]_{n_x \times 1} \left[b_{ij}\right]_{1 \times 1}\right]$$

$$\tilde{\mathbb{E}}\left(X_{t+1} | s_t\right) \equiv \left[\begin{array}{ccc} \mathbb{E}\left(X_{t+1} | s_t = 1\right) & \cdots & \mathbb{E}\left(X_{t+1} | s_t = n_s\right) \end{array}\right]_{n_x \times n_s}$$

$$\tilde{\mu}(s_t) \equiv \left[\begin{array}{ccc} \mu(1) & \cdots & \mu(n_s) \end{array}\right]_{n_x \times n_s}$$

$$\left[\tilde{\mathbb{E}}\left(X_t|\, s_t\right)\right]_{n_x\times n_s}=\left[\tilde{\mu}\left(s_t\right)\right]_{n_x\times n_s}+\left[\Phi\right]_{n_x\times n_x}\left[\tilde{\mathbb{E}}\left(X_t|\, s_t\right)\right]_{n_x\times n_s}\left[B'\right]_{n_s\times n_s}$$

$$\begin{aligned}\left[\begin{array}{ccc}\mathbb{E}\left(X_t|\,1\right) & \cdots & \mathbb{E}\left(X_t|\,n_s\right)\end{array}\right] &= \left[\begin{array}{ccc}\mu\left(1\right) & \cdots & \mu\left(n_s\right)\end{array}\right] \\ &+\Phi\left[\begin{array}{ccc}\mathbb{E}\left(X_t|\,1\right) & \cdots & \mathbb{E}\left(X_t|\,n_s\right)\end{array}\right]\left[\begin{array}{ccccc} b_{11} & \cdots & b_{i1} & \cdots & b_{n_s1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ b_{1j} & \cdots & b_{ij} & \cdots & b_{n_sj} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ b_{1n_s} & \cdots & b_{in_s} & \cdots & b_{n_sn_s}\end{array}\right]\end{aligned}$$

$$\tilde{\mathbb{E}}\left(X_t|\, s_t\right)=\tilde{\mu}\left(s_t\right)+\Phi\tilde{\mathbb{E}}\left(X_t|\, s_t\right)B'$$

$$B\equiv\left[b_{ij}\right]$$

$$vec\left[\tilde{\mathbb{E}}\left(X_t|\, s_t\right)\right]=vec\left[\tilde{\mu}\left(s_t\right)\right]+vec\left[\Phi\tilde{\mathbb{E}}\left(X_t|\, s_t\right)B'\right]$$

$$vec\left[\tilde{\mathbb{E}}\left(X_t|\, s_t\right)\right]=vec\left[\tilde{\mu}\left(s_t\right)\right]+\left[B\otimes\Phi\right]vec\left[\tilde{\mathbb{E}}\left(X_t|\, s_t\right)\right]$$

$$vec\left[\tilde{\mathbb{E}}\left(X_t|\, s_t\right)\right]=\left[I-B\otimes\Phi\right]^{-1}vec\left[\tilde{\mu}\left(s_t\right)\right]$$

B.3.2 Conditional variances

$$X_{t+1}X_{t+1}'=\left[\mu\left(s_{t+1}\right)+\Phi\left(s_{t+1}\right)X_t+\Sigma\left(s_{t+1}\right)\varepsilon_{t+1}\right]\left[\mu\left(s_{t+1}\right)+\Phi\left(s_{t+1}\right)X_t+\Sigma\left(s_{t+1}\right)\varepsilon_{t+1}\right]'$$

