Learning and the Yield Curve

Job Market Paper

Arunima Sinha[1]
Columbia University

November 21, 2009

Abstract

Two central implications of the Expectations Hypothesis under rational expectations are inconsistent with yield curve data: (i) future expected long yields fall, instead of rising, when the yield spread rises, and (ii) long yields are excessively volatile with respect to short yields. I document these puzzles in the U.S. and the U.K. data, for different sub-samples, and for both real and nominal yields. I then propose a micro-founded optimization framework in which boundedly rational agents use adaptive learning to form expectations. The model is successful on both dimensions. First, the belief structure rationalizes the pattern of yields observed in the data so that the first puzzle does not hold with subjective instead of rational expectations. In particular, intertemporal income and substitution effects are amplified relative to the rational expectations case, causing expected long yields to rise when the yield spread falls. The second puzzle is partly accounted for by the extra volatility due to parameter uncertainty. These results suggest that it is the assumption of rational expectations that is at odds with the data, not the (subjective) Expectations Hypothesis. In addition, I find that: the model generates systematic forecast errors in yields and inflation that are consistent with survey data; higher yield volatilities during different monetary policy regimes match the data; and fiscal policy affects the yield spread because Ricardian Equivalence no longer holds.

[1] I am grateful to Bruce Preston, Ricardo Reis and Michael Woodford for their advice and support. Discussions with John Donaldson, Gauti Eggertsson, Stefano Eusepi, Jaromir Nosal, Argia Sbordone, Jón Steinsson and Andrea Tambalotti were extremely helpful, as were the comments of seminar participants at Columbia University and the Federal Reserve Bank of New York. A part of this paper was completed while I was visiting the New York Fed, and their hospitality is gratefully acknowledged. All errors are mine. An updated version of this paper will be available at http://www.columbia.edu/~as2801/research/jmpaper.pdf. Email: as2801@columbia.edu.

1 Introduction

"To preserve the theoretical relationship between long term and future short term interest rates, the 'yields' of bonds of the highest grades should fall during a period in which short term rates are higher than yields on bonds and rise during a period in which short term rates are lower. Now experience is more nearly the opposite."

- Frederick R. Macaulay, 1938

The Expectations Hypothesis is the classic theory used to explain the term structure of yields. It states that long yields are an average of expected future short yields.[2] The implications of the Expectations Hypothesis for the yield spread[3] and the volatility of long yields relative to short yields have been the focus of much of the term structure literature, as they are found to be at odds with the empirical data on the nominal yield curves.

With respect to the yield spread, the Hypothesis implies that the when the spread is rising, expected future long yields must rise as well. This has been extensively tested using the Campbell-Shiller (1991) regression:

$$y_{n-1,t+1} - y_{n,t} = \alpha + \frac{\gamma}{n-1}\left[y_{n,t} - y_{1,t}\right] + \varepsilon_t. \qquad (1)$$

Assuming rational expectations are used to construct expected future yields and the Expectations Hypothesis holds, then when the yield spread is high, the yield on the long bond should rise. In the context of (1), the slope coefficient γ should not be different from one. However, it has been established using empirical data for the nominal yield curves that long yields fall when the yield spread is high. That is, γ is found to be negative. As Macaulay's (1938) quote indicates, this is not simply a feature of yields in recent data. For yield variances, the Hypothesis implies that long yields must be less volatile than shorter yields.

In this paper, I explore the following question: can the pattern of yields, at odds with the above implications of the Expectations Hypothesis, be rationalized by a model in which the beliefs structure of optimizing agents is boundedly rational? To answer it, I use a framework in which expectations are formed using a constant-gain adaptive learning process in a micro-founded model and the Expectations Hypothesis holds (given the subjective beliefs of optimizing agents) but the assumption of rational expectations does not. Agents understand their decision problem, but they must make forecasts of aggregate variables which are exogenous to their decisions by using a probability distribution different from the distribution implied by the true model. The use of adaptive learning to explore these empirical anomalies is motivated by the hypothesis that expectations of optimizing agents do not, in fact, correspond to rational expectations. Froot (1989)

[2] This is the log pure form of the Expectations Hypothesis.

[3] $y_{n,t}$ is the yield on an n-period zero-coupon bond at time t. The yield spread is the difference between the long and short yields: $y_{n,t} - y_{1,t}$.

uses survey data on yields to argue that the rejections of the Expectations Hypothesis may be due to the failure of rational expectations, not the Hypothesis itself. Piazzesi and Schneider (2009) show that expected excess returns can be partly accounted for by subjective expectations, and not just time-varying risk premia.

I find that the model with adaptive learning generates slope coefficients in (1) that are less than one at shorter maturities, and negative at longer maturities (for both the real and nominal yields), as found in the data. This is because the yields in the regression in (1) are constructed using time-varying beliefs from the learning model and the least squares regression is misspecified. The result suggests that it is the assumption of rational expectations, not the Expectations Hypothesis, that is rejected. This is in contrast to other empirical investigations of the Campbell-Shiller regression in the literature, which presume agents have rational expectations, and use ordinary least squares to test the Expectations Hypothesis (which is consistent with the assumption of rational expectations). Then the negative γ in (1) is interpreted as a rejection of the Hypothesis itself.

The adaptive learning framework also generates long yields that are more volatile relative to the short yields, compared to the rational expectations analog. The level of yield volatilities are also higher across the maturity structure, and closer to the levels observed in the data due to parameter uncertainty.

Model and Intuition This paper uses a micro-founded New-Keynesian optimization framework, consisting of households, firms and a monetary authority, with adaptive learning as a theory of expectations formation, to characterize the term structure.

In the adaptive learning model, optimizing households and firms run a vector autoregression of the relevant variables on observed past data to form their conditional expectations. They update their beliefs and form new conditional expectations in successive time periods, revising their estimates in order to reduce the observed error between the forecasted variable and its realized value. This implies that the formation of beliefs is self-referential, and the paths of the asset yields generated by the learning model are more persistent and more volatile relative to those of the rational expectations model.

This has implications for both predictions of the Expectations Hypothesis discussed above. In reference to (1), as the expectations of short yields are formed using observed current yields, the forecast error is no longer orthogonal to the yield spread, and the slope coefficient is biased downwards from one. That is, the use of ordinary least squares by the econometrician is misspecified in tests of the Expectations Hypothesis that use (1). Under the rational expectations case, which is nested in this framework, the ordinary least squares regression in (1) is valid, and the slope coefficient γ is insignificantly different from one.

The intuition for the negative bias in γ with respect to one can be explained as follows: when agents in the model observe the short yields rising, under rational expectations, they (correctly) expect the long yields to rise as well, albeit by less than the rise in short rates. As the agents use the correct probability distribution to make future forecasts of the one-period yield, they do not expect a rise in the average expected yields for the infinite horizon decision problem. Thus, actual future yields rise relative to the current long yield.

However, with adaptive learning, when agents observe the rise in current yields, they attribute this to an increase in average yields in the infinite horizon. This amplifies the income and substitution effects of the increase in expected returns. Optimizing households choose to consume less and save more, demanding larger bond holdings, relative to the rational expectations substitution effect. They also forecast a fall in future income, and the income effect lowers both consumption and desired savings. The income effect dominates, and as the net supply of bonds is fixed, the one-period yield rises more relative to rational expectations under the learning model. Consequently, while the expected long yield falls under rational expectations, it rises under learning as the yield spread falls. This generates a negative bias in the slope coefficient γ.

The variance of the belief parameters in the yields processes contributes to the higher level of variances in the yields. Under rational expectations, without a time-varying risk premia term, yields variances are only functions of the variance of the state variables, and calibrated New Keynesian DSGE models are able to generate only a part of the excess volatility of long yields with respect to the short yields (and the level of yield variances are smaller than in the data). With learning, however, the variance of yields are also a function of the variability in beliefs. Not only does this increase the level of variability, it also implies a non-monotonic decline in variances across the maturity structure, so that the long yields are more volatile that they are under rational expectations. Therefore, learning can help to resolve a part of the excess volatility of long yields relative to short yields, as well as generate a higher level of yield variances, relative to the rational expectations model.

Roadmap Following a brief overview of the literature in section two, I document facts about the real and nominal yield curves, using U.S. and U.K. data in section three. Two data sets are used to evaluate the empirical results on the real term structures: U.S. Treasury-Inflation-Protected-Securities (TIPS) and the U.K. Index-Linked bonds. The pattern of the real term structures are mostly consistent across the two countries for different sampling periods: **(a)** expected future long yields decline as the yield spread rises, resulting in the negative bias (with respect to one) in γ for the Campbell-Shiller regression in (1)[4]; and **(b)** the decline in variances of yields across the maturity structure is similar to the prediction of the Expectations Hypothesis for the U.S., but not for the U.K.. For the nominal term structure for the U.S. and U.K., I find that **(a)** the γ coefficients for (1) are smaller than one, and the deviations get larger as the maturity horizon increases; and **(b)** long

[4] As the short two-year maturity in the U.S. TIPS series is available from 2004 only, I use the U.K. Index-Linked series to run the Campbell-Shiller regression.

yields are excessively volatile relative to short yields compared to the predictions of the standard DSGE model, and the level of yield variances is much smaller across the maturity structure.

The DSGE model, with adaptive learning is discussed in section four. I discuss the quantitative performance of this model with respect to the Campbell-Shiller regression and variance of yields, and develop intuition for the model mechanics in section five. Analytical results are derived in the limiting case of the model with flexible prices and an exogenous endowment process, and this may be interpreted as the Lucas (1978) model generalized to subjective expectations. For real yields, the learning model does well at qualitatively matching the rejections of the Campbell-Shiller test of the Expectations Hypothesis, as documented below for U.K. Index-Linked bonds. The slope coefficients are less than one, and become more negative as the forecast horizon increases. Parameter uncertainty increases the variance of the yields generated by the learning model. For the nominal yield curve, the slope coefficients of the Campbell-Shiller regression match the empirical findings reported below, and with higher gain (implying larger deviations from rational expectations), the point estimates of the slope coefficients are further away from one. The model is also able to explain a part of the excess volatility of long yields relative to short yields, as well as the level of yield variances across the maturity structure compared to the rational expectations analog of the benchmark model.

Since the constant gain learning model generates systematic forecasting errors, I compare these with the forecast errors obtained from survey data. I find that the autocorrelation of the model implied forecast errors for inflation and the one-quarter nominal interest rates are close to those found in surveys. Additionally, I examine the properties of the inflation and yields generated by the learning model, and find that they compare favorably to the data.

In section six, I analyze the implications of different monetary and fiscal regimes in the adaptive learning framework. A more aggressive response to inflation in the Taylor rule (which has been documented for the U.S. since the mid 1980s) is found to generate smaller deviations in γ with respect to one, and increases the variances of yields. A fiscal authority which issues riskless debt in non-zero supply is introduced to analyze the impact of a deficit shock on yields. The wealth effects generated due to the violation of Ricardian equivalence result in a rise in the short yield in the model, and a fall in the yield spread in response to a positive deficit shock. Section seven concludes.

