Limited Asset Market Participation and the Variation of Real Risk Premia

Francisco Rivadeneyra*
Bank of Canada
February 15, 2010

Abstract
This paper studies the contribution of limited asset market participation to the variability of real interest rates in the UK. Using a quadratic term structure model of real interest rates in an economy with endogenously segmented asset markets, I estimate the contribution of money and real shocks to the volatility of interest rates. The estimates suggest that the consumption of households participating in the asset market has time-varying conditional variances and varies more in response to money shocks when inflation is close to its unconditional mean. Asset market segmentation is an uninsurable risk which varies more at lower levels of inflation because more households are ready to participate at low levels than at high levels of inflation. Consequently, participating households bear a larger portion of aggregate consumption variability. The endogenous asset market segmentation generates the rejection of the expectations hypothesis in real interest rates.

JEL classification numbers: E43, E44, G12.

Keywords: general equilibrium asset pricing, limited asset market participation, risk premia variation, expectations hypothesis.

1 Introduction

There is a growing consensus in the literature that varying risk premia is a feature of asset price dynamics. For instance, in US nominal bond yields, Fama and Bliss (1987) document the presence

*frivadeneyra@bankofcanada.ca. I thank Fernando Alvarez, Lars Hansen, Monika Piazzesi and Harald Uhlig for valuable comments. Thanks to the participants of the Economic Dynamics Working Group at the University of Chicago and specially to Nina Boyarchenko, Santiago Garcia-Verdu, Tiago Pinheiro and Francisco Vazquez-Grande. Thanks to participants at seminars of the Bank of Canada, Bank of England, Federal Reserve Bank Boston, BBVA and Goldman Sachs. All errors remain mine.
of time-varying risk premia and Cochrane and Piazzesi (2005) show evidence of excess return predictability. However, there is still some debate on the sources of such variability. Some studies argue for long-run risk (Bansal and Yaron, 2004), habit formation (Campbell and Cochrane, 1999), time-varying disaster risk (Gabaix, 2009), or limited participation in asset markets (Guvenen, 2009). This paper argues for limited participation in asset markets as the main source of time-varying risk. In particular it investigates the effects of money and endowment shocks on the term structure of real interest rates using a general equilibrium model when participation in the asset markets is endogenously limited. I focus on the variation over time of real risk premia. First, I examine the UK real yields from 1985 to 2007 and show the presence of time-varying real risk premia. To explain this variation I propose an extension to the general equilibrium model of Alvarez, Atkeson, and Kehoe (2008). In this model the non-neutrality of money originates from the fixed transaction costs that limit financial participation. The first finding is that money shocks and asset market frictions are necessary to explain the variation over time of real risk premia. The second finding is that risk premia is more responsive to money shocks than to endowment shocks when inflation is close to its unconditional mean.

This paper contributes to the literature in two ways. First it uses a general equilibrium model to understand the variation of risk premia in the term structure of real interest rates. To account for the time-varying volatility of asset prices in a consumption-based model, the pricing kernel has to be highly volatile and display protracted responses to small and homoskedastic shocks to consumption. The endogeneity of the asset market segmentation in the model produces a responsive and heteroskedastic pricing kernel. The second complementary contribution is to provide microfoundations to the quadratic term structure model (QTSM). Quadratic term structure models are rarely motivated by general equilibrium asset pricing models. I provide economic intuition to the dynamics of the term structure of real interest rates using a general equilibrium model of endogenously segmented asset markets. The model of segmented asset markets induces a liquidity effect of money: real interest rates fall temporarily in response to money injections. Short-term real interest rates fall because in the model only a fraction of households are active in the asset market when the money injection occurs therefore reducing their marginal utility of consumption. By no arbitrage, movements in the short end of the curve caused by money shocks are transmitted to the long end.

In the model of segmented asset markets used in this paper, transfers of money between the asset market (where the government carries out open market operations) and the goods market require incurring a real fixed cost, which is heterogeneous across households. At any given moment only

---

1For example Constantinides (1992) stresses that his approach is to explore directly the time-series properties of the quadratic pricing kernel without assuming a representative consumer real economy. Ahn, Dittmar, and Gallant (2002) follow a similar approach. Longstaff (1989) is an exception.
a fraction of households will find it optimal to pay the transaction cost to participate in the asset market and buy assets to avoid the inflationary tax. The marginal household is the one that equates its individual costs with the benefits of participating in the asset market. These benefits vary over time depending on the prevailing level of inflation. The key feature of the model is that changes in money supply have non-linear effects on the marginal utility of active households; marginal utility is more sensitive at low levels of inflation than at large ones. At low levels of inflation money shocks induce larger changes in the benefits of participation because the marginal household has relatively low transaction costs. On the other hand, when inflation is high, the marginal household has larger transaction costs and money shocks of identical size provide smaller changes in the net benefit to participate.

To capture the non-linear effects of money I take the second order approximation of the marginal utility of active households. The coefficients of the second order approximation provide an economic interpretation of the mechanics that drive the dynamics of risk premia. To a first degree, changes in the inflation rate generate direct changes in the consumption of active households. To a second degree, active households’ consumption varies from changes in the endogenous degree of asset market segmentation. These changes drive the variation of risk premia with inflation. When inflation is high, a large fraction of households participate and the risk of asset market segmentation falls.

To compute asset prices I use the ratio of the approximated marginal utility to construct the stochastic discount factor. In this setting, prices of zero coupon bonds are exponential quadratic functions of the money growth rate. This pricing model fits in the general class of quadratic term structure models (QTSM) characterized by Ahn, Dittmar, and Gallant (2002). Instead of estimating the parameters of the model using an observed measure of money growth, I approach the estimation as a filtering problem and invert for the hidden monetary state. Therefore, I use the Extended Kalman Filter (EKF) that allows for estimating the unobserved state in a non-linear environment. The benefits of this strategy are to avoid estimating the QTSM based on arbitrary moment conditions of the data (Longstaff, 1989, Constantinides, 1992, and Leippold and Wu, 2003), as well as to derive implications for consumption from the filtered monetary state. I also include in the estimation a linear observed factor to account for the level of yields in the data. The variable used in the estimation is the GDP growth splined into monthly frequency. In the second order approximation of the marginal utility of active households, the coefficient multiplying GDP growth is the coefficient of relative risk aversion.

The estimates of the parameters of term structure model and the estimated unobserved state imply that the standard deviation of the log consumption of active households is less than 1% in annualized terms. I also compute the volatility of the implied degree of households’ asset market participation in the model. I find that participation varies less than 0.5% annually but decreased during the sample.
The consumption result is consistent with the empirical observation that consumption growth is highly persistent. The asset market participation result is consistent with the finding that the opportunity costs of not participating in asset markets have fallen during the sample time period as macroeconomic volatility, of inflation, or output growth, has as fallen as well (Gali and Gambetti, 2009). Since the model of segmented asset markets provides a precise interpretation of unobserved state, I can decompose the importance of macroeconomic shocks to consumption over time. The unobserved state is labeled in the model as the money growth rate. In the early part of the sample, the growth rate of the unobserved factor is large. In this part of the sample, most of the consumption volatility is due to output growth shocks because when money growth is high, consumption responds less to innovations of money growth. As the implied money growth rate falls towards the end of the sample, consumption volatility responds less to output growth innovations and more to money growth innovations.

After estimating the coefficients of the second order approximation of the marginal utility and the coefficients of the dynamics of the state, I generate samples of artificial yield data. With these data I run the forecasting regressions of Campbell and Shiller (1991) on each sample and compute the histograms of the slope coefficient. The distributions of these coefficients fall around the mean of the coefficients estimated in the UK real yield data. However, the dispersion of the coefficients of the artificial data is larger than the standard deviations of the coefficients estimated in the UK yield data. These results suggest that the estimated term structure model generates the variation in risk premia observed in the data.

This paper aims to bridge the gap between two strands of literature: the reduced form empirical asset pricing and the general equilibrium asset pricing. With respect to the empirical asset pricing literature, this paper is related to Boivin, Dong, and Ang (2009). They estimate a quadratic term structure model for US postwar data with latent and observed macro factors to quantify the effects of monetary policy shifts on the risk premia of long-term nominal bonds. Another recent reduced form term structure model is Campbell, Sunderam, and Viceira (2009). This last paper estimates a quadratic term structure model in which the real and the nominal part are separately modeled. However, neither paper has a general equilibrium macroeconomic model behind the derivation of the term structure. Perhaps closer to the spirit of the present paper is the work of Buraschi and Jiltsov (2005). They derive the asset pricing implications of a general equilibrium model in which money enters in the utility function and estimate the structural parameters of the model. With respect to the literature of asset pricing in general equilibrium models, this paper is closely related to Alvarez, Atkeson, and Kehoe (2002). With a log linear approximation, the authors generate persistence in interest rates and a liquidity effect of money shocks. The shortcoming in that case is that the log linear approximation by definition verifies the expectations hypothesis with a constant risk premia. In this paper, I extend this exercise to a second order approximation to allow for a time-varying risk premia.
premia while maintaining the simplicity of the model on segmented asset markets and add a second source of uncertainty from endowment shocks. This paper is also related to the literature of limited participation in asset markets. Guvenen (2009), Malloy, Moskowitz, and Vissing-Jorgensen (2008) and Vissing-Jorgensen (2002) have suggested that asset market segmentation and transaction costs can help explain empirical asset pricing regularities like the equity premium puzzle. Related to work on real interest rates, this paper is rather novel because, to my knowledge, there is little work on estimating structural parameters of a macro model using the cross-equation restrictions placed by the term structure of real interest rates.

The paper is divided into 7 sections. Section 2 presents the evidence of time-varying real risk premia for the UK yields. Section 3 describes the model of endogenously segmented asset markets. Section 4 derives the term structure of real interest rates from the model of segmented asset markets. Section 5 presents the estimation of the term structure model using the Extended Kalman Filter. Section 6 shows the macroeconomic and asset pricing implications of the estimation. Section 7 concludes. The appendices cover the estimation procedure of the yields model and the derivation of asset prices in the macro model.

2 Empirical Evidence of Real Risk Premia

First I review the Bank of England’s methodology used to derive real yields from the index linked bond data (Anderson and Sleath, 2001) and show some descriptive statistics. The principal components analysis of Litterman and Sheinkman (1991) (not shown), unlike the decomposition of US nominal data, support that two factors explain most of the variance in UK real yields. Second, I show the forecasting regressions used to test for time-varying risk premia.

2.1 Data

I use the publicly available zero-coupon real bond yield data from the Bank of England from 1985.01 to 2007.12. These real yields are derived from the UK Index Linked (IL) bonds first issued in 1981. This market rapidly became highly liquid and grew to be a sizeable portion of the value of government securities. Between 1981 and 1992, 14 maturities were issued up to 30 years and by that time constituted a fifth of the market value of all government debt (Robertson and Symons, 1997). IL bonds are the closest approximation to real bonds available. Bond payments are linked to the Retail Price Index (RPI) but have an 8 month lag (in 2005 the lag was reduced to 3 months). The indexation lag potentially carries an inflation risk because during the last 8 months the IL bond becomes purely nominal. However, Evans (1998) estimated that the inflation premium from indexation lag contributed 1.5 basis points to the annualized yield across maturities. Thus the
methodology of the Bank of England to derive real yields from IL bonds assumes that inflation risk premia from the indexation lag is zero. This is described by Anderson and Sleath (2001). The real yield curve is computed in the following way. First, they obtain the zero-coupon yield curves for nominal and IL bonds using the variable roughness penalty model (Waggoner, 1997). Second, they subtract the proportional nominal yield to the IL yield to account for the indexation lag.

I use end-of-month observations to convert from daily to monthly frequency. The shortest available maturity is 2 years. The reason is twofold. First, the methodology to compute real yields discards IL bonds with maturities shorter than 16 months; second, it further restricts maturities if their estimated values are too unstable to small perturbations to the underlying IL bond data. Moreover, maturities are provided at monthly intervals up to 5 years, and at 6-month intervals up to 25-years. In total there are 76 constant-maturity series and 252 months with some missing observations in the long and short ends of the curve. For purposes of the estimation of the yields model, I will reduce the set of maturities used to estimate the model. Arguably, only a limited number of maturities contribute additional information to the estimation procedure because the procedure to derive the real yield data from the original IL data introduces smoothing and measurement error.

To get a feel of the data Figure (1) plots the time series of selected maturities of the real yields data. Over this period the UK, as the US, observed a period of disinflation and reduction in the macroeconomic volatility (Rudebusch, Sack, and Swanson, 2006). The level of real yields for the long and short maturities falls from approximately 3.5 to 1.5%. Its easy to see that the slope of the yield curve is most of the time upward sloping or flat but downward sloping is not uncommon and is observed at the beginning of the sample and around 2000. In this sample, the mean yield curve is 64% of the time upward sloping and the rest downward sloping. Figure (2) shows the unconditional mean with the standard deviation bands. The flat mean yield curve is robust to the choice of maturities and sample period but, as in the literature on nominal yields, the standard deviation is large. Finally, Figure (3) plots the unconditional volatility of real yields across the term structure computed as the standard deviation of yield changes. Consistent with the findings in the nominal yields literature, the short end displays much larger volatility than the middle and long end of the curve (Piazzesi, 2003).

It is also interesting to look at the autocorrelation structure of real yields. Table (1) summarizes the 1- to 24-month autocorrelations for several maturities. Real yields of any maturity are highly autocorrelated at one month; autocorrelations of larger order for short maturities drop almost to

---

2This zero-coupon curves are piecewise cubic polynomials joined so that the curve itself and its first derivative are continuous at every point. The number of joint points (maturities) to use is a choice of the econometrician but generally results in a large number of parameters compared to the popular 6-parameter cubic spline of Svensson (1995). The objective is to minimize the residual sum of squares plus a roughness measure which penalizes too much curvature around joint points measured by the square of the second derivative of the curve. For details see Anderson and Sleath (2001).
Figure 1: Time series for selected UK real yields derived from the Index Linked bonds. Several features of the time series are interesting. First, the level of yields falls towards the end of the sample. Second, the slope of the yield curve is upward sloping 64% of the time in this sample.