$$\begin{aligned}\mathbb{E}\left[X_{t+1}X_{t+1}'\middle|\, s_{t+1}\right] &= \mu\left(s_{t+1}\right)\mu\left(s_{t+1}\right)'+\Sigma\left(s_{t+1}\right)\Sigma\left(s_{t+1}\right)'+\Phi\left(s_{t+1}\right)\mathbb{E}\left[X_tX_t'|\, s_{t+1}\right]\Phi\left(s_{t+1}\right)' \\ &+\mu\left(s_{t+1}\right)\mathbb{E}\left[X_t|\, s_{t+1}\right]'\Phi\left(s_{t+1}\right)'+\Phi\left(s_{t+1}\right)\mathbb{E}\left[X_t|\, s_{t+1}\right]\mu\left(s_{t+1}\right)'\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[X_{t+1}X_{t+1}'\middle|\, i\right] &= \Phi\left(s_{t+1}\right)\mathbb{E}\left[X_tX_t'|\, s_{t+1}=i\right]\Phi\left(s_{t+1}\right)'+\mu\left(i\right)\mu\left(i\right)'+\Sigma\left(i\right)\Sigma\left(i\right)' \\ &+\mu\left(i\right)\mathbb{E}\left[X_t|\, s_{t+1}=i\right]'\Phi\left(i\right)'+\Phi\left(i\right)\mathbb{E}\left[X_t|\, s_{t+1}=i\right]\mu\left(i\right)'\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[X_{t+1}X_{t+1}'\middle|\, i\right] &= \Phi\left(i\right)\mathbb{E}\left[X_tX_t'|\, s_{t+1}=i\right]\Phi\left(i\right)'+G\left(i\right) \\ G\left(i\right) &\equiv \mu\left(i\right)\mu\left(i\right)'+\Sigma\left(i\right)\Sigma\left(i\right)'+\mu\left(i\right)\mathbb{E}\left[X_t|\, s_{t+1}=i\right]'\Phi\left(i\right)'+\Phi\left(i\right)\mathbb{E}\left[X_t|\, s_{t+1}=i\right]\mu\left(i\right)'\end{aligned}$$

$$\mathbb{E}\left[X_{t+1}X_{t+1}'\middle| i\right] = G\left(i\right) + \Phi\left(i\right)\textstyle\sum_j \mathbb{E}\left[X_tX_t'\middle| j\right]\mathbb{P}\left(s_t = j\middle| s_{t+1} = i\right)\Phi\left(i\right)'$$

$$\mathbb{E}\left[X_tX_t'\middle| i\right] = G\left(i\right) + \textstyle\sum_j b_{ij}\Phi\left(i\right)\mathbb{E}\left[X_tX_t'\middle| j\right]\Phi\left(i\right)'$$

$$vec\left(\mathbb{E}\left[X_tX_t'\middle| i\right]\right) = vec\left(G\left(i\right)\right) + \textstyle\sum_j b_{ij} vec\left[\Phi\left(i\right)\mathbb{E}\left[X_tX_t'\middle| j\right]\Phi\left(i\right)'\right]$$

$$vec\left(\mathbb{E}\left[X_tX_t'\middle| i\right]\right) = vec\left(G\left(i\right)\right) + \textstyle\sum_j b_{ij}\left(\Phi\left(i\right)\otimes\Phi\left(i\right)\right) vec\left(\mathbb{E}\left[X_tX_t'\middle| j\right]\right)$$

$$vec\left(\mathbb{E}\left[X_tX_t'\middle| i\right]\right) = vec\left(G\left(i\right)\right) + \left(\Phi\left(i\right)\otimes\Phi\left(i\right)\right) vec\left(\textstyle\sum_j b_{ij}\mathbb{E}\left[X_tX_t'\middle| j\right]\right)$$

$$\mathbb{E}\left[X_tX_t'\middle| i\right] \equiv V\left(i\right)$$

$$vec\left(V\left(i\right)\right) = vec\left(G\left(i\right)\right) + \left(\Phi\left(i\right)\otimes\Phi\left(i\right)\right) vec\left(\textstyle\sum_j b_{ij}V\left(j\right)\right)$$

$$vec\left(V\right) = \begin{bmatrix} \vdots \\ vec\left(V\left(i\right)\right) \\ \vdots \end{bmatrix}$$

$$\begin{aligned} \tilde{V} &\equiv \begin{bmatrix} \cdots & vec\left(V\left(i\right)\right) & \cdots \end{bmatrix} \\ vec\left(\tilde{V}\right) &= vec\left(V\right) \end{aligned}$$

$$vec\left(G\right) = \begin{bmatrix} \vdots \\ vec\left(G\left(i\right)\right) \\ \vdots \end{bmatrix}$$

$$\mathbf{\Phi} \equiv \begin{bmatrix} \ddots & 0 & 0 \\ 0 & \Phi\left(i\right)\otimes\Phi\left(i\right) & 0 \\ 0 & 0 & \ddots \end{bmatrix}$$

$$vec\left(V\right) = vec\left(G\right) + \mathbf{\Phi} vec\left(\tilde{V}B'\right)$$

$$vec\left(V\right) = vec\left(G\right) + \mathbf{\Phi} vec\left(I_{n_x^2}\tilde{V}B'\right)$$