2 Connections to Existing Literature

There is some consensus in existing literature that the negative slope coefficient in (1) is due to the omission of a time-varying risk premia term from the regression. Wachter (2006) uses habit formation preferences, as in Campbell and Cochrane (1999), to show that the significant time-varying premia generated can explain the negative bias in γ. Affine models with observable and latent factors are also used for characterizing the term structure, introducing the required time variation in the covariance risk premia term. Dai and Singleton (2002) show that a statistical model of the stochastic discount factor can explain the puzzle, and several theories have been proposed to capture the corresponding economic mechanism. Seppälä (2003) shows that limited risk sharing due to risk of default generates enough time variation in risk premia to account for rejections of the expectations hypothesis for the real term structure.

Risk based explanations of term structure anomalies have also been explored in DSGE frameworks. Rudebusch and Swanson (2009) introduce Epstein-Zin preferences in a DSGE model, and find that it can generate empirically consistent results with respect to the term premium, while fitting other macroeconomic variables such as inflation also. Hördahl, Tristani and Vestin (2007) use a DSGE model, with an affine term structure model to characterize the yield curve.

The main difference in the approach pursued in the present paper is that I analyze the predictions of the New Keynesian model using adaptive learning as a means of explaining the term structure anomalies considered. I abstract from the risk premia based explanation of the anomalies - the purpose of this exercise is not to offer a single, learning based explanation to characterize all properties of the term structure that are not in accordance with observed data. Rather, it offers an alternative approach to the risk premia explanation for the two implications of the Hypothesis explored here by using an optimization model to specify a theory of expectations formation that will help explain the anomalies to a quantitatively plausible degree, without making extreme assumptions about parameter values such as the risk aversion coefficient. This has been a limitation of a purely risk based approach. For instance, when habit formation preferences are used to generate a term premia, they are much less successful in a DSGE framework as extremely large movements in prices or wages are required to generate consumption volatility that is consistent with the data. Rudebusch and Swanson (2008) show that introducing habits in a DSGE model to match the term premia in the data, diminish the model's ability to match other macroeconomic variables such as inflation and the short-term nominal interest rates.

A growing literature uses learning to address empirical anomalies in the yield curve. Dewachter and Lyrio (2006) show that a factor model with only observable factors[5] can explain the long yields as well as the behavior of the yield spread when a stochastic unobserved trend is incorporated in the observables, and must be extracted using a Kalman filter. Laubach, Tetlow and Williams (2007) use a factor model with only observables to explain the yield curve. These models have limited success at explaining the long end of the yield curve with a constant coefficient VAR. Instead, the authors use a discounted recursive learning scheme, and find that it significantly improves the performance of the model with respect to longer yields. The main difference with this line of the literature is that the general equilibrium framework used here implies that the processes for inflation and output gap

[5] The macro-factor models used in the literature, such as Ang and Piazzesi (2003), require latent factors to explain the long end of the yield curve.

need not be exogenous, as assumed in factor models of the term structure. The results obtained in the framework used here indicate that the assumption of rational expectations is rejected, and adaptive learning is an empirically consistent way to describe the formation of expectations by optimizing agents.

A number of papers also address other asset pricing puzzles in the context of learning models. Evans and Chakraborty (2007) illustrate that the forward premium puzzle can resolved using an adaptive learning model, where the agents form a one-period ahead forecast of future consumption using the Euler equation. Adam, Marcet and Nicolini (2007) use a non-linear learning scheme to explain the excess volatility in the price-dividend ratio. Carceles-Poveda and Giannitsarou (2007, 2008) evaluate the implications of adaptive learning in an endowment economy, as well as a production model with capital. They find that with production, the results of a decreasing gain learning algorithm are quantitatively not very different from the rational expectations case.

Non-rational expectations based analyses are also being increasingly used in DSGE frameworks to address a wide range of issues. Milani (2005) finds that when the New Keynesian DSGE model is augmented using learning, assumptions such as inflation indexation and habit persistence are no longer required to match persistence of aggregate output and inflation. Eusepi and Preston (2008a) consider the constraints placed on stabilization policy in a New Keynesian framework with monetary and fiscal policies when agents form expectations using adaptive learning. The authors also consider a real business cycle model with adaptive learning in Eusepi and Preston (2008b), and find that the model implied persistence in the output, hours and investment growth matches U.S. data better than the rational expectations analog.

3 Empirical Estimates

In section 3.1 below, I discuss the data on the U.S. and U.K. real and nominal yield curves. The performance of the yield curves with respect to the implications of the Expectations Hypothesis is shown in section 3.2.

3.1 Yield Curves

3.1.1 Real Yield Curves

Treasury-Inflation-Protected-Securities (TIPS) were first issued in the U.S. in 1997, and currently constitute approximately 10% of the outstanding Treasury debt. In contrast to nominal debt securities, the coupon and principal payments for these debt securities are indexed to the Consumer Price Index (CPI) in the following manner: the principal or coupon payment is multiplied with the reference CPI on the date of maturity to the reference CPI on the date of issue.[6] When the ratio of CPIs is less than one, no adjustment is made. There is an indexation lag of approximately 2.5 months on TIPS because the Bureau of Labor Statistics publishes the CPI data with a lag - the index of a given month is released in the middle of the next month. At present, there are 24 outstanding TIPS with maturity dates ranging from 2010 to 2034. The estimates of the TIPS yield curve are taken from the data available at the Federal Reserve Board website, which was constructed for the work by Gürkaynak, Sack and Wright (2007a). This dataset is updated daily.

The data on zero-coupon yields needs to be constructed, and cannot be derived simply from the existing yields, as the Treasury issues a limited number of securities with different maturities and coupons. In this case, the maturity structure of the zero-coupon yields must be estimated. Gürkaynak et al. use a parametric yield curve specification, the Nelson-Siegel-Svensson functional form, which is found the fit the data very well. Under this parameterization, a forward rate $f_t(n,0)$,[7] can be expressed as:

$$f_t(n,0) = \beta_0 + \beta_1 e^{-n/\tau_1} + \beta_2 \frac{n}{\tau_1} e^{-n/\tau_1} + \beta_3 \frac{n}{\tau_2} e^{-n/\tau_2}. \tag{2}$$

Here, the instantaneous forward rate starts at the zero time horizon at $(\beta_0+\beta_1)$ and asymptotically approaches β_0. The estimation of the yield curve requires that the parameters τ_1 and τ_2 be picked so as to minimize the weighted[8] sum of squared deviations between actual prices of Treasury securities and predicted prices.

For the U.S. TIPS data, the shortest maturity available in the Gürkaynak et. al. estimation is the two year maturity. Additionally, the shorter maturities (2, 3 and 4 years) are only available from January 2004. As the data sample is short, I do not consider any subsamples.

Since the TIPS data set is relatively short, I also use the U.K. Index-Linked bonds to check the robustness of the facts relating to the real yield curves. Estimates of the real yield curve in the U.K. are derived using data on the Index-Linked bonds and are available from the Bank of England. The estimates are available from 1984 onwards, and are constructed using prices of the traded securities using a spline method. This is different from parametric estimation - forward rates are modelled as piecewise cubic polynomials, with segments joined at knot-points.[9] The inflation

[6] If the maturity or issue date falls on day d_t of a month with d_n days, then the reference CPI is:

$$CPI(-2)\frac{d_t-1}{d_n} + CPI(-3)\frac{d_n-d_t+1}{d_n}$$

where $CPI(-2)$ and $CPI(-3)$ are the non seasonally adjusted U.S. City Average All Items CPI for the second and third months prior to the month in which the maturity or issue date falls respectively.

[7] A yield $y_{n,t}$ can be equivalently expressed in the form of a forward rate $f_t(n,0)$. This is the yield that an investor agrees to at time t to make an investment for a specified period in the future, in this case, zero periods.

[8] The weights are the inverse of the duration of the individual securities.

[9] The main difference between parametric and spline estimation is that when there is a change in data at any point of the yield curve, the entire curve will be affected under parametric estimation, but not under the spline method.

indexed yields are adjusted for inflation lag and seasonality. Figures 1(a) and 1(b) show the U.S. TIPS yields, along with the slope of the curve. The real yield curve is upward sloping for the U.S. data, for the entire sample period.

For the U.K. Index-Linked bonds, the data set I use consists of maturities 2.5 to 20 years. The entire sample period for which the data is available is January 1985 to the present. I also consider two subsample periods: from January 1985 to October 1992, and from November 1992 to the present. This break is meant to approximate the change in the yield curves that occurred after the U.K. exited from the Exchange Rate Management (ERM) in September 1992. This change in yield curve behavior is a central feature of Pasaogullari and Tsonev (2009). The yields for the Index-Linked bonds in the U.K. are shown in figure 1(c), and the slopes are shown in figure 1(d). The short and long yields along with the yield spread in figure 1(c) show that there are periods in which the spread is falling, even as the long yield is rising. For the entire sample, January 1985 to June 2009, the real yield curve is positively sloped. However, while for the first subsample the slope continues to be positive, for the second sample, October 1992 to June 2009, the yield curve is slightly negatively sloped.

3.1.2 Nominal Yield Curves

I obtain estimates of the U.S. nominal yield curve using the yields constructed by Gürkaynak, Sack and Wright (2007b). The Nelson-Siegel-Svensson parametric estimation is used here as well. This data set uses yields available from the U.S. Treasury, and is updated daily. As the U.S. data on the nominal term structure is a long enough data sample, I use only the estimated term structure data to establish the properties of the nominal yield curve. The sample period is January 1972 to the present. I also check the robustness of the facts by considering two subsamples, January 1972 to December 1979; and January 1984 to the present. The maturity structure estimated for one to ten years is used to establish the following results. Figure 2(a) shows the plots of the short yield (one-year), the ten-year yield and the spread. It can be seen that the longer rate moves more slowly and persistently with respect to the shorter yield. Figure 2(b) shows that the nominal yield curve is upward sloping and this fact is robust across the different subsamples.

The estimates of the nominal yield curve for the U.K. are obtained from the Bank of England. Figures 2(c) and 2(d) show the yields for the U.K. for the period 1972 to the present.

3.2 Data Analysis

I analyze the data first for the real and then the nominal yield curves. I am specifically interested, as mentioned above, in the implications of the log pure Expectations Hypothesis for term structure, which states that the yield on longer maturity bonds can be determined entirely using the expected future yields on shorter bonds, over the maturity of the long bond:

$$y_{n,t} = \frac{1}{n}\left[y_{1,t} + E_t y_{1,t+1} + E_t y_{1,t+2} + ... + E_t y_{1,t+(n-1)}\right]. \tag{3}$$

Thus, the Hypothesis has predictions both for the slope of the term structure as well as the change in the slope: if the Hypothesis is not rejected, then the slope of the yield curve will be flat. Additionally, a steeper slope of the yield curve implies that the long rate will be rising.