Figure 2: Unconditional mean of real yields with two standard deviation bands (1985.01 to 2004.12). The flat mean yield curve is robust to the choice of maturities and sample period. Maturities longer than 20 years are left out since they appear sparsely in the sample. Between the 2- and 3-year maturities, there are some missing observations. On this end, bonds are observed at monthly intervals up to the 5-year maturity.
Figure 3: Volatility curve of real yields in the UK computed as the standard deviation of yield changes (1985.01 to 2004.12). Missing observations have some effect on the computed values. A fully observed sub-sample slightly reduces the volatility near the short end.

zero while the longer yields maintain a high level of time dependence. For example, the one-month autocorrelation of the 2- and the 15-year real yields are 0.928 and 0.985 respectively. On the other hand, the 24-month autocorrelation for the same maturities is 0.026 and 0.826 respectively.

2.2 Forecasting Regressions

Real risk premia is the compensation that investors demand for bearing real risk. To test for the existence of time-varying real risk premia I run the forecasting regressions of Barr and Campbell (1997) which test the log expectations hypothesis (EH). There are several versions of the EH and the results can be interpreted in several ways depending on the estimation performed. I focus on the expectations hypothesis (in logs) which states that expected log holding period returns are equal to the riskless return over the holding period (Piazzesi, 2003). Rejecting the EH implies that risk premia are time-varying or that excess returns are predictable. Since maturity of bonds and length of holding period can vary widely, empirical results testing the EH vary as well. However there are two empirical regularities regarding the EH. First, several studies for the US postwar nominal data are consistent in rejecting the EH for holding periods of one year and maturity of bonds of up to 5

---

3See Cox, Ingersoll and Ross (1981) and Campbell (1986) for a discussion on the different versions and their implications.
Table 1: Descriptive Statistics of Real Yields in the UK, 1985.01-2007.12

<table>
<thead>
<tr>
<th></th>
<th>2y</th>
<th>3y</th>
<th>5y</th>
<th>10y</th>
<th>15y</th>
<th>20y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.84</td>
<td>2.86</td>
<td>2.92</td>
<td>3.00</td>
<td>3.02</td>
<td>2.72</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.931</td>
<td>0.808</td>
<td>0.788</td>
<td>0.918</td>
<td>1.009</td>
<td>1.031</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.16</td>
<td>2.22</td>
<td>2.01</td>
<td>1.68</td>
<td>1.66</td>
<td>1.58</td>
</tr>
</tbody>
</table>

Autocorrelations

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>6</th>
<th>12</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.928</td>
<td>0.938</td>
<td>0.961</td>
<td>0.982</td>
<td>0.988</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td>0.867</td>
<td>0.880</td>
<td>0.922</td>
<td>0.963</td>
<td>0.976</td>
<td>0.970</td>
</tr>
<tr>
<td></td>
<td>0.801</td>
<td>0.839</td>
<td>0.895</td>
<td>0.950</td>
<td>0.968</td>
<td>0.957</td>
</tr>
<tr>
<td></td>
<td>0.701</td>
<td>0.699</td>
<td>0.812</td>
<td>0.908</td>
<td>0.940</td>
<td>0.907</td>
</tr>
<tr>
<td></td>
<td>0.562</td>
<td>0.598</td>
<td>0.730</td>
<td>0.854</td>
<td>0.900</td>
<td>0.851</td>
</tr>
<tr>
<td></td>
<td>0.026</td>
<td>0.260</td>
<td>0.564</td>
<td>0.765</td>
<td>0.826</td>
<td>0.709</td>
</tr>
</tbody>
</table>

years (Fama and Bliss, 1987, Campbell and Shiller, 1991, and Cochrane and Piazzesi, 2005). Second, the forecasting power increases with the maturity of bonds. In contrast, for longer holding periods (2- and 5-years) the EH tends to hold in the data.

Regarding the evidence of the EH in real yields there are mixed results. This arises in part because data on real yields is not easy to come by and usually they are derived indirectly from bonds linked to a price level index. Two studies use the real yields derived from UK Index Linked bonds. Barr and Campbell (1997) find no evidence against the EH using a sample from 1985 to 1995. On the contrary, Evans (1998) finds strong statistical evidence against it with a slightly longer sample. The difference between these two studies is their treatment of risk premia when deriving real rates from the index linked yields.

Specifically Barr and Campbell (1997) estimate the regression:

\[ y_{t+j}^{(\tau-j)} - y_t^{(\tau)} = \alpha(\tau, j) + \beta(\tau, j) \cdot \frac{j}{(\tau-j)} \left[ y_t^{(\tau)} - y_t^{(j)} \right] + \varepsilon_{t+j} \tag{1} \]

where

\[ y_t^{(\tau)} = -\log P_t^{(\tau)}/\tau \]

and \( P_t^{(\tau)} \) is the price at time \( t \) of a zero-coupon bond maturing at \( \tau \) months in the future and \( j \) is the length of the holding period also measured in months.\(^4\)

\(^4\)This regression can be derived for \( \beta = 1 \) from the equation:

\[ E_t \log P_{t+j}^{(\tau-j)} - \log P_t^{(\tau)} = \log P_t^{(j)}/j + \alpha(\tau, j), \]

which states that excess log returns are non-zero but constant over time.
The data on real yields from the Bank of England (derived from the UK Index Linked yields) uses a flexible specification to adjust for the nominal risk of the last months of the bonds (Evans, 1998). Unfortunately the resulting dataset restricts the shortest maturity to 25 months. This fact determines the shortest length of the regressions’ holding-period. Moreover, after 60 months bond maturities appear in intervals of 6 months, for example 60, 66, 72 months and so forth; therefore to perform the regressions I use yield changes over two years which equal a holding period of 24 months. This approximation should induce negligible measurement error into the regression.

The results of the regressions are shown in Table 2 and the slope coefficients plotted in Figure 4. The results suggest that this version of the EH is rejected for holding periods of 25 months with coefficients ranging from 0.4 to -1 the longer the maturity of the bonds. The coefficients are statistically different from 1 (with Newey-West standard errors that correct for overlapping forecasting errors and heteroskedasticity). For the 36-month holding period regression the coefficients are closer to zero for all maturities. This is in accordance to the evidence found in the US nominal data (Campbell and Shiller, 1991, pg. 502). The coefficients can be seen graphically in Figure 4 which plots the coefficients $\beta(\tau,j)$, with $\tau$ in the $x$-axis and $j = 25$ in the left hand side panel and $j = 36$ in the right hand side panel. This shape shows that for longer maturity bonds the forecasting power is stronger as argued by Fama and Bliss (1987) and others. At the same time, the evidence for the US points that longer holding periods tend to verify the EH. For real yields this is the case too: as the coefficients in the 36-month holding period regression are smaller than the 25-month regression for all long bonds. The coefficients of the regressions performed on 48-month holding periods (not shown) are all indistinguishable from 0 or 1, therefore verifying this claim. The percentage of variation explained by the regressions is small yet in line with the numbers found in other studies with similar data. Interestingly, the level coefficients of the regressions imply that log returns of holding real bonds over these particular period have been negative on average. The 24-month log holding period on a 10-year bond averaged $-20$ basis points per annum.

As a comparison, Barr and Campbell (1997) estimate equation (1) for $j = 3, 12$ and $\tau = 12, 60$; that is changes in yields over holding periods of 3 and 12 months onto the yield spreads of 1- and 5-year bonds in their sample from 1985 to 1994. The coefficients of the 3-month holding period onto the 1- and 5-year spreads are 0.66 and 0.06 respectively. The coefficient of the 12-month holding period onto the 5-year spread is 0.25. All the estimates are poorly determined and there is no evidence to accept or reject either one nor zero. They conclude there is no evidence against the expectations hypothesis in real rates for the UK. A possible problem is their treatment of the indexation lag to extract real yields from the IL data. They assume the log pure expectations hypothesis on the index linked bonds to extract the underlying real yields. Evans (1998) allowed more flexible inflation risk premia in IL bonds. He rejects the expectations hypothesis for real yields implying the presence of
Figure 4: Slope coefficient of the Campbell and Shiller regressions: \( y_t^{(\tau-j)} - y_t^{(\tau)} = \alpha + \beta(\tau,j) \cdot \left( \frac{j}{\tau-j} \right) \left[ y_t^{(\tau)} - y_t^{(j)} \right] + \epsilon_{t+j}^{(\tau)} \) for the period 1985.01-2007.12. \( \tau \) is the maturity of the bond in the \( x \)-axis. \( j \) is the length of the holding-period. For panel (a) \( j = 25 \) and panel (b) \( j = 36 \). Both graphs plot the point estimate and its 95\% confidence interval using Newey-West standard errors.

Table 2: Forecasting Regressions for the UK real yield data using log returns from 1985.01-2007.12. \( \tau \) is the maturity in months of the bond, \( j \) is the length in months of the holding period. Below the coefficients are Newey-West standard errors.

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>( j = 25 )</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( R^2 )</th>
<th>Obs.</th>
<th>( j = 36 )</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( R^2 )</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>-0.3867</td>
<td>0.4775</td>
<td>0.063</td>
<td>121</td>
<td>72</td>
<td>-0.4355</td>
<td>0.2858</td>
<td>0.112</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1945)</td>
<td>(0.4335)</td>
<td></td>
<td></td>
<td></td>
<td>(0.1147)</td>
<td>(0.2154)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>-0.3604</td>
<td>0.2096</td>
<td>0.012</td>
<td>152</td>
<td>120</td>
<td>-0.3622</td>
<td>-0.0010</td>
<td>0.010</td>
<td>167</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1641)</td>
<td>(0.3969)</td>
<td></td>
<td></td>
<td></td>
<td>(0.0719)</td>
<td>(0.4283)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>-0.1912</td>
<td>-0.5876</td>
<td>0.058</td>
<td>168</td>
<td>180</td>
<td>-0.3492</td>
<td>-0.1643</td>
<td>0.011</td>
<td>167</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0677)</td>
<td>(0.4645)</td>
<td></td>
<td></td>
<td></td>
<td>(0.0945)</td>
<td>(0.6945)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
time-varying inflation risk premia throughout the term structure. Unfortunately, he does not report
the exact comparable estimates to Barr and Campbell. The closest comparable regression coefficient
is the 5-year bond over a holding period of 6 months (instead of 1 year) which equals -0.54 (compared
to 0.06). The 3-year spread coefficients on 6-, 12- and 24-month holding periods resemble the ones
found in the nominal regressions: negative and decreasing with maturity (-0.62, -0.76 and -1.06
respectively). For all holding periods, the estimates are statistically significant.\(^5\)

3 Model of Segmented Asset Markets

In this section I present the model of endogenously segmented asset markets with stochastic endow-
ment. This model is a variant of the model of segmented asset markets of Alvarez, Atkeson, and
Kehoe (2008) and also their earlier work (2002). First, I describe the economy and compute the
equilibrium. Then, I characterize the consumption of active households with respect to changes in
money growth. Consumption of households active in the asset market is a concave function of money
growth. This result is central for asset prices.

3.1 Description of the Model

Consider an economy with a continuum of infinitely lived households. There are two locations to trade:
the goods market and the asset market. To carry out transactions in either market, households need
money. In the goods market they exchange cash for consumption goods while in the asset market they
trade cash for nominal assets. The first source of uncertainty in this economy is the growth rate of
money. Money enters the economy via open market operations carried out by the government in the
asset market. The second and last source of uncertainty is the shocks to the aggregate endowment in
this economy. Every household receives an identical share of the aggregate endowment. As in other
cash-in-advance models, households cannot consume their own endowment. Consequently households
have to trade their endowment for cash, which they can use to buy consumption goods subject to
the standard cash-in-advance constraint.

The key point of the model is that transfers of cash between the asset market and the goods
market are costly and that households are heterogeneous in terms of the transaction cost. For clarity,
think of a brokerage account that provides access to the asset market and interest-bearing securities;
the goods market requires a checking account that pays no interests on balances. The transaction
cost is the fee charged by the broker for transfers between these two accounts. To model household

\(^5\)There are well documented biases in the coefficients of the expectations hypothesis regressions because, basically,
the forecasting coefficient is a transformation of serial correlation coefficients of interest rates. See more details of this
interpretation in Bekaert, Hodrick, and Marshall (1997). The authors suggest using Monte-Carlo experiments to asses
the validity any hypothesis testing.
heterogeneity, let $\gamma$ be the per period fixed cost in real terms of transactions between the goods and the asset market. Furthermore, let $F(\gamma)$ be the distribution of the transaction costs and $\gamma$ to index households hence assuming a unit measure of households.

The potential sources of uncertainty in this economy are money growth and endowment. Let $\mu_t = M_t / M_{t-1}$ be the money growth rate and $M_t$ the money stock in period $t$. For the money growth process let $g(\mu^t)$ be the density at time $t$ of history of shocks $\mu^t = (\mu_1, ..., \mu_t)$. Regarding the process of endowment, assume each household receives an identical amount $\omega_t$ of consumption goods. Then the period $t \geq 1$ aggregate endowment is $\Omega_t = \int \omega_t f(\gamma) d\gamma$, where $f(\gamma)$ is the density associated with the distribution of households. (The notation distinguishes between the aggregate and the individual endowment although they are identical from the assumption of the unit mass of households). Finally, let $h(\Omega^t)$ be the density at time $t$ of history of endowment realizations $\Omega^t = (\Omega_1, ..., \Omega_t)$. I assume shocks are independent from each other.

The timing of the model follows Lucas (1982). At the beginning of a period $t \geq 1$, the money growth rate, $\mu_t$, and the aggregate endowment, $\Omega_t$, are realized and observed, which determines the current price level $P(\mu^t, \Omega^t)$. Each household $\gamma$ comes into the period $t \geq 1$ with an amount of cash $P(\mu^{t-1}, \Omega^{t-1}) \cdot \omega_{t-1}$ from selling its endowment in the previous period. Consequently the real value of money balances is given by $n(\mu^t, \Omega^t) = P(\mu^{t-1}, \Omega^{t-1}) \cdot \omega_{t-1} / P(\mu^t, \Omega^t)$. After the realization of money growth and endowment, the household splits into a worker and a shopper-trader. The worker receives this period’s realization of endowment $\omega_t$ which is sold for the cash amount $P(\mu^t, \Omega^t) \cdot \omega_t$ in the goods market. This amount will only be used in the next period because within periods transfers between the worker and the shopper-trader are not allowed. On the other hand, the shopper-trader takes the real money balances of the current period into the asset market. In the asset market, the household $\gamma$ has an amount $B(\mu^t, \Omega^t, \gamma)$ of bonds that pay cash into their asset market account contingent on the realization of current period money growth rate and endowment realization. With the state-contingent bonds redeemed for cash in the asset market, the shopper-trader can chose to reinvest in bonds for the next period or transfer cash to the goods market. Also, the household can transfer cash in the other direction, from the goods market account to the asset market account to buy new bonds. To carry out the cash transfer between accounts, the household has to pay the real cost $\gamma$. After rebalancing the portfolio, payment of the fixed cost and cash transfers, the shopper-trader uses the cash in the goods market to consume. Consumption $c(\mu^t, \omega^t, \gamma)$ equals to the real money balances, $n(\mu^t, \omega^t)$, plus (minus) the amount of cash transferred from (to) the asset market account denoted by $x(\mu^t, \omega^t, \gamma)$.