$$vec\left(I_{n_x}\tilde{V}B'\right) = \left(B \otimes I_{n_x^2}\right) vec\left(V\right)$$

$$vec\left(V\right) = vec\left(G\right) + \mathbf{\Phi}\left(B \otimes I_{n_x^2}\right) vec\left(V\right)$$

$$vec\left(V\right) - \mathbf{\Phi}\left(B \otimes I_{n_x^2}\right) vec\left(V\right) = vec\left(G\right)$$

$$vec\left(V\right) = \left[I_{n_s n_x^2} - \mathbf{\Phi}\left(B \otimes I_{n_x^2}\right)\right]^{-1} vec\left(G\right)$$

B.4 Model estimation with the Hamilton-Kim filter

Here we present our likelihood computation algorithm, adapted from Kim and Nelson (1999).

B.4.1 Model specification

Measurement equation

$$Y_t = A\left(s_t\right) + B\left(s_t\right) X_t + e_t$$

Transition equation We here present the general case where there is regime switching in all of the model's parameters.

$$X_t = \mu\left(s_t\right) + \Phi\left(s_t\right) X_{t-1} + \Sigma\left(s_t\right) \upsilon_t$$
$$\begin{pmatrix} e_t \\ \upsilon_t \end{pmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} R & 0 \\ 0 & I \end{bmatrix}\right)$$

The parameters are dependent on unobserved discrete-valued S-state Markov switching variable s_t $\left(s_t = 1, 2, ..., S\right)$ with transition probability matrix

$$\Pi = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1S} \\ p_{21} & p_{22} & \cdots & p_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ p_{S1} & p_{S2} & \cdots & p_{SS} \end{pmatrix}$$

where $p_{ij} = \mathbb{P}\left[s_t = i \middle| s_{t-1} = j\right]$ with $\sum_{i=1}^{S} p_{ij} = 1$ for all j. In what follows we denote $\Phi\left(s_t = j\right)$ with Φ_j, $A\left(s_t = j\right)$ with A_j and $B\left(s_t = j\right)$ with B_j.

B.4.2 Algorithm

To compute $\mathcal{L}\left(Y_t | Y_{t-1}, \tilde{\Theta}\right)$ we need to make an inference about the unobserved states X_t. The proposed algorithm is based on the Kalman filter and calculates S^2 forecasts for each date t corresponding to every possible combination of past and future states i and j, as well as S^2 different mean squared error matrices. Let $X^{i,j}_{t|t-1}$ denote the predicted values for the states given the information at time $t-1$ and given that $s_t = j$ and $s_{t-1} = i$.

$$X^{(i,j)}_{t|t-1} = \mathbb{E}\left[X_t | Y^{t-1}, s_t = j, s_{t-1} = i\right]$$

Let $P^{i,j}_{t|t-1}$ denote their mean square error:

$$P^{(i,j)}_{t|t-1} = E\left[\left(X_t - X_{t|t-1}\right)\left(X_t - X_{t|t-1}\right)' \middle| Y^{t-1}, s_t = j, s_{t-1} = i\right]$$

Conditional on $s_{t-1} = i$ and $s_t = j$ the Kalman filter algorithm works as follows

Prediction

$$X^{(i,j)}_{t|t-1} = \mu_j + \Phi_j X^i_{t-1|t-1}$$

$$P^{(i,j)}_{t|t-1} = \Phi_j P^i_{t-1|t-1}\Phi'_j + \Sigma_j \Sigma'_j$$

The conditional forecast errors of the observations are

$$\eta^{(i,j)}_{t|t-1} \equiv Y_t - A_j - B_j X^{(i,j)}_{t|t-1}$$

and the conditional variance of forecast errors is

$$\sigma^{(i,j)}_{t|t-1} \equiv B_j P^{(i,j)}_{t|t-1} B'_j + R$$

Updating We denote the updated state by

$$X^{i,j}_{t|t} = \mathbb{E}\left[X_t | Y^t, s_t = j, s_{t-1} = i\right]$$

and its mean square error by

$$P^{i,j}_{t|t} = \mathbb{E}\left[\left(X_t - X^{i,j}_{t|t}\right)\left(X_t - X^{i,j}_{t|t}\right)' \middle| Y^t, s_t = j, s_t = i\right]$$

