The yield curve and macroeconomic dynamics*

Peter Hördahl, Oreste Tristani and David Vestin**
European Central Bank
First draft: 6 July 2004
This draft: 31 May 2005

1 Introduction

Macroeconomic models are generally considered to be little useful in matching the stylised facts of yield data. On the one hand, the finance literature has lost interest in grounding term structure models on a macroeconomic framework and has chosen to rely on unobservable factor models (e.g. Duffie and Kan, 1996; Dai and Singleton, 2000). On the other hand, micofounded macroeconomic models with flexible prices have not proven to perform well in matching basic features of yield data (den Haan, 1995, representing the most recent statement of this conclusion).

Some recent empirical applications, however, have started to move towards a merger between the two fields. While providing a more intuitive explanation of term structure dynamics, such attempts at a joint modelling of yields and the macroeconomy can improve

---

*We wish to thank Federico Ravenna and Frank Schorfeide for useful comments and suggestions. A previous version of this paper, entitled “Monetary policy and the yield curve,” was presented at the Society for Computational Economics 2004 Conference in Amsterdam. The opinions expressed are personal and should not be attributed to the European Central Bank.

** European Central Bank, DG Research, Kaiserstrasse 29, D - 60311, Frankfurt am Main, Germany. E-mail: peter.hordahl@ecb.int, oreste.tristani@ecb.int, david.vestin@ecb.int.
yields forecasts (see Hördahl, Tristani and Vestin, 2005). These results are based on macroeconomic relations that are broadly consistent with the aggregate dynamics of well-known microfounded models. In this strand of literature, however, the merger between the macroeconomic and term structure models is not complete. The market prices of risk needed to price yields are in fact assumed exogenously, rather than being derived consistently with the microfounded structure.

In this paper we redress this shortcoming and analyse more closely the term structure implications of a relatively standard microfounded DSGE model with Calvo-style nominal rigidities, inflation indexation and habit persistence. In order to allow for a meaningful role of risk, the model is solved using a second order approximation following Schmitt-Grohé and Uribe (2004). We then study whether the model can match the key unconditional features of yield data characterised as puzzling in previous contributions: the positive and large slope of the curve, on average; and the similar volatility of yields across maturities.

The broad flavour of our results is to show that modern DSGE models are much closer to yield data than what one would think based on the aforementioned previous analyses. Various versions of the model can replicate the key features of yield data, while remaining broadly consistent with the unconditional second moments of macroeconomic data. Importantly, two persistent sources of exogenous uncertainty, the main one being a standard technology shock, are sufficient to make the model quite successful in matching the data. It is therefore conceivable that richer and more elaborate versions of the model could be brought even closer to the data.

Our general equilibrium model also allows us to derive a complete Fisher-type decomposition of nominal excess holding period returns – i.e. the excess return which can be earned holding a long term nominal bond for one period compared to the return on a 1-period nominal bond – into a real excess return and an inflation risk premium. Armed with such decomposition, we analyse the relationship between monetary policy and inflation risk premia. Our most striking result is that, compared to the case of an empirically plausible simple rule, optimal monetary policy reduces inflation risk premia, but results at the same time in a dramatic increase in real excess returns or, equivalently, in the slope of the yield curve. While based on the assumption of full monetary policy credibility and

---

1 Other attempts to merge macroeconomic and term structure modes are Rudebusch and Wu (2004) and Bekaert, Cho and Moreno (2003).
of a given and commonly known inflation target, this result suggests that caution should be used when interpreting the yield curve for macroeconomic purposes.

From a technical viewpoint, we derive analytical second order approximate solutions for bond prices consistent with the Schmitt-Grohé and Uribe (2004) formulation. This allows us to solve the macroeconomic model first, and compute yield prices in a second stage. Compared to the alternative of solving jointly for yields and macroeconomic variables using only numerical methods, our solution generates huge savings in terms of computational time. Our expressions also allow one to easily derive risk premia in any DSGE model, based on the sole knowledge of the loglinearised solution.

The paper is organised as follows. Section 2 summarises a few stylised facts on the term structure of interest rate, which we use as a benchmark for the evaluation of the theoretical model. The model is described in Section 3, while Section 4 presents the solution method and the analytical second order approximation to bond prices. Section 5 discusses the quantitative performance of the model and shows that matching the size of excess holding period returns is the most demanding empirical test. The decomposition of excess holding period returns and the analysis of the relationship between monetary policy and term premia are presented in Section 6. Section 7 concludes.

2 Stylized facts of bond yields: a quick review

Ideally, a microfounded model, such as the one we employ here, should be able to capture not only the dynamics of key macro variables, but it should also be able to capture salient features of the term structure of interest rates. These are well known from previous studies, but we briefly summarise them again in this section.

Looking at US nominal bond yields since the beginning of the 1960s, a number of key features become apparent (see Table 1a). First, on average, the term structure has been upward-sloping, with the mean of the 10-year yield exceeding the mean of the one-month interest rate by 120 basis points over the 1960Q1-1997Q2 period.\(^2\) Second, the term structure of yield volatilities has tended to be only slightly downward-sloping; the standard deviation of short-term interest rates (up to about one year) was around 2.7% while the corresponding value for the 10-year yield was around 2.4%. Third, yields have been highly

\(^2\)We stop in 1997 for consistency with our macroeconomic dataset.
persistent. The first-order sample autocorrelations of yields were above 0.9 across all maturities. Fourth, realized excess holding returns of bonds tended to be increasing with the maturity of the bonds. For example, while the average return earned from investing in a 6-month bond for one quarter, in excess of the 3-month interest rate, was around 30 basis points (in annualized terms), this excess return increased up to 76 basis points for 5-year bonds (also annualized). This would suggest that expected excess returns should also be increasing with maturity.

Table 1 (a): Summary yields statistics: full sample.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>3m</th>
<th>6m</th>
<th>1y</th>
<th>3y</th>
<th>5y</th>
<th>10y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>6.16</td>
<td>6.44</td>
<td>6.64</td>
<td>7.04</td>
<td>7.20</td>
<td>7.36</td>
</tr>
<tr>
<td>Std.Dev.</td>
<td>2.76</td>
<td>2.8</td>
<td>2.72</td>
<td>2.60</td>
<td>2.52</td>
<td>2.44</td>
</tr>
<tr>
<td>Autocorr.</td>
<td>0.91</td>
<td>0.92</td>
<td>0.92</td>
<td>0.94</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>ex-post xhpr</td>
<td>0.24</td>
<td>0.44</td>
<td>0.68</td>
<td>0.76</td>
<td>0.60</td>
<td></td>
</tr>
</tbody>
</table>

Quarterly US data. Source: ...

Table 1 (b): Summary yields statistics: subsample breakdown.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>3m</th>
<th>6m</th>
<th>1y</th>
<th>3y</th>
<th>5y</th>
<th>10y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.72</td>
<td>1.76</td>
<td>1.76</td>
<td>1.60</td>
<td>1.56</td>
<td>1.48</td>
</tr>
<tr>
<td>Std.Dev.</td>
<td>1.76</td>
<td>1.76</td>
<td>1.76</td>
<td>1.60</td>
<td>1.56</td>
<td>1.48</td>
</tr>
<tr>
<td>ex-post xhpr</td>
<td>0.16</td>
<td>0.20</td>
<td>0.08</td>
<td>-0.04</td>
<td>-0.04</td>
<td>-0.68</td>
</tr>
</tbody>
</table>

Quarterly US data. Source: ...

It could be argued that these unconditional moments should be calculated over a shorter sample period, because of the impact on the yield curve of possible structural breaks. Many authors, for example, have argued that a break in the Fed’s monetary policy rule took place most probably in 1979 (see e.g. Clarida, Galí and Gertler, 2000, and Orphanides, 2004). The so-called “monetarist experiment” of the initial years of Paul Volcker’s tenure as Chairman of the FOMC is also assumed to represent a specific regime.
by some authors. We analyse these subperiods in Table 1(b). The broad qualitative features of the term structure data are robust, even if quantitative variations are evident in the subsample analysis especially concerning ex-post excess returns.\(^3\)

Given these results, Table 1 focuses on full sample statistics because of the need for many observations to identify unconditional means of highly persistent variables.

### 3 The model

The model can be seen as a simplified version of that proposed in the Christiano, Eichenbaum and Evans (2001) model. It also borrows heavily from Woodford (2003).

The central feature of these models is the assumption of nominal rigidities, which allows for real effects of nominal shocks. The specification of the monetary policy rule becomes crucial in affecting economic dynamics.

The real side of the model will feature a continuum of households consuming a basked of goods, and providing labour services to firms. The firms are monopolistic competitors and only use labour for production.

The solution (i.e. how aggregate output and inflation respond to the state variables and the shocks in the economy) will pin down the stochastic behavior of the pricing kernel (SDF). This will then be used in a second step to price nominal and real bonds, hence allowing us to recover the term structure. We will show that only two shocks are sufficient to replicate the basic undonditional features of yield data.

#### 3.1 Consumers

Consumers maximize the discounted sum of the period utility

\[
U(C_t, H_t, L_t) = \frac{(C_t - hC_{t-1})^{1-\gamma}}{1-\gamma} - \int_0^1 \chi L_t (i)^\phi \, di
\]

where \(C\) is a consumption index satisfying

\[
C = \left( \int_0^1 C(i)^{\frac{\phi-1}{\phi}} \right)^{\frac{\phi}{\phi-1}}.
\]

\(^3\)The large time variability of \(xhpr\)’s is well known in finance-type studies – see e.g. Dai and Singleton () – and a puzzle by itself puzzle for microfounded models. We do not explore this aspect of the data here, since second order approximations yield constant excess returns by construction.
H is the habit stock and, by assumption, workers provide \( L(i) \) hours of labor to firm \( i \).

The elasticity of intratemporal substitution between goods, \( \theta \), should be strictly greater than 1. For consistency with Smets and Wouters (2003) and Christiano, Eichenbaum and Evans (2001), habit formation is modelled in difference form. Moreover, habit is internal, so that households care about their own lagged consumption. Habit persistence has been used to improve the models’ performance both in macroeconomics (Fuhrer, 2000, Amato and Laubach, 2004) and in finance (e.g. Constantinides, 1990, Campbell and Cochrane, 2003, Dai, 2003).

The households’ budget constraint is given by

\[
P_tC_t + B_t \leq \left(1 - \frac{\tau_t}{1+\tau_t}\right) \left(\int_0^1 w_t(i) L_t(i) \, di + \int_0^1 \Xi_t(i) \, di\right) + W_t
\]

with the price level \( P_t \) defined as the minimal cost of buying one unit of \( C_t \), hence equal to

\[
P_t = \left(\int_0^1 p(i)^{1-\theta_t}\right)^{-\frac{1}{1-\theta_t}}.
\]

In the budget constraint, \( B_t \) denotes end of period holdings of a complete portfolio of state contingent assets. \( W_t \) denotes the beginning of period value of the assets, \( w_t(i) \) is the nominal wage rate, \( L_t(i) \) is the supply of labor to firm \( i \) and \( \Xi_t(i) \) are the profits received from investment in firm \( i \). Following Steinsson (2003), we also introduce a stochastic income tax, which in equilibrium will lead to a trade-off between inflation and the output gap. We write the tax rate as \( \frac{\tau_t}{1+\tau_t} \) to ensure that the total tax is bounded between 0 and 1, given that we assume that

\[
\log \tau_t = \rho \log \tau_{t-1} + \varepsilon_t, \quad \varepsilon_t^T \sim N \left(0, \sigma_\tau^2\right).
\]

Finally, complete markets imply the existence of a unique pricing kernel, or stochastic discount factor, or state price deflator, denoted \( Q_{t,t+1} \). This implies that

\[
B_t = E_t \left(Q_{t,t+1}W_{t+1}\right).
\]

The first order conditions w.r.t intertemporal aggregate consumption allocation and labour supply are

\[
\left(1 - \frac{\tau_t}{1+\tau_t}\right) \frac{w_t(i)}{P_t} = \frac{v_{t,t}}{\Lambda_t}
\]

\[
\Lambda_t = (C_t - hC_{t-1})^{-\gamma} - \beta hE_t \left[(C_{t+1} - hC_t)^{-\gamma}\right]
\]

\[
Q_{t,t+1} = \beta \frac{P_t}{P_{t+1}} \frac{\Lambda_t}{\Lambda_{t+1}}.
\]
3.2 Firms

Turning to the firms, the production function is given by

\[ Y_t(i) = A_t L^\alpha \]

\[ A_t = A^\alpha_{t-1} e^{\varepsilon_t} \]

where \( A_t \) is a technology shock and \( \varepsilon_t^\alpha \) is a normally distributed shock with constant variance \( \sigma^2 \).

We assume Calvo (1983) contracts, so that firms face a constant probability \( \zeta \) of being unable to change their price at each time \( t \). Firms will take this constraint into account when trying to maximize expected profits, namely

\[ \max_{P^i_t} \mathbb{E}_t \sum_{s=t}^{\infty} \zeta^{s-t} Q_{t,s} \left( P^i_s Y^i_s - TC_s \right), \]

where \( TC \) denotes total costs and, as in Smets and Wouters (2003), firms not changing prices optimally are assumed to modify them using a rule of thumb that indexes them partly to lagged inflation and partly to steady-state inflation, namely \( P^i_t (\Pi)^{1-i} \left( \frac{P_{s-1}}{P_{s-1}} \right)^{i} \), where \( 0 \leq i \leq 1 \).

Indexation will turn out to play a negligible role on our results. We introduce it for two reasons. First, inflation will be driven to some extent by lagged inflation, which is an empirically plausible hypothesis. Second, the firms not allowed to update their prices optimally will nevertheless change based on aggregate past inflation, which means that the welfare consequences of the suboptimal policy shocks we will consider (shocks to the target rate of inflation) will be limited, as the distribution prices will be more compressed. We also introduce the inflation target indexation component to ensure that the steady-state welfare loss arising from temporary deviations from zero in the inflation target is zero for any value of \( \iota \) in the \( 0 - 1 \) range.

Under the assumption that firms are perfectly symmetric in all other respects than the ability to change prices, all firms that do get to change price will set it at the same optimal level \( P_t^* \). Furthermore, the average level of prices in the group that does not change prices is equal to the average price level from the last period (since, by a law of large number type of argument, those firms are drawn from the pool of all firms). We can therefore
characterise firms’ decisions as implying

\[
\left( \frac{P_t^*}{P_t} \right)^{1-\theta(1-\frac{\phi}{\alpha})} = \frac{\phi \chi_0}{\alpha (\theta - 1)} K_{2,t}^{\frac{1}{\alpha}} \frac{\Pi^{1-i_1} \Pi_{t+1}^{1-i_2}}{\Pi_t} \frac{1-\theta}{(1-\zeta)} \frac{1}{\sigma_t}
\]

\[
P_t^* = \left( 1 - \zeta \frac{\Pi^{1-i_1} \Pi_{t+1}^{1-i_2}}{\Pi_t} \right)^{1-\theta} \frac{1}{(1-\zeta)}
\]

\[
K_{2,t} = \frac{A_t \pi_t^{\phi}}{1-\frac{\tau_t}{1+\tau_t}} Y_t^\phi + E_t \xi \Pi^{1-i_1} \Pi_{t+1}^{1+i_2} Q_{t,t+1} K_{2,t+1} \Pi_t^{1-\theta} \Pi_{t+1}^{\phi}
\]

\[
K_{1,t} = Y_t + E_t \xi \Pi^{1-i_1} \Pi_{t+1}^{1+i_2} Q_{t,t+1} K_{1,t+1} \Pi_t^{1-\theta} \Pi_{t+1}^{\phi}
\]

where \(\Pi_t\) is the inflation rate defined as \(\Pi_t = \frac{P_t}{P_{t-1}}\) (similar expressions are derived in Ascari, 2004, for the case without indexation).

### 3.3 Monetary policy

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

\[
i_t = \frac{1 - \rho_t}{\beta} + \delta (\pi_t - \pi_t^*) + \kappa (y_t - y_t^n) + \rho_t i_{t-1} + \eta_t \tag{1}
\]

\[
\pi_t^* = \rho_t \pi_t^{\pi_t^*} + \varepsilon_t^{\pi_t^*} \tag{2}
\]

where \(i_t\) is the logarithm of the gross nominal interest rate (which is riskless in nominal terms), \(\pi_t^*\) is the inflation target, \(\eta_t\) is a policy shock, \(y_t^n\) is the logarithm of the level of output that would emerge in the equilibrium characterized by flexible prices and \(\eta_t\) and \(\varepsilon_t^{\pi_t^*}\) are white noise shocks with variances \(\sigma_{\eta_t}^2\) and \(\sigma_{\pi_t^*}^2\), respectively.