Asset market segmentation is a consequence of the heterogeneity of households in this economy. All households face the same cost of inflation since their endowment is identical, while the heterogeneity of transaction costs generates heterogeneous incentives to carry out transfers between the asset
market and the goods market. In equilibrium only a fraction of agents will incur the cost to exchange
 cash for financial assets. Dub households that incur the cost as active and the rest non-active. To
 keep track of households let \( z(\mu^t, \omega^t, \gamma) \) be an indicator function equal to one if household \( \gamma \) is active
 in period \( t \) and zero otherwise.

Now let’s turn to the constraints in this economy and the problem households solve. The amount
of real resources in this economy is the sum of all households’ endowment which must equal their
consumption, \( c(\mu^t, \omega^t, \gamma) \), plus the cost in real terms of carrying out the cash transfers, or in equations:

\[
\int [c(\mu^t, \omega^t, \gamma) + \gamma z(\mu^t, \omega^t, \gamma)] f(\gamma) d\gamma = \omega_t,
\]

where \( f(\gamma) \) is the density associated with the distribution of households. The real money balances
are given by:

\[
n(\mu^t, \omega^t) = \frac{\omega_{t-1} \cdot P(\mu^{t-1}, \omega^{t-1})}{P(\mu^t, \omega^t)},
\]

where \( P(\mu^t, \omega^t) \) is the price level at time \( t \). Hence, at any period \( t \geq 1 \), the money market has the
following clearing condition:

\[
\frac{M(\mu^t)}{P(\mu^t, \omega^t)} = \int \{n(\mu^t, \omega^t) + \gamma z(\mu^t, \omega^t, \gamma)\} f(\gamma) d\gamma.
\]

The last equation shows the uses of money in this economy: to buy consumption goods in the goods
market, to transfer resources between their asset market and cash accounts, and to pay the cost of such
transfers. Finally, money is introduced by the government via open market operations exchanging
money for bonds in the asset market. Thus the budget constraint of the government at time \( t \geq 1 \) is
given by:

\[
B(\mu^t, \omega^t) = M(\mu^t) - M(\mu^{t-1}) + \int_{\omega_{t+1}} \int_{\mu_{t+1}} Q(\mu^t, \mu_{t+1}, \omega^t, \omega_{t+1}) B(\mu^t, \mu_{t+1}, \omega^t, \omega_{t+1}) d\mu_{t+1} d\omega_{t+1},
\]

where \( B(\mu^t, \omega^t) \) is a one-period bond or claim to cash in the aggregate state \( (\mu^t, \omega^t) \) and \( Q(\mu^t, \mu_{t+1}, \omega^t, \omega_{t+1}) \)
is the price of a bond paying in state \( (\mu_{t+1}, \omega_{t+1}) \) given history of shocks \( (\mu^t, \omega^t) \). Each period \( t \geq 1 \)
the government redeems the outstanding state-contingent bonds by introducing additional stock of
money and by selling a full set of bonds contingent on the next period’s state. Complete markets are
assumed.

Given this structure of markets and information, the problem of typical household \( \gamma \) is to choose
a sequence of consumption that maximizes the expected discounted utility:

\[
\sum_{t=0}^{\infty} \beta^t \int U(c(\mu^t, \omega^t, \gamma)) g(\mu^t) h(\omega^t) d\mu^t d\omega^t,
\]
subject to the cash-in-advance constraint:

\[ c(\mu^t, \omega^t, \gamma) = n(\mu^t, \omega^t) + x(\mu^t, \omega^t, \gamma)z(\mu^t, \omega^t, \gamma), \]

and sequence of asset market constraints \( t \geq 1 \):

\[
B(\mu^t, \omega^t, \gamma) = \int_{\omega_{t+1}}^{\mu_{t+1}} \int_{\mu_{t+1}}^{\mu_{t+1}} Q(\mu^t, \mu_{t+1}, \omega^t, \omega_{t+1}) B(\mu^t, \mu_{t+1}, \omega^t, \omega_{t+1}, \gamma) d\mu_{t+1} d\omega_{t+1} + P(\mu^t, \omega^t) [x(\mu^t, \omega^t, \gamma) + \gamma] z(\mu^t, \omega^t, \gamma),
\]

where \( B(\mu^t, \omega^t, \gamma) \) is the specific position of household \( \gamma \) in period \( t \) bonds. Once the transaction cost to participate in asset market has been paid, agents can make transfers to the goods market and access the complete set of state contingent cash claims \( B(\mu^t, \mu_{t+1}, \omega^t, \omega_{t+1}, \gamma) \).

In equilibrium, if inflation is positive, some proportion of households will incur the fixed cost and become active in the asset market. The next section characterizes the level of consumption of active households and the market segmentation in this economy.

### 3.2 Equilibrium

The concept of equilibrium in this economy is standard: a collection of sequences \( \{c, x, n, z, B\} \) and prices \( \{P, Q\} \) such that: i) for each household \( \gamma \), the allocation solves the household utility maximization problem, and (ii) the government budget constraint, resource constraint and the money market-clearing condition hold. Notice that corresponding to this equilibrium is an equilibrium level of transaction cost, which I denote by \( \bar{\gamma} \). This level of transaction cost is also the name of the marginal household that finds optimal to pay the fixed cost to participate in the asset market. I will use these two concepts interchangeably.

Computing the equilibrium requires several steps, which I describe intuitively below.\(^6\) Given the price level, each household chooses to participate if the benefit, net of its idiosyncratic transaction cost, is larger or equal to the inflation tax to their real money balances. The cost of inflation is identical for any household that does not participate in the asset market. On the other hand, the benefits of participating in the asset market vary across households depending on the particular level of transaction cost \( \gamma \). Households with a low transaction cost, benefits of participation will be larger. Households that do not participate must find the cost too high. Hence, the household with equilibrium level \( \bar{\gamma} \) is the marginal household, which finds that the costs equal the benefits of participating.

\(^6\)The solution to the model, equilibrium consumption levels and degree of market segmentation, can be easily computed by decentralizing the planner’s problem using the appropriate transfers (see Alvarez, Atkeson, and Kehoe, 2008).
To compute the equilibrium, first I calculate the price level. Second I determine the level of consumption of non-active households, which is independent of the degree of segmentation in equilibrium. Finally I compute the level of consumption of active households and the equilibrium transaction costs such that all households maximize their utility and the goods and asset markets clear. In the remainder of this section I provide intuition on the responses of the active households to changes in money growth rate and endowment, and the variability of the asset market segmentation.

1. **Price Level.** Use the cash-in-advance constraint:

\[
c(\mu^t, \omega^t, \gamma) = n(\mu^t, \omega^t) + x(\mu^t, \omega^t, \gamma)z(\mu^t, \omega^t, \gamma),
\]

to substitute for the real balances, \( n \), in the money market clearing condition

\[
\frac{M(\mu^t)}{P(\mu^t, \Omega^t)} = \int \left\{ n(\mu^t, \omega^t) + [x(\mu^t, \omega^t, \gamma) + \gamma]z(\mu^t, \omega^t, \gamma) \right\} f(\gamma) d\gamma
\]

\[
= \int \left\{ c(\mu^t, \omega^t, \gamma) - x(\mu^t, \omega^t, \gamma)z(\mu^t, \omega^t, \gamma) + [x(\mu^t, \omega^t, \gamma) + \gamma]z(\mu^t, \omega^t, \gamma) \right\} f(\gamma) d\gamma
\]

which in the last line equals the resource constraint. Therefore the price level is given by:

\[
P(\mu^t, \Omega^t) = \frac{M(\mu^t)}{\Omega_t},
\]

thus it becomes obvious that velocity is constant in this economy. Contrary to the original model of *Alvarez, Atkeson, and Kehoe* (2008), inflation in this environment is not equal to money growth but depends on the realization of endowment in the current period. If goods are more plentiful this period, inflation will be lower and the opportunity cost for households of not participating in the asset market will decrease.

2. **Non-active Household Consumption.** Households that do not carry out transactions in the asset market consume only the endowment they sold for cash in the previous period. This nominal amount is affected by inflation via the money growth and the current period endowment. Let \( c_{NA}(\mu^t, \omega^t) \) denote the consumption of the non-active households. This function does not depend on the household’s transaction cost because the endowment and real money balances do not depend on transaction costs. Since non-active households do not have any other source of cash to carry out purchases in the goods market, their consumption is given by the amount of real money balances:

\[
c_{NA}(\mu^t, \omega^t) = n(\mu^t, \omega^t) = \frac{\omega_{t-1} \cdot P(\mu^{t-1}, \Omega^t)}{P(\mu^t, \Omega^t)} = \frac{\Omega_t}{\mu_t}.
\]
Clearly all non-active households are identical in their consumption regardless of the heterogeneity in transaction costs.

3. **Period-by-Period Solution.** Under the assumption that households cannot store cash in the goods market the cash-in-advance constraint always binds. Moreover, whenever inflation is positive, bonds will dominate cash in terms of return, therefore households will not store cash in the asset market. Given these conditions, the infinite sequence problem of the households becomes a series of period-by-period static problems. This is the case because, given the timing of the shocks and information in the model, consumption in the period is perfectly forecastable when the shopper-trader is at the asset market rebalancing the portfolio of the household. Therefore, the choice of cash transfer is purely to meet the transactions demand for cash of the households. There is no precautionary demand for cash in the model. Hence, the choice of transfers has no effect on the future period’s consumption of the household.\(^7\)

4. **Active Household Consumption and Segmentation.** Given the form of the solution, I can drop the time subscripts. Now let \(c_A(\mu, \omega)\) be the consumption level of active households. As with non-active households, all active households have the same consumption level. This is true because by participating in the asset market and trading in a complete set of securities they insure all idiosyncratic risk by pooling their endowment and sharing a common level of consumption. To recap, we have so far that non-active households consume \(c_{NA}\). The marginal household, \(\gamma\), is the one that, given the shock \((\mu, \omega)\), equates the utility gain from cash transfers between the asset and goods market to the cost of carrying out such transactions. Formally:

\[
[U(c_A(\mu, \omega)) - U(c_{NA}(\mu, \omega))] = U'(c_A(\mu, \omega))(c_A(\mu, \omega) + \gamma(\mu, \omega) - c_{NA}(\mu, \omega)).
\]  

(2)

This is the first order condition of the planner’s problem when all households are given equal weights. Using the planner’s problem instead of the decentralized problem works because the welfare theorems apply in this economy (see Alvarez and Jermann, 2001). Then given the equilibrium level of transaction cost, \(\gamma\), and the level of consumption of the active households, \(c_A\), the resource constraint is given by:

\[
\int_0^{\gamma(\mu, \omega)} \gamma f(\gamma)d\gamma + c_{NA}(\mu, \omega) \cdot [1 - F(\gamma(\mu, \omega))] + c_{A}(\mu, \omega)F(\gamma(\mu, \omega)) = \Omega.
\]  

(3)

This equation shows that households above the cutoff transaction cost \( \bar{\gamma} \) are non-active and vice versa. Notice from the resource constraint that the equilibrium level of transaction cost, \( \bar{\gamma} \), and consumption of active agents, \( c_A \), adjust for the goods market to clear. Therefore, the solution to the model is characterized by functions \( c_A(\mu, \omega) \) and \( \bar{\gamma}(\mu, \omega) \) such that the net gain equation and the resource constraint hold.

The uniqueness of the equilibrium can be argued as follows. Consider the equilibrium pair \((c_A, \bar{\gamma})\) for \( \mu > 0 \). Since it is an equilibrium, the allocation is feasible. Now consider an alternative equilibrium pair for the same level of money growth and aggregate endowment but for a larger consumption level of active households \( \tilde{c}_A > c_A \). Notice that real money balances for non-active households are the same and, consequently, for equation 3 to hold implies that the equilibrium segmentation level \( \tilde{\gamma} < \bar{\gamma} \) is smaller. But this contradicts the optimality condition stated in equation 2 because to a larger direct utility gain from engaging in the asset market corresponds a larger transaction cost.

### 3.3 Characterizing the Consumption of Active Households

In the last section, I argued that the equilibrium of this economy can be characterized by a pair of functions \( c_A(\mu, \omega) \) and \( \bar{\gamma}(\mu, \omega) \). This section explains the attributes of the equilibrium functions; in particular it shows how consumption of active households changes in response to money growth and endowment shocks. The main focus is on the dynamics of consumption of active households because households attending the asset market are the ones pricing assets in this economy. First I explain the intuition of the consumption of active households. Second a plot of a parametrized version of the model shows graphically the characteristics of the model. Third, a proposition formalizes the results.

To gain intuition of the mechanics of the equilibrium, note that when money growth is zero the economy reduces to a representative consumer economy with an asset market segmentation equal to zero regardless of endowment. All households are identical because without a decision to participate in the asset market, there is no heterogeneity in this economy. Departing from this trivial case, consider how changes in the growth rate of money affect the consumption of active households and the decision to participate in the asset market. Take the equilibrium pair \( (c_A(\mu, \omega), \bar{\gamma}(\mu, \omega)) \). Holding the endowment fixed, an increase in the money growth rate to \( \mu' > \mu \), increases inflation one-for-one and reduces the real money balances that households hold at the beginning of the period. In the margin the next household \( \hat{\gamma} > \bar{\gamma} \) will find it profitable to engage in the asset market thus increasing the degree of market segmentation. However, this next household brings to the asset market smaller net real money balances because it has to pay a larger transaction cost \( \bar{\gamma} \) to perform transfers to the goods market.