The updated state conditional on the current regime only is denoted by

$$X^j_{t|t} = \mathbb{E}\left[X_t | Y^t, s_t = j\right]$$

and its mean square error by

$$P^j_{t|t} = \mathbb{E}\left[\left(X_t - X^{i,j}_{t|t}\right)\left(X_t - X^{i,j}_{t|t}\right)' \middle| Y^t, s_t = j\right]$$

Then we can compute those quantities as follows:

$$X_{t|t}^{(i,j)} = X_{t|t-1}^{(i,j)} + P_{t|t-1}^{(i,j)} B_j' \left[\sigma_{t|t-1}^{(i,j)} \right]^{-1} \eta_{t|t-1}^{(i,j)}$$

$$P_{t|t}^{(i,j)} = P_{t|t-1}^{(i,j)} \left(I - B_j' \left[\sigma_{t|t-1}^{(i,j)} \right]^{-1} B_j \left[P_{t|t-1}^{(i,j)} \right]' \right)$$

As can be seen each iteration produces an S-fold increase in the number of cases to consider. The key is to collapse terms in the right way at the right time. It remains to somehow reduce the $(S \times S)$ posteriors $\left(X_{t|t}^{(i,j)}, P_{t|t}^{(i,j)} \right)$ into S posteriors $\left(X_{t|t}^{j}, P_{t|t}^{j} \right)$ Consider the following approximation. If $X_{t|t}^{(i,j)}$ represented $\mathbb{E}\left[X_t | Y_t, s_{t-1} = i, s_t = j \right]$, then

$$X_{t|t}^{j} = \frac{\sum_{i=1}^{S} \mathbb{P}\left[s_{t-1} = i, s_t = j | Y^t \right] X_{t|t}^{(i,j)}}{\mathbb{P}\left[s_t = j | Y^t \right]}$$

where $X_{t|t}^{j}$ would represent $\mathbb{E}\left[X_t | Y^t, s_t = j \right]$. In this case denote

$$\Delta_t = \frac{\mathbb{P}\left[s_{t-1} = i, s_t = j | Y^t \right]}{\mathbb{P}\left[s_t = j | Y^t \right]}$$

$P_{t|t}^{j}$, the mean-squared error matrix of X_t conditional on Y^t and $s_t = j$ could be derived in the following way:

$$\begin{aligned}
P_{t|t}^{j} &= \mathbb{E}\left[\left(X_t - \mathbb{E}\left[X_t | Y^t, s_t = j \right] \right) \left(X_t - \mathbb{E}\left[X_t | Y^t, s_t = j \right] \right)' \middle| Y^t, s_t = j \right] \\
&= \mathbb{E}\left[\left(X_t - X_{t|t}^{j} \right) \left(X_t - X_{t|t}^{j} \right)' \middle| Y^t, s_t = j \right] \\
&= \sum_{i=1}^{S} \Delta_t \mathbb{E}\left[\left(X_t - X_{t|t}^{j} \right) \left(X_t - X_{t|t}^{j} \right)' \middle| Y^t, s_{t-1} = i, s_t = j \right] \\
&= \sum_{i=1}^{S} \Delta_t \mathbb{E}\left[\left(X_t - X_{t|t}^{(i,j)} + X_{t|t}^{(i,j)} - X_{t|t}^{j} \right) \left(X_t - X_{t|t}^{(i,j)} + X_{t|t}^{(i,j)} - X_{t|t}^{j} \right)' \middle| Y^t, s_{t-1} = i, s_t = j \right]
\end{aligned}$$

Note that given the information set $(Y^t, s_{t-1} = i, s_t = j)$ the only unknown is X_t. Then we have