With respect to variances, the Hypothesis implies that the variance of the long yields should be smaller than the variance of the shorter yields.

3.2.1 Performance of the Real Yield Curves

Campbell-Shiller Regression As the TIPS data series is very short, and the shortest maturity yield (the two year) is only available from 2004, I illustrate the slope coefficients of the Campbell - Shiller regression in (1) only using the U.K. Index-Linked bonds.[10] The regression in (1) is constructed using the shortest maturity available for the U.K. data, the Index-Linked bond for 2.5 years.

As shown in figure 3, I find that for the full sample period, while the point estimates are smaller than one, they are not statistically different from one. This is also the case for the first sub-sample I consider, between January 1985 and September 1992. It is only for the second sample, from October 1992 to June 2009, that a familiar pattern is observed in the slope coefficients - the point estimates are smaller than one, declining across maturities, and are statistically different from one.

Variances The empirical yield curves for the U.S. show the same qualitative pattern (figure 4(a)) as predicted by the Expectations Hypothesis. For the U.K. data, as shown in figure 4(b), only the first sample (January 1985 to September 1992) matches the predictions of the Hypothesis qualitatively. For the full sample period, as well as for the second sample period, the variances of the longer yields are *higher* than the shorter yields.

3.2.2 Performance of the Nominal Yield Curves

Campbell-Shiller Regression The slope coefficients of the Campbell-Shiller (1991) regression are reported in figure 5(a) for the U.S. nominal yield curve. The short rate used is the three-month Treasury bill rate for the period under consideration. These are all statistically different from one at conventional levels of significance. Figure 5(b) shows the regression coefficients for U.K. nominal yields.

[10] Since the shortest maturity of 2 years is only available from 2004 for TIPS data, and (1) requires the one-period to be 2 years, the data series is not long enough to construct the regression.

Variances The term structure of variances is shown in figure 6(a). The qualitative behavior matches the predictions of the Hypothesis - the variance of the longer yields is smaller than the variance of the shorter yields. The predictions of the DSGE rational expectations model for the variances are shown in figure 6(a). For U.S. data, the ratio of the standard deviation of the three year to the one year yield is 0.93. For the calibrated model, the ratio is 0.88.[11] Thus, yields in the data are slightly more volatile with respect to the DSGE model described in section three below. The other striking fact about variances of nominal yields is that the level of volatilities generated by the model are significantly smaller. For the full sample period, the level of variances are almost three times smaller than the those seen the data. For the U.K., as figure 6(b) shows, the variance of long yields is as large as the variance at the short end of the yield curve. Although the calibration of the DSGE model is for U.S. data, the variance of the rational expectations model is also shown in figure 6(b) to illustrate that U.K. nominal long yields are excessively volatile with respect to short yields.

4 Benchmark Model with Adaptive Learning

A continuum of households $i \in [0,1]$ consume a consumption index consisting of $k \in [0,1]$ products. They supply labor hours to k monopolistically competitive firms which face do not choose their optimal prices every period, following Calvo (1983). The monetary authority is assumed to follow the Taylor rule for specifying the short interest rate, responding to the output gap and inflation. The fiscal authority is assumed to issue riskless bonds of different maturities in zero net supply. The model is based on the cashless version of the DSGE model (Clarida, Gali and Gertler, 1999 and Woodford, 2003), and adaptive learning is introduced directly into the primitives of the model following Preston (2005).

4.1 Intertemporal Optimization by Agents

Agents know their own intertemporal decision problems and form conditional forecasts of the aggregate variables that are exogenous to their individual decisions.

Households The optimization problem of household i is:

$$\max_{\{C_t^i, B_{1,t}^i, B_{2,t}^i, \ldots, B_{n,t}^i\}} \tilde{E}_t \sum_{j=0}^{\infty} \beta^j \left(U(C_{t+j}^i; \xi_{t+j}) - \int_0^1 v(h_{t+j}^i(k); \xi_{t+j}) dk \right). \tag{4}$$

[11] For instance, these are in line with those estimated for the U.S. economy by Rabanal and Rubio-Ramirez (2005).

The consumption index, C_t^i is defined over the consumption of i over the k goods:

$$C_t^i = \left(\int_0^1 c_t^i(k)^{\frac{\theta-1}{\theta}} dk \right)^{\frac{\theta}{\theta-1}}, \tag{5}$$

where θ is the elasticity of substitution, and $c_t^i(k)$ denotes household i's consumption of good k. The aggregate preference shocks are denoted with ξ_t. The household supplies $h_t^i(k)$ hours of work to firm k, and obtains disutility $v(h)$ for doing so. The utility function U is concave and the disutility function v is convex.

The only difference with the standard maximization problem is that here $\tilde{E}_t$ is used to denote subjective expectations. Rational expectations will be denoted using E_t.

Rational expectations, in the sense of Muth [39], imply that the optimizing household's probability distribution over the relevant state variables coincides with the true model. That is, an implication of rational expectations is that, other than random variation, expectations are generated by the same process through which the actual variables are determined.

In this paper, the deviation from rational expectations is that probability distribution used by optimizing household to form conditional forecasts may differ from the distribution under the true model. That is, while $\tilde{E}_t$ constitutes a complete model that can be used to specify joint probabilities, and satisfies all the probability axioms, it will imply different joint probabilities than the rational expectations operator E_t. The specification of $\tilde{E}_t$ will be described in section 4.6 below. In the present framework, all households and firms (described below) are assumed to have the same subjective expectations $\tilde{E}_t$.[12]

Asset markets are incomplete and the households have access to n riskless bonds. Here $B_{1,t}^i$ denotes the net holdings of a bond of maturity one period at time t by household i. Each household optimally chooses its holdings of each n period bonds. As households do not own capital, wealth, denoted by $\tilde{W}_t^i$ can only be held in the form of these riskless bonds.

Using $P_{m,t}^B$ to denote the price of an $m-$period bond at time t (this is the bond that will mature in period $t+m$), the flow budget constraint of household i is:

$$\begin{aligned} C_t^i + \sum\nolimits_{m=1}^n P_{m,t}^B B_{m,t}^i &\leq Y_t^i + \tilde{W}_t^i; \\ \tilde{W}_{t+1}^i &= B_{1,t}^i + \sum\nolimits_{m=2}^n P_{m-1,t+1}^B B_{m,t}^i. \end{aligned} \tag{6}$$

Here P_t is the composite price index and Y_t^i is the nominal income of the household i:

$$P_t = \left(\int_0^1 p_t(k)^{1-\theta} dk \right)^{\frac{1}{1-\theta}}; \ P_t Y_t^i = W_t h_t^i + \int_0^1 \Pi_t(k) dk, \tag{7}$$

[12] The framework can be extended to allow for different subjective expectations among households, that is $\tilde{E}_t^i$.

where W_t is the competitive wage, and $\Pi_t(k)$ denotes the profits from k accruing to the household.

The No-Ponzi condition holds:

$$\lim_{j \longrightarrow \infty} \tilde{E}_t P^B_{1,t,t+j} \tilde{W}^i_{t+j+1} \geq 0, \tag{8}$$

where $P^B_{1,t,t+j} = \prod_{k=0}^{j} P^B_{1,t+k}$.

The optimization problem of household i is to choose $\{c^i_t(k), h^i_t(k), B^i_{m,t}\}$ for all k, m to maximize the present discounted sum of utilities subject to the constraints in (64) and (8), taking as given $\{p_{t+j}(k), w_{t+j}(k), \Pi_{t+j}(k), P^B_{m,t+j}, \xi_{t+j}\}$ for all j, for the subjective expectations operator $\tilde{E}_t$.

Firms There are k differentiated goods, and monopolistic competition among the firms which supply them. The production technology for output $y_t(k)$ of firm k is specified as:

$$y_t(k) = A_t f(h_t(k)), \tag{9}$$

where $A_t > 0$ is the time-dependent, exogenous technology shock, and f is an increasing, concave function. Labor $h_t(.)$ is the variable factor of production. Since the households are assumed to have homogenous beliefs, under monopolistic competition, the demand function faced by each firm k is identical, and the optimal price p^*_t chosen by each firm is the same. The demand function faced by a firm is:

$$y_t(k) = Y_t \left(\frac{p_t(k)}{P_t} \right)^{-\theta}, \tag{10}$$

where Y_t is an index of aggregate demand, and the individual firm's price $p_t(k)$ is determined taking Y_t and P_t as given. Following Calvo (1983), $0 < \alpha < 1$ of the good prices remain unchanged every period and the probability of new prices being set $(1-\alpha)$ every period is assumed to be time-independent. The aggregate price index, for optimal price p^*_t is:

$$P_t = \left[(1-\alpha) p_t^{*1-\theta} + \alpha P_{t-1}^{1-\theta} \right]^{\frac{1}{1-\theta}}. \tag{11}$$

The discount factor used by a firm for discounting future profits is given by:

$$Q_{t,t+j} = \beta^j \frac{P_t}{P_{t+j}} \frac{U_C(C_{t+j}; \xi_{t+j})}{U_C(C_t; \xi_t)}. \tag{12}$$

Here, the assumption of homogenous beliefs among households implies that there will be a symmetric equilibrium, in which all households will get an equal share of the profit and wage income. Then, the discount factor defined in terms of the marginal valuation of an additional unit of aggregate income by the households implies that in the approximation of the model, the average discount rate will be identical.

Finally, the optimization problem of the firm i is:

$$\max_{p_t(k)} \tilde{E}_t \sum\nolimits_{j=0}^{\infty} \alpha^j Q_{t,t+j}(\Pi^i_{t+j}(p_t(i))), \tag{13}$$

where the profit function is $\Pi^i_t(p) = p(Y_t P_t^\theta / p^\theta) - w_t(i) f^{-1}[p(Y_t P_t^\theta / p^\theta)/A_t]$.

In this framework, a firm's problem then is to choose $p_t(i)$ to maximize the discounted sum of profits, taking as given $\{Y_{t+j}, P_{t+j}, w_{t+j}(k), A_{t+j}, Q_{t,t+j}\}$ for all j.

Monetary Authority The central bank is assumed to perfectly observe current inflation and output, and specifies a Taylor rule for the evolution of the one-period interest rate in the economy. A wider set of monetary policy rules can easily be considered in this framework. However, the purpose of the present analysis is to examine the implications of the New-Keynesian DSGE model with adaptive learning for the term structure. Therefore, I restrict my attention to the Taylor rule that has found empirical support (see for instance Taylor, 1999).