Figure (5) shows the consumption level and degree of segmentation for different levels of inflation holding endowment fixed. The consumption of active households and the level of transaction costs
Figure 5: Active households’ log consumption, $\log c_A(\mu, 1)$, and proportion of participating households, $F(\bar{\gamma}(\mu, 1))$, as a function of inflation. These plots are computed by solving equations (2) and (3) for a range of $\log \mu$ with endowment equal to one unit of consumption goods. The utility function is constant relative risk aversion with parameter $\psi = 2$. For this exercise 5% of households are assumed to have zero transaction cost. The rest of households are distributed uniformly with an upper bound in transaction cost of 10% of endowment. From this plot, it is clear that the consumption of active agents is a concave function of inflation.
Figure 6: Active households’ log consumption, log $c_A(\mu, \omega)$, as a function of inflation and endowment. This plot was computed by solving equations (2) and (3) for a range of log $\mu$ and $\omega$. The utility function is constant relative risk aversion with parameter $\psi = 2$. For this exercise 5% of households are assumed to have zero transaction cost. The rest of households are distributed uniformly with an upper bound in transaction cost of 10% of endowment. As in Figure (5), consumption of active agents is concave in money growth but linear in endowment. With proportional transaction costs the curvature of consumption with respect to money growth is independent of endowment.

were computed using a parametrized version of equations (3) and (2). For this example I assume constant relative risk aversion utility $U(c_A(\mu_t, 1)) = c(\mu_t, 1)^{1-\psi}/(1-\psi)$ where $\psi = 2$ is the coefficient of relative risk aversion. Transaction costs are assumed uniformly distributed between 0 and 10% of endowment and a mass of 5% of households with zero transaction cost. The objective of the figure is to show that log consumption of active households is a concave function of log inflation and thus can be closely approximated in the relevant range of money growth by a quadratic function. Notice that the marginal utility is higher at low levels of inflation. For this example, log consumption of active households responds one for one with log inflation at low levels of inflation. Likewise, the share of households participating as inflation becomes positive is the mass of households with zero transaction cost.

What happens when endowment is not fixed? This depends on the assumption of transaction costs. If transaction costs are proportional to endowment, the consumption of active households is proportional to endowment for any level of money growth. Figure (6) shows the surface of the log of consumption of active households for different levels of inflation and different levels of endowment. Using the same parametrization as in Figure (5), the schedule log $c_A(\mu, 1)$ is proportional to
endowment, \( \omega \), preserving the curvature of consumption with respect to the money growth rate.\(^8\)

In summary, the response of consumption of active households to changes in money growth is non-linear given that transaction costs are proportional to endowment. At any level of money growth, a fraction of households participate while some remain inactive. An increase in the rate of money growth increases the inflation tax levied on households that remain inactive and consequently the benefits from participating in the asset market. In equilibrium the increase in money growth increases the proportion of active households. However, as money growth increases, consumption of participating households increases less than proportionally. To see this, notice that all active households equate their marginal rate of substitution by trading securities in the complete asset market. As households with higher transaction cost find it optimal to participate in the asset market, they bring smaller net real money balances. Therefore, consumption of active households will grow less than the increase in money growth. As a result, the marginal utility of active households varies less with higher money growth rates because of the endogeneity of the asset market segmentation. This endogeneity is a non-insurable source of risk for active agents. The implication for asset prices is that the marginal utility of active households is more variable at lower levels of money growth. The changes in the variability of the marginal utility of active households with the level of money growth introduce heteroskedasticity to bond prices.

The next proposition summarizes the result of the form of the equilibrium assuming transaction costs are proportional to endowment.

**Proposition.** Under the assumption of proportional and uniform transaction costs, and homogeneity of the utility function, the equilibrium consumption and segmentation functions are proportional to endowment, \( c_A(\mu, \omega) = c_A(\mu, 1) \cdot \omega \), and \( \bar{\gamma}(\mu, \omega) = \bar{\gamma}(\mu, 1) \cdot \omega \).

**Proof.** The proof is in two parts: (i) proves that the resource constraint holds under the guessed form of the equilibrium functions, while (ii) shows that the optimality condition remains unchanged as well.

Part (i). Use the resource constraint:

\[
\int_{0}^{\bar{\gamma}(\mu, \omega)} \gamma f(\gamma) d\gamma + \left( \frac{\omega}{\mu} \right) \cdot [1 - F(\bar{\gamma}(\mu, \omega))] + c_A(\mu, \omega) \cdot F(\bar{\gamma}(\mu, \omega)) = \Omega_t.
\]

Replace the cdf with the functional form assumption of transaction costs. Let \( f \) be the piece-wise distribution of transaction costs with a positive mass of households with zero cost \( F(0) > 0 \) and the

\(^8\text{When transaction costs are not proportional to endowment, larger endowment realizations reduce the curvature of the consumption function of active households because participating in the asset market becomes cheaper for all households. This setting of transaction costs may be of interest in a time series perspective to reflect improvements in technology in the asset market that have reduced the costs of trading. I leave this task for future research.}\)
rest distributed uniformly $U[0, b]$ where $b$ is the maximum level of transaction costs. Let $s(x)$ be the density corresponding to the uniform part. Then for any transaction cost $x \in (0, b]$ we have that $s(x) = (1/b) \cdot (1 - F(0))$. The cdf for the uniform part is given by $S(x) = U(x \leq X; 0, b) \times (1 - F(0))$ and the cdf for the whole distribution is $F(x) = F(0) + S(x)$. Replacing in the budget constraint:

$$\left(\frac{\bar{\gamma}(\mu, \omega)}{2}\right)^2 S(\bar{\gamma}(\mu, \omega)) + \left(\frac{\omega}{\mu}\right) \cdot [1 - F(0) - S(\bar{\gamma}(\mu, \omega))] \\ c_A(\mu, \omega) \cdot [F(0) + S(\bar{\gamma}(\mu, \omega))] = \Omega.$$ 

Replace with the guessed form of the equilibrium $\bar{\gamma}(\mu, \omega) = \gamma(\mu) \cdot \omega$ and $c_A(\mu, \omega) = c_A(\mu) \cdot \omega$ and notice that $S(x \cdot \alpha) = S(x) \cdot \alpha$, then:

$$\left(\frac{\gamma(\mu) \cdot \omega}{2}\right)^2 S(\bar{\gamma}(\mu, \omega)) + \left(\frac{\omega}{\mu}\right) \cdot [1 - F(0) - S(\bar{\gamma}(\mu, \omega))] \\ c_A(\mu) \cdot \omega \cdot [F(0) + S(\bar{\gamma}(\mu, \omega))] = \Omega.$$ 

Recall that $\Omega = \int \omega f(\gamma) d\gamma$. Then if $b = \alpha \cdot \omega$ for some $0 < \alpha < 1$, is easy to verify that the budget constraint holds.

Part (ii). Now I verify that under the guessed form of the equilibrium the optimality condition holds. Replace the hypothesized functional form of the equilibrium in the optimality condition:

$$[U(c_A(\mu) \cdot \omega) - U(\omega/\mu)] = U'(c_A(\mu) \cdot \omega) (c_A(\mu) \cdot \omega + \bar{\gamma}(\mu) \cdot \omega - \omega/\mu).$$ 

Now since $U$ is homogeneous of degree $k$, then

$$\omega^k [U(c_A(\mu)) - U(1/\mu)] = \omega^{k-1} U'(c_A(\mu)) (\omega [c_A(\mu) + \bar{\gamma}(\mu) - 1/\mu])$$

which yields the optimality condition in the case of unitary endowment. Q.E.D.

This proof shows that endowment is irrelevant for the margin of substitution of money growth and is a scale variable under the assumption that transaction costs are proportional to endowment.

The next section derives the prices for zero-coupon real bonds in the macro model and characterizes the dynamics of risk premia.

4 Bond Yields and Asset Market Segmentation

In this section I derive the term structure of real interest rates in the model of segmented asset markets. In this economy the marginal investors (or highest valuation agents) are the active households. Therefore, pricing of assets can be done using their marginal rate of substitution (Luttmer, 1996 and Alvarez and Jermann, 2001). Notice that once engaged in the complete asset market, the marginal rate of substitution of all participating households are equated state-by-state and period-by-period.
Using the marginal rate of substitution of active agents, the price at period $t$ of a zero-coupon bond that pays one unit of consumption in $\tau$ periods is:

$$P_t(\tau) = E_t \left[ \beta^\tau \frac{U'(c_A(\mu_{t+\tau}, \omega_{t+\tau}))}{U'(c_A(\mu_t, \omega_t))} \right].$$  \hspace{1cm} (4)

To derive closed form formulas for bond prices, I approximate the marginal rate of substitution of active agents. With the second order approximation, the conditional expectation of the marginal rate of substitution can be computed in closed form and fits under the QTSM class. With this structure I also obtain closed form formulas for real risk premia and different moments on real interest rates. In addition, the coefficients in this approximation clearly map the coefficients in the pricing kernel $m_t$. Such mapping will provide economic content to the kernel coefficients $\phi$ and $\eta$.

The idea behind the order of the approximation is the following. Mathematically, log consumption of active households is concave function with respect to money growth (see Figure 5). Intuitively, for a given level of money growth and endowment, each household decides to participate in the asset market by comparing the utility gains to its idiosyncratic transaction cost. To a first order, this is an individual decision. However, to a second order, changes in the money growth rate change the equilibrium degree of asset market segmentation and, in turn, the consumption level of active households. Since log consumption of active households is linear in endowment, a first order approximation in this direction will suffice. Subsection 4.1 explains this intuition, subsection 4.2 computes prices of real bonds.

### 4.1 Second Order Approximation

The following is the second order approximation to the log marginal utility of consumption of active agents, $\log U'(c_A)$. The utility function is of the form $U(c_A(\mu_t, \omega_t)) = c(\mu_t, \omega_t)^{1-\psi}/(1-\psi)$ and $\psi$ is the coefficient of relative risk aversion. I will use the property of proportionality of consumption of active households to consider a second order approximation on the margin of money growth. Specifically, I will only approximate to a second order the term $c_A(\mu_t, 1)$ around the unconditional growth rate of money, $\log \bar{\mu}$. Therefore define the variable $\hat{\mu}_t = \log \mu_t - \log \bar{\mu}$. The units of this variable can be interpreted in percentage points. The idea behind the approximation is to capture the response of consumption to shocks along the trends of money growth.

Then consider the second order approximation around $\bar{\mu}$ of the log marginal utility of the form:

$$\log U'(c_A(\mu_t, \omega_t)) = \log U'(c_A(\bar{\mu}, 1)) - \phi(\bar{\mu}, 1) \cdot \hat{\mu}_t + \frac{1}{2} \eta(\bar{\mu}, 1) \cdot \hat{\mu}_t^2 - \psi \log \omega_t.$$  \hspace{1cm} (5)

Using the result on the proportionality of consumption of active households and the assumption of CRRA, the first and second order coefficients of the approximation with respect to money growth
are:
\[
\phi(\mu, \omega) \equiv -\frac{\partial \log U'(c_A(\mu, \omega))}{\partial \log \mu} = \psi \frac{\partial \log c_A(\mu, \omega)}{\partial \log \mu}
\]
and
\[
\eta(\mu, \omega) \equiv \frac{d^2 \log U'(c_A(\mu, \omega))}{(d \log \mu)^2} = -\psi \frac{d^2 \log c_A(\mu, \omega)}{(d \log \mu)^2}.
\]

The first order coefficient with respect to endowment is given simply by the coefficient of relative risk aversion:
\[
\frac{\partial \log U'(c_A(\mu, \omega))}{\partial \log \mu} = -\psi.
\]

Given the proportionality of the consumption of active households, the coefficients of the approximation with respect to money growth are independent of endowment. Furthermore, given the form of utility function, the term \(d \log c_A(\mu)/d \log \mu\) is positive and the term \(d^2 \log c_A(\mu)/(d \log \mu)^2\) is negative. Evidently \(\phi(\mu)\) is the elasticity of active households consumption to money growth and \(\eta(\mu) = \phi'(\mu)\). This implies that the elasticity of active households’ consumption decreases with money growth. As money growth increases, the variability of the pricing kernel decreases. This is the source of time-varying risk premia. The same intuition can be seen in Figure (5), where the log consumption of active households is increasing and concave in the log money growth rate.

### 4.2 Real Bond Prices

To compute real bond prices, write equation (4) using the second order approximation of the marginal utility of active households described in equation (5). With this approximation bond prices can be written as:

\[
P_t^{(r)} = \beta^t \mathbb{E}_t \exp \left[ -\phi \hat{\mu}_{t+\tau} + \frac{\eta}{2} \hat{\mu}_{t+\tau}^2 - \psi \log \omega_{t+\tau} + \phi \hat{\mu}_t - \frac{\eta}{2} \hat{\mu}_t^2 + \psi \log \omega_t \right].
\]

Now assume that the deviations from the mean money growth follow an AR(1) process:

\[
\hat{\mu}_t = \rho \hat{\mu}_{t-1} + \varepsilon_t
\]

where \(0 < \rho < 1\) is the autocorrelation coefficient and the innovations are \(\varepsilon \sim N(0, \sigma^2_\varepsilon)\) i.i.d. With respect to endowment, assume that the growth rate of endowment is an AR(1) with drift:

\[
\log \frac{\omega_{t+1}}{\omega_t} = \delta (1 - \theta) + \theta \log \frac{\omega_t}{\omega_{t-1}} + \iota_t,
\]

with \(\iota_t \sim N(0, \sigma^2_\iota)\) i.i.d. and \(0 < \theta < 1\). Then we can define \(\hat{\omega}_{t+1} = \log(\omega_{t+1}/\omega_t) - \delta\) as the period \(t + 1\) deviation of the growth rate of endowment from its long run mean. Formula (9) says that bond prices are given by the conditional expectation of two independent functions of the shocks: quadratic for money growth shocks and linear for endowment shocks. Notice, I have assumed that endowment
growth shocks are log normal, therefore in this respect, it is identical to the standard consumption based asset pricing formula (Cochrane, 2001).