$$\begin{aligned}
P_{t|t}^{j} &= \sum_{i=1}^{S} \Delta_t \mathbb{E}\left[\left(X_t - X_{t|t}^{(i,j)} + X_{t|t}^{(i,j)} - X_{t|t}^{j} \right) \left(X_t - X_{t|t}^{(i,j)} + X_{t|t}^{(i,j)} - X_{t|t}^{j} \right)' \middle| Y^t, s_{t-1} = i, s_t = j \right] \\
&= \sum_{i=1}^{S} \Delta_t \left\{ \left(\mathbb{E}\left[\left(X_t - X_{t|t}^{(i,j)} \right) \left(X_t - X_{t|t}^{(i,j)} \right)' \middle| Y^t, s_{t-1} = i, s_t = j \right] \right) + \left(X_{t|t}^{j} - X_{t|t}^{(i,j)} \right) \left(X_{t|t}^{j} - X_{t|t}^{(i,j)} \right)' \right\} \\
&+ \sum_{i=1}^{S} \Delta_t \left(\mathbb{E}\left[\left(X_t - X_{t|t}^{(i,j)} \right) \middle| Y^t, s_{t-1} = i, s_t = j \right] \right) \left(X_{t|t}^{(i,j)} - X_{t|t}^{j} \right)' \\
&+ \sum_{i=1}^{S} \Delta_t \left(X_{t|t}^{(i,j)} - X_{t|t}^{j} \right) \left(\mathbb{E}\left[\left(X_t - X_{t|t}^{(i,j)} \right) \middle| Y^t, s_{t-1} = i, s_t = j \right] \right)' \\
&= \sum_{i=1}^{S} \Delta_t \left\{ \left(\mathbb{E}\left[\left(X_t - X_{t|t}^{(i,j)} \right) \left(X_t - X_{t|t}^{(i,j)} \right)' \middle| Y^t, s_{t-1} = i, s_t = j \right] \right) + \left(X_{t|t}^{j} - X_{t|t}^{(i,j)} \right) \left(X_{t|t}^{j} - X_{t|t}^{(i,j)} \right)' \right\}
\end{aligned}$$

$$\mathbb{E}\left[\left(X_t - X_{t|t}^{(i,j)}\right)\left(X_t - X_{t|t}^{(i,j)}\right)'\middle| Y^t, s_{t-1} = i, s_t = j\right] = P_{t|t}^{(i,j)}$$

$$P_{t|t}^{j} = \sum_{i=1}^{S} \Delta_t \left[P_{t|t}^{(i,j)} + \left(X_{t|t}^{j} - X_{t|t}^{(i,j)}\right)\left(X_{t|t}^{j} - X_{t|t}^{(i,j)}\right)'\right]$$

$$P_{t|t}^{j} = \frac{\sum_{i=1}^{S} \mathbb{P}\left[s_{t-1} = i, s_t = j \middle| Y^t\right]\left[P_{t|t}^{(i,j)} + \left(X_{t|t}^{j} - X_{t|t}^{(i,j)}\right)\left(X_{t|t}^{j} - X_{t|t}^{(i,j)}\right)'\right]}{\mathbb{P}\left[s_t = j \middle| Y^t\right]}$$

So at the end of each iteration we collapse $S \times S$ posteriors into S posteriors using this approximation to make the filter operable. Notice however, that these collapsed posteriors involve approximations, as $X_{t|t}^{(i,j)}$ and $P_{t|t}^{(i,j)}$ (in the updating equations above) are not exactly equal to $\mathbb{E}\left[X_t \middle| Y^t, s_t = j, s_{t-1} = i\right]$ and $\mathbb{E}\left[\left(X_t - X_{t|t}^{j}\right)\left(X_t - X_{t|t}^{j}\right)'\middle| Y^t, s_t = j, s_{t-1} = i\right]$ since X_t conditional on Y^{t-1}, $s_t = j$, and $s_{t-1} = i$ is a mixture of normals for $t > 2$. However, this updating equation is still the linear projection of X_t on Y^t and $X_{t-1|t-1}^{i}$ given s_{t-1} and s_t. However, this updating equation is still not a linear projection of X_t on $Y^t \equiv \{Y_t, Y_{t-1}, ...\}$. since $X_{t-1|t-1}^{i}$ is a non-linear function of $Y^{t-1} \equiv \{Y_{t-1}, Y_{t-2}, ..\}$.

Inference on the probability terms via the Hamilton filter

STEP 1: At the beginning of t-th iteration, given $\mathbb{P}\left[s_{t-1} = i \middle| Y^{t-1}\right]$ for $i = 1, 2, ...S$, we can calculate

$$\mathbb{P}\left[s_t = j, s_{t-1} = i \middle| Y^{t-1}\right] = \mathbb{P}\left[s_t = j \middle| s_{t-1} = i\right] \mathbb{P}\left[s_{t-1} = i \middle| Y^{t-1}\right]$$

for $i, j = 1, 2, ...S$. where $\mathbb{P}\left[s_t = j \middle| s_{t-1} = i\right]$ is of course the transition probability.