Variants of this rule have been found to be empirically-plausible by a number of authors, such as Clarida, Gali and Gertler (2000), Smets and Wouters (2005) and Hördahl, Tristani and Vestin (2005). The particular formulation we adopt has the advantage of allowing us to explore easily some interesting alternatives proposed in the literature. One is a superinertial rule, namely a rule with \(\rho_t > 1\), which has desirable theoretical features (see Woorford, 2003). Another variant of the rule is a version which implements optimal monetary policy within the model, when only technology shocks are assumed to affect the economy. Woodford (2003) has demonstrated that optimal policy involves tracking the interest rate that would prevail in the equilibrium characterized by flexible prices. If we denote such rate as \(r_t^n\), the rule becomes simply \(i_t = r_t^n\).\(^4\)

\(^4\)The natural rate is the solution of \(\frac{1}{\exp(i_t)} = E_t \left( \beta \frac{Y_t^n - h Y_{t+1}^n}{Y_t^n - h Y_{t+1}^n} \right) \gamma - \beta \eta_t \left( \frac{Y_t^n - h Y_{t+1}^n}{Y_t^n - h Y_{t+1}^n} \right) \gamma \right) \) where \(Y_t^n\)
Incidentally, we assume that the long run inflation target is zero in order to better isolate the properties of the model in terms of risk premia. A non-zero long-run inflation target would obviously have an impact, in particular, on the spread between nominal and real yields. However, it would not modify the risk properties of the model.

4 Solving the model

We solve the model using a second order approximation around the non-stochastic steady state. The model dynamics will then be described by two equations: a quadratic law of motion for the predetermined variables of the model and a quadratic relationship linking each non-predetermined variable to the predetermined variables. Note that the predetermined variables include lagged inflation, lagged output (consumption), lagged natural output and the lagged nominal interest rate.

4.1 Macro-model solution

The solution of the macro-model is obtained numerically using the approach (and the routines) described in Schmitt-Grohé and Uribe (2002). For the vector $\tilde{x}_t$ of predetermined variables, we obtain the (partly endogenous) law of motion

$$\tilde{x}_{t+1} = c_1 \tilde{x}_t + \frac{1}{2} \begin{bmatrix} \tilde{x}'_t c_2 [i] \tilde{x}_t \end{bmatrix} + \frac{1}{2} c_0 \sigma^2 + \eta \sigma \varepsilon_{t+1}$$

where $c_1$, $c_0$ and $\eta$ are $n_x \times n_x$, $n_x \times 1$, and $n_x \times n_x$ matrices ($n_x$ the size of the shock vector $\varepsilon_{t+1}$), respectively, and $c_2 [i]$ is the $i$-th page of the tensor $c_2$ (a symmetric matrix). For each non-predetermined variable, we obtain a similar expression. More specifically, for inflation and the marginal utility of consumption we obtain

$$\tilde{\lambda}_t = \lambda'_x \tilde{x}_t + \frac{1}{2} \tilde{x}'_t \lambda_x \tilde{x}_t + \frac{1}{2} \lambda_{\sigma \sigma} \sigma^2$$

$$\tilde{\pi}_t = \pi'_x \tilde{x}_t + \frac{1}{2} \tilde{x}'_t \pi_x \tilde{x}_t + \frac{1}{2} \pi_{\sigma \sigma} \sigma^2$$

where $\lambda_x$, $\pi_x$, $\lambda_{\sigma \sigma}$ and $\pi_{\sigma \sigma}$ are vectors, $\lambda_{xx}$ and $\pi_{xx}$ are symmetric matrices, $\sigma$ is the perturbation parameter and $\tilde{x}_t$ is the vector of $n_x$ predetermined variables.

Note that the nominal discount factor can be written as $Q_{t,t+1} = (\beta / \Pi_{t+1}) (\Lambda_{t+1} / \Lambda_t)$, or, in log-deviation from the deterministic steady state, $\tilde{q}_{t,t+1} = \Delta \tilde{\lambda}_{t+1} - \tilde{\pi}_{t+1}$, where is natural output.
Δ is the first difference operator such that $Δ\hat{λ}_{t+1} = \hat{λ}_{t+1} - \hat{λ}_t$. Hence, the above solution equations complete the description of the dynamics of the stochastic discount factor.

4.2 Solving for bond prices

Given the stochastic discount factor, bond prices can be derived recursively. The price of an $n$-period nominal bond, $B_{t,n+1}$, such that $B_{t+n+1,n+1} = 1$, can be written as $B_{t,n+1} = E_t [Q_{t,t+1} B_{t+1,n}]$. In the appendix (see section A.4), we show that a second order approximate solution for bond prices, in log-deviation from their deterministic steady state, can also be written as

$$\hat{b}_{t,n} = \hat{b}_{t,n}^0 + \frac{1}{2} x_t \hat{b}_{n,xx} \hat{x}_t + \frac{1}{2} \sigma^2 \hat{b}_{n,\sigma \sigma}$$

where $\hat{b}_{t,1} = -\hat{h}_t$, or

$$\hat{b}_{1,x} = (\lambda'_x - \pi'_x) c_1 - \lambda'_x$$

$$\hat{b}_{1,\sigma \sigma} = (\lambda'_x - \pi'_x) c_0 - \pi_{\sigma \sigma} + tr (\eta' (\lambda'_{xx} - \pi_{xx}) \eta) + (\lambda'_x - \pi'_x) \eta \eta' (\lambda'_x - \pi'_x)'$$

$$\hat{b}_{1,xx} = c'_1 (\lambda_{xx} - \pi_{xx}) c_1 - \lambda_{xx} + \sum_{i=1}^{n_x} (\lambda'_x [i] - \pi'_x [i]) c_2 [i]$$

and for $n > 1$

$$\hat{b}_{n,x} = \hat{b}_{1,x} + \hat{b}_{n-1,x} c_1$$

$$\hat{b}_{n,xx} = \hat{b}_{1,xx} + c'_1 \hat{b}_{n-1,xx} c_1 + \sum_{i=1}^{n_x} \hat{b}_{n-1,x} [i] c_2 [i]$$

$$\hat{b}_{n,\sigma \sigma} = \hat{b}_{1,\sigma \sigma} + \hat{b}_{n-1,\sigma \sigma} + \hat{b}_{n-1,x} c_0 + tr (\eta \hat{b}_{n-1,xx} \eta) + \hat{b}_{n-1,x} \eta \eta' (\lambda_x - \pi_x)$$

where $\lambda'_x [i]$, $\pi'_x [i]$ and $\hat{b}_{n-1,x} [i]$ are the $i$-th (scalar) elements of the $\lambda'_x$, $\pi'_x$ and $\hat{b}_{n-1,x}$ vectors, and $c_2 [i]$ is the $i$-th page of the $c_2$ tensor. In order to derive these expressions we rely on the suggestion in Sutherland (2002) that first order approximate solutions are sufficient to derive second order approximate solution to second moments.

The bond solutions are useful for two reasons. First, they allow us to develop a better intuition for the determinants of risk premia, because we can derive yield moments in terms of their macroeconomic driving forces. Second, they produce huge savings in terms of computational time. The alternative would in fact be that of solving for bond prices recursively, maturity by maturity, jointly with the rest of the model. The macroeconomic
model used in this paper adds up to 15 equations. Solving numerically for 10-year, namely 40-quarter, bond prices would amount to adding 39 equations to the model. Given the curse of dimensionality, the 15-equation model can be solved in a couple of minutes on a Pentium 4 personal computer, while a 54-equation model takes approximately two hours.

5 The term structure of interest rates

In this section, we analyse in turn the main stylised facts on the term structure of interest rates that are considered puzzling in the macroeconomic literature. As already emphasised, we mainly focus on the "large" volatility of long term yields and on the size of term premia. Before tackling these puzzles, however, we also discuss briefly the alleged puzzle on the sign of term premia highlighted by Backus, Gregory and Zin (1989).

In order to provide a better intuition for our results, we structure our discussion around analytical expressions for the variance of yields and holding period returns. Such expressions are based on our analytical second-order approximation to bond prices. Details on all derivations are provided in the appendix.

5.1 The (positive) sign of term premia is not a puzzle

If we define \( \tilde{ytm}_n \) and \( \tilde{ytm}_n^{\text{real}} \) as the yield to maturity on an \( n \)-period nominal and real bond, respectively, \( \hat{i} \) as the short term nominal interest rate and \( \hat{r} \) as the real interest rate (all in deviation from the nonstochastic steady state), the appendix proves that the unconditional slope of the term structure of interest rates can be written as

\[
E \left[ \tilde{ytm}_n \right] - E \left[ \hat{r} \right] = \left( E \left[ \tilde{ytm}_n^{\text{real}} \right] - E \left[ \hat{r} \right] \right) - \frac{1}{2} \left( \frac{\Var \left[ \sum_{i=1}^{n} \hat{\pi}_i \right]}{n} - \Var \left[ \hat{\pi} \right] \right) + \left( \frac{\Cov \left[ \sum_{i=1}^{n} \hat{\pi}_i, \Delta^n \lambda \right]}{n} - \Cov \left[ \hat{\pi}, \Delta \lambda \right] \right)
\]

where \( E \left[ \tilde{ytm}_n^{\text{real}} \right] - E \left[ \hat{r} \right] = -\frac{1}{2} \left( \frac{1}{n} \Var \left[ \Delta^n \lambda_n \right] - \Var \left[ \Delta \lambda \right] \right) \) is the slope of the real term structure. Note that, in the special case with CRRA utility, the real slope becomes

\[
E \left[ \tilde{ytm}_n^{\text{real}} \right] - E \left[ \hat{r} \right] = -\frac{1}{2} \gamma^2 \left( \frac{1}{n} \Var \left[ \Delta^n \gamma_l \right] - \Var \left[ \Delta \gamma \right] \right).
\]

The last expression is used by den Haan (1995, equation (3)) to illustrate an alleged puzzle first highlighted by Backus, Gregory and Zin (1989). The problem is that, in the data, the rate of growth of consumption is positively serially correlated. This should
intuitively imply that the variance of the rate of growth of consumption over long periods will be larger than the variance of the 1-steap ahead growth rate. As a result, the difference \( \frac{1}{n} \text{Var} [\Delta^n \hat{y}_k] - \text{Var} [\Delta \hat{y}] \) should be positive, which implies that the model predicts a negative slope for the real yield curve. In the data, however, the average yield curve is upward sloping. The puzzle is then that the model appears to be incapable of generating a positive term structure slope and a positive serial correlation of consumption at the same time.

A first contribution of this paper is to show how this puzzle is not robust. This conclusion is based on two arguments. First, we demonstrate that the positive serial correlation of consumption growth does not necessarily imply that the variance of consumption growth over long periods will be larger than the 1-steap ahead variance. Second, we show that inflation risk premia can also contribute to dispel the puzzle. We analyse these two arguments in turn.

### 5.1.1 Real term structure

We first demonstrate that a stochastic process of the form which consumption follows in the equilibrium of a DSGE model is capable of producing both a positive serial correlation for consumption growth and a positive slope of the yield curve. In order to illustrate this argument, we start from a case where the puzzle does emerge in our model. If our vector of predetermined variables were a scalar, and the serial correlation of this variable were \( \rho \), the serial correlation of consumption growth and the slope of the yield curve would be given by \( \rho [\Delta \hat{y}_{t+1} | \Delta \hat{y}] = -\frac{1}{2} (1 - \rho) \) and \( ytm_n - r = \gamma^2 y_x^2 \left( 1 - \frac{1 - \rho^2 n}{1 - \rho^2} \right) \sigma_x^2 \), respectively. Since \( \frac{1}{n} \frac{1 - \rho^2 n}{1 - \rho^2} < 1 \) for any \(-1 < \rho < 1\), the model always predicts that the slope is positive and the serial correlation of output growth negative.

In the vector case, however, the first order serial correlation of output growth is (see section A.8 in the appendix)

\[
\rho [\Delta \hat{y} | \Delta \hat{y}_{t+1}] = \frac{y_x' \left( 2 \text{E} \left[ \hat{x'}_{t+1} - \text{E} \left[ \hat{x'}_{t+2} \right] - \text{E} \left[ \hat{x'}_{t+1} \right] \right] \right) y_x}{2y_x' \left( \text{E} \left[ \hat{x'}_{t+1} \right] - \text{E} \left[ \hat{x'}_{t+2} \right] \right) y_x} \tag{4}
\]

while the slope of the term structure is

\[
ytm_n - r = -\gamma^2 \sigma^2 y_x' \left( (I - c_1)^{-1} (c_1 - c_1^n) \eta' + \eta' (c_1 - (c_1^n))' (I - c_1)^{-1} \right) y_x
\]

\[
-\gamma^2 \sigma^2 \frac{1}{n} y_x' \left( \sum_{i=1}^{n-1} (I - c_i) \eta' (I - c_i)' - 2 (n - 1) \eta' \right) y_x
\]
The signs of these two variables are now not obvious. For example, it is clear that correlations across different predetermined variables will also play an important role. To illustrate how easy it is to break the puzzle within a more general model, consider an example based on two states, one exogenous and one endogenous. If we assume that the endogenous state is non-stochastic (e.g., lagged consumption), then consumption will follow an ARMA(1,1) process

$$y_t = \phi y_{t-1} + x_t,$$

where $$x_t = \theta x_{t-1} + \sigma \varepsilon_t$$ and $$\varepsilon_t$$ a standard normal process (this assumption is adopted, e.g., in Wachter, 2005). The autocovariances of such process are well known (see e.g. Hamilton, 1994). In the appendix we show that this implies

$$E [\Delta y_{t+1} \Delta y_t] = \frac{\sigma^2}{1 + \phi} \left( 2\theta - (1 - \phi) \left( 1 + \theta^2 + \phi\theta \right) \right)$$

Note that the serial correlation is always negative when $$\theta = 0$$. When $$\theta$$ is positive, however, a large $$\phi$$ will tend to reduce the negative component of the above polynomial and allow for a positive serial correlation. As $$\theta$$ increases, the lower the value of $$\phi$$ necessary to induce a positive correlation. For example, when $$\theta = 0.5$$ it is sufficient that $$\phi > 0.28$$.

Now consider the real yield curve slope and, for simplicity, focus on the two-period maturity. The general expression above simplifies to

$$y_{tm_2} - r = \frac{1}{4} \sigma^2 \phi^3 - \frac{(1 + 2\theta) \phi^2 - (5 - 6\theta - \theta^2) \phi + (1 - \theta^2)}{1 - \phi}$$

which is obviously positive when $$\phi^3 - (1 + 2\theta) \phi^2 - (5 - 6\theta - \theta^2) \phi + (1 - \theta^2) > 0$$. Once again, note that for $$\theta = 0$$ this simplifies to $$(1 - \phi) (5 - \phi^2) > 0$$, which is only positive for a low $$\phi$$ (approximately $$\phi < 0.2$$). As $$\theta$$ increases, however, a positive slope can be obtained under different values of $$\phi$$. In the above example of $$\theta = 0.5$$, we obtain $$\phi < 0.33$$, so that both a positive slope of the yield curve and a positive serial correlation of consumption can be achieved, for example, when $$\theta = 0.5$$ and $$\phi = 0.3$$.

Intuitively, the more complicated driving processes that would arise from a model with more state variables would further alleviate the restrictions of the simple AR(1) model. Table 2 shows the results of simulations based on various simplified versions of our baseline model. More specifically, we focus on cases where technology shocks are the only source of exogenous disturbance to the system and analyse the consequences of introducing various sources of endogenous persistence in the model.

The puzzle is replicated in the plain-vanilla, flex-price case, where the model generates a positive slope of the yield curve only at the cost of also generating a negative serial cor-
relation for consumption growth. However, the puzzle disappears as soon as we introduce habit formation. In this case, the marginal rate of intertemporal substitution is not simply proportional to the rate of consumption growth. The former can then be negative, thus producing a positive term structure slope, even if the latter is positive.

In sticky-price versions of the model, we see that the puzzle can be solved even without breaking the proportionality between marginal utility of consumption and consumption growth. Any source of endogenous persistence is in fact sufficient to make the model consistent with the signs of both the yield curve slope and consumption growth. The intuition here is simply that when consumption growth follows a vector process, there is no longer a direct link between its first order serial correlation and its cumulative variance.

Table 2: Simulation results: the serial correlation of consumption growth

<table>
<thead>
<tr>
<th></th>
<th>$\rho[c_{t},c_{t+1}]$</th>
<th>$\rho[\Delta c_{t},\Delta c_{t+1}]$</th>
<th>$\rho[\hat{\lambda}<em>{t},\hat{\lambda}</em>{t+1}]$</th>
<th>$\rho[\Delta \hat{\lambda}<em>{t},\Delta \hat{\lambda}</em>{t+1}]$</th>
<th>$\hat{ym}_{t^n}$ - $\hat{\pi}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>US 1960:1-1997:2</td>
<td>0.98</td>
<td>0.29</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>$h = \nu = \rho_I = 0$</td>
<td>0.95</td>
<td>-0.025</td>
<td>0.95</td>
<td>-0.025</td>
<td>+</td>
</tr>
<tr>
<td>$h = 0.68; \nu = \rho_I = 0$</td>
<td>0.99</td>
<td>0.53</td>
<td>0.89</td>
<td>-0.14</td>
<td>+</td>
</tr>
<tr>
<td>$h = \nu = \rho_I = 0$</td>
<td>0.95</td>
<td>-0.025</td>
<td>0.95</td>
<td>-0.025</td>
<td>+</td>
</tr>
<tr>
<td>$h = 0.68; \nu = \rho_I = 0$</td>
<td>0.99</td>
<td>0.64</td>
<td>0.95</td>
<td>-0.03</td>
<td>+</td>
</tr>
<tr>
<td>$\nu = 0.66; h = \rho_I = 0$</td>
<td>0.96</td>
<td>0.03</td>
<td>0.96</td>
<td>0.03</td>
<td>+</td>
</tr>
<tr>
<td>$\rho_I = 0.87; h = \nu = 0$</td>
<td>0.97</td>
<td>0.02</td>
<td>0.97</td>
<td>0.02</td>
<td>+</td>
</tr>
<tr>
<td>Flex prices: $\zeta = 0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sticky prices: $\zeta = 0.8$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The parameter values are $\beta = 0.99$, $\alpha = 0.76$, $\gamma = 2.6$, $\rho_a = 0.95$, $\sigma_a = 0.02$, $\sigma_{\pi} = 0$, $\sigma_{\tau} = 0$, $\sigma_{\eta} = 0$.