To compute the conditional expectation above, we can exploit the assumption of normal innovations on both processes. Given the independence between the two shocks the problem can be split into two expectations: one of a quadratic random variable and one of a log normal random variable. To compute the expectation of the quadratic random variable, one completes the square and makes a change of variable. With respect to endowment, I solve forward the AR process for endowment growth similar to the standard model with consumption growth. The appendix shows the derivation in detail. The solution to the expectation of the quadratic part is:

\[
E_t \exp \left[ -\phi \hat{\mu}_{t+\tau} + \frac{\eta}{2} \hat{\mu}_{t+\tau}^2 \right] = (1 - \eta \kappa_1^2(\tau))^{-1/2} \\
\times \exp \left\{ \frac{1}{2\kappa_1^2(\tau)} \left( \frac{(\kappa_1(t, \tau) - \phi \kappa_1^2(t, \tau))^2}{(1 - \eta \kappa_1^2(t, \tau))} - \kappa_1^2(t, \tau) \right) \right\},
\]

where \( \kappa_1(t, \tau) = \rho \hat{\mu}_t \), and \( \kappa_1^2(\tau) = \sigma_\varepsilon^2(1 - \rho^2)/(1 - \rho^2) \) are the conditional mean and variance for the money process. The solution requires \( 1 - \eta \kappa_1^2(\tau) > 0 \) for any \( \tau \geq 1 \). I will focus on the limiting case when maturity tends to infinity, \( \tau \to \infty \), because the condition becomes more restrictive. Substituting for the conditional variance and taking the limit I get:

\[
\eta \sigma_\varepsilon^2 < (1 - \rho^2).
\]

The units of the state, \( \mu \), are chosen by the model, therefore we can give an intuitive interpretation to the restriction. Using the parametrization of Alvarez, Atkeson, and Kehoe (2008), \( \log \bar{\mu} = 5\% \) in annualized terms, \( \eta = 1,000 \) and \( \rho = 0.9 \), which corresponds to a speed of mean reversion of money growth of 7 months. In this case the corresponding upper bound of the standard deviation of money growth innovations is \( \sigma_\varepsilon < 0.014 \) or 17% in annualized terms. Larger shocks require faster mean reversion.

Using the previous result from the expectation of the quadratic random variable of money growth and the standard result of the expectation of log normal variables for the endowment, real bond prices
are:

\[
P_t^{(\tau)} = (\beta^\tau) (1 - \eta_2^2(\tau))^{-1/2} \\
\times \exp \left\{-\frac{\eta}{2} \mu_t^2 \left(1 - \frac{\rho^{2\tau}}{1 - \eta_2^2(\tau)}\right)\right\} \\
\times \exp \left\{+\phi_1 \mu_t \left(1 - \frac{\rho^\tau}{1 - \eta_2^2(\tau)}\right) + \frac{\phi^2_2 \xi_1^2(\tau)}{2 (1 - \eta_2^2(\tau))}\right\} \\
\times \exp \left\{-\psi \left(\hat{\omega}_t - \frac{\theta (1 - \theta^\tau)}{(1 - \theta)} + \delta \sum_{j=1}^{\tau} (1 - \theta^{\tau-j+1})\right)\right\} \\
\times \exp \left\{+\frac{\psi^2 \sigma_\nu^2}{2} \sum_{j=1}^{\tau} \left(\sum_{k=1}^{j} \theta^{j-k}\right)^2\right\}. \tag{12}
\]

4.3 Bond Yields

Real bond yields are quadratic functions in money growth but linear in endowment growth:

\[
y_t^{(\tau)} = -\log \frac{P_t^{(\tau)}}{\tau} = a(\tau) + b_1(\tau) \mu_t + q(\tau) \mu_t^2 + b_2(\tau) \hat{\omega}_t, \tag{13}
\]

where the coefficients are given by:

\[
a(\tau) = -\log \beta + \frac{1}{2\tau} \left\{\log (1 - \eta_2^2(\tau)) - \frac{\phi^2_2 \xi_1^2(\tau)}{1 - \eta_2^2(\tau)}\right\} \\
+ \frac{1}{\tau} \left(\psi \delta \sum_{j=1}^{\tau} (1 - \theta^{\tau-j+1}) - \psi^2 \sigma_\nu^2 \sum_{j=1}^{\tau} \left(\sum_{k=1}^{j} \theta^{j-k}\right)^2\right) \right\} \\
b_1(\tau) = -\frac{\phi}{\tau} \left(1 - \frac{\rho^\tau}{1 - \eta_2^2(\tau)}\right), \\
b_2(\tau) = \frac{\psi (1 - \theta^\tau)}{\tau} \frac{\theta}{(1 - \theta)}, \\
q(\tau) = \frac{\eta}{2\tau} \left(1 - \frac{\rho^{2\tau}}{1 - \eta_2^2(\tau)}\right). \]

Given the functional form of the marginal utility of active households and the assumptions on the processes of money and endowment growth, the term structure model of real interest rates is given by the system of coefficients (14). This system fits under the class of quadratic term structure models of Ahn, Dittmar, and Gallant (2002).

It is useful to have intuition on the sign and shape of the coefficients in the cross section because they will help us to understand the movements of the yield curve and risk premia from a shock to
money growth. The sign of the quadratic coefficient, \( q(\tau) \), is positive for any maturity given the restrictions on the parameters for the existence of the expectation. On the other hand, the sign of the coefficient of the first order term, \( b_1(\tau) \), depends on the maturity and parameter values. When \( \rho \) is close to one, the restriction imposed for existence of the expectation on \( \eta \sigma^2_\varepsilon \) becomes more stringent and \( b_1 \) becomes positive at any maturity.

To clearly illustrate the role of segmented asset markets, consider three special cases. First note that in the standard model with no asset market segmentation, the pricing kernel is constant and \( \eta = 0 \) and \( \phi = 0 \). In this case money is neutral and the term structure or real yields is equal to \(- \log \beta\) minus the precautionary savings factor attributed to the risk of endowment growth shocks. Second, consider the case where there are no general equilibrium effects of asset market segmentation. In this case \( \eta = 0 \), therefore the pricing kernel varies linearly with money growth thus yields are a linear function of money and endowment growth:

\[
y_t^{(\tau)} = a(\tau) + b_1(\tau)\hat{\mu}_t + b_2(\tau)\hat{\omega}_t, \quad \text{(log linear)}
\]

and real interest rates are mean reverting. I will call this the log linear model because in this case the approximation to the marginal utility of the marginal household is log linear. Alvarez, Atkeson, and Kehoe (2002) analyze this case for nominal yields when endowment is constant and show that risk premia are constant. For the case of real yields, the real short rate reverts to the mean at rate \((1 - \rho)\) because the coefficient \( b_1(1) = -\phi(1 - \rho) \). In this case real risk premia are constant over time, although they may vary over the term structure. This case illustrates the role of endogeneity.

The third case illustrates the importance of the persistence of money growth rate for the predictions of time-varying risk premia. Let the serial correlation of money be \( \rho = 0 \). Also shut down shocks to endowment growth and call this the i.i.d. model. In this case risk premia varies over time but less than it would because shocks to yields last one period. To see this set \( \rho = 0 \) and get:

\[
y_t^{(\tau)} = -\log \beta + \frac{1}{2\tau} \kappa - \frac{\phi}{\tau} \varepsilon_t + \frac{\eta}{2\tau} \varepsilon^2_t, \quad \text{(i.i.d. shocks)}
\]

where \( \kappa \) is a constant that depends on the parameters.

The next section shows the estimation procedure of the term structure model.

5 Estimation of the Term Structure Model

This section describes the quadratic term structure model (QTSM) of real interest rates derived from the model of segmented asset markets. I present a discussion of the characteristics of QTSM. Then I describe the Extended Kalman Filter used to estimate the coefficients of the term structure model and analyze the results.
5.1 Quadratic Term Structure Models

I present the squared-autoregressive-independent-variable nominal term structure (SAINTS) model of Constantinides (1992) which, as the name suggests, assumes that all the state variables are independent and autoregressive. The exposition here ignores the linear term of endowment but it would be trivial to consider it given the assumption of the independence of the two random variables. Ahn, Dittmar, and Gallant (2002) fully characterize this class of models, for example, when the variables are not independent.

Quadratic term structure models assume that the pricing kernel is a quadratic function of the state $X$:

$$ m_t = \exp \left\{ \zeta \cdot t - \phi X_t + \frac{\eta}{2} X_t^2 \right\} $$

(15)

where $\zeta$ is a pure time-discount term, and $\phi$ and $\eta$ are referred to respectively as the first and second order coefficients of the pricing kernel.\(^9\) I have chosen the signs and fraction for the second order term to match the second order approximation of $\log U'(c_A(\mu_t))$ in (5). The pricing kernel is always positive and prices any claim to future consumption. The stochastic discount factor that prices at time $t$ a claim to a unit of consumption to be delivered at $t + \tau$ is given by $m_{t+\tau}/m_t$. Therefore, zero-coupon real bond prices are given by $P_t(\tau) = \mathbb{E}_t[m_{t+\tau}/m_t]$. Let the process for $X_t$ follow a AR(1) process:

$$ X_{t+1} = \rho X_t + \varepsilon_{t+1}, $$

(16)

where the innovations $\varepsilon$ are i.i.d. normal with zero mean and variance $\sigma^2_\varepsilon$. Solving this conditional expectation (see appendix) gives the following yields model. The yield to maturity at time $t$ of a zero-coupon real bond maturing at $t + \tau$ is a quadratic function of the state:

$$ y_t(\tau) = a(\tau) + b(\tau)X_t + q(\tau)X_t^2, $$

(17)

where the coefficients are:

$$ a(\tau) = \zeta + \frac{1}{2\tau} \log \left( 1 - \eta \varsigma^2(\tau) \right) - \frac{\eta^2 \varsigma^2(\tau)}{1 - \eta \varsigma^2(\tau)}, $$

(18)

$$ b(\tau) = -\frac{\phi}{\tau} \left( \frac{\rho^2}{1 - \eta \varsigma^2(\tau)} \right), $$

$$ q(\tau) = \frac{\eta}{2\tau} \left( \frac{1 - \rho^2}{1 - \eta \varsigma^2(\tau)} \right), $$

and $\varsigma^2(\tau)$ is the conditional variance of $X_t$.\(^{10}\) The existence of the solution to the expectation relies

---

\(^9\)A trend variable independent from $X$ could be added to avoid assuming that the kernel is stationary. The pricing results would follow identically by modifying the term $\zeta$ to pick up the per-period variance of the trend variable.

\(^{10}\)Specifically:

\[ \varsigma^2(\tau) = \sigma^2_\varepsilon \left( \frac{1 - \rho^2}{1 - \rho^2} \right). \]
on the following condition: $1 - \eta \kappa^2(\tau) > 0$ for all $\tau > 1$. Section 4 provides an economic interpretation of each of the coefficients and of the restriction in terms of the model of segmented asset markets. The more traditional characterization of the coefficients, in terms of a system of ordinary difference equations, is shown in the appendix.

The key point of the term structure model is that the assumed absence of arbitrage and the functional form of the pricing kernel impose cross-equation restrictions into the coefficients $a(\tau), b(\tau)$ and $q(\tau)$ of the yields model, which help in the estimation of the filter.

The state $X_t$ should be understood to be an unobservable (latent) state for the purpose of the estimation outlined below. The model of segmented asset markets provides economic content, or a label, to this unobserved state. The filtered series of the unobservable state is the implied money growth rate that drive the changes in consumption of the marginal household in the model of segmented asset markets.

Quadratic term structure models, compared to affine models, have more flexibility in accounting for the patterns of time varying volatility of yields without imposing a particular correlation structure to the factors (Dai and Singleton, 2000). However, as shown by Ahn, Dittmar, and Gallant (2002) this flexibility is lost in the case of independent state variables. Another attractive feature for empirical work on nominal yields is that QTSM can rule out negative interest rates.

In spite of the convenient features of quadratic term structure models, they are less prominent than affine models in empirical work. The main reason is that their estimation is difficult. Unlike affine models, the short rate is not a sufficient statistic of the state vector: yields in QTSM cannot be written as functions of a single state variable. The same level of the short rate can generate different yield curves depending on the sign of the state variable. Therefore, to estimate these models, the econometrician needs to filter the latent state variables or to make strong assumptions on which observed variables determine yields. The traditional maximum likelihood approach is not convenient in this case because writing a likelihood function is, in general, not feasible. Ahn, Dittmar, and Gallant (2002) show that the distribution of interest rates are infinite sums of non central chi-square variates. Thus to estimate the term structure model Ahn, Dittmar, and Gallant (2002) suggest using the EMM to overcome the difficulties of writing the likelihood function. Another approach is to use moment conditions from a particular parametrization of the quadratic model and use GMM to test them. Leippold and Wu (2003) follow this approach to estimate a two-factor quadratic term structure model using moment conditions derived from stylized facts in the US treasury data. The downside to this procedure is that their latent variables remain unobserved.

The focus of this paper is the macroeconomic interpretation of real risk premia. Therefore the
objective of the estimation is to invert the implied state variable $X_t$, estimate the coefficients $\rho$ and $\sigma_\varepsilon$, and the coefficients of the quadratic pricing kernel $\phi$ and $\eta$. To filter the latent state variable I use the Extended Kalman Filter (EKF), which linearizes the yields equation around the predicted level of the state (Claessens and Pennacchi, 1996). This filtering method provides simpler implementations than non-linear filters and under certain conditions is a close approximation.\textsuperscript{12} Moreover, the simplicity of the pricing kernel (15) allows me to write the quasi-likelihood function of the filtering problem. Other papers using non-linear filters in term structure models are Campbell, Sunderam, and Viceira (2009) and Lund (1997). Duffee and Stanton (2004) praise the use of the Extended Kalman Filter as a reasonable estimation technique for “second-generation term structure models”. They argue that its implementation is simple and its finite-sample biases are similar to the ones of maximum likelihood.

5.2 Extended Kalman Filter Procedure

I use the Extended Kalman Filter to estimate the coefficients of the quadratic term structure model. Filtering the state implies estimating the coefficients of the dynamics of the state. I will also be mapping the resulting coefficients to the pricing kernel derived from the model of segmented asset market and will draw implications for market segmentation, volatility of consumption and risk premia. From the perspective of the estimation, I will interpret the state as a latent macroeconomic (money) factor that affects the consumption of marginal households and therefore the decision of asset market participation.