4.2 Optimality Conditions and Equilibrium

The first order conditions from the household's optimization problem yield the Euler equations:

$$\begin{aligned} P^B_{1,t} &= \beta \tilde{E}_t \left[\frac{U_C(C^i_{t+1}; \xi_{t+1})}{U_C(C^i_t; \xi_t)} \frac{P_t}{P_{t+1}} \right], \\ P^B_{m,t} &= \beta \tilde{E}_t \left[\frac{U_C(C^i_{t+1}; \xi_{t+1})}{U_C(C^i_t; \xi_t)} \frac{P_t}{P_{t+1}} P^B_{m-1,t+1} \right], \forall m > 1. \end{aligned} \tag{14}$$

Two asset pricing implications of the model can be analyzed: first, since the expectation of future marginal utility of consumption is distinct from the rational expectations analog, the perceived path of the one-period price will be different from the corresponding price under rational expectations. Second, the $m-$period price is a function of the subjective expectations of future bond prices of shorter maturity (from the Euler equation for the longer period bonds). The bond prices can also be written in terms of the nominal stochastic discount factor (one-period ahead analog of (12)):

$$Q_{t,t+1} = \beta \frac{U_C(C^i_{t+1}; \xi_{t+1})}{U_C(C^i_t; \xi_t)} \frac{P_t}{P_{t+1}}. \tag{15}$$

For instance, the $m-$period nominal bond is written as $P^B_{m,t} = \left[\prod_{j=1}^m Q_{t+j} \right]$.

Additionally, for pricing the real bonds in this framework, the real stochastic discount factor is used:

$$q_{t+1} = \beta \frac{U_C(C^i_{t+1}; \xi_{t+1})}{U_C(C^i_t; \xi_t)}, \tag{16}$$

such that

$$
\begin{aligned}
P_{1,t}^{B,R} &= \beta \tilde{E}_t \left[\frac{U_C(C_{t+1}^i; \xi_{t+1})}{U_C(C_t^i; \xi_t)} \right], \\
P_{m,t}^{B,R} &= \beta \tilde{E}_t \left[\frac{U_C(C_{t+1}^i; \xi_{t+1})}{U_C(C_t^i; \xi_t)} P_{m-1,t+1}^{B,R} \right], \forall m > 1,
\end{aligned}
\tag{17}
$$

where $P_{1,t}^{B,R}$ denotes the one-period price of a real bond at time t. The $m-$period real can be analogously expressed as $P_{m,t}^{B,R} = \left[\prod_{j=1}^{m} q_{t+j}\right]$.

The household will also choose the labor hours supplied:

$$
\frac{v_h(h_t(k); \xi_t)}{U_C(C_t^i; \xi_t)} = \frac{w_t(k)}{P_t}. \tag{18}
$$

The intertemporal budget constraint of household i is satisfied, and the total expenditures are optimally allocated by i as:

$$
c_t^i(k) = C_t^i \left(\frac{p_t(k)}{P_t} \right)^{-\theta}. \tag{19}
$$

For the firm, the optimality conditions entails choosing the optimal price p_t^* :

$$
\tilde{E}_t \sum\nolimits_{j=0}^{\infty} \alpha^j Q_{t,t+j} Y_{t+j} P_{t+j}^{\theta} \left(p_t^*(i) - \frac{\theta}{\theta - 1} P_{t+j} s_{t,t+j}(i) \right) = 0, \tag{20}
$$

where $s = (v_h(f^{-1}(y/A; \bar{\xi})/U_C(Y; \bar{\xi})A)(1/f'(f^{-1}(y)))$ and $\bar{\xi}_t \equiv (\xi_t, A_t)'$.

Finally, the goods and asset market clearing conditions must hold. As the fiscal authority is assumed to issue riskless bonds in zero net supply, the aggregate wealth holdings of all i households must be zero.

$$
\begin{aligned}
\int_0^1 C_t^i di &= Y_t; \\
\int_0^1 B_{m,t}^i di &= B_{m,t}^s \ \forall m,
\end{aligned}
\tag{21}
$$

where $B_{m,t}^s$ is the supply of bond of maturity m.[13] Here, it is assumed that the fiscal authority follows a balanced budget policy, and the net supply of bonds of all maturities is zero. In section 6.2 below, I analyze the case where the government issues riskless debt in non-zero net supply.

For the model described above, an equilibrium is the path of the endogenous processes such that the optimality conditions hold and the market clearing conditions are satisfied, for a given expectations operator $\tilde{E}_t$.

[13] Here, aggregate wealth takes the form in (6), where the bond holdings are the total holdings of all households.

4.3 Approximation

I consider a first order log-linear approximation around the steady state output level $\bar{Y}$, and the one-period bond price β. It is assumed that the fiscal authority issues riskless bonds of all maturities in zero net supply.

Households The optimal consumption decision rule for household i, derived in Appendix A.1., is:

$$\hat{C}_t^i = (1-\beta)\hat{W}_t^i + (1-\beta)\tilde{E}_t \sum\nolimits_{j=0}^{\infty} \beta^j \left[\hat{Y}_{t+j}^i - \sigma\beta(\hat{\imath}_{1,t+j} - \hat{\pi}_{t+j+1}) + \beta(g_{t+j} - g_{t+j+1})\right], \quad (22)$$

where $\hat{W}_t^i = W_t^i/(P_t\bar{Y})$ is the net real wealth of the household in time t relative to steady state income $\bar{Y}$. The intertemporal elasticity of substitution is denoted by $\sigma = -U_c/\bar{C}U_{cc}$, and the one-period interest rate is $1/(1+i_{1,t}) = P_{1,t}^B$. Inflation is $\pi_t = P_t/P_{t-1}$ and $g_t = -U_{c\xi}\xi_t/\bar{C}U_c$ is an exogenous disturbance term. The hat variables denote the log deviations of the respective variable from its steady state value. The consumption decision rule shows that the deviations in current consumption from its steady state value depend on the current wealth, and discounted values of income, as well as expected real interest rates. An increase in the expected real interest rates will lower consumption, and increase savings.

In the present framework, households are forming forecasts of total income, which is composed of wage and profit income. Alternatively, in a more general formulation, they could form separate forecasts of their wage and profit incomes, since labor supply is an endogenous decision of the households.

The first term in the consumption decision rule captures the effect of current asset prices on consumption, and is a part of the permanent income of the household. The second term shows how the remaining permanent income affects current consumption. An increase in income (through an increase in wages or profits) will have a positive effect on current consumption - both income and substitution effects of an increase in either component will be to increase consumption. An increase in the real interest rate, will have a negative substitution effect - the household will postpone current consumption, and choose to save more by holding more riskless bonds (these are the only means of saving available to the household in this framework).

Summing consumption and wealth holdings over the i households, imposing the market clearing conditions in (65), and rewriting the consumption decision rule in terms of the output gap[14] yields:

$$\hat{x}_t = \tilde{E}_t \sum\nolimits_{j=0}^{\infty} \beta^j \left[(1-\beta)\hat{x}_{t+j+1} - \sigma\beta(\hat{\imath}_{1,t+j} - \tilde{E}_t\hat{\pi}_{t+j+1}) + \hat{r}_{t+j+1}^n\right]. \quad (23)$$

[14] Please refer to Appendix A.1. for details of the derivation.

Here, $\hat{x}_t = \log(Y_t/Y_t^n)$, Y_t^n is the natural rate of output and $\hat{r}_t^n = (\hat{Y}_{t+1}^n - g_{t+1}) - (\hat{Y}_t^n - g_t)$ is the vector of exogenous disturbances. It can be seen that not only is the current real one-period interest relevant for determining the output gap at t, but expected future one-period rates matter as well. This equation will determine aggregate dynamics, as market clearing conditions have been imposed. Additionally, as noted by Preston (2005), under rational expectations, the expectation that (23) will hold in $t+1$ and other future periods implies that the relation $\hat{x}_t = E_t\hat{x}_{t+1} - \sigma(\hat{\imath}_{1,t} - E_t\hat{\pi}_{t+1}) + \hat{r}_t^n$ will hold as well, and vice versa. However, under subjective beliefs, this is no longer true.

Term Structure Using the Euler equation with respect to longer bond prices, the prices of an $n-$period bond is rewritten in the linearized version as:

$$\hat{P}_{n,t}^B = \left[\hat{P}_{1,t}^B + \tilde{E}_t\hat{P}_{n-1,t+1}^B\right]. \tag{24}$$

This can be rewritten in terms of the one-period bond prices as:

$$\hat{P}_{n,t}^B = \left[\hat{P}_{1,t}^B + \tilde{E}_t\hat{P}_{1,t+1}^B + ... + \tilde{E}_t\hat{P}_{1,t+(n-1)}^B\right]. \tag{25}$$

The corresponding $n-$period interest rates are:

$$\hat{\imath}_{n,t} = \frac{1}{n}\left[\hat{\imath}_{1,t} + \tilde{E}_t\hat{\imath}_{1,t+1} + ... + \tilde{E}_t\hat{\imath}_{1,t+(n-1)}\right]. \tag{26}$$

as $\hat{\imath}_{n,t} = -\hat{P}_{n,t}^B/n$. This is log pure version of the Expectations Hypothesis shown in (3), with the subjective expectations operator $\tilde{E}_t$.

The corresponding real interest rates of maturity n, denoted by $\hat{\imath}_{n,t}^R$, are:

$$\hat{\imath}_{n,t}^R = \hat{\imath}_{n,t} - \frac{1}{n}\left[\tilde{E}_t\hat{\pi}_{t+1} + \tilde{E}_t\hat{\pi}_{t+2} + ... + \tilde{E}_t\hat{\pi}_{t+n}\right]. \tag{27}$$

Firms The full linearization of the optimization problem of the firm is derived in appendix A.1. The optimal price p_t^* can be written as:

$$\hat{p}_t^* = \tilde{E}_t\sum\nolimits_{j=0}^{\infty}(\alpha\beta)^j\left[\frac{1-\alpha\beta}{1+\omega\theta}(\omega+\sigma^{-1})\hat{x}_{t+j} + \hat{\pi}_{t+j}\right]. \tag{28}$$

This is rewritten using the price index as:

$$\hat{\pi}_t = \kappa\hat{x}_t + \tilde{E}_t\sum\nolimits_{j=0}^{\infty}(\alpha\beta)^j\left[\kappa\alpha\beta\hat{x}_{t+j+1} + (1-\alpha)\beta\hat{\pi}_{t+j+1}\right], \tag{29}$$

where $\kappa = ((1-\alpha)/\alpha)((1-\alpha\beta)/(1+\omega\theta))(\omega+\sigma^{-1}) > 0$, and ω is the elasticity of the marginal cost of production to the output (also defined in A.1.). As before, under the assumption, (29) will imply that $\hat{\pi}_t = \kappa\hat{x}_t + \beta E_t\hat{\pi}_{t+1}$ holds and vice versa, but not under the subjective expectations operator $\tilde{E}_t$.