### 5.1.2 Nominal term structure

So far we focused on the real term structure. Even if the real term structure were negatively sloped, the nominal term structure could still slope positively if inflation risk premia were positive and large.

Equation (3) shows that two terms matter in this respect. The first one on the second line is a convexity term and its role is analogous to the real term, since it involves the difference between the variance of inflation over long periods and the 1-step ahead. Since inflation is even more positively serially correlated than consumption growth in the data, this term may well weigh negatively on the nominal term structure. The last term in equation (3) is the true inflation risk premium. Given the persistence of both the
marginal utility of consumption and inflation, this term should be positive, at least in simple models, when the rate of change in the marginal utility of consumption is positively correlated with the inflation rate. In the CRRA case, this term boils down to 
\[-\gamma \left( \frac{1}{n} \text{Cov} \left[ \sum_{i=1}^{n} \hat{\pi}_i, \Delta^n \hat{y} \right] - \text{Cov} \left[ \hat{\pi}, \Delta \hat{y} \right] \right) .\]
Given the persistence of consumption growth and inflation, the inflation risk premium will be positive, at least in a simple model, when consumption growth and output are negatively correlated.

We underline that this is true only in simple models because one could repeat here the argument used concerning the real term structure. In general, it is not necessarily the case that a negative correlation between consumption growth and inflation implies that the inflation risk premium is positive. In the more general case with habits, for example, outcomes are more uncertain and will depend on parameter values.

5.2 Matching the large variance of yields

As already emphasised, a stylised fact from Table 1(a) is that the standard deviation of yields at all maturities is not too different from that of the short term rate. The standard deviation of 5-year and 10-year yields is approximately 90% of the standard deviation of the short term rate. Den Haan (1995) argues that matching the standard deviation of long term yields is particularly difficult within a simple flexible-price general equilibrium model. Intuitively, it is difficult to generate sufficient persistence in the short term rate to ensure that its variability is transmitted almost one-to-one to long term rates.

The problem of inducing sufficiently high volatility in yields can be seen most clearly in the case with 1 state variable, where (see section A.4.4 in the appendix)

\[
\frac{\text{Std} \left[ \hat{y}T_m \right]}{\text{Std} \left[ \hat{i} \right]} = \frac{1}{n} \frac{1 - \rho^n}{1 - \rho}
\]

where \( \rho \) is the serial correlation of the simple state variable of the model and \( n \) is the maturity of the yield. The only way to boost the volatility of long term yields is really by inducing a near-random walk behaviour in the state variable. More precisely, note that \( \frac{1 - \rho^n}{1 - \rho} \) is increasing in \( \rho \) and it is therefore maximised for \( \rho = 1 \). In practice, \( \rho \) needs to be very close to 1. The ratio between the standard deviations of the 5-year yield and the 3-month rate would only equal 0.64 for shocks with a serial correlation equal to 0.95. Even with a persistence of the shocks equal to 0.99, the ratio would reach 0.91 for 5-year yields, but only 0.83 at the 10-year horizon.
The high volatility of yields can therefore be considered a puzzle or not, depending on whether highly persistent shocks are deemed to be realistic. This intuition carries through to the multivariate case (see section A.4.4 of the appendix), taking into account that persistence can be endogenous or exogenous. In general, however, various sources of persistence can reinforce each other, so that very extreme degrees of persistence are not necessary to match the data. Intuitively, the variance of long term rates – relative to that of short term rates – will be higher when the covariance between the state variables is higher (due to the the full $c_1$) and when inflation and/or the marginal utility of consumption are affected strongly (with coefficients larger than 1) by many state variables (so that $\pi_\tau \pi_\tau'$, $\lambda_x \lambda_x'$ and $\pi_\tau \lambda_x'$ will have large off-diagonal elements).

We illustrate in Table 3 the fact that very persistent shocks, possibly combined with each other or with other sources of persistence, can replicate a large variance of long term yields. In the table, we focus on the ability of the model to reproduce a large ratio between the variance of 5-year yields and that of the short term rate. For policy, we rely on the benchmark rule originally suggested by Taylor (1993), namely a rule without interest rate smoothing and with coefficients of 1.5 and 0.5 on inflation and output, respectively.

Table 3: Simulation results: variance of yields

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>US 1960:1-1997:2</td>
<td>83%</td>
<td>47%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_{\Pi} = 0.999, \rho_{\eta,\tau} = 0, h = 0.68$</td>
<td>97%</td>
<td>0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_{\Pi} = 0.999, \rho_{\eta} = 0.987, \rho_{\eta,\tau,I} = 0, h = 0$</td>
<td>96%</td>
<td>4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_{\eta} = 0.987, \rho_{\Pi,\eta,\tau} = 0, h = 0$</td>
<td>81%</td>
<td>18%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The parameter values are $\beta = 0.99$, $\alpha = 0.76$, $\gamma = 2.6$, $\zeta = 0.8$, $\sigma_r = 0$, $\sigma_n = 0$, $\iota = 0.66$, $\delta = 1.5$, $\kappa = 0.5$, $\rho_I = 0$.

The examples are related of three simulations where we experiment with three alternative sources of persistence: in technology shocks, in the shocks to the inflation target, and in consumption habits. The table shows that pairs of these sources of persistence are sufficient to generate a realistic variance of long term yields relative to that of the short term rate. Clearly, the more general model where these various sources of persistence are allowed to play a role simultaneously would find it even easier to generate realistic results.

We also emphasise in the table, however, that xhpr’s relative to the variance of yields are remarkably small, compared to the data. The real difficulty of general equilibrium models lies with the large term premia observed in the data.
5.3 Large term premia

Table 1(a) showed that both the slope of the term structure and \(xhpr\)’s – namely the difference between the expected return from holding a bond of maturity \(n\) for only 1 quarter and the 3-month rate in the same quarter – are upward sloping. In this section, we analyse risk premia focusing on the latter measure, namely \(xhpr\)’s. The advantage of holding period returns over yields is that the former are not affected by first moments. While the slope in yields can be caused purely by (risk-neutral) expectations of future policy changes over the horizon of the yields, holding period returns have all the same time horizon as the short-term rate. They are the most direct measure of risk premia on bonds, as emphasised by the fact that they are all zero in the first order approximation of a model.

In this section, we first perform a simple calibration exercise in the spirit of Mehra and Prescott (1985). This shows that fitting average \(xhpr\)’s over long periods with a DSGE model is significantly less puzzling than trying to explain the equity premium. This is confirmed in an exercise using a highly simplified real model, which can match both the variance of consumption and \(xhpr\)’s. The problem of simple models, however, is that they have other unrealistic implications – for example, the aforementioned simple model generates a negative serial correlation in consumption growth. The real test of a model, therefore, must come from a joint analysis of its macro and term-structure implications. We do this in the final subsection, which presents results based on our general model.

5.3.1 A Mehra-Prescott-type exercise

A second order approximation to this expression (see section A.2 in the appendix) is

\[
B_{t,n+1} = E_t [Q_{t,t+1} B_{t+1,n}]
\]

\[
\tilde{b}_{n+1,t} = \tilde{b}_{1,t} + E_t [\tilde{b}_{n,t+1}] + \frac{1}{2} \text{Var}_t [\tilde{b}_{n,t+1}] + \text{Cov}_t [\tilde{q}_{1,t}, \tilde{b}_{n,t+1}]
\] (5)

Note that the first two terms of the latter expression represent the “pure expectations hypothesis”, which postulates that the price of an \(n\)-period bond is the simple average of the expected future prices of a 1-period bond up to \(n\) periods ahead. Deviations from the pure expectations hypothesis – which implies that all bonds yield the same return in steady state – are going to be generated by the last two terms in the latter expression. Note that, as is well-known, a second order approximation will only generate constant
Holding period returns will be defined recursively as $HPR_{n,t} = \frac{E_t \left[ B_{t+1}^{n-1} \right]}{B_t^n}$. The other variable on which we will focus in the empirical evaluation of the model is the $xhpr$’s on bonds of any maturity, i.e. the holding period return on a $n$-period bond in excess of the short term rate. These excess returns represent the true compensation for the risk of holding a long-term bond over a short-term one. A second order approximation of $xhpr$’s in the special case of the model without habit persistence is

$$
\hat{hpr}_{t,n} - \hat{i}_t = -(n - 1) \left( \gamma \text{Cov}_t [\Delta \hat{y}_{t+1}, ytm_{t+1,n-1}] + \text{Cov}_t [\hat{\pi}_{t+1}, ytm_{t+1,n-1}] \right)
$$

The first component of expression (6) is the covariance between yields and output growth. This increases the $xhpr$ when it is negative, namely for shocks that create a negative correlation between output growth and long term yields, or else shocks that generate a negative correlation between output growth and the future path of policy rates. This negative sign is likely to emerge for shocks that generate a trade-off between output growth and inflation. This, in turn, will obviously be affected by the monetary policy rule. For example, rules with a relatively high weight on inflation imply that the whole term structure of interest rates will move up, when inflation increases and output (growth) falls. The second covariance in the $xhpr$ is likely to be positive under more general circumstances and it should therefore weigh negatively on excess holding period returns. However, this is also likely to be smaller than the first covariance as risk aversion increases.

Note also that the absolute value of both covariances is likely to be higher for rules with a high degree of interest rate smoothing. Ceteris paribus, high interest rate smoothing implies high persistence in short rates and therefore a larger movement in long rates after an inflationary shock. Viceversa for low interest rate smoothing.

Large $xhpr$’s will therefore tend to arise for policy rules characterised by high interest rate smoothing on shocks that generate a trade-off between inflation and output growth.

Based on the results above, we can now plug in the second moments of output and inflation to derive model consistent implications for the $xhpr$. The covariances between yield to maturities and inflation or output are reported in Table 2. Based on these values, we can compute the $xhpr$’s in Table 4 using equation (6).

Table 4 shows that the simplest CRRA-utility model is unlikely to generate large $xhpr$’s because the covariances which enter equation (6) are very small in the data. Even with a
relatively high risk-aversion of 6, the model only generates an \( xhpr \) of 30 basis points at the 5-year horizon, compared to the 76 basis points observed in the data (see Table 1a). Nevertheless, these two figures are within the same order of magnitude. It is therefore conceivable that habit persistence can easily bring the model closer to the data.

Table 4: Theoretical excess holding period returns (annualised values)

<table>
<thead>
<tr>
<th>maturity</th>
<th>( \gamma = 2 )</th>
<th>( \gamma = 6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 4 )</td>
<td>( hpr_{t,n} - i_t )</td>
<td>(-0.005)</td>
</tr>
<tr>
<td>( n = 20 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n = 20 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Based on equation (6) and sample covariances of quarterly US data over the 1960Q1-1997Q2 period. The inputs are (all \( \times 10^4 \)): for \( n = 4 \), \( \text{Cov}(\pi_{t+1}, ytm_{t+1,n-1}) = \frac{.26}{\epsilon} \) and \( \text{Cov}(\Delta y_{t+1}, ytm_{t+1,n-1}) = -.11 \); for \( n = 20 \), \( \text{Cov}(\pi_{t+1}, ytm_{t+1,n-1}) = .21 \) and \( \text{Cov}(\Delta y_{t+1}, ytm_{t+1,n-1}) = -.10 \).

Data source: for inflation and output, den Haan (2000); for yields,.....

5.3.2 Special case: flexible prices and no habit formation

In this section, we analyse if a highly simplified version of the model is capable of generating plausible \( xhpr \)'s. This is useful to illustrate the tension between fitting yields-related moments and macroeconomic moments.

We start from a simple case of a model without habit persistence nor price rigidities and only technological shocks (this is equivalent to considering the term structure of natural rates in our model, when \( h = 0 \)). In this case it is easy to derive analytically the yield to maturity of a bond with maturity \( n \) at time \( t \) is (see section A.5 in the appendix)

\[
\hat{ytm}_{n,t}^{\text{nat}} = a_n + b_n \ln A_t
\]

where

\[
a_n \equiv -\ln \beta - \frac{1}{2n} \frac{1 - \rho^{2n}}{1 - \rho^2} \gamma^2 \left( \frac{\phi}{\phi + \alpha (\gamma - 1)} \right)^2 \sigma_e^2 \]

\[
b_n \equiv \frac{1}{n} (1 - \rho^n) \gamma \frac{\phi}{\phi + \alpha (\gamma - 1)}
\]

The continuously compounded yield is therefore affine in \( \ln A_t \). This model represents a special case of the general class of defined by Dai and Singleton (xxx). Note also that \( \frac{\partial a_n}{\partial n} > 0 \), hence term premia are increasing in maturity.
In this simple model, term premia arise as a result of precautionary savings. Compared to a world of certainty, technology shock imply that the future consumption stream is uncertain. Since utility is concave in output (consumption), expected output at any future period is always smaller than its certainty equivalent. As a result, all households try to save more (the precautionary savings effect) and delay consumption to the future. Since this is infeasible in equilibrium, the return on real bonds must fall to discourage savings.

The positive slope of the term structure is due to the fact that short term rates fall more, when compared to the certainty equivalent world, than long term yields. The reason is that technology shocks are known to be stationary and therefore die out over time. Consumption far out in the future is, in this sense, less risky than consumption in the near future and the precautionary savings motive is therefore less strong. Long term yields are therefore higher, namely closer to the certainty equivalent level. They actually approach that level in the limit of an infinite maturity bond ($\lim_{n \to \infty} a_n \equiv -\ln \beta$ and $\lim_{n \to \infty} b_n \equiv 0$).

Note that the term $\phi / (\alpha (\gamma - 1) + \phi)$ represents the elasticity of output (and consumption) to the technology shock, so that we could define $\sigma_c \equiv \phi / (\alpha (\gamma - 1) + \phi) \sigma_\varepsilon$. Using this definition, $xhpr$’s can be written as

$$hpr_{n,t}^{nat} - i_t^{nat} = (1 - \rho^{n-1}) \gamma^2 \sigma_c^2$$ (7)

Holding premium returns are therefore higher when risk aversion or the volatility of output are higher.\(^{5}\)

Now note the dependence on the persistence of technology shocks. Highly persistent technology shocks tend to also make holding period returns very similar across maturities: integrated shocks ($\rho = 1$), in particular, annull $xhpr$’s and make yields completely constant over time. The $xhpr$’s is in fact decreasing in the persistence of the shocks and it becomes largest (but constant across maturities, for $n > 1$) when shocks are white noise ($\rho = 0$). To match both the large values of $xhpr$’s and the large variance of the long term yields, intermediate values of persistence of the underlying shock are therefore needed. This

\(^{5}\)In turn, the volatility of output is affected by risk aversion $\gamma$, by the curvature of the disutility of labour $\phi$, and by the weight of labour in the production function, $\alpha$. A higher disutility of labour, $\phi$, matters because it increases households’ aversion to changing their hours worked. The labour share in the production and risk aversion interact because they contribute to affect labour demand and labour supply, thus employment and the real wage.
makes holding period returns upward-sloping functions of maturity, with a curvature that is decreasing in the persistence of the underlying shocks. At the same time, however, we know that intermediate values of the persistence of underlying shocks also generate a quickly decreasing volatility of yields across maturities.

Note that equation (7) corresponds to the simplest case considered by den Haan (1995) (see section 2.4). Using den Haan’s calibration parameters ($\rho = 0.95$, $\gamma = 6$ and $\sigma_c = 0.007$), the $xhpr$ on a 5-year bond is equal to 44 basis points, a value not too different from what is observed in the data. Clearly, adding habit formation will easily help to match the data. In this case, a back of the envelope computation of the $xhpr$ is not possible, but we can solve the model numerically to obtain a 5-year $xhpr$ of 1.45% already for a habit parameter $h = 0.3$.\(^6\) Now move on to consider what happens when the variance of consumption is derived endogenously. Consider again the natural rate case, but and use the standard values for the structural parameters and calibrate, still $\rho = 0.95$ and then solve for the variance of technology shocks which yields the same variance of consumption growth as above (namely 0.0071). The $xhpr$ drops to 65 basis points for $h = 0.3$, but this is still easily made consistent with the data.