The EKF is a traditional Kalman filter with the addition of the observation equation being linearized in the update stage and evaluated around the level of the state predicated in the last period (Anderson and Moore, 1979). Therefore, the quadratic term structure model is naturally expressed in a state space form that fits the EKF by adding measurement error to the yields equation (17) and letting equation (16) be the process of the unobserved state (Hamilton, 1994). The observation equation relates the yields with the unobserved state $X_t$ and the predetermined variable for endowment growth $\hat{\omega}_t$ with the added assumption of measurement error:

$$y_t(\tau) = a(\tau) + b_1(\tau)X_t + b_2(\tau)\hat{\omega}_t + q(\tau)X_t^2 + \nu_t(\tau),$$

where $\nu(\tau)$ is assumed to be normally distributed with zero mean. The contemporaneous variance-covariance of innovations in the term structure is given by the diagonal matrix $E_t[\nu\nu'] = R$. I will interpret the innovations to the yields equation as measurement error, either from mistakes in data recording, or noise introduced from the procedure to compute constant maturity yields. For simplicity I assume that the measurement errors are uncorrelated across yields or over time.

Notice that part of the coefficients that determine \( a(\tau) \) are known as well as all the coefficients inside \( b_2(\tau) \) if the parameters of the process of endowment growth are known. This can be seen clearly by inspecting the functional form of the coefficients in system (14). Thus I will estimate the process of \( \hat{\omega}_t \) from an observed variable and treat coefficients \((\theta, \delta, \sigma_\epsilon)\) as known. This will leave the coefficient \( \psi \) unknown in \( b_2 \) as well as the coefficients inside \( a, b_1 \) and \( q \). With state equation (16), observation equation (19), and an initial value of \( X_0 \) the EKF recursively extracts the time series \( \hat{X}_t \) for a given set of parameters \( \Theta = \{ \rho, \sigma_\epsilon, \phi, \eta, \psi, R \} \). The number of parameters to estimate is 2 from the unobserved state, 3 from the pricing kernel and the diagonal size of \( R \). Therefore the number of parameters to estimate depends on the number of yields used in the estimation. In the next section I describe which set of yields I use in the estimation. With the parameters and time series of the latent state and observed factor of endowment, the (quasi-) maximum likelihood function is computed. Then a numerical optimization procedure maximizes the log likelihood function in terms of \( \Theta \). The following diagram clarifies the estimation algorithm:

\[
\Theta_0 \rightarrow \{ \hat{X}_t \}_0 \rightarrow \{ \hat{y}_t(\tau) \}_0 \rightarrow L(\Theta_0),
\]

or i) choose a initial set of coefficients; ii) using the Extended Kalman Filter, compute a filtered series for the unobserved state; iii) compute fitted yields given the set of parameters and filtered state; iv) compute the log likelihood for these set of coefficients. Then using a simulated annealing maximizer algorithm I search for the global maximum until \( \Theta_j = \Theta_{j+1} \) for some large \( j \).

To be more precise about the filter, let \( \hat{X}_{t|t-1} \) be the predicted state variable at time \( t \) with information up to time \( (t - 1) \). Then, let the error of the estimation of the observation equation be:

\[
\Psi_{t|t-1} = y_t - \hat{y}_{t|t-1},
\]

where \( y_t \) is the stacked vector of available maturities of yields at time \( t \) and

\[
\hat{y}_{t|t-1} = a + b_1 \hat{X}_{t|t-1} + q \hat{X}_{t|t-1}^2 + b_2 \hat{\omega}_{t-1},
\]

is the predicted level of yields with information up to time \( (t - 1) \). The EKF differs from the traditional Kalman filter in the update stage. The mean square error of the forecast of yields used to update the state with information up to time \( t|t \) is computed replacing observation equation (20) with its linearized version evaluated at the predicted value \( \hat{X}_{t|t-1} \). The Appendix shows the filtering procedure in detail.

In the absence of restrictions the parameters that enter the pricing kernel and state are not uniquely identified. To ensure identification of the state, I normalize the scale of the state variable and allow the variance of the innovation \( \sigma_\epsilon^2 \) to be treated as a free parameter (see Dai and Singleton, 2000 and Sandell and Yared, 1978). Also I choose an arbitrary level of the state and use the sign
restrictions imposed by the model of segmented asset markets on the coefficients $\phi$ and $\eta$ to identify the sign of the state. Moreover, I use the theoretical restriction on the sign of $\psi$, the coefficient of relative risk aversion. Finally, I can identify the time discount factor $\beta$ from the long term value of the yields, $E y_t^{(\infty)} = -\log \beta$.

5.3 Results of the Estimation

As a measure of endowment I use a quarterly series of real GDP interpolated into a monthly frequency using a cubic spline. Figure (7) shows this time series. The process for GDP growth to be estimated on the resulting monthly series is

$$\log \frac{\omega_{t+1}}{\omega_t} = \delta(1 - \theta) + \theta \log \frac{\omega_t}{\omega_{t-1}} + \iota_{t+1},$$

with $\iota \sim N(0, \sigma^2_\iota)$. I use a sample from January 1985 to December 2007; the estimates are shown in Table (4). The reported estimated standard deviation is in annualized terms in percent. The estimates are not very sensitive to the sample used to estimate the regression. The unconditional mean of the process, $\delta$, does increase a bit if the sample includes the 1960s and 1970s. The implied unconditional mean of GDP growth is $0.88\%$ per month. Possibly the spline introduces smoothing into the series. For further research I could use the series of consumption at quarterly frequency since the filter can easily accommodate the variable $\hat{\omega}$ being observed once every quarter. Another problem with the estimates of the process is that in finite samples it is hard to distinguish between a random walk with trend and a unit root (Hamilton, 1994).

**Table 3:** Estimates of the Endowment Process. AR(1) process for splined GDP growth from 1985.01 to 2007.12. Standard errors under the coefficients.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>$\delta(1 - \theta)$</th>
<th>$\theta$</th>
<th>$\sigma_\iota$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.0010$</td>
<td>$0.8820$</td>
<td>$0.3212$</td>
<td>$0.77$</td>
<td></td>
</tr>
<tr>
<td>(0.0003)</td>
<td>(0.0286)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I will use these estimates as known parameters in the EKF. With the endowment process estimates, the loadings of the yields model on endowment, $b_2(\tau)$, and the second line of $a(\tau)$ depend only on the coefficient of relative risk aversion, $\psi$. For $\beta$ I use the unconditional mean of the long term bond, $E y_t^{(\infty)} = -\log \beta$. Therefore I set $\beta = \exp \left\{ -0.031/12 \right\}$ to match the monthly frequency of the other parameters. In summary, I have estimates or priors of the following parameters: $\delta, \theta, \sigma_\iota$ and $\beta$. It
Figure 7: UK GDP growth rate from 1985.01 to 2007.12. This series is computed interpolating the quarterly level of GDP using a cubic spline and then computing monthly growth rates.
remains to estimate the coefficients of the quadratic part of the pricing kernel, \( \eta \) and \( \phi \), and the coefficients of the process of money growth \( \rho \) and \( \sigma_\varepsilon \) and the coefficient of relative risk aversion, \( \psi \). The set of parameters to estimate is 17: from the state dynamics \( \rho \) and \( \sigma_\varepsilon^2 \), from the pricing kernel \( \eta \) and \( \phi \), and the 13 diagonal elements of \( R \). I use the following fully observed 13 maturities: 4, 4.5, and yearly from 5 to 15 years. The original IL data contains a set of maturities close to this selection. I do not use the short or long end of the term structure because these have many missing observations that complicate the convergence of the estimation.

Maximizing the quasi-likelihood function with a mean zero state and a quadratic yields equation is a delicate procedure. In addition to the approximation error of the filtering procedure, the initial condition for the state has to be chosen carefully because the unconditional mean of the state is zero but in the data the yields show a clear downward trend. To ensure convergence I have to estimate the model in several steps. First I estimate the coefficients of the state and pricing kernel, fixing the measurement error variance covariance matrix equal to the variance of yields. Once I attain convergence I fix the coefficients of the state and re-estimate the coefficients in the kernel and measurement errors. In each step I use the simulated annealing algorithm to find a global maximum.

To compute standard errors of the estimates I run a bootstrap of 1,000 samples. The bootstrap is estimated by generating samples of \( \{ \hat{X}_t \}_{t=1}^T \) using \( \hat{\rho}, \hat{\sigma}_\varepsilon \), samples of \( \{ \hat{\omega}_t \}_{t=1}^T \) using \( \hat{\theta}, \hat{\delta}, \hat{\sigma}_\iota \), and samples of \( \hat{R} \) to form simulated yields treating the estimated parameters as population parameters. Then I run simulated annealing on each set of artificial yields and compute the standard errors on the resulting set of coefficients. Next, I discuss the results of the estimation of the state dynamics and report the estimates of the yields equation.

The estimated autocorrelation coefficient of the state variable is \( \hat{\rho} = 0.9952 \) with a standard error 0.004; the estimated standard deviation of the state is \( \hat{\sigma}_\varepsilon = 3.017\% \) and standard error of 0.15\% both in annualized terms. Given that the frequency is monthly, this implies a highly persistent dynamics of the state. The half-life of a one standard deviation shock is 50 months. This result implies that, in the model of segmented asset markets, the state variable identified as a money growth shock will have protracted effects in the movements of the marginal utility of active households. Figure (8) plots the filtered state variable. This variable closely follows the first principal component (see appendix) and captures the downward trend of rates in the sample as well as the volatility periods in mid 1992, 1999 and 2003. The correlation between the implied state and the first principal component is 0.98. This is not surprising since filtering this single variable amounts to inverting the state from a single yield if assumed without measurement error. Therefore, the filtered state closely mimics the behavior of yields.
Table 4: Estimates of the Extended Kalman Filter. Estimates of the parameters of the model using a sample from 1985.01 to 2007.12. Standard errors calculated from a bootstrap of 1,000 samples are shown under the coefficients.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>ρ</th>
<th>σ_ε</th>
<th>η</th>
<th>φ</th>
<th>ψ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.9952</td>
<td>3.0175</td>
<td>971.27</td>
<td>15.194</td>
<td>6.063</td>
</tr>
<tr>
<td>(Standard Error)</td>
<td>(0.004)</td>
<td>(0.1546)</td>
<td>(37.671)</td>
<td>(3.5989)</td>
<td>(0.3602)</td>
</tr>
</tbody>
</table>

The coefficients of the pricing kernel η and φ are identified up to the scale of the state variable. To interpret the results, I present the ratio of the coefficients. The ratio $\sqrt{\hat{\eta}/\hat{\phi}}$ equals 2.195 which can be interpreted as the relative loading of the state on the kernel. The interpretation in terms of the second order approximation of the log kernel $\log m_t$ is: a 1% increase in the level of the state reduces to a first order the kernel while to a second order it increases it. The overall effect is an increase of $\exp\{0.75\%\}$ in the marginal utility in the current period. Moreover, a further increase of the state to 2% will have a smaller increase of $\exp\{0.65\%\}$ from computing $\exp\{-\phi X + \eta X^2\}$. These incremental changes in the state have smaller increments in the log kernel and thus generate time-varying risk premia.

I evaluate the performance of the yields model computing the pricing errors of the fitted yields. I construct the fitted yields using the filtered state variable and the estimated coefficients of the yields model. I use two measures of pricing errors: the mean absolute deviation (MAD) and the root mean square errors (RMSE). The pricing errors are on average 20 and 1 basis points for the 5-year and 10-year respectively, under both measures. The performance of the maturities at the long and short ends of the curve is worse than in the middle of the curve: the 4- and 15-year bond pricing errors are on average 25 basis points. Considering the time series performance, Figure (9) shows the 5- and 10-year fitted and observed yields. Another aspect of the model (not shown) is the close fit of the estimated mean term structure. However, the volatility across the term structure in the model is flat while the data shows a steep decline (see Figure 3).

6 Model Implications

6.1 Implications Asset Market Segmentation and Consumption

With the filtered state I show that the macro model implies a very small variation over time in the degree of market segmentation as well as in the consumption of active households. This implies high variability of the pricing kernel in the macro model due to the endogeneity of the asset market.
Figure 8: Filtered state variable $\hat{X}$ using the EKF in the quadratic term structure model in UK real yield data (1985.01-2004.12). Notice that it closely tracks the level of yields or equivalently the first principal component of the variance-covariance of yields (shown in the appendix).

Table 5: Model pricing errors. Fitted yields are constructed using the filtered state variable and the coefficients of the term structure model. The units are percentages, therefore 0.20 means 20 basis points.

<table>
<thead>
<tr>
<th>Yields (years)</th>
<th>RMSE</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.2075</td>
<td>0.1639</td>
</tr>
<tr>
<td>10</td>
<td>0.0899</td>
<td>0.0674</td>
</tr>
<tr>
<td>15</td>
<td>0.2020</td>
<td>0.1538</td>
</tr>
</tbody>
</table>
segmentation. The segmentation of asset markets provides the economic mechanism that generates time-varying risk premia.

To derive the implications of the estimation for the consumption of active households, I will use the estimated parameters $\hat{\Theta}$ and the filtered state $\hat{X}$ to compute the implied dynamics of consumption of active households which I will call $\hat{c}_A$. Second, I ask which parametrization of the model of segmented asset markets generates such dynamics of consumption. The parametrization is in terms of the distribution of transaction costs, $F$, the fraction with zero transaction cost, $F(0)$, and maximum level of transaction cost $\gamma_{\text{max}}$. Finally, since I use the filtered state $\hat{X}$ as the input in the model of segmented markets, I show measures of money growth in the UK to support my claim.

In the model of segmented asset markets, endowment is fixed while transaction costs are increasing. This amounts to an assumption that the relative costs of transfers are smaller for wealthier households than for low-income households.\footnote{This is supported by US survey data. See Vissering-Jorgensen (2002).} The costs may include, apart from the actual brokerage account fees, costs of financial information or the degree of financial sophistication of households. Therefore, I will take the upper bound of transaction costs, $\gamma_{\text{max}}$, to be 10% of prevailing endowment. This reflects the case of some low-income households for which the costs of participating in the asset market may be prohibitively high. Regarding, $F$, the distribution of households along the line of transaction costs, I will take an agnostic stand. The uniform distribution is the one that makes the effects of money shocks smaller around the mean level of money growth, since, at any given level of transaction costs, the mass of households is the same. In other words, the hazard rate, the mass of households ready to

Figure 9: Fitted 5- and 10-year yields compared to the observed data. The fitted yields are constructed using the filtered state variable and the estimated coefficients of the yields model.
participate in the asset market, is the lowest. The remaining assumption is the mass of zero-cost households, \( F(0) \). The curvature of the approximation of the marginal utility depends crucially on this parameter.