STEP 2: Consider the joint density of Y_t, s_t and s_{t-1} :

$$f\left(Y_t, s_t = j, s_{t-1} = i \middle| Y^{t-1}\right) = f\left(Y_t \middle| s_t = j, s_{t-1} = i, Y^{t-1}\right) \mathbb{P}\left[s_t = j, s_{t-1} = i \middle| Y^{t-1}\right]$$

for $i, j = 1, 2, ...S$ from which the marginal density of Y_t is obtained by

$$\begin{aligned} f\left(Y_t \middle| Y^{t-1}\right) &= \sum_{j=1}^{S}\sum_{i=1}^{S} f\left(Y_t, s_t = j, s_{t-1} = i \middle| Y^{t-1}\right) \\ &= \sum_{j=1}^{S}\sum_{i=1}^{S} f\left(Y_t \middle| s_t = j, s_{t-1} = i, Y^{t-1}\right) \mathbb{P}\left[s_t = j, s_{t-1} = i \middle| Y^{t-1}\right] \end{aligned}$$

where the conditional density $f\left(Y_t \middle| s_t = j, s_{t-1} = i, Y^{t-1}\right)$ is obtained based on the prediction error decomposition

$$f\left(Y_t \middle| s_t = j, s_{t-1} = i, Y^{t-1}\right) = (2\pi)^{-\frac{N}{2}} \left|\sigma_{t|t-1}^{(i,j)}\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\eta_{t|t-1}^{(i,j)}\right)'\left(\sigma_{t|t-1}^{(i,j)}\right)^{-1}\eta_{t|t-1}^{(i,j)}\right\}$$

for $i, j = 1, 2, ...S$.where $\eta_{t|t-1}^{(i,j)}$ are computed from the above prediction equations.

STEP 3: Once Y_t is observed at the end of time t,we can update the probability term $\mathbb{P}\left[s_t = j, s_{t-1} = i\right| Y^{t-1}]$ to get

$$\begin{aligned}
\mathbb{P}\left[s_t = j, s_{t-1} = i\right| Y^t\right] &= \mathbb{P}\left[s_t = j, s_{t-1} = i\right| Y^{t-1}, Y_t\right] \\
&= \frac{f\left(Y_t, s_t = j, s_{t-1} = i\right| Y^{t-1})}{f\left(Y_t\right| Y^{t-1})} \\
&= \frac{f\left(Y_t\right| s_t = j, s_{t-1} = i, Y^{t-1}) f\left[s_t = j, s_{t-1} = i\right| Y^{t-1}]}{f\left(Y_t\right| Y^{t-1})} \\
&= \frac{f\left(Y_t\right| s_t = j, s_{t-1} = i, Y^{t-1}) \mathbb{P}\left[s_t = j, s_{t-1} = i\right| Y^{t-1}]}{f\left(Y_t\right| Y^{t-1})}
\end{aligned}$$

for $i, j = 1, 2, ...S$ and

$$\mathbb{P}\left[s_t = j\right| Y^{t-1}\right] = \sum_{i=1}^{S} \mathbb{P}\left[s_t = j, s_{t-1} = i\right| Y^{t-1}\right]$$

Full procedure

Then the likelihood computation procedure works as follows. Given the parameter vector $\tilde{\Theta}$, we start with an initial guess for the state vector $X^j_{0|0}$ and its mean square error $P^j_{0|0}$. Then, from period 1 to period T, we run the filter, which consists of three stages:

STAGE 1: Compute $X^{i,j}_{t|t-1}, P^{i,j}_{t|t-1}, \eta^{i,j}_{t|t-1}$ and $\sigma^{i,j}_{t|t-1}$ as well as the updates $X^{i,j}_{t|t}, P^{i,j}_{t|t}$ by running the Kalman filter conditional on the regimes $s_t = j$ and $s_{t-1} = i$ for each i and j.

STAGE 2: Compute $\mathbb{P}\left(s_t, s_{t-1}\right| Y^{t-1})$, the likelihood of the current observation y_t, $\mathcal{L}\left(y_t\right| Y^{t-1}, \Theta)$, $\mathbb{P}\left(s_t, s_{t-1}\right| Y^t)$ and $\mathbb{P}\left(s_t\right| Y^t)$ using the Hamilton filter.

STAGE 3: Use the approximations suggested by Kim (1994) to get $X^j_{t|t}, P^j_{t|t}$ and go back to Step 1.

Finally, we sum up the log-likelihood for each period to get the log-likelihood for the whole sample.