Monetary Authority The one-period interest rate evolves according to the rule:

$$\hat{\imath}_{1,t} = \bar{\imath}_{1,t} + \phi_x \hat{x}_t + \phi_\pi \hat{\pi}_t, \tag{30}$$

where $\bar{\imath}_{1,t}$ is stochastic intercept term, and is denoted as the monetary policy shock.

4.4 Adaptive Learning

The complete description of the model requires specifying a forecasting model for the optimizing agents, which can be used to construct forecasts of the variables that are exogenous to the optimization problems of households and firms: output gap, inflation, one-period interest rate, the longer interest rates, and the vector of exogenous disturbances $r_t = (\hat{r}_t^n, \bar{\imath}_{1,t})'$.

The structural relation in (25) states that, under the subjective beliefs of the household, the Expectations Hypothesis of the term structure holds. That is, the price of the longer maturity bond is determined by subjective expectations of future one-period bond prices, over the maturity of the long bond. I assume that (25) will be used by the household to make conditional forecasts of longer interest rates. Consider the case where such an assumption is not made: a household may believe that there are large arbitrage opportunities possible in the future, either from selling short or holding long. In this case, the budget constraint must reflect any arbitrage opportunities that arise from the household's beliefs, given its own subjective probability distribution over future state variables.[15] The first order approximation of the wealth accumulation equation may be invalid in case the households perceive arbitrage opportunities that cause shifts in their portfolio holdings between short and long term riskless bonds. Under this assumption, the information set used the household only contains the one-period price (in addition to the output gap and the inflation) as (25) can be used to form forecasts of the longer interest rates.

Then, the set of variables that must be forecasted by the optimizing agents is denoted by the vector $z_t = \{\hat{x}_t, \hat{\pi}_t, \hat{\imath}_{1,t}\}$. Following the Minimum State Variable specification, z_t is a function of the exogenous disturbances in r_t.

Formation of Expectations Following Evans and Honkapohja (2001) and Preston (2005), beliefs are formed using least squares learning dynamics: agents run a linear regression of the past observed variables to be forecasted on the corresponding history of the vector of variables that can be used as the basis for a forecast in the future.

[15] An example of what would happen if the Expectations Hypothesis is not assumed to hold for household i: Suppose $P_{2,t} < P_{1,t} + \hat{E}_t P_{1,t+1}$.Then, if the household buys one unit each of the one and two period riskless bonds in period t at prices $P_{1,t}$ and $P_{2,t}$, subsequently selling the two period bond in time $t+1$ at the (expected) price of $\hat{E}_t P_{1,t+1}$, it will make a profit.

For the monetary policy rule considered, under the rational expectations equilibrium, the variables in z_t are a function of the disturbances $r_t = (\hat{r}_t^n, \bar{\imath}_t)'$. Here, the dynamics of the disturbances are assumed to follow the state-space representation:

$$r_t = Hr_{t-1} + \varepsilon_{r,t}, \tag{31}$$

Here H is a matrix with eigenvalues within the unit circle, so that the processes in $\{r_t\}$ are stationary. $\varepsilon_{r,t}$ is a vector of i.i.d. disturbances.

The perceived data generating process for the variables to be forecasted can be represented as (in the Minimum State Variable form):

$$z_t = a_t + b_t r_{t-1} + \eta_t, \tag{32}$$

where $a_t = [a_t^x, a_t^\pi, a_t^{i_1}]'$ is used to denote the households uncertainty about the average of the aggregate variables. The b_t matrix denotes these variables depend on the vector of states r_{t-1}.The η_t matrix is a vector of i.i.d shocks, and η_{t+1} is assumed to be unforecastable in period t.

Updating of Beliefs Given the perceived data generating process in (32), after observing current data, households update their estimates of $\Omega_t = \{a_t, b_t\}$ using a recursive least squares estimator, following Marcet and Sargent (1989). The algorithm is written as:

$$\begin{aligned} \Omega_t &= \Omega_{t-1} + g^{-1}\Upsilon_{t-1}^{-1} q_{t-1}[z_t - \Omega_{t-1}' q_{t-1}]'; \\ \Upsilon_t &= \Upsilon_{t-1} + g^{-1}[q_{t-1} q_{t-1}' - \Upsilon_{t-1}], \end{aligned} \tag{33}$$

where $q_{t-1} = [1, r_t]_{t=0}^{t-1}$, and Υ_t is the variance-covariance matrix of the coefficients in Ω_t.

The single degree of freedom in the least squares formulation of the learning model, that is allowed to differ from the rational expectations case, is the updating or gain coefficient, g which controls the rate at which new information affects beliefs. If g was a decreasing function over time, such as $g_t = 1/t$, the system of equations in (33) would be recursive representations of an ordinary least squares technique.

A constant gain parameter g implies that the household puts greater weight on the more recent observations in the updating procedure; declining gain means that equal weight is put on each available historical observation. As the constant gain algorithm has found empirical support (Branch and Evans, 2006), I will use this for analysis. It is also a natural way to allow households to consider the possibility of structural change in the data.

Using the recursive estimator, subjective forecasts of z_t are formed. For instance, in the case of the $n-$period ahead forecast:

$$\tilde{E}_t z_{t+n} = a_{t-1} + b_{t-1} H^{n-1} r_t,\ \forall n \geq 1. \tag{34}$$

Here a_{t-1} and b_{t-1} are the previous period's belief parameters, and the beliefs about the future variables in z_t are not predetermined. This corresponds to the households running a constant coefficient vector autoregression to form their beliefs: at time t, they do not take into account the fact that their belief coefficients will be updated in the future. This modelling strategy can be justified using an anticipated utility argument as in Kreps (1989), and has formed much of the basis of the learning literature (Evans and Honkapohja, 2001 and Sargent, 1993). As Cogley and Sargent (2005) discuss, beliefs (a_t, b_t) treated as random variables when they are estimated, but as constants when optimizing decisions are made. While agents can observe the past data to see how their beliefs have evolved, they are assumed to believe that future beliefs will remain constant in the infinite future.

I assume that the optimizing agents know the parameters of the H matrix with probability one, and this is a standard assumption in the adaptive learning literature (see for instance, Evans and Chakraborty, 2007). This reduces the degree of uncertainty generated in the model.

Actual Data Generating Process Substituting forecasts of the vector z_t in (34) into the structural relations determining the aggregate dynamics of the output gap and inflation, yields the actual data generating process for z_t, consistent with the process perceived by the households. The actual data generating process is:

$$z_t = T^0(a_{t-1}) + T^z(b_{t-1}) r_{t-1} + T^{\varepsilon}(b_{t-1}) \varepsilon_{r,t}, \tag{35}$$

which implies:

$$E_t z_{t+n} = T^0(a_{t-1}) + T^z(b_{t-1}) H^{n-1} r_t, \forall n \geq 1. \tag{36}$$

The T matrices are functions of the model parameters.

Expectational Stability The fixed point of the $T-$mappings in (36) is a self-consistent equilibrium: beliefs generating the data must confirm those beliefs. This corresponds to the rational expectations equilibrium when the class of forecasting models is such that the optimal forecasting rule given subjective beliefs (such as in (34)) belong to this class.

The self consistent equilibrium is Expectationally Stable (E-stable) if it is the locally stable rest point of the dynamics defined by the ordinary differential equation:

$$\dot{\Omega} = T(\Omega) - \Omega. \tag{37}$$

This ensures that the households' beliefs about the right forecasting model evolve over time to correct the discrepancy between their current beliefs given by (a_t, b_t) and the actual data that is generated as a result of their beliefs given by $T(a_t, b_t)$. Thus, conditions can be determined so that households will asymptotically converge to the rational expectations equilibrium.

For the model described above, with only one-period riskless bonds being issued by the government, Preston (2005) shows that the condition for the determinacy of the rational expectations equilibrium, the Taylor principle discussed in Woodford (2003), is also necessary and sufficient for E-stability.[16] The following proposition presents this result:

Proposition 1 *For the model described by (23), (29) and (30), combined with the forecasting rule in (33) and (34), the minimum state variable rational expectations equilibrium is linear in the state variable r_t, and the Taylor principle*

$$\kappa(\phi_\pi - 1) + (1 - \beta)\phi_x > 0, \tag{38}$$

is necessary and sufficient for determinacy of the rational expectations equilibrium and E-stability under adaptive learning.

In the case of longer period bond issuances, as considered here, since from (26) it is evident that the long interest rates are linear functions of the one-period rate, and its expected future realizations, only $\hat{\imath}_{1,t}$ is the relevant variable to be forecasted. Therefore, the conditions for determinacy of the rational expectations equilibrium as well as E-stability under adaptive learning apply here as well. Additionally, the dynamics of the longer term interest rates are stable under the Taylor principle.

Finally, the model of the economy can be summarized: equilibrium dynamics determined by (23) and (29), the determination of asset prices in (26) and (27), the interest rate rule in (30), along with the system of equations used for forecasting in (33) and (34).

4.5 Learning Dynamics and the Term Structure

Given the perceived process for the state variables, the one-period asset price is constructed as:

$$\hat{\imath}_{1,t} = T^0_{\hat{\imath}_1}(a_{t-1}) + T^z_{\hat{\imath}_1,r}(b_{t-1})r_{t-1} + T^\varepsilon_{\hat{\imath}_1}(b_{t-1})\varepsilon^r_t. \tag{39}$$

The $T-$mappings are functions of the beliefs, and this is used to derive the entire term structure of asset prices, and exhibit self referential behavior. $T^0_{\hat{\imath}_1}, T^z_{\hat{\imath}_1,r}$ and $T^\varepsilon_{\hat{\imath}_1}$ denote the coefficients in the $T-$ mappings corresponding the one-period interest rate. The longer term interest rates are constructed using (26). For instance, the two-period interest rate is:

$$\hat{\imath}_{2,t} = \frac{1}{2}\left[\hat{\imath}_{1,t} + (a_{t-1} + b_{t-1}r_t)\right]. \tag{40}$$

[16]This result is obtained for decreasing gain adaptive learning.

5 Results

Before I interpret the predictions of the adaptive learning framework for real and nominal yields with respect to the Campbell-Shiller test, and variances of yields, I discuss values of the different model parameters.

5.1 Model Parameters and Solution

Two set of parameters need to be fixed to generate quantitative results for the learning model - the constant gain, and the parameters of the New Keynesian model. The implications of the benchmark model for high and low gain are different: the higher the gain, the greater are deviations from rational expectations. The difference is large - a gain of 0.02 places a weight of 0.83 on an observation ten quarters ago, and the corresponding weight placed by a gain of 0.002 is 0.98. Table 1 shows the weights for different gain parameters. The literature is, unfortunately, mixed on the value of the gain parameter. For instance, Orphanides and Williams (2005) use a gain of 0.02, while Eusepi and Preston (2008b) estimate a value of 0.002 for a real business cycle model with adaptive learning.