Fitting $xhpr$’s, therefore, is not very difficult within a simple model, provided that only a limited set of moments is matched in the calibration. In the example above, only the variance of consumption is matched with the data. As emphasised in section 5.2, a clear unrealistic feature of this model is however that it generates a negative autocorrelation in consumption growth.

The truly puzzling feature of $xhpr$’s is that they are difficult to generate while fitting moments of other macroeconomic variables at the same time.

5.3.3 Fitting yields and macro data simultaneously

In this section we simulate our model using relatively standard parameter values and check the implications of those simulations in terms of a number of unconditional moments in the data. More precisely, we focus on: the variance of output growth, inflation and the short term interest rate; the three covariances between these three variables; the variance of yields and the excess holding period returns on bonds at various maturities.

\(^6\)These values are obtained solving for the equilibrium natural rate of the general model assuming $\alpha = 0$, $\rho = 0.95$ and $\sigma_c = 0.007$. 

21
For comparability, macro data are taken from Den Haan (2000). Sample covariances are broadly in lines with the results presented in that paper based on other estimates. Yield data are taken from the McCulloch-Kwon dataset.\(^7\)

The parameters values used in the calibration are relatively standard. We use the results of Smets and Wouters (2005), who estimate them on US data over the 1974:1-2002:2 and 1983:1-2002:2 periods and find them to remain relatively stable. Many parameters are consistent with those used by Woodford (2003). More specifically, the weight of labour in the production function is \(\alpha = 0.76\), the probability of firms not changing their price is \(\zeta = 0.8\), the elasticity \(\theta = 7.88\) corresponding to a steady state mark-up of about 15%, and the degree of habit persistence is relatively high, \(h = 0.68\). We also set \(\iota = 0.66\) implying a relatively high degree of indexation to past inflation. We also set \(\beta = 0.99\) as is common in the literature. In principle, we could use changes in \(\beta\) and a non-zero steady state inflation target to match the sample means of inflation and the short term rate, but this would have no change on the higher moments of our interest. Finally, we experiment with two values of the risk aversion parameter: \(\gamma = 2.6\), which is Smets and Wouters’ estimated value, and \(\gamma = 6\), which is the valued used by Fuhrer (2000). Since habits modify the relationship between the intertemporal rate of substitution of consumption and risk aversion, we take Fuhrer’s estimate as an example of high but still reasonable value of \(\gamma\).\(^8\)

In our benchmark simulation, we try to approximate the aforementioned features of the data with the smallest possible set of shocks. It turns out that we only need two shocks: technology and inflation target shocks. We therefore set to zero tax and monetary policy shocks. For the persistence of the shocks, we choose Smets and Wouters’ estimated values of \(\rho_A = 0.986\) and \(\rho_{\Pi} = 0.999\) (Smets and Wouters actually allow for a unit root in the inflation target). The standard deviations of these two shocks are chosen so as to maximise the match between observed and calibrated moments. They turn out to be \(\sigma_A = 0.0237\) and \(\sigma_{\Pi} = 0.0002\). Alternatively, we substitute the inflation target shock with a highly persistence tax shock, such that \(\rho_{\tau} = 0.95\) and \(\sigma_{\tau} = 0.0910\).

Finally, the policy rule parameters are slightly adjusted due to the fact that we use a

\(^7\)The McCulloch-Kwon dataset can be downloaded from http://www.econ.ohio-state.edu/jhm/ts/mcckwon/mccull.htm

\(^8\)The main unattractive implications of a high degree of risk aversion is that, with standard preferences, the elasticity of intertemporal substitution of consumption becomes unreasonably low. With habit persistence, however, intertemporal substitution is larger than \(1/\gamma\).
slightly different rule. We use Smets and Wouters’ estimate for the interest rate smoothing coefficient \( \rho_I = 0.85 \) and the reaction to deviations in the output gap, \( \kappa = 0.06 \). The inflation reaction parameter is slightly higher than in Smets and Wouters in order to improve the match between data and model: \( \delta = 1.7 \).

Table 5: Main calibration results

<table>
<thead>
<tr>
<th>Unconditional moments ((\times 10^4))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var ( \hat{y}_{t=20} )</td>
</tr>
<tr>
<td>US 1960:1-1997:2</td>
</tr>
<tr>
<td>( \gamma = 2.6, \sigma_r = 0 )</td>
</tr>
<tr>
<td>( \gamma = 6, \sigma_r = 0 )</td>
</tr>
<tr>
<td>( \gamma = 2.6, \sigma_\pi = 0 )</td>
</tr>
<tr>
<td>( \gamma = 6, \sigma_\pi = 0 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Var ( [\Delta \hat{y}] )</th>
<th>Cov ( [\Delta \hat{y}, \hat{\pi}] )</th>
<th>Cov ( [\Delta \hat{y}, \hat{i}] )</th>
<th>Cov ( [\hat{i}, \hat{\pi}] )</th>
<th>( \rho [\Delta \hat{y}, \Delta \hat{y}_{t+1}] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>US 1960:1-1997:2</td>
<td>0.88</td>
<td>-0.15</td>
<td>-0.12</td>
<td>0.27</td>
</tr>
<tr>
<td>( \gamma = 2.6, \sigma_r = 0 )</td>
<td>0.90</td>
<td>-0.03</td>
<td>-0.09</td>
<td>0.25</td>
</tr>
<tr>
<td>( \gamma = 6, \sigma_r = 0 )</td>
<td>0.68</td>
<td>-0.02</td>
<td>-0.04</td>
<td>0.27</td>
</tr>
<tr>
<td>( \gamma = 2.6, \sigma_\pi = 0 )</td>
<td>0.93</td>
<td>-0.03</td>
<td>-0.09</td>
<td>0.03</td>
</tr>
<tr>
<td>( \gamma = 6, \sigma_\pi = 0 )</td>
<td>0.70</td>
<td>-0.02</td>
<td>-0.04</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Note: Other parameter values are \( \beta = 0.99, \alpha = 0.76, \zeta = 0.8, \sigma_a = 0.024, \sigma_\eta = 0, \nu = 0.66, \delta = 1.7, \kappa = 0.06, \rho_I = 0.85 \). When non-zero, \( \sigma_r = 0.09, \sigma_x = 0.0002 \).

The results of these simulations are summarised in Table 5. It is clear that the data moments, including the variance of yields, are matched better when \( \gamma = 2.6 \). The higher value of \( \gamma \), however, is necessary to generate more realistic excess holding period returns (although at the cost of some deterioration in the matching of moments). These results appear consistent with the widespread belief in the literature that allowing for a separation between intertemporal substitution and risk aversion in preferences is useful.

The most notable feature of these figures is that two exogenous shocks are sufficient to make the model broadly consistent with most of the unconditional features of the data analysed here. One shock has to be the technology shock, while tax shocks or inflation target shocks can both play the additional role.

Figure 2 shows the impulse response functions in the \( \gamma = 6 \) case with inflation target shocks. Clearly, the sole technology shock generates most of the dynamics in output and the nominal interest rate. Under the hypothesised policy rule, however, inflation would
move very little following the technology shock (Figure 2). The main role of inflation target shocks is to generate some volatility in inflation. In the end, however, inflation volatility remains lower than in the data, while interest rate volatility is a bit higher than in the data. Introducing additional shocks, in particular tax shocks, helps to increase the volatility in inflation only at the cost of increasing further the volatility of the short term rate (this is even more directly the case for policy shocks).

Figure 2: Impulse response functions: $\gamma = 6$ case

![Impulse response functions: $\gamma = 6$ case](image)

The variance of yields and the xhpr’s on bonds at various maturities are shown in Figure 3 (for the same case considered in Figure 2). The variance of the short term rate is a bit larger than in the data (2.6% compared to 1.9% in the data), but the ratio between the standard deviations of long term bonds and the short term rate is correct (92% for 5-year bonds, for example). Correspondingly, excess holding period returns are even a few basis points too large (in annualised values).

Figure 3: Variance of yields and excess holding period returns: $\gamma = 6$ case.

![Variance of yields and excess holding period returns: $\gamma = 6$ case](image)
6 Term premia, inflation risk premia and monetary policy

In this section we investigate the determinants of the relatively good performance of the model, especially in terms of excess holding period returns. For this reason, we derive a full decomposition of excess holding period returns into real and inflation-related components. We then go on to translate this decomposition in terms of yield premia. Finally, we highlight the important role of monetary policy in determining nominal yield premia and inflation risk premia.

6.1 The market prices of risk

In order to understand the determinants of the excess holding period returns shown in Figure 2, it is first useful to analyse separately the real and nominal excess holding period returns. We show in the appendix (see section A.6) that these can be written as

\[
\bar{hpr}_{t,n}^{\text{real}} - \bar{r}_t = -\sigma \lambda_x (I - c_1^{n-1}) \eta \xi
\]

\[
\bar{hpr}_{t,n} - \bar{r}_t = -\sigma \left( \lambda_x (I - c_1) + \pi_x c_1 \right) (I - c_1)^{-1} (I - c_1^{n-1}) \eta \xi
\]

where \( \xi \equiv -\sigma \eta' \lambda_x \) and \( \xi \equiv \sigma \eta' \left( \pi_x - \lambda_x \right) \).

Intuitively, \( \lambda_x \) measures the impact of any exogenous shock on the marginal utility of consumption (to a first order approximation), while \( \sigma \eta' \) is the variance-covariance matrix of the exogenous shocks. The product \( \xi \) represents therefore a standard error shock to the marginal utility of consumption (with a negative sign). This is a measure of the relevance of any standard shock for the representative household and it is in fact named "(real) market price of risk" in the finance literature. The (real) risk premium on any asset can be written as the product of the price of risk and the "amount of risk" of the asset. For real long term bonds, the amount of risk is the standard deviation of a real bond price, which from section 5.2 can be written as \( -\sigma \lambda_x (I - c_1^{n-1}) \eta \).

We have already shown that replicating the variance of long term bonds within a general equilibrium model is possible. Note also that the variance of bond prices is linearly increasing in maturity (see appendix). The problem of generating a large \( \chi hpr \) on long term bonds is therefore strictly that of generating large market prices of risk. For real bonds, this amounts to generating a large variance in the marginal utility of consumption, which is well known to be possible through habit persistence. The only question is whether
the increase in the price of risk can be achieved without increasing the volatility of short
term rates unrealistically at the same time.

It is important to underline that, for the same reasons highlighted in the context of
a simpler model, real $x_{hpr}$'s are decreasing in the persistence of shocks. The deviation
of the real interest rate from the certainty equivalent case is not due to a risk premium,
but to a precautionary demand motive which leads to a fall in returns. Since the same
shocks have an effect on short rates and long yields, the latter will tend to fall as much
as short rates when shocks are very persistent. They fall less, thus generating a steeper
yield curve, when the persistence of the shocks is small.

For nominal bonds, matters are more intricate because of the presence of inflation risk.
From the definition of $\xi$, the "nominal market price of risk", it is clear that the market
price of inflation risk is $\xi_\pi \equiv \sigma\eta'\pi_x$. To understand the determinants of nominal excess
holding period returns we need to decompose them into their various sources.

6.2 A Fisher-type decomposition for holding premia

In the appendix (section A.6), we show that the excess holding period return on nominal
bonds can be written as

$$\tilde{h}_{n,t} - \hat{i}_t = (\tilde{h}_{n,t} - \hat{r}_t) + (\hat{r}_{t,n} - \hat{r}_{t,1}) + conv_{t,n}$$

where

$$\tilde{h}_{n,t} - \hat{r}_t = \sigma^2 \lambda'_x (I - c_1^{n-1}) \eta'\lambda_x$$

$$\hat{r}_{t,n} - \hat{r}_{t,1} = \sigma^2 \lambda'_x (c_1^{n-1} \eta' + \eta'c'_1 (I - c_1^{n-1})' (I - c_1')^{-1}) \pi_x$$

$$= \sigma \lambda'_x c_1^{n-1} \eta^{\xi_\pi} - \sigma \pi_x' (I - c_1^{-1}) (I - c_1) c_1 \eta^{\xi_\lambda}$$

$$conv_{t,n} = -\sigma^2 \pi_x c_1 (I - c_1)^{-1} (I - c_1^{n-1}) \eta' \pi_x$$

Hence, nominal $x_{hpr}$'s are equal to real $x_{hpr}$'s plus a term which arises because of
inflation risk. Note that long and short term bonds are both subject to inflation risk in the
same manner over a 1-period horizon. Hence this inflation risk premium only comprises
the component due to the fact that the price of long term bonds reacts to the inflation
news, while the price of a 1-period bond stays constant by construction. The inflation risk
premium component common also to 1-period nominal bonds, $\hat{r}_{t,1} = \sigma^2 \lambda'_x \eta' \pi_x$, cancels
out from this expression.
Such "excess" inflation risk premium, in turn, includes two components. The first,

\[-\sigma \pi_x (I - c_1)^{-1} (I - c_1^{n-1}) c_1 \eta \xi_\lambda,\]

is due to the risk of a surprisingly low real return on long term bonds because of the possibility of future inflation surprises. This risk is valued using the market price of real risk, \( \xi_\lambda \), and it is such that positive premia are associated to a positive correlation between inflation and the marginal utility of consumption. The intuition is straightforward. Inflationary shocks are relevant to the household only to the extent that they have an impact on marginal utility. A positive covariance implies that an inflationary shock, which generates a surprise reduction of the real return on nominal bonds, produces, at the same time, a fall in the current marginal utility of consumption with respect to expected future marginal utility. In this sense, inflationary shocks have adverse effects and households try to ensure against their occurrence by requiring a risk premium on nominal bonds. The second component of the inflation risk premium, 

\[\sigma \lambda_x c_1^{n-1} \eta \xi_\pi,\]

represents instead the possibility that future surprises in the marginal utility of consumption will affect inflation over and above their direct effect on the real excess holding period return. Since the negative inflation consequences of future real shocks are what matters to determine the premium, this sort of risk is valued using the market price of inflation risk.

Finally, the nominal excess holding period return includes a convexity term, \( conv_{t,n} \), which is linked to the variance of inflation. When shocks tend to be positively serially correlated (formally, \( c_1 (I - c_1)^{-1} (I - c_1^{n-1}) \eta \eta' \) is positive definite), the convexity term is always negative and tends to reduce the nominal \( xhpr \).

Figure 4 shows the above decomposition for the case where only technology and inflation target shocks affect the system and \( \gamma = 6 \) (the case where tax shocks are used in place of inflation target shocks yields very similar results). Clearly, the real \( xhpr \) plays a vastly predominant role in shaping the nominal \( xhpr \) at all maturities. At the 10-year horizon, for example, 86 out of a total of 96 basis points of excess holding period returns are real excess returns, while only 10 basis points are an inflation risk premium. The convexity term is negligible across the whole maturity spectrum.

According to this decomposition, bonds with real returns such as index-linked bonds should display excess holding period returns very similar to those of nominal bonds. There is very little inflation risk premium in the yield curve at any maturity.

While these results are only suggestive, TIPS data appear to provide some support to
the hypothesis that a large component of nominal excess holding period returns reflects excess real returns, rather than inflation premia. Roll (2004, Table 2) reports average daily excess holding period returns on TIPS and nominal securities over a period between January 1997 and September 2003. Even if the periods are not exactly identical for all securities, it appears that the excess return on nominal and index-linked bonds is broadly comparable. More specifically, 10-year bonds appear to deliver an annualised excess return over 3-month bonds of just over 5 percentage points, both in the nominal and index-linked case. The evidence obtained by Ang and Bekaert (2004) – which does not incorporate information from TIPS – appears to uncover larger inflation risk premia.

Figure 4: Decomposition of excess holding period returns: $\gamma = 6$ case.

It should be emphasised here that this result relies on a model where there is full credibility of monetary policy and the inflation target is, thought time-varying, well-anchored in the long run and perfectly understood and known by all agents. Relaxing these assumptions could obviously lead to different conclusions on inflation risk premia.

The decomposition of excess holding period returns is useful to understand why the relatively high level of the average excess holding period returns has tended to be puzzling for DSGE models. More specifically, we can analyse the decomposition in the 1-state
variable case, where
\[
\begin{align*}
\hat{hpr}_{t,n} - \hat{r}_t &= \lambda_x^2 (1 - \rho^{n-1}) \sigma^2 \\
\hat{hpr}_{t,n} - \hat{r}_t &= \hat{hpr}_{t,n} - \hat{r}_t + (2\rho - 1) \pi_x \lambda_x \frac{1 - \rho^{n-1}}{1 - \rho} \sigma^2 - \rho \pi_x^2 \frac{1 - \rho^{n-1}}{1 - \rho} \sigma^2
\end{align*}
\]

Note first that, as emphasised above, excess real returns are higher when the serial correlation of the shock is lower (they are also increasing in maturity).\(^9\)

As to the inflation risk premium component of the spread, note first that \((1 - \rho^{n-1}) / (1 - \rho)\) is increasing in \(\rho\). Differently from real excess holding period returns, the size of inflation risk premia is therefore increasing in the persistence of exogenous shocks. The sign of the inflation risk premium depends on the covariance between the marginal utility of consumption and inflation, as determined by \(\pi_x \lambda_x\), and on the serial correlation of the shock, the \((2\rho - 1)\) term. If shocks are highly serially correlated \((\rho > 1/2)\), the inflation risk premium is positive when inflation is positively cross-correlated with the marginal utility of consumption. Typically, this implies that output and inflation shocks must be negatively correlated. If, instead, shocks are mildly serially correlated \((\rho < 1/2)\), then the risk premium tends to be positive when output and inflation are positively cross-correlated. When \(\rho = 1/2\), the two components of the inflation risk premium cancel each other out.