I compute two series for consumption, ignoring endowment, to show how consumption solved numerically in the model and consumption computed with the approximated version track each other. Choosing an arbitrary scale, I compute \( \log \hat{c}_A \). Second, I feed \( \hat{X}_t \) and compute the equilibrium using equations (3) and (2) with different values of \( F(0) \) such that the implied consumption \( \log c_A \) matches \( \log \hat{c}_A \). The two series of consumption are presented in Figure (10). The exact series is computed using a \( F(0) = .07 \), or 7% mass of zero-cost households. In the time series, the exact model is tracked very closely by the estimated approximation. This serves as evidence that the asset prices derived from the approximation to the marginal utility of active households are a good approximation (at least locally) to asset prices derived from the exact equilibrium.

Now I put together the two factors of the model, the implied unobserved state and GDP growth, to compute the series of consumption of active households. The estimates of the parameters of term structure model, the estimated unobserved state, and the observed GDP growth imply that

---

14 Another possibility is the truncated log normal distribution. However, distributions that concentrate the mass of households would overstate (understate) the effects of money shocks around (away) the mean level of inflation.
Figure 11: Time series of: i) implied consumption of active households in percent above non-active households; ii) implied money growth rate in percent per month; and iii) GDP growth in percent per month. In the first part of the sample when money growth rate is above its long-run mean, the variability of consumption is mostly determined by GDP growth. In the latter part of the sample, when money growth is smaller, consumption of active households is mostly determined by money growth.

The standard deviation of the log consumption of active households is less than 1% in annualized terms. I also compute the volatility of the implied degree of households’ asset market participation in the model. I find that participation varies less than 0.5% annually but decreased over time. The result for consumption is consistent with the empirical observation that consumption growth is highly persistent. The result for asset market participation is during sample period as macroeconomic volatility, of inflation or output growth, has also fallen in the sample (Galí and Gambetti, 2009). Since the macroeconomic model provides a precise interpretation of unobserved state, I can decompose the importance of macroeconomic shocks to consumption over time. The unobserved state is labeled by the model as the money growth rate. In the early part of the sample, the growth rate of the unobserved factor is large. In this part of the sample, most of the consumption volatility is due to output growth shocks because when money growth is high, consumption responds less to innovations of money growth. As the implied money growth rate falls towards the end of the sample, consumption volatility responds less to output growth innovations and more to money growth innovations. Figure (11) shows this argument graphically.

Interpreting the QTSM estimates using the segmented asset market model implies that the es-
Figure 12: UK seasonally adjusted money aggregates at yearly growth rates. M0 includes only notes and coins and financial institutions’ deposits at the Bank of England. M2 is M0 plus retail deposits less than 1 month while M3 includes deposits up to one year. After 1997, the correlation among these series decreases significantly, due probably to the increase in financial intermediation after the disinflation period and improvements in retail banking that offered small returns on checking accounts.

timated state variable is the money aggregate used by the government to carry out open market operations. Base money or M0 are too narrow empirical counterparts compared to the theoretical monetary aggregate of the segmented asset market model. The key point of the definition of money used in the goods market is that it either pays no or less interest than other assets that permit households insuring against inflation shocks. Thus M2 is the appropriate measure because it includes notes and all retail deposits of less than one month. In the UK, retail deposits are defined as deposits where the customer accepted an advertised rate (including zero) at a branch of a financial institution (Bank of England, 2008). Figure (12) shows the seasonally adjusted aggregates M0, M2 and M3 at yearly growth rates. M0 includes only notes and coins and financial institutions’ deposits at the Bank of England. M3 is the same M2 except that it only includes deposits up to one year. After 1997, the correlation among these series decreases significantly, probably due to the increase in financial intermediation after the disinflation period and improvements in retail banking that offered small returns on checking accounts. Finally, Figure (13) shows the plot of the estimated state variable and (demeaned) M2 yearly growth. The two series move broadly in cycles of 3 years and have a correlation of 0.31.
6.2 Implications for Risk Premia

In this subsection I analyze the cross section and dynamics of risk premia in the model. I show how the model generates risk premia in the cross section. Then I run the forecasting regressions of Barr and Campbell using the bootstrapped samples of data to show the existence of time-varying risk premia.

6.2.1 Cross Section of Risk Premia

Define the one-period log holding return of a long bond of maturity $\tau$ as

$$r_{t+1}^{(\tau)} \equiv \log P_{t+1}^{(\tau-1)} - \log P_t^{(\tau)}.$$ (21)

Now define real risk premia as:

$$rpr_t^{(\tau)} \equiv E_t[r_{t+1}^{(\tau)}] - y_t^{(1)},$$ (22)

which is the expected excess return on a long bond. This measures the risk of a one-period investment strategy of holding a $\tau$-period bond funded by short selling the real short rate. In the case of the pure log expectations hypothesis, the return risk premium is constant (Cox, Ingersoll, and Ross, 1981). The quadratic term structure model allows risk premia to be time-varying because the pricing kernel has time-varying volatility. Let $\tilde{a}(\tau) = -a(\tau) \cdot \tau$, $\tilde{b}(\tau) = -b(\tau) \cdot \tau$, and $\tilde{q}(\tau) = -q(\tau) \cdot \tau$, thus the

![Comparison of filtered state variable and money aggregate UK](image)

**Figure 13**: Estimated state variable $\hat{X}$ (RHS) and M2 yearly growth rate (LHS). The two series move broadly in cycles of 3 years. The correlation in changes is 0.31.
explicit form for the quadratic model is:

\[
    rpr_{i}^{(\tau)} = \tilde{a}(\tau - 1) - \tilde{a}(\tau) - a(1) + \tilde{q}(\tau - 1)\sigma_{\varepsilon}^2 \\
    + \left[ b(\tau - 1)\rho - \tilde{b}(\tau) - b(1) \right] \tilde{\mu}_t \\
    + \left[ \tilde{q}(\tau - 1)\rho - \tilde{q}(\tau) - q(1) \right] \tilde{\mu}_1^2,
\]

which is quadratic on the state variable. Risk premia can be positive or negative and can accommodate rich shapes over the term structure depending on the particular realization of \( \tilde{\mu} \). To see the source of time variation more clearly I will show two cases, first the log linear model and second the model where the state is i.i.d. The log linear model verifies the expectations hypothesis, or, equivalently, has constant risk premia. First, note that when \( \eta = 0 \) the coefficient \( q(\tau) \) is zero as well. Also compute the coefficient \( b \) using \( \eta = 0 \). Then

\[
    rpr_{\eta=0}^{(\tau)} = \tilde{a}(\tau - 1) - \tilde{a}(\tau) - a(1) \\
    = -\frac{\phi^2}{2} \frac{\sigma_{\varepsilon}^2}{1 - \rho^2} \left[ (1 - \rho^2(\tau - 1)) - (1 - \rho^2) - (1 - \rho) \right],
\]

which is constant over time since the state variable does not appear in the equation. In general, risk premia in the log linear case is positive and decreasing with maturity. Using the estimated values from the quadratic model, at the short end of the curve the risk premium is 55bp and decreases smoothly toward 20bp in the long end. Since money growth rate is mean reverting, the variability of long bonds is smaller than of short bonds therefore investors require a smaller compensation to hold them. Now consider the case when the state variable is a random walk (\( \rho = 0 \)), then excess returns are constant across maturities and over time:

\[
    rpr_{\rho=0} = -\frac{1}{2} \left[ \log \left( 1 - \eta \sigma_{\varepsilon}^2 \right) - \frac{\phi^2}{1 - \eta \sigma_{\varepsilon}^2} + \eta \sigma_{\varepsilon}^2 \right],
\]

and equals 40bp. This shows that variation over time of risk premia is given by the persistence of money growth and the endogeneity of the asset market segmentation. In the model, when money growth deviates from the mean, the volatility of the pricing kernel changes through the second order coefficient \( \eta \). Thus, even if the kernel is quadratic, if money growth shocks are not persistent, then the pricing kernel is homoskedastic.

6.2.2 Variation Over Time of Risk Premia

To show that the model of segmented asset markets accounts for the variation over time of risk premia, I will run the Campbell and Shiller (1991) forecasting regressions with the bootstrapped yield data. The bootstrapped yield data was generated in the following way. First I generated samples of \( X \) using the estimates \( \hat{\Theta} \). Then I formed yields from equation (19) by computing the coefficients \( \hat{a}, \hat{b} \) and \( \hat{q} \) and drawing the corresponding measurement error from the estimated variance covariance.
Now I derive the forecasting regression. Write the excess return equation (22) in terms of yields (substitute prices for yields in the expected return). I obtain:

\[ r_{pr}^{(τ)}(t) = \left[ y^{(τ)}(t) - y^{(1)}(t) \right] - (τ - 1)E_t \left[ y^{(τ-1)}(t+1) - y^{(1)}(t) \right]. \]

This expression is derived solely from the definition of yields and returns. Thus far no assumption has been taken on the form or dynamics of expected excess returns. Rearrange to express the expected changes in future interest rate in terms of the spread and the excess returns:

\[ E_t \left[ y^{(τ-1)}(t+1) - y^{(1)}(t) \right] = \frac{1}{(τ - 1)} \left[ y^{(τ)}(t) - y^{(1)}(t) \right] - \frac{1}{(τ - 1)} r_{pr}^{(τ)}(t), \]

and recall that the last term is the expected excess return. If return risk premia is constant (zero) then we call this equation the (Pure) Expectations Hypothesis.\(^{15}\) To test the model, Campbell and Shiller (1991) suggest regressing the changes in yields onto the spread and test if the coefficient \( \beta_τ \) equals one in the following equation:

\[ y^{(τ)}_{t+1} - y^{(1)}(t) = \text{constant} + \beta(τ) \frac{1}{(τ - 1)} \left[ y^{(τ)}(t) - y^{(1)}(t) \right] + \text{error}. \]

More generally for holding period \( j \) the regression is

\[ y^{(τ-j)}_{t+j} - y^{(τ)}(t) = \alpha(τ, j) + \beta(τ, j) \cdot \frac{j}{(τ - j)} \left[ y^{(τ)}(t) - y^{(j)}(t) \right] + \varepsilon^{(τ)}_{t+j}, \]

which makes explicit that the constant in the regression is a function of maturity and the holding period. As stressed by Campbell and Shiller (1991) “apart from testing the model [for constant risk premia], we can evaluate its usefulness by checking to what extent the spread resembles an optimal forecast of the changes in interest rates.”

With the estimated coefficients of the second order approximation of the marginal utility and the coefficients of the dynamics of the state, I generate samples of artificial yield data. With these data I run the forecasting regressions of Campbell and Shiller (1991) on each sample and compute the histograms of the slope coefficient. The distributions of these coefficients fall around the mean of the coefficients estimated in the UK real yield data. However, the dispersion of the coefficients of the artificial data is larger than the standard deviations of the coefficients estimated for the UK yield data. These results suggest that the estimated term structure model generates the variation in risk premia observed in the data.

I run the forecasting regression for each one of the 1,000 samples. The histograms for the regression coefficient \( \beta(τ, j) \) for four sets of regressions are presented in Figure (6.2.2). The first panel shows the regression for a holding period of 24-months and the second panel shows the regression for a holding period of 36-months. For each holding period I focus on the bond at 48- and 120-months. The

\(^{15}\)The appendix generalizes this equation to longer holding periods.
Figure 14: Histograms of the forecasting regressions in bootstrapped yield data. Left panel: regressions on 24-month holding periods for bonds of 48- and 120-months. Right panel: regressions on 36-month holding periods for bonds of 48- and 120-months. The histograms show the distribution of the coefficients in the bootstrapped data. The coefficients are significantly different from one therefore providing strong evidence against the expectations hypothesis in real yields.

Histograms show the distribution of the coefficients in the bootstrapped data. The coefficients are significantly different from one therefore providing strong evidence against the expectations hypothesis in real yields. The strength of the rejection is surprising but this may be due to the fact that the short end of the yields where not used in the estimation of the EKF thus I have no estimate for their measurement error. Therefore, the short end of the yields display a smaller variance than the long end strengthening the rejection of the EH. The regression in the data shows a smaller coefficient between 0 and -1.

7 Concluding Remarks

This paper has two contributions. First it highlights the importance of limited asset market participation for macroeconomic volatility. The segmentation that arises from transaction costs can account for the variability of real bond risk premia in the UK data. I derive the pricing implications for real bonds in the model of segmented asset markets. The second contribution is to perform a structural estimation of a non-linear term structure model. I estimate the parameters of the marginal utility of active households in this model using the cross-equation restrictions from quadratic term structure model. I approach the estimation as a filtering problem and back out the implied state for the UK real yields data from 1985 to 2007. Given that the model is non-linear, the filtering procedure is an extension of the traditional Kalman Filter. Therefore, I obtain quasi-maximum likelihood estimates
of the structural parameters of the model of segmented asset markets.

I use the estimated state variable to derive implications in the time series of the market segmentation and consumption of the marginal agents. The implied variability of consumption is less that 1%; asset market segmentation varies less than 0.5%. The pricing kernel is highly volatile and showcases time-varying conditional variances. I decompose the variability of consumption of active households, and show that most of the variability in the early part of the sample is due to GDP shocks while in the latter part of the sample is due to money shocks. This is consistent with the prediction of the model of segmented asset markets: when money growth is large, the variability of consumption of active households is small with respect to money. Moreover, I show how the filtered state matches roughly its empirical counterpart: the correlation of the filtered state and M2 is 0.31. Finally, I show how the model accounts for the dynamics of risk premia. Using the parameter estimates I generate samples of artificial data and I run the forecasting regressions of Campbell and Shiller (1991). I prove the rejection of the log pure expectations hypothesis as shown in the data. The source of time-varying risk premia is the heteroskedasticity of the pricing kernel. This in turn varies over time because volatility of consumption of active households changes over time with money shocks. Times of high inflation help forecastability and reduce risk premia because risk of asset market segmentation is low.