I present results for the gain parameter for $g = 0.05$, which minimizes the difference between the model implied Campbell-Shiller slope coefficient for the nominal yield curve at the longest maturity, and the coefficient found in the data.[17] Robustness of the model results for different gains will be discussed below. For the rest of the parameters, I follow values used in the New-Keynesian literature. In particular: α, the frequency of price adjustment is 0.75, corresponding to an yearly price adjustment; β, the discount factor is 0.99, implying a quarterly real interest rate of approximately 4% and σ, the intertemporal elasticity of substitution is 0.1. The remaining parameters are the autoregressive parameters of the shock processes. For the preference, technology and monetary policy shocks, these are set at 0.99, 0.98 and 0.98 respectively. The corresponding standard deviations of the errors are 0.05, 0.01 and 0.02.

To numerically analyze the model, I initialize beliefs at their rational expectations values, and simulate 3000 draws of the model for 3000 time periods. The analysis is conducted in the region where the beliefs have converged to an ergodic distribution around the rational expectations beliefs, and the effect of initial conditions has been eliminated. I do this by reporting the quantitative results for the last 100 periods of the simulations. This ensures that the results of the learning model are not simply an artifact of the transitional dynamics.

Additionally, I also discuss the analytics for the mechanics of the learning model in a limiting

[17] This does not imply that the gain of 0.05 minimizes the deviation between the model implied γ coefficients at every maturity or for the real yield curve with respect to the empirical estimates. It also does not minimize the distance between the model implied yield variances and the the volatilities found in the data for real or nominal yields. Therefore, using this value of the parameter to obtain results is valid. In addition, Milani (2005) reports a gain of 0.05.

case of the benchmark model, by abstracting from nominal rigidities and therefore, from the real effects of monetary policy. This is the case when all prices are reoptimized every period, that is $\alpha \longrightarrow 0$ and output remains at its natural level every period, so that $Y_t = Y_t^n$. The model may be considered as the Lucas model, generalized to subjective expectations. The exogenous process for the evolution of the log deviations of the output level from its steady state follows an AR(1) process, specified as $\hat{y}_t = \rho \hat{y}_{t-1} + \nu_t$,where $0 < \rho < 1$, and ν_t is an identically and independently drawn shock process. This can be interpreted as the technology shock in the benchmark model with nominal rigidities. In Appendix A.2., I show that the limiting case of the benchmark model satisfies the E-stability criterion.

Figure 7 shows that the beliefs are distributed around the rational expectations values of the parameters, and are approximately normal for a small gain.[18] The dispersion of beliefs under the learning model around the time-invariant beliefs is a central feature of the learning model. For the flexible price limit of the benchmark model, the distribution of the belief parameters can be characterized as follows:

Proposition 2 *With flexible prices, under constant gain learning for a small gain* $g > 0$, *and large enough* gt, b_t *is approximately normal:*

$$b_t \sim N(\bar{b}, gC),$$

where

$$\bar{b} = \frac{\rho(1-\rho)}{\sigma}, C = \frac{(1-\rho)^2(1-\beta\rho)(1-\rho^2)}{2\sigma^2}.$$

Proof. Appendix A.2. ■

5.2 Campbell-Shiller Regression

I first present numerical estimates of the Campbell-Shiller regression coefficients for the real and nominal yield curves. In section 5.2.2, I discuss the analytical result for the flexible price limit of the benchmark model, and in 5.2.3, I discuss the results obtained in the adaptive learning framework in context of the existing literature.

5.2.1 Numerical Analysis

I construct the Campbell-Shiller regression from the yields generated by the learning model as they are reported for the data. The short rate is the one-year interest rate, and the results are reported for the same forecasting horizon, between two and ten years.

The second column in table 2 reports the slope coefficients for the gain parameter $g_1 = 0.05$ that matches the γ coefficients for the U.S. nominal yield curve data at the ten-year maturity.

[18] The distribution of beliefs are shown for the constant coefficients. The slope coefficients are similarly well behaved.

This gain places a weight of 0.22 on an observation 30 quarters ago and will be the benchmark gain parameter in for the rest of the analysis. As can be seen from the table, the slope coefficients become more negative as the forecast horizon increases. The slope coefficients for shorter horizons are less negative than for the data. In the column adjacent to the slope coefficients, I report the percentage of times the Expectations Hypothesis cannot be rejected at the 95% confidence level.

The rational expectations case is obtained when the gain is zero since beliefs have been initialized at their rational expectations values, and these beliefs are fixed points of the $T-$ mappings. In this case, there is no misspecification in (1), other than sampling error. Then, the slope coefficients must not be statistically different from one. This is indeed the case. The last column in table 2 reports the Campbell-Shiller coefficients for $g = 0$, in case of the nominal yield curve. They are statistically not different from one at the 95% confidence interval. In the rest of the discussion, results are shown for different gains, with the other parameters fixed at their benchmark values. I discuss the implications of different policy parameter values in section six below.

Table 3 shows the results of the benchmark model for the real yield curve, for the same gain parameters. As can be seen, the slope coefficients are smaller than one, and more negative than for the corresponding nominal yields.

The intuition for the negative bias in the slope coefficients can be seen from the dynamic responses of the yields to a monetary policy shock for the rational expectations and learning models. For constructing the dynamic responses, I consider a unit impulse to the monetary policy shock, at the beginning of the period where the distribution of the model has converged to a stationary distribution, at period 2901. The pointwise median response in the difference of the trajectory of the relevant variables (with and without the shock) is then considered.[19]

Figure 8 shows the response of the short (one-period) yield and the long (two-period) yield to a monetary policy shock under rational expectations and learning. The short and long yields rise on impact, as shown in figures 8(a) and 8(b). In the period of the shock, the impact effects under rational expectations and learning are similar, since beliefs under learning are distributed around the rational expectations beliefs for a small gain. As the long yield is constructed using the (subjective) Expectations Hypothesis, it will rise less than the short yield. That is, (negative of the) spread will rise on impact of the shock in this framework, and is plotted in figure 8(c). Thus, on average, both the short and long yields rise to the same extent under learning and rational expectations. It is after the period of the shock that the assumption of less than rational expectations will be important.

After the period of impact, under rational expectations, future expectations of short and long yields coincide exactly with those of the true model - agents will correctly forecast that the transitory monetary policy shock dissipates, and the behavior of the expected long yield (which is the expected

[19] This experiment is repeated for 3000 draws. The construction of the dynamic responses follows Eusepi and Preston (2008b).

one-period yield in the case considered) with respect to the current long yield will be consistent with the Expectations Hypothesis. In terms of the difference between the expected future long yield[20] and the current long yield, the (negative of the) difference between them is positive. That is, the model predicts that when the yield spread rises, the expected future long rate (of maturity one-period) will rise. Or as the yield spread falls, the expected long yield falls as well, and the slope coefficient γ in (1) is statistically not different from one.

Under learning, the households do not recognize that the rise in the yield is due to the transitory monetary policy shock. The increase in one-period yield at time t is perceived to be an increase in the expected average returns, across the infinite horizon decision problem of the optimizing households, and subsequently the yields are more persistent than under rational expectations. In this case, due to the intertemporal substitution effect, households demand less consumption and more savings, relative to the rational expectations case. The only mechanism available to the households to save in this framework is by holding riskless bonds, and their demand for the one-period asset increases relative to the rational expectations case. Since these bonds are only available in zero net supply the higher demand for the assets leads to an increase in bond price of the one-period maturities, with a corresponding decline in its yield. At the same time, the households forecast a fall in all future income levels, which lowers consumption and desired savings for the infinite horizon problem. This is seen from the responses of the output gap and inflation in figures 8(e) and 8(f) - under learning, after the period of impact, the output gap falls more than under rational expectations. In this case, the one-period yield under must rise under learning relative to rational expectations. This effect dominates, and the one-period yield rises. In contrast to rational expectations, the (negative of the) difference between the expected one-period ahead yield and the long yield is negative. This is seen in the dynamic responses in figure 8(d). In case of a decline in the yields due to a negative monetary policy shock, the yield spread will be positive on impact under rational expectations and learning. As households misperceive the fall in the one-period yield to be a decline in expected returns, their demand for savings will fall. Therefore, the amplified intertemporal substitution and income effects in the learning framework result in the negative bias with respect to one in the slope coefficient γ.[21]

5.2.2 Campbell-Shiller Coefficients in the Flexible-Price Model

Under rational expectations, the one-period asset price is determined entirely by the realization of the exogenous endowment process. This can be seen from the fixed point of the $T-$mappings:

$$\hat{P}^B_{1,t} = \bar{T}^0 + \bar{T}^z \hat{y}_{t-1} + \bar{T}^\varepsilon \varepsilon^y_t. \tag{41}$$

[20] Here it is the expected one-period yield.

[21] The analysis here uses the two-period yield as the long yield. As other longer yields are simply a monotonic transformation of this, the qualitative analysis will be identical in that case.

Then, the rational expectations model will also not be able to generate rejections of the Expectations Hypothesis in this first order approximation of the model when tested using the Campbell-Shiller regression. The error term in the regression arises only due to random variation, and is orthogonal to the yield processes. Then the slope estimator γ is unbiased in (1), and statistically not different from one.

Why does the learning model do better? Under a constant gain learning algorithm, the updated coefficients converge to an ergodic normal distribution, centered at the rational expectations beliefs, with a non-zero variance (proposition 2 above). Due to the nature of the $T-$mappings, the paths of the asset prices are more complex: the law of motion in (30) illustrates that the coefficients (a_t, b_t) are used to form conditional forecasts of the asset price, but these are a function of the past realizations of the price itself. This self-referentiality in price determination is a key feature of the learning model.

The following proposition can be shown for the limiting case of the benchmark model:

Proposition 3 *With flexible prices, under constant gain learning for small $g > 0$, and large gT, bias of the slope coefficient in (1), $plim_{T \longrightarrow \infty}(\gamma_T - 1)$ is negative for $\beta, \rho \in (0, 1)$.*

Proof. Appendix A.3. ■

When the asset price is determined using the lagged endowment interacted with a mapping that is endogenously determined and the corresponding error term, the regressor and error will no longer be orthogonal. This misspecification implies that when the asset price determined as above, but tests of the Expectations Hypothesis are constructed assuming beliefs are fully rational, the ordinary least squares estimates will be biased. The source of misspecification in (1) is not the absence of a time varying risk premia, but formation of expectations that are less than fully rational. If tests of the Expectations Hypothesis are constructed by assuming that beliefs are rational, they will yield biased ordinary least square coefficients.

Therefore, the above results indicate that the misspecification in beliefs about the data generating process of yields generates the biases observed in the slope coefficient γ in the data. In this framework, since the subjective Expectations Hypothesis holds, the bias in the coefficients is a result of the fact that the Campbell-Shiller regression in (1) is misspecified, as it is constructed using rational expectations.