In the more general case with multiple state variables, inflation premia generated by the various shocks can be positive or negative and therefore cancel each other out. Based on the above intuition, we can in fact analyse a taxonomy of the inflation risk consequences of various shocks. Our 4 shocks are in fact such as to generate all possible combinations of signs of \(\pi_x\) and \(\lambda_x\).

1. Technology shocks tend to generate an increase in output and a fall in inflation and they therefore imply \(\pi_x < 0\) and \(\lambda_x < 0\). They will contribute to produce a positive inflation risk premium if their serial correlation is high (above 0.5).

2. Tax shocks imply \(\pi_x > 0\) and \(\lambda_x > 0\). As for technology shocks, they generate a negative correlation between output and inflation. They will also produce a positive inflation risk premium if they are highly serially correlated.

\(^9\)Note that \(\frac{\partial \hat{hpr}_{t,n}}{\partial \rho} = -(n - 1) \rho^{n-2} \lambda_x^2 \sigma^2\) and \(\frac{\partial (1 - \rho^{n-1})}{\partial n} = -\lambda_x^2 (\ln \rho) \rho^{n-1} \sigma^2\), so that \(\frac{\partial \hat{hpr}_{t,n}}{\partial \rho} < 0\) and \(\frac{\partial (1 - \rho^{n-1})}{\partial n} > 0\) for \(\rho > 0\).
3. Inflation target shocks imply $\pi_x > 0$ and $\lambda_x < 0$. In isolation, they therefore tend to contribute positively to the inflation risk premium only if their serial correlation is small (less than 0.5). They also tend to produce a positive output-inflation correlation.

4. Policy shocks, finally, imply $\pi_x < 0$ and $\lambda_x > 0$. They should also be little correlated to generate a positive inflation risk premium. They also tend to produce a positive output-inflation correlation.

The puzzling feature of excess holding period returns can therefore be seen from two viewpoints.

On the one hand, we can look at the decomposition of nominal term premia into real and inflation-induced components. In simple models, these two components tend to offset, rather than reinforce, each other, because of the way they are affected by the persistence of exogenous shocks. Real premia tend to be higher when the persistence of the shocks is lower, so that the precautionary demand motive is less important for long bonds. When the data imply a negative correlation between output and inflation, however, we are in cases 1 and 2 of the above taxonomy. Little persistent shocks, therefore, tend to weigh negatively on the inflation risk premium, thus to reduce the total risk premium.

On the other hand, the puzzle can be viewed from the angle of market prices of risk, and their link to the volatilities of inflation and the marginal utility of consumption. For inflation risk premia to affect long term bonds, shocks must be very persistent. Very persistent shocks, however, must have a very small variance to avoid generating an unrealistically large volatility in output and inflation. A small variance of shocks, however, also implies small market prices of risk and therefore weighs down on risk premia.

### 6.3 Monetary policy and term/risk premia

In this section, we investigate the effects on term and inflation risk premia of changes in the monetary policy rule. We focus, more precisely, on whether a more hawkish central bank, modelled here as a central bank which responds more aggressively to inflation deviations from target, is rewarded by financial markets with lower long term rates and a flatter yield curve. We test, in other words, the idea that a more determined inflation fighter will induce smaller inflation risk premia in the yield curve. Our main result is to emphasise again the
difference between inflation risk premia and total holding premia. In our simulations, the more hawkish central bank does induce a smaller inflation risk premium, but this does not translate into smaller $xhpr$’s.

Throughout the section, we focus on the $\gamma = 6$ case.

We first analyse the consequence of changes in the policy rule coefficient attached to inflation deviations from target. With respect to the benchmark results of the previous section, we consider two cases. The first is that of a milder inflation reaction coefficient, namely $\delta = 0.5$. The second case is that of a much higher $\delta$ coefficient, such that no actual deviations of inflation from target ever occur. It is well known that this happens under optimal policy within our simple model. Woodford (2003) has in fact shown that one way to characterise optimal policy is to replicate the flex-price allocation and that this can be achieved through a rule where the intercept is equal to the interest rate which would prevail if there were no nominal rigidities in the economy. It goes without saying that optimal policy also implies absence of inflation target shocks.

The implications on holding premia of the $\delta = 0.5$ and the optimal rule cases are shown in figure 5. The figure confirms, first, that a weak response to inflation does tend to boost inflation risk premia. In the figure, this becomes approximately three times larger than in the benchmark case, partly because of the parallel increase in the volatility of inflation and the nominal interest rate. In spite of the larger inflation risk premium, however, the total excess holding period return remains broadly unchanged, because of a compensating fall in the real excess holding period return associated. This, in turn, is associated to a reduction in the variance of output growth – a natural by-product of the increased volatility of inflation.

The most striking results of the figure, however, is that, while indeed reducing to zero inflation risk premia at all maturities, optimal policy does not lead to a flat yield curve or zero excess holding period returns. On the contrary, excess holding period returns increase dramatically over the whole maturity structure. At the 10-year horizon, excess holding period returns soar to over 2 percentage points. The main reason is a dramatic increase in the volatility of the short term interest rate, which is necessary to drive to zero the variance of inflation. The variance of the short term rate soars to over 10 percentage points, while the term structure of volatilities declines much faster with maturity because of the absence of interest rate smoothing motives.
This result shows that caution is needed when trying to draw conclusion on the inflationary determination of a central bank based on the slope of the nominal yield curve. The relationship between the aggressiveness of the monetary policy response to inflation and the size of excess holding period returns is likely to be nonlinear. It is larger both for very mild inflation response parameters, mainly due to larger inflation risk premia, and for very aggressive policy responses, because of the increase in interest rate volatility.

Figure 5: Excess returns decompositions under different policy rules.

Since these conclusions are partly affected by the fact that optimal policy entails absence of interest rate smoothing, as well as (implicitly) a very strong inflation response coefficient, we investigate next the effect on holding premia of different degrees of interest rate smoothing. Again, we consider only two illustrative cases: lower interest rate smoothing ($\rho_I = 0.5$) and superinertial policy (see Woorford, 2003), namely an interest rate smoothing coefficient above 1 ($\rho_I = 1.3$ in the simulation).

Lower interest rate smoothing implies a much more volatile interest rate and, due to the smaller persistence of interest rate changes, more protracted deviations of inflation from target. Consequently, inflation volatility is higher and output growth volatility lower. As a result, the real component of the excess holding period return becomes somewhat smaller, but there is almost a threefold increase in the inflation risk premium, as well as a marked increase in the volatility of yields at all maturities. At the 10-year horizon, the total excess holding period return increases to over 1 percentage points compared to the baseline, and the inflation risk premium component goes from 10 to 30 basis points.

With a superinertial rule, the real excess holding period return remains broadly unchanged compared to the benchmark case. The main difference is a doubling of the inflation
risk premium (20 basis points at the 10-year maturity). Intuitively, the superinertial rule implies that upward (downward) inflation deviations from target will be followed at some point by a downward (upward) correction. This induces more inflation variability in spite of a lower variability of the interest rate. The result is a higher inflation risk premium in spite of a lower volatility of yields across all maturities.

To summarise, excess holding period returns are nonlinearly related to the anti-inflationary determination of the central banks, namely the bank’s reaction coefficient to deviations of inflation from target. Compared to empirically estimated values of such coefficient, both a higher and a lower coefficient would tends to increase the slope of the yield curve. In the first case, this would happen because of a higher real risk premium; in the second case, because of an increase in the inflation risk premium.

Changes in the interest rate smoothing coefficient also have a nonlinear impact on the excess holding premium. A lower interest rate smoothing increases the inflation risk premium through a higher volatility of both interest rates and inflation. A larger smoothing coefficient tends to reduce the inflation risk premium. A superinertial rule, however, goes too far. In spite of the lower interest rate volatility, it increases the inflation risk premium – though by a lesser amount – because it induces higher volatility in inflation.

7 Conclusions

This paper analyses the term structure implications of a relatively standard microfounded DSGE model with nominal rigidities. The model also includes features traditionally employed to help explain the equity premium puzzle, such as a high degree of habit persistence in consumption. We use second order approximate solutions to allow for a meaningful role for risk.

Our results show that modern DSGE models are much closer to term structure data than previously thought. A relatively simple version of the model with only two persistent exogenous shocks can replicate quite well the sign and size of average excess holding period returns and the variance of yields across the term structure without unrealistic repercussions on the moments of macroeconomic variables. The model also shows that a hawkish monetary policy rule does not necessarily lead to a flatter yield curve – even if it does reduce inflation risk premia.
While matching relatively well the unconditional features of term structure data over a long period average, our approach is unable to generate time variations in risk premia. As already mentioned, this happens by construction when using second order approximations (unless shocks were assumed to be conditionally heteroskedastic, an assumption which is not standard in the macroeconomic literature). The hypothesis of constant risk premia, however, is rejected by many empirical applications in the so-called “essentially affine” term structure literature (see Duffee, 2002). In order to investigate the ability of DSGE models to yield state dependent prices of risk, higher-order or global approximation methods must be employed.
A Appendix

A.1 Equation (3)

By definition,

\[ B_{t,t+n} = E_t [Q_{t,t+n}] \]

and note that the non-stochastic steady state is such that \( B = Q \) and \( \ln B = \ln Q \). Now note that for any variable \( X \), a second order approximation around the log-steady state is

\[ X = x^1 + b x^2 + \frac{1}{2} b^2 \]

so that (simplifying the notation from \( B_{t,t+n} \) to \( B_{t;n} \))

\[ b^2 \left( 1 + \hat{b}_{t,n} + \frac{1}{2} \hat{b}^2_{t,n} \right) = E_t \left[ \hat{q} \left( 1 + \hat{q}_{t+n} + \frac{1}{2} \hat{q}^2_{t+n} \right) \right] \]

\[ = \hat{q} E_t \left[ 1 + \hat{q}_{t+n} + \frac{1}{2} \hat{q}^2_{t+n} \right] \]

or

\[ \hat{b}_{t,n} = E_t \left[ \hat{q}_{t+n} + \frac{1}{2} \hat{q}^2_{t+n} \right] - \frac{1}{2} \hat{b}^2_{t,n} \]

An expression for \( \hat{b}^2_{t,n} \) is needed, which can be obtained from its first order approximation. The latter can be denoted as

\[ \hat{b}_{t,n} = E_t \left[ \hat{q}_{t+n} \right] \]

so that

\[ \hat{b}^2_{t,n} = \left( E_t \left[ \hat{q}_{t+n} \right] \right)^2 \]

\[ \hat{b}_{t,n} = E_t \left[ \hat{q}_{t+n} \right] + \frac{1}{2} E_t \left[ \hat{q}^2_{t+n} \right] - \frac{1}{2} \left( E_t \left[ \hat{q}_{t+n} \right] \right)^2 \]

The price of a 1-period bond is therefore

\[ \hat{b}_{t,1} = -i_t = E_t \left[ \hat{q}_{t+1} \right] + \frac{1}{2} \text{Var} \left[ \hat{q}_{t+1} \right] \]

Now note that the log-stochastic discount factor is \( \hat{q}_{t,t+1} = \Delta \hat{\lambda}_{t+1} - \hat{\pi}_{t+1} \), so that

\[ \hat{q}_{t,t+n} = \hat{\lambda}_{t+n} - \hat{\lambda}_{t+1} - \sum_{i=1}^{n} \hat{\pi}_{t+i} \equiv \Delta^n \hat{\lambda}_{t+n} - \sum_{i=1}^{n} \hat{\pi}_{t+i} \]

and

\[ \hat{b}_{t,n} = E_t \left[ \Delta^n \hat{\lambda}_{t+n} \right] - \sum_{i=1}^{n} E_t \left[ \hat{\pi}_{t+i} \right] + \frac{1}{2} \text{Var}_t \left[ \Delta^n \hat{\lambda}_{t+n} - \sum_{i=1}^{n} \hat{\pi}_{t+i} \right] \]

35
Since \( \bar{ym}_{t,n} = -1/\hat{h}_{t,n} \), we also obtain

\[
\bar{ym}_{t,n} = - \frac{E_t \left[ \Delta^n \hat{\lambda}_{t+n} \right]}{n} + \sum_{i=1}^{n} \frac{E_t \left[ \hat{\pi}_{t+i} \right]}{n} - \frac{1}{2} \frac{n}{n} \text{Var}_t \left[ \Delta^n \hat{\lambda}_{t+n} - \sum_{i=1}^{n} \hat{\pi}_{t+i} \right]
\]

\[
= - \frac{E_t \left[ \Delta^n \hat{\lambda}_{t+n} \right]}{n} + \sum_{i=1}^{n} \frac{E_t \left[ \hat{\pi}_{t+i} \right]}{n} - \frac{1}{2} \frac{n}{n} \text{Var}_t \left[ \Delta^n \hat{\lambda}_{t+n} \right]
\]

\[
- \frac{1}{2} \frac{n}{n} \text{Var}_t \left[ \sum_{i=1}^{n} \hat{\pi}_{t+i} \right] + \frac{\text{Cov}_t \left[ \sum_{i=1}^{n} \hat{\pi}_{t+i}, \Delta^n \hat{\lambda}_{t+n} \right]}{n}
\]

and note that

\[
\hat{\pi}_t = \bar{ym}_{t,1} = -E_t \left[ \Delta^n \hat{\lambda}_{t+1} \right] + \frac{1}{2} \text{Var}_t \left[ \Delta^n \hat{\lambda}_{t+1} \right] - \frac{1}{2} \text{Var}_t \left[ \hat{\pi}_{t+1} \right] + \text{Cov}_t \left[ \hat{\pi}_{t+1}, \Delta^n \hat{\lambda}_{t+1} \right]
\]

It follows that

\[
\bar{ym}_{t,n} - \hat{\pi}_t = - \left( \frac{1}{n} E_t \left[ \Delta^n \hat{\lambda}_{t+n} \right] - E_t \left[ \Delta^n \hat{\lambda}_{t+1} \right] \right) - \frac{1}{2} \left( \frac{1}{n} \text{Var}_t \left[ \Delta^n \hat{\lambda}_{t+n} \right] - \text{Var}_t \left[ \Delta^n \hat{\lambda}_{t+1} \right] \right)
\]

\[
+ \left( \frac{1}{n} \text{Var}_t \left[ \sum_{i=1}^{n} \hat{\pi}_{t+i} \right] - \text{Var}_t \left[ \hat{\pi}_{t+1} \right] \right) - \frac{1}{2} \frac{n}{n} \text{Var}_t \left[ \sum_{i=1}^{n} \hat{\pi}_{t+i} \right] - \text{Cov}_t \left[ \hat{\pi}_{t+1}, \Delta^n \hat{\lambda}_{t+1} \right]
\]

and, in steady state,

\[
E \left[ \bar{ym}_{n} - \hat{\pi} \right] = - \frac{1}{2} \left( \frac{1}{n} \text{Var} \left[ \Delta^n \hat{\lambda} \right] - \text{Var} \left[ \Delta \hat{\lambda} \right] \right) - \frac{1}{2} \left( \frac{1}{n} \text{Var} \left[ \sum_{i=1}^{n} \hat{\pi}_i \right] - \text{Var} \left[ \hat{\pi} \right] \right)
\]

Real returns can be obtained similarly as

\[
\bar{ym}_{t,n} = - \frac{E_t \left[ \Delta^n \hat{\lambda}_{t+n} \right]}{n} - \frac{1}{2} \frac{n}{n} \text{Var}_t \left[ \Delta^n \hat{\lambda}_{t+n} \right]
\]

where again note that

\[
\hat{\pi}_t = \bar{ym}_{t,1} = -E_t \left[ \Delta^n \hat{\lambda}_{t+1} \right] - \frac{1}{2} \text{Var}_t \left[ \Delta^n \hat{\lambda}_{t+1} \right]
\]

Hence

\[
\bar{ym}_{t,n} - \hat{\pi}_t = - \left( \frac{1}{n} E_t \left[ \Delta^n \hat{\lambda}_{t+n} \right] - E_t \left[ \Delta^n \hat{\lambda}_{t+1} \right] \right) - \frac{1}{2} \left( \frac{1}{n} \text{Var}_t \left[ \Delta^n \hat{\lambda}_{t+n} \right] - \text{Var}_t \left[ \Delta^n \hat{\lambda}_{t+1} \right] \right)
\]

and

\[
E \left[ \bar{ym}_{n} - \hat{\pi} \right] = - \frac{1}{2} \left( \frac{1}{n} \text{Var} \left[ \Delta^n \hat{\lambda}_n \right] - \text{Var} \left[ \Delta \hat{\lambda} \right] \right)
\]