References


Appendix A: Computation of Bond Prices in the Segmented Asset Market Model

Prices

Using the independence between the money growth and the endowment shocks, bond prices can be computed in two separate parts:

\[ P^*_t = \beta^* \times PQ^*_t(\hat{\mu}) \times PL^*_t(\hat{\omega}) \]

Let the quadratic part be:

\[ PQ^*_t(\hat{\mu}) = \mathbb{E}_t \left[ \exp \left\{ -\phi^\tau + \frac{\eta}{2} \hat{\mu}^2 + \eta \hat{\mu} \right\} \right] \]

and the linear part be:

\[ PL^*_t(\hat{\omega}) = \mathbb{E}_t \left[ \exp \left\{ -\psi \log \omega t \right\} \right] \]

I solve for each part below.

Conditional Expectation of a Quadratic Random Variable

I need compute the conditional expectation of \( \exp(ax + bx^2) \). I assume normal innovations, \( x \sim N(\kappa, \varsigma^2) \), then use the definition of the normal density, collect terms and complete the quadratic form:

\[
\mathbb{E} \exp(ax + bx^2) = \frac{1}{\varsigma \sqrt{2\pi}} \int \exp \left\{ ax + bx^2 \right\} \exp \left\{ -\frac{(x - \kappa)^2}{2\varsigma^2} \right\} dx
\]

\[
= \frac{1}{\varsigma \sqrt{2\pi}} \int \exp \left\{ \frac{1}{2\varsigma^2} \left( -x^2 (1 - 2\varsigma^2 b) + 2x \left( \varsigma^2 a + \kappa \right) - \kappa^2 \right) \right\} dx
\]

\[
= \exp \left( -\frac{\kappa^2}{2\varsigma^2} + \frac{(\varsigma^2 a + \kappa)^2}{2\varsigma^2 (1 - 2\varsigma^2 b)} \right)
\]

and define the variable \( z = (x - m)/s \) with \( m = (\varsigma^2 a + \kappa) / (1 - 2\varsigma^2 b) \) and \( s = \varsigma (1 - 2\varsigma^2 b)^{-1/2} \), and note \( z \sim N(0, 1) \). Then we can write:

\[
\mathbb{E} \exp(ax + bx^2) = \exp \left( -\frac{\kappa^2}{2\varsigma^2} + \frac{(\varsigma^2 a + \kappa)^2}{2\varsigma^2 (1 - 2\varsigma^2 b)} \right) \frac{1}{\varsigma \sqrt{2\pi}} \int \exp \left\{ -\frac{z^2}{2} \right\} dz
\]

\[
= \exp \left( -\frac{\kappa^2}{2\varsigma^2} + \frac{(\varsigma^2 a + \kappa)^2}{2\varsigma^2 (1 - 2\varsigma^2 b)} \right) (1 - 2\varsigma^2 b)^{-1/2}.
\]
I iterate forward the AR(1) process for endowment growth. Start with a simple example:

\[ \omega_{t+1} + \delta = \log((\omega_{t+1}/\omega_t)) \]

Therefore, rewrite the bond price in terms of the autoregressive process:

\[ P_t \exp(ax_{t+\tau} + bx_{t+\tau}^2) = \exp \left( -\frac{\kappa(t, \tau)^2}{2\varsigma^2(\tau)} + \frac{(a\varsigma^2(\tau) + \kappa(\tau))^2}{2\varsigma^2(\tau)(1 - 2b\varsigma^2(\tau))} \right) (1 - 2b\varsigma^2(\tau))^{-1/2}. \]

Note that so far the only assumption on the process is conditional normality.

Use this result setting \( a = -\phi \) and \( b = \eta/2 \), then:

\[ PQ_t^\tau = (1 - \eta\varsigma^2(\tau))^{-1/2} \cdot \exp \left\{ \frac{1}{2\varsigma^2(\tau)} \left[ \frac{(\kappa(t, \tau) - \phi\varsigma^2(\tau))^2}{(1 - \eta\varsigma^2(\tau))} - \kappa^2(t, \tau) \right] + \phi \hat{\mu}_t - \frac{\eta}{2} \hat{\mu}_t^2 \right\}. \]

Specializing for \( \hat{\mu}_t \) autoregressive:

\[ \hat{\mu}_t = \rho \hat{\mu}_{t-1} + \varepsilon_t, \]

with \( \varepsilon_t \sim N(0, \sigma^2) \) i.i.d. and \( \rho < 1 \), the conditional mean and variance for this process are \( \kappa(t, \tau) = \mathbb{E}_t[\hat{\mu}_{t+\tau}] = \rho^\tau \hat{\mu}_t \) and \( \varsigma^2(\tau) = \sigma^2(1 - \rho^2)/(1 - \rho^2) \). Substitute for the conditional mean, develop the quadratic term and rearrange and express in terms of the degree of the state variable:

\[ PQ_t^\tau = (1 - \eta\varsigma^2(\tau))^{-1/2} \cdot \exp \left\{ -\frac{\eta}{2} \hat{\mu}_t^2 \left( 1 - \frac{\rho^\tau}{1 - \eta\varsigma^2(\tau)} \right) + \phi \hat{\mu}_t \left( 1 - \frac{\rho^\tau}{1 - \eta\varsigma^2(\tau)} \right) + \frac{\phi^2\varsigma^2(\tau)}{2(1 - \eta\varsigma^2(\tau))} \right\}. \]

which is the equation in the text.

**Conditional Expectation of the Linear Variable**

I took a stand on the process of the growth rate of endowment. Therefore, rewrite the bond price in terms of \( \hat{\omega}_{t+1} + \delta = \log((\omega_{t+1}/\omega_t)) \):

\[ PL_t^\tau = \mathbb{E}_t[\exp\left\{ -\psi (\log(\omega_{t+\tau} + \log(\omega_t)) \right\}] \]

\[ = \mathbb{E}_t[\exp\left\{ -\psi (\log(\omega_{t+\tau} - \log(\omega_{t+\tau-1} + \log(\omega_{t+\tau-1} + \ldots + \log(\omega_1)) \right\}] \]

\[ = \mathbb{E}_t[\exp\left\{ -\psi ((\hat{\omega}_{t+\tau} + \delta) + (\hat{\omega}_{t+\tau-1} + \delta) + \ldots + (\hat{\omega}_{t+1} + \delta)) \right\}] \]

I iterate forward the AR(1) process for endowment growth. Start with a simple example:

\[ \hat{\omega}_{t+1} + \delta = (1 - \theta) \delta + (\hat{\omega}_t + \delta) + \nu_{t+1} \]

\[ (\hat{\omega}_{t+2} + \delta) + (\hat{\omega}_{t+1} + \delta) = (\theta + \theta^2) \hat{\omega}_t + \delta + (1 + \theta) \nu_{t+1} + \nu_{t+2} \]

\[ + (1 + \theta) \nu_{t+2} + (1 + \theta) \nu_{t+3} + (1 - \theta) \delta \]

\[ (\hat{\omega}_{t+3} + \delta) + (\hat{\omega}_{t+2} + \delta) + (\hat{\omega}_{t+1} + \delta) = (\theta + \theta^2 + \theta^3) \hat{\omega}_t + \delta + (1 + \theta + \theta^2) \nu_{t+1} + (1 + \theta) \nu_{t+2} + \nu_{t+3} \]

\[ + [(1 - \theta) \delta](1 + \theta + \theta^2) + [(1 - \theta) \delta](1 + \theta) + [(1 - \theta) \delta], \]
and in general:
\[
\sum_{j=1}^{\tau} (\omega_t + j + \delta) = \omega_t \left( 1 - \theta^\tau \right) + \sum_{j=1}^{\tau} (\nu_t + \tau + j + 1 + (1 - \theta) \delta) \sum_{k=1}^{j} \theta^{j-k}.
\]

This summation is distributed normal because all innovations are independent from each other. Then use the well known result of log normal random variables \(\mathbb{E}e^x = \exp \{ \mu_x + \sigma_x^2/2 \}\), to get:
\[
PL_t^{(\tau)} = \mathbb{E}_t \left[ \exp \left\{ -\psi \left( \frac{\omega_t (1 - \theta^\tau)}{1 - \theta} + \sum_{j=1}^{\tau} \frac{1 - \theta^{\tau-j+1}}{1 - \theta} (\nu_t + j + 1 + (1 - \theta) \delta) \right) \right\} \right]
= \exp \left\{ -\psi \left( \frac{\theta (1 - \theta^\tau)}{1 - \theta} + \delta \sum_{j=1}^{\tau} (1 - \theta^{\tau-j+1}) \right) \right\} \exp \left\{ \frac{\psi^2 \sigma^2}{2} \sum_{j=1}^{\tau} \left( \sum_{k=1}^{j} \theta^{j-k} \right)^2 \right\}.
\]

**Risk Premia**

Start defining the one-period log holding period return:
\[
r_{t+1}^{(\tau)} = \log P_{t+1}^{(\tau-1)} - \log P_t^{(\tau)}.
\]

Now define real risk premia as:
\[
rpr_t^{(\tau)} = \mathbb{E}_t [r_{t+1}^{(\tau)}] - y_t^{(1)},
\]
which measures the real risk of an investment strategy of a one-period holding return funded by short selling the real short rate. In terms of the model:
\[
rpr_t^{(\tau)} = \tilde{a}(\tau - 1) - \tilde{a}(\tau) - a(1) + \tilde{q}(\tau - 1) \sigma_\epsilon^2
+ \left[ \tilde{b}(\tau - 1) \rho - \tilde{b}(\tau) - b(1) \right] \mu_t
+ \left[ \tilde{q}(\tau - 1) \rho^2 - \tilde{q}(\tau) - q(1) \right] \mu_t^2.
\]

Extending to holding periods of arbitrary length \(j\) let:
\[
r_{t+1}^{(\tau,j)} \equiv \log P_{t+1}^{(\tau-j)} - \log P_t^{(\tau)},
\]

where \(j\) identifies the length of the investment. Now define real risk premia at maturity \(\tau\) for holding period of length \(j\) as:
\[
rpr_t^{(\tau,j)} \equiv \mathbb{E}_t [r_{t+1}^{(\tau,j)}] - y_t^{(j)},
= \tilde{a}(\tau - j) - \tilde{a}(\tau) - a(j) + \tilde{q}(\tau - j) \sigma_\epsilon^2(j)
+ \left[ \tilde{b}(\tau - 1) j \rho^j - \tilde{b}(\tau) - b(j) \right] \mu_t
+ \left[ \tilde{q}(\tau - j) \rho^{2j} - \tilde{q}(\tau) - q(j) \right] \mu_t^2.
\]
Appendix B: Extended Kalman Filter

This section derives the EKF in detail. For further reference see Anderson and Moore (1979). Given the assumptions in the observation and state equations, the recursive predict and update stages are initiated with some prior values for the state and covariance of the state innovations. The state is predicted using the state equation and used to evaluate the observation equation. Mean square errors in the observation equation are used to update the prediction of the state and compute the next period predicted state. The difference with the traditional Kalman filter is that the mean square error used in the update stage is the linearized observation equation evaluated at the predicted state.

Several problems have been identified in the control literature with the EKF (Julier and Uhlmann, 1997) because the system has to be almost linear on the time dimension, otherwise the estimates may diverge quickly when iterating the filter. A traditional sign of this problem is the difficulty computing the inverse of the Jacobian matrix in the likelihood function. Also the estimates may not be efficient nor consistent but can be corrected using the methodology suggested by Gourieroux, Monfort, and Renault (1993). To ensure reasonable starting values of $\rho$, $a$, $b$, and $q$ for the numerical optimization procedure, I estimate the model via GMM using 10 moment conditions derived from the segmented asset market model. A similar approach is taken by Leippold and Wu (2003) who estimate a 2-factor QTSM using several moment conditions to match some stylized facts in their data. The moment conditions I use are the unconditional mean and standard deviation of a short and long bond, the 1- and 24- month autocorrelations of a short and long bonds and the return risk premia on a 2-year holding period return. I use these estimates as starting values.

The algorithm of the EKF is reproduced in the following equations:

1. Start recursion. The initial value of the state is:
   \[
   \hat{X}_{1|0} = X_0,
   \]
   and of the MSE of the forecast of the state is given by:
   \[
   P_{1|0} = \sigma^2_{\epsilon}.
   \]

2. Update state. With these starting values, the state is updated:
   \[
   \hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t \Psi_t
   \]
   where $\Psi_t = y_t - \hat{y}_t$, and
   \[
   \hat{y}_t = a(\tau) + b(\tau)\hat{X}_{t|t-1} + q(\tau)\hat{X}^2_{t|t-1},
   \]
   is the vector of errors in the prediction of the observed variables using the fitted yields. The Kalman gain is defined in terms of the linearized equation
   \[
   K_t = P_{t|t-1}H_t(H'_tP_{t|t-1}H_t + R)^{-1}
   \]
where
\[
H_t = \left. \frac{\partial y(X_t)}{\partial X} \right|_{X_t = \hat{X}_{t|t-1}}.
\]

3. Update the state forecast’s MSE:
\[
P_t = P_{t|t-1} - K_t H_t' P_{t|t-1}.
\]

4. In the forecast stage, use the equation of the dynamics of the state and compute:
\[
\hat{X}_{t+1|t} = \rho \hat{X}_{t|t}.
\]

The mean square error associated with this forecast is
\[
P_{t+1|t} = \rho^2 P_{t|t} + \sigma^2 \varepsilon.
\]

This recursion is used to obtain a filtered time series \( \hat{X} \). The log likelihood function is defined as
\[
\mathcal{L} = \sum_{t=1}^{T} \frac{1}{2} \left( \log |H_t' P_{t|t-1} H_t + \mathbf{R}| + \Psi_t' (H_t' P_{t|t-1} H_t + \mathbf{R})^{-1} \Psi_t \right).
\]

With the likelihood function in hand, the numerical procedure optimizes the parameters given the implied filtered state. The algorithm to maximize the likelihood function is the following. First I estimate all the parameters simultaneously without achieving convergence to get a first estimate of the state variable and its dynamics as well as the cross section of coefficients. Second, I check if the condition for existence of the solution of the expectation that defines bond process is satisfied. If this test fails, I restart the numerical procedure perturbing the initial value of \( \eta \) and \( \sigma \). The perturbation of the initial values is directed by the fitted mean term structure and volatility curves. It is clear that this procedure may be path dependent but I found no other procedure to achieve convergence and sensible fitted yield curves. This procedure also guarantees that the cross-section of coefficients is consistent with the pricing kernel.