I show the dynamic responses of yields analytically, for an endowment shock for the Lucas economy in Appendix A.4. In response to the positive technology shock, the one-period (and two-period bond prices from the subjective Expectations Hypothesis) rise, or the corresponding yields fall. Under rational expectations, the household correctly attributes the decline in the one-period yield in the current period to the transitory technology shock, since its conditional forecast of the future yield (relevant for the permanent income of the household) is the same as under the true model. There is a substitution effect where the household lowers its savings, and increases consumption. Under learning, however, the household does not correctly perceive the decline in yields to be due to the transitory technology shock. It therefore expects average returns to be lower in the future than it would under rational expectations. This amplifies the substitution effect relative to rational expectations - the household lowers its savings even more, demanding lesser riskless bonds. However, as the net supply of bonds is fixed, the one-period yield falls much less relative to the rational expectations case, as the future yields must rise to encourage households to save more. This implies that the difference between the expected future one-period yield and the current long yield is negative. Under rational expectations, this difference is positive.

Therefore, in the benchmark model, learning is a necessary ingredient for generating a negative bias in γ with respect to one. As I show in Proposition three above, the negative bias will be generated even in limiting case of flexible prices ($\kappa \longrightarrow \infty$), where the interest rate rule must only ensure asset and goods market clearing.

5.2.3 Connections to the Literature

The result that γ is negatively biased with respect to one in the benchmark model used here can also be interpreted as the under-reaction of expected future yields of maturity $(n-1)$ to changes in the current short yield. This is discussed by Froot (1989) and Mankiw and Summers (1984). Froot (1989) rejects the alternative hypothesis that expected future yields are excessively sensitive to changes in the contemporaneous short yield, as well as the Expectations Hypothesis. He also cannot reject the hypothesis that the change in the long rate in excess of the spread can be attributed exclusively to expectational errors. In his analysis, the test of the Expectations Hypothesis in (1) is decomposed into two slope coefficients, one corresponding to the expectational errors and the other to a term premium. The first is found to be negative, that is, a portion of the deviation of γ from one can be attributed to expectational errors. It is also found that at longer maturities, the slope coefficient corresponding to the term premium becomes quantitatively less important. In the adaptive learning model used here, the first order approximation of the model implies that time-varying risk premia is not present, and cannot explain for the rejections of the Expectations Hypothesis.[22] Thus, in the spirit of the Froot's (1989) exercise, expectational errors entirely account for the rejections of the rational Expectations Hypothesis. In addition, the adaptive learning approach pursued here provides a theory of expectations that generates systematic expectational errors. These errors yield biased slope coefficients in (1), emphasized by the Froot's (1989) empirical work using survey expectations data.

Mankiw and Summers (1984) reject the hypothesis that expected future yields are excessively sensitive to changes in the contemporaneous short yield, along with the Expectations Hypothesis. They test if myopic expectations can justify the rejections of the Expectations Hypothesis, but this is rejected as well - that is, financial markets are 'hyperopic', giving lesser weight to contemporaneous

[22] However, time-variation in beliefs due to the learning process is a feature of the model.

fundamentals than to future fundamentals. The authors use a term premia is used to explain the rejections of the Expectations Hypothesis. The benchmark model used in the present paper shows that adaptive learning generates persistence in longer yields relative to shorter yields, supporting the finding that expected future yields do not overreact to short yield changes. As discussed in the context of the dynamic response of yields to the monetary policy shock, the misperception of the transitory change in the short yield as a permanent change, translates into greater persistence of yields under learning

Piazzesi and Schneider (2009) find, using survey data, that before 1980, when the level of yields were rising and the yield spread was small, survey forecasters predicted lower long yields than those which would be predicted by a statistical model. Since the forecasters update their information about high long yields slowly, they predict lower excess returns than were observed in the data. Thus, when the yield spread was low, and yield levels were high, the survey forecasters predicted that long rates would fall, as seen in the empirical data. In the benchmark model used here, the fact that optimizing agents misperceive the current increase in the short yield (due to a monetary policy shock) as an increase in yields for decisions they face over the infinite horizon results in a fall in the actual expected future yields. Therefore, as found by Piazzesi and Schnieder (2009), in a partial equilibrium framework, the fact that the adaptive learners update their beliefs about yield processes slowly, leads them to predict different paths of yields than under the true model. In the general equilibrium model used here, the effects operate through intertemporal consumption and savings decisions.

The analysis can be connected to the findings of Laubach, Tetlow and Williams (2007). The authors use an affine factor model with only observables and time-varying coefficients that are re-estimated over time, and find that the deviations from the Expectations Hypothesis' implication are significantly smaller. The use of adaptive learning as a theory of expectations formation gives a theoretical foundation to this finding in the literature.

It is also useful here to compare the performance of the benchmark model, with other approaches that have attempted to fit the yield curve in a DSGE framework with rational expectations. Ravenna and Seppälä (2007) characterize the term premia using a third order approximation in a New Keynesian DSGE framework, and find that a high degree of habit formation, and a persistent monetary policy rule are necessary to obtain rejections of the Expectations Hypothesis (tested using (1)) for the nominal term structure. They also find that the number of rejections of the hypothesis of γ different from one in (1) fall significantly in case of the real yield curve.[23]

Rudebusch and Swanson (2009) find that when recursive Epstein-Zin preferences are introduced in an otherwise standard New Keynesian model, long run inflation risk must be introduced to fit the nominal yield curve, to avoid using a very high degree of risk aversion. However, while they

[23] Therefore, the authors argue that the monetary policy specification is also important, other than habit formation in preferences.

are able to successfully fit the nominal yield curve, the model would potentially have difficulty in generating the right slope of the real term structure observed in U.S. TIPS data. In their model the expectation of higher inflation reduces the value of nominal bonds, leading to a positive term premia, and an upward sloping yield curve. However, this would not be the case for real bonds - higher expected inflation will increase the demand for inflation indexed bonds, such as TIPS, leading to a negatively sloped yield curve, due to a negative risk premia.

Unlike the Ravenna and Seppälä (2007) and Rudebusch and Swanson (2009) analyses, the learning model used here is successful with respect to the real yield curve as well. The Campbell-Shiller coefficients are negative, as found in the data.

5.3 Volatility of Yields

As for the Campbell-Shiller regression coefficients, I discuss the numerical result, followed by the results for the flexible price limit of the benchmark model.

5.3.1 Numerical Analysis

Before the variances of yields under the benchmark model are discussed, it is useful to analyze the role of other factors in the model. Specifically, what is the effect of price rigidity on yield variances? This can be understood in the context of the rational expectations model.

As can be seen from figure 6(b), when the rational expectations model is calibrated to U.S. data, the level of yield variances generated is much smaller relative to the data. The relative variance of the long yield to the short yield is also slightly smaller than the data. For instance, the ratio of the three year to the one year volatility is 0.96 in U.S. nominal yields data, and for the rational expectations model, it is 0.94. Reducing the degree of price stickiness lowers the relative volatility of the long yield with respect to the short yield, even as it increases the level of yield volatilities across the maturity structure. Therefore, the assumption of price stickiness is integral to explaining the excess volatility in long yields, and keeping the level of volatilities within empirically consistent ranges. In the rational expectations analog of the model considered here, as the degree of nominal rigidities becomes smaller, tending to the flexible price limit, the variance of the one-period interest rate increases by approximately four times. This is far greater than the variance of the nominal three month Treasury bill rate for the U.S. data for the 1972-2009 period.

Learning affects both dimensions of volatilities. The self-referential nature of the adaptive learning process implies that the beliefs are dispersed around the rational expectations beliefs. As can be seen from tables 4 and 5 , the variances of yields are larger than their rational expectations counterparts, for the different values of the gain considered. For higher gains, the variances of the corresponding yields are larger under learning. This is expected - as the misspecification becomes larger, the dispersion of beliefs around the rational expectations beliefs will be greater, and the implied volatility of the yields under learning will also be higher. For the benchmark gain parameter ($g = 0.05$), the volatility of the ten-year yield under learning is approximately 8% higher than the rational expectations case.

Additionally, the long yields are also more volatile relative to the short yields, than under rational expectations. For the benchmark gain, the ratio of the three-year volatility to the one-year volatility is 0.97.

The variances of real yields are higher than those observed in the data, both for the U.S. TIPS and U.K. Index-Linked bonds. The model is therefore inconsistent on this dimension - while the variances of the nominal yields are closer to those observed in the data, they are significantly higher than the volatilities of yields observed in the data on real bonds.

5.3.2 Volatilities in the Flexible-Price Model

For the limiting case with flexible prices, the distribution of beliefs is derived analytically in proposition three. This can be used to show that the variance of yields under the learning model can be decomposed into the variance of yields under the rational expectations model, and a function of the variance of the belief parameters:

Proposition 4 *With flexible prices, under constant gain learning for small $g > 0$, and large gt:*

$$Var(y^L) > Var(y^{RE}) \tag{42}$$

where $Var(y^L)$ is the yield of the one-period yield under the learning model and $Var(y^{RE})$ is the variance of the corresponding yield under rational expectations

Proof. Appendix A.5. ■

As beliefs are centered around the rational expectations beliefs with a non zero variance, the variance of the yields in the learning model will be higher. The dispersion of beliefs around the rational expectations parameters is due to the self-referential nature of the belief formation process. The assumption of price stickiness is necessary to ensure that the variance of yields at the short end of the nominal yield curve does not become very large, and for explaining excess volatility in long yields relative to short yields.

5.3.3 Connections to the Literature

The implications of the learning model for variances can be compared to the predictions of the DSGE model with rational expectations. Hördahl, Tristani and Vestin (2007) analyze the predictions of the New-Keynesian model for the term structure, and introduce habit formation preferences, a difference rule in monetary policy, and inflation indexation in addition to Calvo pricing for the firms. They can generate 94% of the volatility at the short end of the yield curve (for the three-month interest rate). However, when price stickiness is eliminated, the volatility of the the short, one-period interest rate increases, and far exceeds the volatility in U.S. data.[24] Thus, price stickiness helps to lower the volatility of yields at the short end of the curve and prevents a very steep decline in the variances across the maturity structure, as would be predicted by the Expectations Hypothesis. Therefore, the assumption of nominal rigidities helps on both dimensions with respect to variances - keeping the level of yield volatilities within an empirically consistent range, and tempering the decline across the maturity structure.

The fact that adaptive learning can generate larger volatilities has been discussed elsewhere in the literature as well. For instance, Piazzesi and Schneider (2007) show that when adaptive learning is introduced on the intercept terms of the inflation and consumption processes (these are exogenous processes, in a partial equilibrium model), the level of volatilities increases across the maturity structure, and the long yields are more volatile compared to the implication of the Expectations Hypothesis.