36
Finally, in the special case with CRRA utility we obtain that \( \Delta_n \hat{\lambda}_n = -\gamma \Delta_n \hat{y}_n \), so that

\[
E \left[ \hat{y}_{tn} - \bar{y} \right] = -\frac{1}{2} \gamma^2 \left( \frac{1}{n} \text{Var} \left[ \Delta_n \hat{y}_n \right] - \text{Var} \left[ \Delta \bar{y} \right] \right) - \frac{1}{2} \left( \frac{1}{n} \text{Var} \left[ \sum_{i=1}^{n} \hat{\pi}_i \right] - \text{Var} \left[ \hat{\pi} \right] \right) + \gamma \left( -\frac{1}{n} \text{Cov} \left[ \sum_{i=1}^{n} \hat{\pi}_i, \Delta_n \hat{y}_n \right] + \text{Cov} \left[ \hat{\pi}, \Delta \bar{y} \right] \right)
\]

and

\[
E \left[ \hat{y}_{tn} - \bar{y} \right] = -\frac{1}{2} \gamma^2 \left( \frac{1}{n} \text{Var} \left[ \Delta_n \hat{y}_n \right] - \text{Var} \left[ \Delta \bar{y} \right] \right)
\]

Using the above results, we can write

\[
\hat{y}_{tn} = \hat{y}_{tn} \text{real} + \sum_{i=1}^{n} E_t \left[ \hat{\pi}_{t+i} \right] - \frac{1}{2} \text{Var}_t \left[ \sum_{i=1}^{n} \hat{\pi}_{t+i} \right] + \frac{\text{Cov}_t \left[ \sum_{i=1}^{n} \hat{\pi}_{t+i}, \Delta_n \hat{\lambda}_{t+n} \right]}{n}
\]

or, in steady state,

\[
E \left[ \hat{y}_{tn} \right] = E \left[ \hat{y}_{tn} \right] + E \left[ \hat{\pi} \right] - \frac{1}{2} \text{Var} \left[ \sum_{i=1}^{n} \hat{\pi}_i \right] + \frac{1}{n} \text{Cov} \left[ \sum_{i=1}^{n} \hat{\pi}_i, \Delta \hat{\lambda} \right]
\]

where

\[
E \left[ \hat{y}_{tn} \text{real} \right] = E_t \left[ \Delta_n \hat{\lambda} / n \right] - \frac{1}{2} \text{Var} \left[ \Delta_n \hat{\lambda} \right] = -\Delta \hat{\lambda} - \frac{1}{2} \text{Var} \left[ \Delta_n \hat{\lambda} \right]
\]

so that

\[
E \left[ \hat{y}_{tn} \right] = -\Delta \hat{\lambda} - \frac{1}{2} \text{Var} \left[ \Delta_n \hat{\lambda} \right] + E \left[ \hat{\pi} \right] - \frac{1}{2} \text{Var} \left[ \sum_{i=1}^{n} \hat{\pi}_i \right] + \frac{1}{n} \text{Cov} \left[ \sum_{i=1}^{n} \hat{\pi}_i, \Delta_n \hat{\lambda} \right]
\]

Once again, under CRRA utility we obtain

\[
E \left[ \hat{y}_{tn} \right] = \gamma \Delta \bar{y} + E \left[ \hat{\pi} \right] - \frac{1}{2} \gamma^2 \frac{1}{n} \text{Var} \left[ \Delta \bar{y} \right] - \frac{1}{2} \text{Var} \left[ \sum_{i=1}^{n} \hat{\pi}_i \right] - \gamma \frac{1}{n} \text{Cov} \left[ \sum_{i=1}^{n} \hat{\pi}_i, \Delta \bar{y} \right]
\]

### A.2 Equation (5)

To obtain a recursive formulation, start from

\[
B_{t,n+1} = E_t \left[ Q_{t+1} B_{t+1,n} \right]
\]

and note that the non-stochastic steady state is such that \( B_{n+1} = Q_1 B_n \) and \( \ln B_{n+1} = \ln Q_1 + \ln B_n \). Following the same reasoning as above, a second order approximation
around the log-steady state is

\[
\hat{b}_{t,n+1} = E_t [\hat{q}_{t+1}] + E_t [\hat{b}_{t+1,n}] \\
+ \frac{1}{2} \left( E_t [\hat{q}_{t+1}^2] - (E_t [\hat{q}_{t+1}])^2 \right) \\
+ \frac{1}{2} \left( E_t [\hat{b}_{t+1,n}^2] - (E_t [\hat{b}_{t+1,n}])^2 \right) \\
+ E_t [\hat{q}_{t+1} \hat{b}_{t+1,n}] - E_t [\hat{q}_{t+1}] E_t [\hat{b}_{t+1,n}]
\]

or, equivalently,

\[
\hat{b}_{t,n+1} = E_t [\hat{q}_{t+1}] + E_t [\hat{b}_{t+1,n}] + \frac{1}{2} \text{Var}_t [\hat{q}_{t+1} + \hat{b}_{t+1,n}]
\]

Alternatively, these can be written as

\[
\hat{b}_{t,n+1} = \hat{b}_{t,1} + E_t [\hat{b}_{t+1,n}] \\
+ \frac{1}{2} \left( E_t [\hat{b}_{t+1,n}^2] - (E_t [\hat{b}_{t+1,n}])^2 \right) \\
+ E_t [\hat{q}_{t+1} \hat{b}_{t+1,n}] - E_t [\hat{q}_{t+1}] E_t [\hat{b}_{t+1,n}]
\]

which can in turn be written as

\[
\hat{b}_{t,n+1} = \hat{b}_{t,1} + E_t [\hat{b}_{t+1,n}] + \frac{1}{2} \text{Var}_t [\hat{b}_{t+1,n} + \text{Cov}_t [\hat{q}_{t+1}, \hat{b}_{t+1,n}]]
\]

A.3 Equation (6)

From \(HPR_{t,n} = E_t [B_{t+1,n-1}/B_{t,n}]\), we can apply the usual approximation to obtain

\[
\hat{hpr}_{t,n} = E_t [\hat{b}_{t+1,n-1}] - \hat{b}_{t,n} + \frac{1}{2} \left( E_t [\hat{b}_{t+1,n-1}^2] - (E_t [\hat{b}_{t+1,n-1}])^2 \right)
\]

or

\[
\hat{hpr}_{t,n} = E_t [b_{t+1,n-1}] - \hat{b}_{t,n} + \frac{1}{2} \text{Var} (\hat{b}_{t+1,n-1})
\]

Using equation (5), we can rewrite this as

\[
\hat{hpr}_{t,n} = -\hat{b}_{t,1} - \text{Cov}_t [\hat{q}_{t+1}, \hat{b}_{t+1,n-1}]
\]

Given that, for the short term holding period return, we get

\[
\hat{hpr}_{t,1} = -\hat{b}_{t,1} = \hat{i}_t
\]
the \( x_{hpr} \) on a \( n \)-period bond is

\[
\hat{hpr}_{t,n} - \hat{\iota}_t = -Cov\left( \hat{q}_{t,1}, \hat{b}_{t+1,n-1} \right)
\]

or, using the expression for the stochastic discount factor,

\[
\hat{hpr}_{t,n} - \hat{\iota}_t = Cov\left( \hat{\pi}_{t+1}, \hat{b}_{t+1,n-1} \right) - Cov\left( \Delta \hat{\lambda}_{t+1}, \hat{b}_{t+1,n-1} \right)
\]

Finally, notice that this could also be rewritten in terms of yields to maturity \( ytm_{t+1,n-1} \), which are defined as

\[
ytm_{t+1,n-1} = -\frac{1}{n-1} \hat{b}_{t+1,n-1}
\]

We obtain

\[
\hat{hpr}_{t,n} - \hat{\iota}_t = (n-1) \left( Cov\left( \hat{\lambda}_{t+1}, ytm_{t+1,n-1} \right) - Cov\left( \hat{\pi}_{t+1}, ytm_{t+1,n-1} \right) \right)
\]

### A.4 Solution for bond prices

Recall that

\[
\hat{x}_{t+1} = c_1 \hat{x}_t + \frac{1}{2} \begin{bmatrix} \cdots \end{bmatrix} + \frac{1}{2} c_0 \sigma^2 + \eta \sigma \varepsilon_{t+1}
\]

\[
\hat{\lambda}_t = \lambda'_{x} \hat{x}_t + \frac{1}{2} \hat{x}_t \lambda_{xx} \hat{x}_t + \frac{1}{2} \lambda_{\sigma \sigma} \sigma^2
\]

\[
\hat{\pi}_t = \pi'_{x} \hat{x}_t + \frac{1}{2} \hat{x}_t \pi_{xx} \hat{x}_t + \frac{1}{2} \pi_{\sigma \sigma} \sigma^2
\]

and stack the state variables to write

\[
\hat{x}_{t+1} = c_1 \hat{x}_t + \frac{1}{2} \begin{bmatrix} \cdots \end{bmatrix} + \frac{1}{2} \sigma^2 c_0 + \sigma \eta \varepsilon_{t+1}
\]

where now \( c_1 \) is a \( n_x \times n_x \) matrix, \( c_0 \) is a \( n_x \times 1 \) vector and \( c_i \) is the \( i \)-th page (of dimension \( n_x \times n_x \)) of a \( n_x \)-pages tensor and finally \( \eta \) is a \( n_x \times n_x \) matrix.

#### A.4.1 1-period bonds

Even if this is known from the solution for the short term rate, it is instructive to derive first the price of a 1-period bond. For this purpose, note first that a second order approximation to the stochastic discount factor is

\[
\hat{q}_{t+1} = (\lambda'_{x} - \pi'_{x}) \hat{x}_{t+1} + \frac{1}{2} \hat{x}_{t+1} (\lambda_{xx} - \pi_{xx}) \hat{x}_{t+1} - \lambda'_{x} \hat{x}_t - \frac{1}{2} \hat{x}_t \lambda_{xx} \hat{x}_t - \frac{1}{2} \pi_{\sigma \sigma} \sigma^2
\]
\[
\hat{q}_{t,t+1} = ((\lambda'_x - \pi'_x) c_1 - \lambda'_t) \hat{x}_t + \frac{1}{2} (\lambda'_x - \pi'_x) \begin{bmatrix} \ldots \ldots \\ \ldots \ldots \end{bmatrix} \\
+ \frac{1}{2} \hat{x}'_t c'_1 (\lambda_{xx} - \pi_{xx}) c_1 \hat{x}_t - \frac{1}{2} \hat{x}'_t \lambda_{xx} \hat{x}_t \\
+ \frac{1}{2} \sigma^2 (\lambda'_x - \pi'_x) c_0 - \frac{1}{2} \pi \sigma \sigma^2 \\
+ \sigma (\lambda'_x - \pi'_x) \eta \varepsilon_{t+1} + \frac{1}{2} \sigma \hat{x}'_t c'_1 (\lambda_{xx} - \pi_{xx}) \eta \varepsilon_{t+1} \\
+ \frac{1}{2} \sigma \varepsilon'_{t+1} \eta' (\lambda_{xx} - \pi_{xx}) c_1 \hat{x}_t + \frac{1}{2} \sigma^2 \varepsilon'_{t+1} \eta' (\lambda_{xx} - \pi_{xx}) \eta \varepsilon_{t+1}
\]

Now recall that the price of a 1-period bond is

\[
\hat{b}_{t,1} = -i_t = E_t [\hat{q}_{t+1}] + \frac{1}{2} \left( E_t [\hat{q}_{t+1}] - (E_t [\hat{q}_{t+1}])^2 \right)
\]

for which we need

\[
E_t [\hat{q}_{t+1}] - (E_t [\hat{q}_{t+1}])^2 = \sigma^2 (\lambda'_x - \pi'_x) \eta \eta' (\lambda'_x - \pi'_x)'
\]

and

\[
E_t [\hat{q}_{t,t+1}] = (\lambda'_x - \pi'_x) c_1 \hat{x}_t + \frac{1}{2} (\lambda'_x - \pi'_x) \begin{bmatrix} \ldots \ldots \\ \ldots \ldots \end{bmatrix} \\
+ \frac{1}{2} \hat{x}'_t c'_1 (\lambda_{xx} - \pi_{xx}) c_1 \hat{x}_t + \frac{1}{2} \sigma^2 E_t [\varepsilon'_{t+1} \eta' (\lambda_{xx} - \pi_{xx}) \eta \varepsilon_{t+1}]
\]

\[
-\lambda'_x \hat{x}_t - \frac{1}{2} \hat{x}'_t \lambda_{xx} \hat{x}_t - \frac{1}{2} \pi \sigma \sigma^2
\]

Now note that, for any matrix \( A \) and vector \( x \),

\[
E [x' Ax] = E [\text{vec} (x' Ax)] \\
= E [x' \otimes x'] E (A) \\
= \text{vec} (E [xx'])' \text{vec} (A)
\]

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

\[
E_t [\varepsilon'_{t+1} \eta' (\lambda_{xx} - \pi_{xx}) \eta \varepsilon_{t+1}] = (\text{vec} (I)) \text{vec} (\eta' (\lambda_{xx} - \pi_{xx}) \eta) \\
= \text{tr} (\eta' (\lambda_{xx} - \pi_{xx}) \eta)
\]

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

$$\begin{align*}
\tilde{b}_{t,1} &= ((\lambda'_x - \pi'_x) c_1 - \lambda'_x) \tilde{x}_t + \frac{1}{2} \sigma^2 ((\lambda'_x - \pi'_x) c_0 - \pi_{\sigma \sigma}) \\
&\quad + \frac{1}{2} \sigma^2 \text{tr} (\eta' (\lambda_{xx} - \pi_{xx}) \eta) + \frac{1}{2} \sigma^2 (\lambda'_x - \pi'_x) \eta \eta' (\lambda'_x - \pi'_x)'
\end{align*}$$

$$+ \frac{1}{2} \tilde{x}'_t (c'_1 (\lambda_{xx} - \pi_{xx}) c_1 - \lambda_{xx}) \tilde{x}_t + \frac{1}{2} (\lambda'_x - \pi'_x) \left[ \tilde{x}'_t c_2 \tilde{x}_t \right]$$

Finally, note that

$$\begin{align*}
(\lambda'_x - \pi'_x) \left[ \begin{array}{c} ... \\ \tilde{x}'_t c_2 \tilde{x}_t \\ ... \end{array} \right] &= \sum_{i=1}^{n_x} (\lambda'_x [i] - \pi'_x [i]) \tilde{x}'_t c_2 [i] \tilde{x}_t \\
&= \sum_{i=1}^{n_x} \tilde{x}'_t (\lambda'_x [i] - \pi'_x [i]) c_2 [i] \tilde{x}_t \\
&= \tilde{x}'_t \left( \sum_{i=1}^{n_x} (\lambda'_x [i] - \pi'_x [i]) c_2 [i] \right) \tilde{x}_t
\end{align*}$$

where the second equality follows from the fact that $\lambda'_x [i]$ and $\pi'_x [i]$ are scalar elements of $\lambda'_x$ and $\pi'_x$, respectively. We can therefore rewrite the 1-period bond as

$$\tilde{b}_{t,1} = \tilde{b}'_{1,x} \tilde{x}_t + \frac{1}{2} \sigma^2 \tilde{b}_{1,\sigma \sigma} + \frac{1}{2} \tilde{x}'_t \tilde{b}_{1,xx} \tilde{x}_t$$

where

$$\begin{align*}
\tilde{b}'_{1,x} &= (\lambda'_x - \pi'_x) c_1 - \lambda'_x \\
\tilde{b}_{1,\sigma \sigma} &= (\lambda'_x - \pi'_x) c_0 - \pi_{\sigma \sigma} + \text{tr} (\eta' (\lambda_{xx} - \pi_{xx}) \eta) + (\lambda'_x - \pi'_x) \eta \eta' (\lambda'_x - \pi'_x)'
\end{align*}$$

$$\tilde{b}_{1,xx} = c'_1 (\lambda_{xx} - \pi_{xx}) c_1 - \lambda_{xx} + \sum_{i=1}^{n_x} (\lambda'_x [i] - \pi'_x [i]) c_2 [i]$$

Note also that, by construction, $E \left[ \tilde{b}_1 \right] = -E \left[ \tilde{i} \right]$, so $\tilde{b}_{1,x} = -\tilde{i}_x$, $\tilde{b}_{1,xx} = -\tilde{i}_{xx}$ and $\tilde{b}_{1,\sigma \sigma} = -\tilde{i}_{\sigma \sigma}$.

Note that this definition also allows us to rewrite $\tilde{q}_{t,t+1}$ as

$$\begin{align*}
\tilde{q}_{t,t+1} &= \tilde{b}'_{1,x} \tilde{x}_t + \frac{1}{2} \tilde{x}'_t \tilde{b}_{1,xx} \tilde{x}_t + \frac{1}{2} \sigma^2 ((\lambda'_x - \pi'_x) c_0 - \pi_{\sigma \sigma}) \\
&\quad + \sigma (\lambda'_x - \pi'_x) \eta \tilde{\varepsilon}_{t+1} + \sigma \tilde{x}'_t c'_1 (\lambda_{xx} - \pi_{xx}) \eta \tilde{\varepsilon}_{t+1} + \frac{1}{2} \sigma^2 \tilde{x}'_t \eta' (\lambda_{xx} - \pi_{xx}) \eta \tilde{\varepsilon}_{t+1}
\end{align*}$$

A.4.2 2-period bonds

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

$$\tilde{b}_{t,2} = \tilde{b}_{t,1} + E_t \left[ \tilde{b}_{t+1,1} \right] + \frac{1}{2} \text{Var}_t \left[ \tilde{b}_{t+1,1} \right] + \text{Cov}_t \left[ \tilde{q}_{t+1, \tilde{b}_{t+1,1}} \right]$$
Based on 1-period prices, we can derive

\[ E_t \left[ \hat{b}_{t+1,1} \right] = \hat{b}_{1,x} c_1 \hat{x}_t + \frac{1}{2} \hat{b}_{1,x}' \begin{bmatrix} \hat{x}_t' \hat{x}_t & \cdots & \hat{x}_t' \hat{x}_t \end{bmatrix} + \frac{1}{2} \hat{x}_t' c_1' \hat{b}_{1,xx} c_1 \hat{x}_t \]

\[ + \frac{1}{2} \hat{b}_{1,x} c_0 \sigma^2 + \frac{1}{2} \sigma^2 \hat{b}_{1,\sigma} + \frac{1}{2} \sigma^2 \text{tr} \left[ \eta \hat{b}_{1,xx} \eta \right] \]

and

\[ E_t \left[ \hat{b}_{t+1,1} \hat{b}_{t+1,1}' \right] - E_t \left[ \hat{b}_{t+1,1}' \right] E_t \left[ \hat{b}_{t+1,1} \right] = \sigma^2 \hat{b}_{1,x} \eta' \hat{b}_{1,x} \]

\[ E_t \left[ \hat{b}_{t+1,1} \hat{b}_{t+1}' \right] - E_t \left[ \hat{b}_{t+1,1}' \right] E_t \left[ \hat{b}_{t+1} \right] = \hat{b}_{1,x} \eta' (\lambda_x - \pi_x) \]

It follows that

\[ \hat{b}_{t,2} = \hat{b}_{2,x} \hat{x}_t + \frac{1}{2} \hat{x}_t' \hat{b}_{2,xx} \hat{x}_t + \frac{1}{2} \sigma^2 \hat{b}_{2,\sigma} \]

where

\[ \hat{b}_{2,x} = \hat{b}_{1,x} (I + c_1) \]

\[ \hat{b}_{2,xx} = \hat{b}_{1,xx} + c_1' \hat{b}_{1,xx} c_1 + \sum_{i=1}^{n_x} \hat{b}_{i,x} [i] c_2 [i] \]

\[ \hat{b}_{2,\sigma} = 2 \hat{b}_{1,\sigma} + \hat{b}_{1,x} c_0 + \text{tr} \left( \eta' \hat{b}_{1,xx} \eta \right) + \hat{b}_{1,x} \eta' \hat{b}_{1,x} + 2 \hat{b}_{1,x} \eta' (\lambda_x - \pi_x) \]

### A.4.3 n-period bonds

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

\[ \hat{b}_{t,n} = \hat{b}_{n,x} \hat{x}_t + \frac{1}{2} \hat{x}_t' \hat{b}_{n,xx} \hat{x}_t + \frac{1}{2} \sigma^2 \hat{b}_{n,\sigma} \]

where for \( n > 1 \)

\[ \hat{b}_{n,x} = \hat{b}_{1,x} + \hat{b}_{n-1,x} c_1 \]

\[ \hat{b}_{n,xx} = \hat{b}_{1,xx} + c_1' \hat{b}_{n-1,xx} c_1 + \sum_{i=1}^{n_x} \hat{b}_{i,x} [i] c_2 [i] \]

\[ \hat{b}_{n,\sigma} = \hat{b}_{1,\sigma} + \hat{b}_{n-1,\sigma} + \hat{b}_{n-1,x} c_0 + \text{tr} \left( \eta' \hat{b}_{n-1,xx} \eta \right) + \hat{b}_{n-1,x} \eta' \hat{b}_{n-1,x} + 2 \hat{b}_{n-1,x} \eta' (\lambda_x - \pi_x) \]

Note that the first order term \( \hat{b}_{n,x} \) can be solved explicitly as

\[ \hat{b}_{n,x} = \hat{b}_{1,x} + \hat{b}_{n-1,x} c_1 \]

\[ = \hat{b}_{1,x} \sum_{j=0}^{n-1} c_1^j \]

\[ = \hat{b}_{1,x} (I - c_1)^{-1} (I - c_1^n) \]
A.4.4 Unconditional moments of n-period bonds

The unconditional mean of an n-period bond is

\[
E[b_{t,n}] = \tilde{b}_{n,x} E[\tilde{x}] + \frac{1}{2} E[\tilde{x} \tilde{b}_{n,x} \tilde{x}] + \frac{1}{2} \sigma^2 \tilde{b}_{n,\sigma}
\]

and the unconditional variance

\[
\text{Var}[b_{t,n}] = E\left[(\tilde{b}_{1,x} (I - c_1)^{-1} (I - c_1^n) \tilde{x}_t) \left(\tilde{b}_{1,x} (I - c_1)^{-1} (I - c_1^n) \tilde{x}_t\right)\right]
\]

For yields, we obtain

\[
\text{Var}[\tilde{y}_{tm,n}] = \frac{1}{n^2} \left(\tilde{b}_{1,x} (I - c_1)^{-1} (I - c_1^n) E[xx'] (I - c_1^n)' (I - c_1)^{-1}' \tilde{b}_{1,x}\right)
\]

Finally note the ratio

\[
\frac{\text{Var}[\tilde{y}_{tm,n}]}{\text{Var}[\tilde{t}]} = \frac{1}{n^2} \frac{\tilde{b}_{1,x} (I - c_1)^{-1} (I - c_1^n) E[xx'] (I - c_1^n)' (I - c_1)^{-1}' \tilde{b}_{1,x}}{\tilde{b}_{1,x} E[xx'] \tilde{b}_{1,x}}
\]

which in the scalar case, redefining \(c_1\) as \(\rho\), becomes

\[
\frac{\text{Var}[\tilde{y}_{tm,n}]}{\text{Var}[\tilde{t}]} = \left(\frac{1 - \rho^n}{n} \frac{1 - \rho}{1 - \rho}\right)^2
\]

A.5 Holding period return in the flex-price no-habit case

The model solution in the flex-price no-habit case is

\[
Y_t = \left(\mu - \frac{\alpha}{\chi}\right) \frac{\phi}{\phi + \alpha(\gamma - 1)} A_t^{\phi} \frac{A_t^{\phi}}{\phi + \alpha(\gamma - 1)}
\]

while bond prices are

\[
N_{n,t} = \beta^n (Y_t)^{\gamma} E_t \left[(Y_{t+n})^{-\gamma}\right]
\]

Substituting out output and evaluating expectations

\[
N_{n,t} = \beta^n A_t^{\gamma(1 - \rho^n)} \frac{1 - \rho^n}{1 - \rho} \left(\frac{\phi}{\phi + \alpha(\gamma - 1)}\right)^2 \sigma_t^2
\]
which characterize the whole term-structure in closed form. Note that, if we focus on continuously compounded returns defined as $\tilde{r}_{n,t} = -\frac{1}{n} \ln N_{n,t}$, we obtain

$$\tilde{ytm}_{n,t} = a_n + b_n \ln A_t$$

where

$$a_n = -\ln \beta - \frac{1}{2} \frac{1 - \rho^{2n}}{1 - \rho^2} \left( \frac{\gamma \phi}{\phi + \alpha (\gamma - 1)} \right)^2 \sigma^2$$

$$b_n = -\frac{1}{n} \frac{\gamma \phi}{\phi + \alpha (\gamma - 1)}$$

Finally, note that the holding period return is

$$HPR_{n,t} = E_t \left[ \frac{N_{t+1}^{n-1}}{N_t^n} \right]$$

$$= \frac{e^{\frac{1}{2} \left( \frac{\gamma \phi}{\gamma \phi + \alpha + \phi} \right)^2 (1-2\rho^{n-1}) \sigma^2}}{\beta} \frac{\gamma \phi (1-\rho)}{\gamma \phi + \alpha + \phi} a_t$$

so that

$$hpr_{t,n} = \ln HPR_{t,n}$$

$$= \frac{\frac{1}{2} \left( \frac{\gamma \phi}{\gamma \phi + \alpha + \phi} \right)^2 (1-2\rho^{n-1}) \sigma^2}{\beta} - \frac{\gamma \phi (1-\rho)}{\gamma \phi + \alpha + \phi} a_t$$

and the $xhpr$ is

$$hpr_{n,t} - hpr_{1,t} = (1-\rho^{n-1}) \left( \frac{\gamma \phi}{\gamma \phi + \alpha + \phi} \right)^2 \sigma^2$$

### A.6 Analytical $xhpr$ and the market prices of risk

Recall that

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

and that, up to a first order approximation, we can write $\tilde{b}_{t+1,n-1} = \tilde{b}'_{n-1,x} \tilde{x}_{t+1}$, $\Delta \tilde{\lambda}_{t+1} = \lambda'_{x'} (\tilde{x}_{t+1} - \tilde{x}_t)$. Since first order approximate expressions are sufficient to evaluate second moments up to a second order approximation, we obtain

$$\tilde{hpr}_{t,n} - \tilde{i}_t = \text{Cov}_t \left[ \pi_{x'} \tilde{x}_{t+1}, \tilde{b}'_{n-1,x} \tilde{x}_{t+1} \right] - \text{Cov}_t \left[ \lambda'_{x'} (\tilde{x}_{t+1} - \tilde{x}_t), \tilde{b}'_{n-1,x} \tilde{x}_{t+1} \right]$$

$$= (\pi_{x} - \lambda_{x})' (E_t \left[ \tilde{x}_{t+1} \tilde{x}_{t+1}' \right]) - E_t \left[ \tilde{x}_{t+1} \right] E_t \left[ \tilde{x}_{t+1}' \right] \tilde{b}_{n-1,x}$$

$$= \sigma^2 (\pi_{x} - \lambda_{x})' \eta_{y} \tilde{b}_{n-1,x}$$

$$= \sigma^2 \tilde{b}'_{n-1,x} \eta_{y} (\pi_{x} - \lambda_{x})$$
where the last transposition can be performed because the whole expression is a scalar.

Using the solution for bond prices, the excess holding period return can equivalently be written as

\[
\tilde{\text{hpr}}_{t,n} - \tilde{\tau}_t = \sigma^2 \hat{b}_{1,x}^t (I - c_1)^{-1} (I - c_1^{n-1}) \eta' (\pi_x - \lambda_x)
\]

\[
= \sigma^2 (\lambda'_x (I - c_1) + \pi'_x c_1) (I - c_1)^{-1} (I - c_1^{n-1}) \eta' (\pi_x - \lambda_x)
\]

This can be rewritten as

\[
\tilde{\text{hpr}}_{t,n} - \tilde{\tau}_t = \sigma \hat{b}_{n-1,x}^t \eta \xi
\]

where \( \xi \equiv \sigma \eta' (\pi_x - \lambda_x) \) is an endogenous vector of "(nominal) market prices of risk". The nominal market prices of risk can be divided into market prices of real risk and of inflation risk, \( \xi_\lambda = -\sigma \eta' \lambda_x \) and \( \xi_\pi = \sigma \eta' \pi_x \), respectively.

It follows that the nominal excess holding period return can be written as

\[
\tilde{\text{hpr}}_{t,n} - \tilde{\tau}_t = \left( \tilde{\text{hpr}}^\text{real}_{t,n} - \tilde{\tau}_t \right) - \sigma \left( \lambda'_x (I - c_1^{n-1}) \eta \xi_\pi + \pi'_x c_1 (I - c_1)^{-1} (I - c_1^{n-1}) \eta \xi_\lambda \right)
\]

\[-\sigma^2 \pi'_x c_1 (I - c_1)^{-1} (I - c_1^{n-1}) \eta' \pi_x
\]

where

\[
\tilde{\text{hpr}}^\text{real}_{t,n} - \tilde{\tau}_t = -\sigma \lambda'_x (I - c_1^{n-1}) \eta \xi_\lambda
\]

and the other two components are an "excess" inflation risk premium and a convexity term. The former, more specifically, can be written as

\[
\tilde{\text{hpr}}_{t,n} - \tilde{\tau}_{t,1} = \sigma \lambda'_x c_1^{n-1} \eta \xi_\pi - \sigma \pi'_x (I - c_1)^{-1} (I - c_1^{n-1}) c_1 \eta \xi_\lambda
\]

In the scalar case (setting \( c_1 = \rho \) and \( \eta = 1 \)), we obtain

\[
\tilde{\text{hpr}}_{t,n} - \tilde{\tau}_{t,1} = (\lambda_x - \pi_x) (\lambda_x (1 - \rho) + \pi_x \rho) \frac{1 - \rho^{n-1}}{1 - \rho} \sigma^2
\]

The decomposition becomes

\[
\tilde{\text{hpr}}_{t,n} - \tilde{\tau}_t = \lambda_x^2 (1 - \rho^{n-1}) \sigma^2 - \pi_x \lambda_x (1 - 2 \rho) \frac{1 - \rho^{n-1}}{1 - \rho} \sigma^2 - \pi_x^2 \rho \frac{1 - \rho^{n-1}}{1 - \rho} \sigma^2
\]

This shows first that real excess return is always positive, while the convexity term is always negative (for positive autocorrelation of the shocks). The sign of the nominal component depends on the sign of the term \( -\lambda_x \pi_x (1 - 2 \rho) / (1 - \rho) \), namely on the sign of \( \lambda_x \) and \( \pi_x \).
A.7 The slope of the real term structure

Note that, to a first order approximation,

\[ x_{t+n} = c_1^n x_t + \sigma \sum_{i=1}^{n} c_1^{n-i} \eta \varepsilon_{t+i} \]

and

\[ \lambda_{t+n} = \lambda'_x c_1^n x_t + \sigma \lambda'_x \sum_{i=1}^{n} c_1^{n-i} \eta \varepsilon_{t+i} \]

while

\[
\Delta \lambda_{t+n} = \lambda_{t+n} - \lambda_{t+n-1} = \lambda'_x c_1^n x_t - \lambda'_x c_1^{n-1} x_t + \sigma \lambda'_x \left( \sum_{i=1}^{n} c_1^{n-i} \eta \varepsilon_{t+i} - \sum_{i=1}^{n-1} c_1^{n-1-i} \eta \varepsilon_{t+i} \right) \\
= -\lambda'_x (I - c_1) c_1^{n-1} x_t - \sigma \lambda'_x (I - c_1) \sum_{i=1}^{n-1} c_1^{n-1-i} \eta \varepsilon_{t+i} + \sigma \lambda'_x \eta \varepsilon_{t+n}
\]

and

\[
\Delta^n \lambda_{t+n} = \sum_{j=1}^{n} \Delta \lambda_{t+j} = -\lambda'_x (I - c_1) c_1^{j-1} x_t - \sigma \lambda'_x (I - c_1) \sum_{j=1}^{n} \sum_{i=1}^{j-1} c_1^{j-1-i} \eta \varepsilon_{t+i} + \sigma \lambda'_x \eta \sum_{j=1}^{n} \varepsilon_{t+j} \\
= -\lambda'_x (1 - c_1^n) x_t - \sigma \lambda'_x (I - c_1) \sum_{j=1}^{n} \sum_{i=1}^{j-1} c_1^{j-1-i} \eta \varepsilon_{t+i} + \sigma \lambda'_x \eta \sum_{j=1}^{n} \varepsilon_{t+j}
\]

Note that

\[ \Delta \lambda_{t+1} = -\lambda'_x (I - c_1) x_t + \sigma \lambda'_x \eta \varepsilon_{t+1} \]

Since

\[
\sum_{j=1}^{n} \sum_{i=1}^{j-1} c_1^{j-1-i} \eta \varepsilon_{t+i} = \eta \varepsilon_{t+1} + (c_1 \eta \varepsilon_{t+1} + \eta \varepsilon_{t+2}) + ... + (c_1^{n-2} \eta \varepsilon_{t+1} + ... + \eta \varepsilon_{t+n-1}) \\
= \sum_{i=0}^{n-2} c_1^i \eta \varepsilon_{t+1} + \sum_{i=0}^{n-3} c_1^i \eta \varepsilon_{t+2} + ... + \sum_{i=0}^{n-n+1} c_1^i \eta \varepsilon_{t+n-2} + \eta \varepsilon_{t+n-1} \\
= (I - c_1)^{-1} \sum_{j=1}^{n-1} \left( I - c_1^{n-j} \right) \eta \varepsilon_{t+j}
\]

we can write

\[
\Delta^n \lambda_{t+n} = -\lambda'_x (I - c_1^n) x_t - \sigma \lambda'_x \sum_{j=1}^{n-1} \left( I - c_1^{n-j} \right) \eta \varepsilon_{t+j} + \sigma \lambda'_x \eta \sum_{j=1}^{n} \varepsilon_{t+j}
\]
Hence

\[
\begin{align*}
&\mathbb{E}_t \left[ (\lambda_{t+n} - \lambda_t) (\lambda'_{t+n} - \lambda'_t) \right] - \mathbb{E}_t \left[ (\lambda_{t+n} - \lambda_t) \right] \mathbb{E}_t \left[ (\lambda'_{t+n} - \lambda'_t) \right] \\
&= \sigma^2 \lambda'_x \mathbb{E}_t \left[ \left( \sum_{j=1}^{n-1} (I - c_1^{n-j}) \eta \bar{e}_{t+j} \right) \left( \sum_{j=1}^{n-1} \bar{e}'_{t+j} \eta' \left( I - c_1^{n-j} \right)' \right) \right] \lambda_x \\
&\quad - \sigma^2 \lambda'_x \mathbb{E}_t \left[ \left( \sum_{j=1}^{n-1} (I - c_1^{n-j}) \eta \bar{e}_{t+j} \right) \sum_{j=1}^{n-1} \bar{e}'_{t+j} \right] \eta' \lambda_x \\
&\quad - \sigma^2 \lambda'_x \eta \mathbb{E}_t \left[ \sum_{j=1}^{n-1} \bar{e}_{t+j} \left( \sum_{j=1}^{n-1} \bar{e}'_{t+j} \eta' \left( I - c_1^{n-j} \right)' \right) \right] \lambda_x \\
&\quad + \sigma^2 \lambda'_x \eta \mathbb{E}_t \left[ \sum_{j=1}^{n} \bar{e}_{t+j} \sum_{j=1}^{n} \bar{e}'_{t+j} \right] \eta' \lambda_x
\end{align*}
\]

where

\[
\mathbb{E}_t \left[ \left( \sum_{j=1}^{n-1} (I - c_1^{n-j}) \eta \bar{e}_{t+j} \right) \left( \sum_{j=1}^{n-1} \bar{e}'_{t+j} \eta' \left( I - c_1^{n-j} \right)' \right) \right] = \sum_{i=1}^{n-1} (I - c_1^{i}) \eta \eta' (I - c_1^{i})'
\]

and

\[
\mathbb{E}_t \left[ \left( \sum_{j=1}^{n-1} (I - c_1^{n-j}) \eta \bar{e}_{t+j} \right) \sum_{j=1}^{n} \bar{e}'_{t+j} \right] = (n - 1) \eta - (I - c_1)^{-1} (c_1 - c_1^n) \eta
\]

and

\[
\mathbb{E}_t \left[ \sum_{j=1}^{n} \bar{e}_{t+j} \left( \sum_{j=1}^{n-1} \bar{e}'_{t+j} \eta' \left( I - c_1^{n-j} \right)' \right) \right] = (n - 1) \eta' - \eta' (c_1' - (c_1^n)') (I - c_1)^{-1}
\]

and

\[
\mathbb{E}_t \left[ \sum_{j=1}^{n} \bar{e}_{t+j} \sum_{j=1}^{n} \bar{e}'_{t+j} \right] = n I
\]

It follows that

\[
\begin{align*}
&\mathbb{E}_t \left[ (\lambda_{t+n} - \lambda_t) (\lambda'_{t+n} - \lambda'_t) \right] - \mathbb{E}_t \left[ (\lambda_{t+n} - \lambda_t) \right] \mathbb{E}_t \left[ (\lambda'_{t+n} - \lambda'_t) \right] \\
&= \sigma^2 \lambda'_x \left( \sum_{i=1}^{n-1} (I - c_1^i) \eta \eta' (I - c_1^i)' \right) \lambda_x + n \sigma^2 \lambda'_x \eta \eta' \lambda_x \\
&\quad - (n - 1) \sigma^2 \lambda'_x \eta \lambda_x + \sigma^2 \lambda'_x (I - c_1)^{-1} (c_1 - c_1^n) \eta \eta' \lambda_x \\
&\quad - (n - 1) \sigma^2 \lambda'_x \eta \lambda_x + \sigma^2 \lambda'_x \eta \eta' (c_1' - (c_1^n)') (I - c_1)^{-1} \lambda_x
\end{align*}
\]
and

$$\text{Var}_t [\Delta \lambda_{t+1}] = \sigma^2 \lambda'_x \eta x' \lambda_x$$

Finally

$$\frac{1}{n} \left( E_t \left[ (\lambda_{t+n} - \lambda_t) (\lambda'_{t+n} - \lambda'_t) \right] - E_t [\lambda_{t+n} - \lambda_t] E_t [\lambda'_{t+n} - \lambda'_t] \right) - E_t \left[ (\lambda_{t+1} - \lambda_t) (\lambda'_{t+1} - \lambda'_t) \right] + E_t [\lambda_{t+1} - \lambda_t] E_t [\lambda'_{t+1} - \lambda'_t] = \frac{1}{n} \sigma^2 \lambda'_x \left( (I - c_1)^{-1} (c_1 - c_1^0) \eta y' + \eta y' (c_1 - (c_1^0))' (I - c_1')^{-1} \right) \lambda_x$$

$$- \frac{1}{n} \sigma^2 \lambda'_x \left( 2 (n - 1) \eta y' - \sum_{i=1}^{n-1} (I - c_i) \eta y' (I - c_i)' \right) \lambda_x$$

and

$$ytm_n - r = - \frac{1}{2 n} \sigma^2 \lambda'_x \left( (I - c_1)^{-1} (c_1 - c_1^0) \eta y' + \eta y' (c_1 - (c_1^0))' (I - c_1')^{-1} \right) \lambda_x$$

$$- \frac{1}{2 n} \sigma^2 \lambda'_x \left( -2 (n - 1) \eta y' + \sum_{i=1}^{n-1} (I - c_i) \eta y' (I - c_i)' \right) \lambda_x$$

In the scalar case

$$\sum_{i=1}^{n-1} (1 - \rho^i)^2 = n - 1 + \sum_{i=11}^{n-1} \rho^i - 2 \sum_{i=1}^{n-1} \rho^i = n - 1 + \frac{\rho^2 - \rho^{2n}}{1 - \rho^2} - 2 \frac{\rho - \rho^n}{1 - \rho}$$

and

$$\frac{1}{n} \left( E_t \left[ (\lambda_{t+n} - \lambda_t) (\lambda'_{t+n} - \lambda'_t) \right] - E_t [\lambda_{t+n} - \lambda_t] E_t [\lambda'_{t+n} - \lambda'_t] \right) - E_t \left[ (\lambda_{t+1} - \lambda_t) (\lambda'_{t+1} - \lambda'_t) \right] + E_t [\lambda_{t+1} - \lambda_t] E_t [\lambda'_{t+1} - \lambda'_t] = - \left( 1 - \frac{11 - \rho^{2n}}{n \left( 1 - \rho^2 \right)} \right) \sigma^2 \lambda_x^2$$

so that the real spread is

$$ytm_n - r = \lambda_x^2 \left( 1 - \frac{11 - \rho^{2n}}{n \left( 1 - \rho^2 \right)} \right) \frac{\sigma^2}{2}$$

or, without habit persistence

$$ytm_n - r = \gamma^2 y_x^2 \left( 1 - \frac{11 - \rho^{2n}}{n \left( 1 - \rho^2 \right)} \right) \frac{\sigma^2}{2}$$

Since $\frac{11 - \rho^{2n}}{n \left( 1 - \rho^2 \right)} < 1$ for any $-1 < \rho < 1$, the slope is positive for any $\rho$. 

48
A.8 Unconditional serial correlations

Note that the first order autocovariances for bonds are

\[
E \left[ \tilde{b}_{t+1,n} \tilde{b}'_{t,n} \right] = E \left[ \tilde{b}'_{n,x} \tilde{x}'_{t+1} \tilde{x}'_{t} \tilde{b}_{n,x} \right] \\
= \tilde{b}'_{n,x} E \left[ (c_1 \tilde{x}_t + \eta \sigma \tilde{\epsilon}_{t+1}) \tilde{x}'_{t} \right] \tilde{b}_{n,x} \\
= \tilde{b}'_{n,x} c_1 E \left[ \tilde{x}' \tilde{x}' \right] \tilde{b}_{n,x}
\]

where the unconditional variance covariance matrix of \( \tilde{x}_t \) is –see Hamilton (1994, p.265) – the solution of

\[
E \left[ \tilde{x}' \tilde{x}' \right] = E \left[ \tilde{x}' c_1 \tilde{x} + \sigma^2 E [\tilde{\epsilon}' \tilde{\epsilon}] \right] \\
= (\text{vec} [E \tilde{x} \tilde{x}'])' \text{vec} [c_1] + \sigma^2 (\text{vec} [I]) \text{vec} [\tilde{\epsilon} \tilde{\epsilon}]
\]

namely

\[
\text{vec} [E \tilde{x} \tilde{x}'] = \sigma^2 (I_{n^2} - c_1 \otimes c_1)^{-1} \text{vec} [\tilde{\epsilon} \tilde{\epsilon}]
\]

Similarly, for the rate of growth of output

\[
\Delta \tilde{y}_{t+1} = y_x' (\tilde{x}_{t+1} - \tilde{x}_t) + \frac{1}{2} \tilde{x}'_{t+1} y_{xx} \tilde{x}_{t+1} - \frac{1}{2} \tilde{x}'_t y_{xx} \tilde{x}_t
\]

we obtain

\[
E \left[ \Delta \tilde{y}_t \Delta \tilde{y}'_{t+1} \right] = E \left[ y_x' (\tilde{x}_{t+1} - \tilde{x}_t) (\tilde{x}_{t+1} - \tilde{x}_t)' y_x \right] \\
= y_x' \left( 2E \left[ \tilde{x}'_{t+1} \right] - E \left[ \tilde{x}'_{t+2} \right] - E \left[ \tilde{x}' \right] \right) y_x
\]

and

\[
E \left[ \Delta \tilde{y}_{t+1} \Delta \tilde{y}'_{t+1} \right] = y_x' E \left[ (\tilde{x}_{t+1} - \tilde{x}_t) (\tilde{x}_{t+1} - \tilde{x}_t)' \right] y_x \\
= 2y_x' \left( E \left[ \tilde{x}' \right] - E \left[ \tilde{x}'_{t+1} \right] \right) y_x \\
= E \left[ \Delta \tilde{y}_t \Delta \tilde{y}'_t \right]
\]

Hence

\[
\rho \left[ \Delta \tilde{y} \Delta \tilde{y}'_{t+1} \right] = \frac{y_x' \left( 2E \left[ \tilde{x}'_{t+1} \right] - E \left[ \tilde{x}'_{t+2} \right] - E \left[ \tilde{x}' \right] \right) y_x}{2y_x' \left( E \left[ \tilde{x}' \right] - E \left[ \tilde{x}'_{t+1} \right] \right) y_x}
\]

In the scalar case, this becomes

\[
\rho \left[ \Delta \tilde{y} \Delta \tilde{y}'_{t+1} \right] = \frac{2E \tilde{x}'_{t+1} - E \tilde{x}'_{t+2} - E \tilde{x}'}{2E \tilde{x}' - 2E \tilde{x}'_{t+1}}
\]
and since

\[
\begin{align*}
\mathbb{E}xx &= \frac{1}{1 - \rho^2}\sigma^2 \\
\mathbb{E}xx_{+1} &= \frac{\rho}{1 - \rho^2}\sigma^2 \\
\mathbb{E}xx_{+2} &= \frac{\rho^2}{1 - \rho^2}\sigma^2
\end{align*}
\]

it follows that

\[
\rho [\Delta \tilde{y}\Delta \tilde{y}'_{+1}] = -\frac{1}{2} (1 - \rho)
\]

A.8.1 Moments of an unrestricted ARMA(1,1)

Consider an example based on two states, one exogenous and one endogenous. If we assume that the endogenous state is lagged consumption, then consumption will follow an ARMA(1,1) process \( \tilde{y}_t = \phi \tilde{y}_{t-1} + x_t \), where \( x_t = \theta \tilde{x}_{t-1} + \sigma \varepsilon_t \) and \( \varepsilon_t \) a standard normal process. The autocovariances of such process are well known (see e.g. Hamilton, 1994). It follows that the first order autocovariance of the rate of growth \( \Delta \tilde{y} \) is

\[
E[\Delta \tilde{y}_{+1}\Delta \tilde{y}] = \frac{\sigma^2 (\theta + \phi)}{1 - \phi^2} ((1 - \phi)(1 + \theta \phi) - \theta) + \frac{\sigma^2 (1 + \theta \phi)}{1 - \phi^2} (\theta + \phi - 1)
\]

\[
= \frac{\sigma^2}{1 - \phi^2} ((\theta + \phi)((1 + \theta \phi)(1 - \phi) - \theta) + (\theta + \phi - 1)(1 + \theta \phi))
\]

\[
= \frac{\sigma^2}{1 - \phi^2} ((\theta + \phi)(1 + (1 - \theta)(1 - \phi))(1 - \phi) - (1 + \theta \phi))
\]

\[
= \frac{\sigma^2}{1 - \phi^2} (-\phi (1 - \phi) \theta^2 + (2 - (1 - \phi) \phi) \theta - (1 - \phi))(1 - \phi)
\]

\[
= \frac{\sigma^2}{1 - \phi^2} (2\theta - (1 - \phi)(1 + \theta^2 + \phi \theta))
\]

Now consider the real yield curve. In the two-period case, the slope of the yield curve is, for \( n = 2 \),

\[
y_{tm} = y'_{2} - r
\]

\[
= -\frac{1}{4}\sigma^2 y' \left( (I - c_1)^{-1} (c_1 - c_1^2) \eta \eta' + \eta (c_1 - (c_1^2))' (I - c_1')^{-1} - 2\eta \eta' + (I - c_1) \eta \eta' (I - c_1)' \right) y_x
\]

Note that we are now assuming the simple model \( \tilde{y}_t = \phi \tilde{y}_{t-1} + x_t \), where \( x_t = \theta \tilde{x}_{t-1} + \sigma \varepsilon_t \), so that

\[
\begin{align*}
y_{t+1} &= \begin{bmatrix} \phi & 1 \end{bmatrix} \begin{bmatrix} y_t \\ x_{t+1} \end{bmatrix} \\
\begin{bmatrix} y_t \\ x_{t+1} \end{bmatrix} &= \begin{bmatrix} \phi & 1 \\ 0 & \theta \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \sigma \end{bmatrix} \begin{bmatrix} 0 \\ \varepsilon_{t+1} \end{bmatrix}
\end{align*}
\]
and
\[
\begin{align*}
c_1 &= \begin{bmatrix} \phi & 1 \\ 0 & \theta \end{bmatrix} \\
y_x &= \begin{bmatrix} \phi & 1 \end{bmatrix}
\end{align*}
\]

It follows that
\[
(I - c_1) \eta '\ (I - c_1)' = \begin{bmatrix} 1 - \phi & 1 \\ 0 & 1 - \theta \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 - \phi & 0 \\ 1 & 1 - \theta \end{bmatrix} = \begin{bmatrix} 1 & 1 - \theta \\ 1 - \theta & (1 - \theta)^2 \end{bmatrix}
\]
\[
2\eta' = 2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}
\]
\[
(I - c_1)^{-1} (c_1 - c_1^2) \eta' = \begin{bmatrix} \frac{1}{1 - \phi} & \frac{1}{\phi + \theta - 1} \\ 0 & \frac{1 - \theta}{1 + \phi - \theta - \phi \theta} \end{bmatrix} \begin{bmatrix} \phi & 1 \\ 0 & \theta \end{bmatrix} \begin{bmatrix} 1 - \phi & 1 \\ 0 & 1 - \theta \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & \phi + \frac{1 - \theta}{1 + \phi - \theta - \phi \theta} \\ 0 & \frac{1}{1 - \phi} + \frac{(1 - \theta)\theta}{\phi + \theta - 1} \end{bmatrix}
\]

and
\[
ytm_2 - r = -\frac{1}{4} \sigma^2 y_x \left( (I - c_1)^{-1} (c_1 - c_1^2) \eta' + \eta' (c_1 - (c_1^2))' (I - c_1')^{-1} - 2\eta' + (I - c_1) \eta' (I - c_1)' \right) y_x
\]
\[
= -\frac{1}{4} \sigma^2 y_x \left[ \frac{1}{2 + \theta - \theta - (1 - \theta)^2} \\ \frac{1 - \theta}{1 - \phi} - \theta \right] y_x
\]
\[
= -\frac{1}{4} \sigma^2 \left( \theta^2 + \phi^2 - 1 + 2\phi \left( \frac{1 - \theta}{1 - \phi} - \theta \right) \right)
\]
\[
= \frac{1}{4} \sigma^2 \frac{\phi^3 - (1 + 2\theta) \phi^2 - (5 - 6\theta - \theta^2) \phi + (1 - \theta^2)}{1 - \phi}
\]

which is obviously positive when \( \phi^3 - (1 + 2\theta) \phi^2 - (5 - 6\theta - \theta^2) \phi + (1 - \theta^2) > 0. \)
References


[12] Constantinides, 1990,


[23] Fuhrer, 2000,