5.4 Model Forecast Errors and Survey Evidence

In the section below, I compare the forecasting errors implied by the model with those obtained from survey data. This is motivated by Froot (1989) and Piazzesi and Schneider (2009) who find that subjective expectations can be reasonably approximated by survey expectations. If the conjecture about the adaptive learning being an empirically consistent characterization of the expectations formation process of agents is true, then the properties of the forecast errors implied by the model, and real-time survey expectations should be close.

The systematic forecast errors generated in this framework are an artifact of the constant gain adaptive learning process, and can be considered a success of the model. The fact that survey data exhibits systematic errors across the variables in the data set, is evidence of the limited rationality of agents. This has been discussed widely in the literature for several data sets. For instance, Bacchetta, Mertens and van Wincoop (2008) discuss the role of expectational errors in explaining empirical anomalies in different markets, such as the exchange rates.

The two survey datasets used are the Survey of Professional Forecasters (SPF) and the Michigan Survey of Consumer Finances (MSCF). The SPF was started in 1968 by the American Statistical Association and the NBER, and its administration was taken over by the Federal Reserve Bank of Philadelphia in 1990. It asks a set of industry economists, their quarterly and annual forecasts for different sets of variables, such as the three month Treasury Bill rate, the CPI Inflation rate and real GDP. The forecasters are required to report their one to five quarter ahead forecasts. The survey data then reports mean and median forecasts of each variable. To ensure a consistent survey

[24] In the Hördahl et. al. (2007) model, the variance of the one-quarter interest rate increases by five times when $\kappa \longrightarrow \infty$.

data series, I only use values reported from 1990Q2.

The MSCF asks a set of households, considered representative of all American households, their forecasts of the changes in variables such as expected changes in inflation, unemployment and interest rates. The survey is conducted monthly, covers approximately 50 questions and atleast 500 interviews. Details of the surveys are shown in table 6.

I use survey forecasts to gauge the performance of the model in the following ways:

1. The autocorrelation in the one-quarter ahead forecast errors of the three month nominal interest rate in mean SPF forecasts is 0.24;

2. The autocorrelation in the one year ahead forecast errors of the inflation rate with the MSCF forecasts is 0.37[25];

The survey forecast error is the difference between the realization of the variable m at time t, and the forecast one period ago, that is $E_{t-1}m$. I compute the forecast errors from the model in the same way.

The benchmark gain $g = 0.05$ matches the the slope coefficient γ in (1) at the longest forecast horizon (ten years) and the autocorrelation in one-quarter ahead nominal interest rate forecast errors is 0.12, and for one-year ahead inflation forecasts is 0.15. Thus, the forecast errors implied by the model are smaller than for survey data for $g = 0.05$. If the gain is calibrated to match the autocorrelation in forecast errors for the nominal one-quarter interest rate, the resulting gain is 0.07.

5.5 Robustness

To check the robustness of the model implications, I consider other values of gain parameter. Orphanides and Williams (2005) find that a gain of 0.02 is consistent with Survey of Professional Forecasters data on inflation. As expected, the slope coefficients are less negative than for $g = 0.05$, for the corresponding maturities. For gain $g = 0.06$, the slope coefficients are smaller (that is, more negative) than for the benchmark gain parameter. These are shown in table 8.

For the benchmark gain parameter ($g = 0.05$), the volatility of the ten-year yield under learning is 8% higher than the rational expectations case and, the corresponding ratio is 10% when the gain is 0.06. Volatilities of nominal yields for different gain parameters are shown in table 9.

5.6 Testable Implications for Inflation and Yield Processes

Given the performance of the learning model with respect to the yield curves, it is also useful to evaluate its implications for the unconditional moments of the inflation process and the one-year

[25] This is reported by Mankiw, Reis and Wolfers (2004).

yield, and compare these with the rational expectations analog of the DSGE model, as well as the data.

Theoretically, as the beliefs are centered at their time-invariant, rational expectations means, the implications of the learning model should not be very different for the autocorrelation processes. The variance of the processes is expected to be higher than the rational expectations analog, given that the beliefs have a non-zero variance.

As table 7 shows, for the benchmark gain parameter, this is indeed the case. In case of inflation, the standard deviation is 1.39 for the learning model, and 1.28 for the rational expectations case. These are both smaller than the standard deviation of inflation for quarterly U.S. data between 1972 and 2009. The autocorrelations in the learning model are higher than the data. However, the autocorrelation in the U.S. inflation series has not been the same for the entire time period - the 1980s were characterized by higher autocorrelation in the inflation series, compared to the 1990s.

The autocorrelation of the one-year yield closely matches the data. This is true for the rational expectations and the learning models, as the calibrated shock processes are very persistent. The standard deviations are smaller than the data, and the learning model generates a slightly higher volatility than the rational expectations case.

6 Policy Effects in the Benchmark Model

Here I explore the implications of the model on two policy dimensions. First, what are the properties of yields under different monetary policy regimes? The 1980s have been characterized by a larger response of the monetary policy coefficient to inflation, and the yields across the maturity structure have been much more volatile. In the benchmark model, when the Taylor parameter ϕ_π increases, how do the variances of yields change?

The effects of fiscal policy on the yield spread are also explored in this framework. For government debt to have a non trivial effect on the consumption decisions of optimizing households, Ricardian equivalence must not hold. As I show below, this will be true when agents are boundedly rational.

6.1 Different Monetary Policy Regimes

Taylor (1999) and Taylor and Smith (2007) discuss the change in the response coefficients in the monetary policy rule response to inflation and output gap. The 1980s and 1990s have been characterized by a more aggressive response to inflation.

In case of the learning model, I consider two different policy experiments. I first increase the Taylor coefficient for inflation to $\phi_\pi = 4$, keeping all other parameters in the model constant. In the second experiment, I consider an inflation targeting rule, by considering a very large value of

ϕ_π.[26]

In terms of the Campbell-Shiller slope coefficients, as can be seen from the first column in table 10, the slope coefficients for U.S. nominal yield curve data are less negative for the period between 1984-2009. This is also a feature of the model - deviations from the Expectations Hypothesis (with rational expectations) are smaller for a more aggressive response of the central bank to inflation.

For the 1980-2009 period, as documented by the first column in table 11, the volatility of yields is greater than the entire sample period. In the case of rational expectations. As can be seen from comparing the last two columns in table 9, the volatility of yields increases here as well. This effect is transmitted throughout the term structure.

Between 1984 and 2009, the volatilities of yields in the U.S. nominal yield curve data are smaller. For instance, the standard deviation of the one-year yield is 2.43 and of the ten-year yield is 2.11. The respective volatilities for the entire period are 2.99 and 2.47. The year 1984 is the commonly used date for the beginning of the Great Moderation, and the lower volatilities in yields during this period could be attributed to this in the sample 1984-2009.

6.2 Fiscal Policy and the Yield Spread

The literature has been mixed about the effects of fiscal variables on interest rates. While Engen and Hubbard (2004) find insignificant effects of government deficits on interest rates, Evans and Marshall (2002) find varying effects, depending on how the identification of fiscal shocks is constructed. Dai and Phillippon (2006) construct a factor model of the yield curve using observable and latent factors. In addition to the standard macro variables as observables, namely the output gap and inflation, they introduce a deficit variable, which is found to be statistically significant in explaining the long end of the yield curve, and the spread. In this framework, a deficit shock is found to affect the yield spread by affecting the expectations of future short rates, as well as the risk premia.

To explore the effect of fiscal policy on the term structure, the fiscal authority is now assumed to issue riskless debt in non-zero supply, as in Eusepi and Preston (2008a). For analytical tractability, I will assume that only the one-period debt is issued in non-zero supply. Bonds of all other maturities are issued in zero net supply.

Government expenditures are financed by issuances of riskless m-period bonds, and lump-sum taxes. The liabilities of the government at time t are given by:

$$W_{t+1}^s = B_{1,t}^s + \sum\nolimits_{m=2}^{n} P_{m-1,t+1}^B B_{m,t}^s, \tag{43}$$

and the flow budget constraint of the government is:

$$\sum\nolimits_{m=1}^{n} P_{m,t}^B B_{m,t}^s \leq W_t^s - T_t. \tag{44}$$

[26] I consider $\phi_\pi = 15$.

Here $B^s_{m,t}$ is the supply of riskless bonds of maturity m, and T_t is the total tax collection of the government. I assume that the government only imposes lump sum taxes.

The full specification of the now model requires that the amount of liabilities which need to be financed every period by the government are specified. Here, I assume that taxes are a proportion of the net liabilities outstanding at time t:

$$\frac{T_t}{\bar{T}} = \sum_{m=1}^{n} \left(\frac{B^s_{m,t-1}}{\bar{B}^s_m} \right)^{\phi_\tau}; \phi_\tau \geq 0. \tag{45}$$

This allows for both passive and active formulations of fiscal policy, although I will explore only the Ricardian or passive formulation of fiscal policy to simplify the analysis.[27]

The optimal consumption decision rule for household i, derived in Appendix A.6., is:

$$\begin{aligned} \hat{C}^i_t &= s_W(1-\beta)\hat{W}^i_t + (1-\beta)\tilde{E}_t \left[\sum_{j=0}^{\infty} \beta^j \hat{Y}_{t+j} - s_T \sum_{j=0}^{\infty} \beta^j \hat{T}^i_{t+j} \right] \\ &\quad + \beta(1-s_T)\hat{E}_t \sum_{j=0}^{\infty} \beta^j \hat{P}^B_{1,t+j}, \end{aligned} \tag{46}$$

where $s_W = (\bar{W}/\bar{Y})$ and $s_T = (\bar{T}/\bar{Y})$. Summing consumption and wealth holdings over the the i households, imposing the market clearing conditions , and rewriting the consumption decision rule in terms of the output gap[28], I derive:

$$\begin{aligned} \hat{x}_t &= \tilde{E}_t \sum_{j=0}^{\infty} \beta^j \begin{bmatrix} (1-\beta)\hat{x}_{t+j+1} - \sigma\beta(\hat{\imath}_{1,t+j} - \tilde{E}_t \hat{\pi}_{t+j+1}) \\ + \hat{r}^n_{t+j+1} \end{bmatrix} \\ &\quad + s_T \begin{bmatrix} \frac{(\hat{b}_{1,t} - \hat{\pi}_t)}{\beta} - \frac{\hat{\tau}_t}{\beta} \\ + \tilde{E}_t \sum_{j=0}^{\infty} \left[(\hat{\imath}_{1,t+j} - \tilde{E}_t \hat{\pi}_{t+j+1}) - (1-\beta)\hat{\tau}_{t+j+1} \right] \end{bmatrix}. \end{aligned} \tag{47}$$

Here, since the one-period bonds are issued in non-zero net supply, it appears in the determination of the current output gap.

Under rational expectations, Ricardian equivalence holds: the households recognize that intertemporal budget constraint of the government must be satisfied. That is, the present value of debt is equal to the discounted sum of future tax collections. Then for the optimal consumption decision rule in (67), the first term corresponds to a permanent income effect, and the second term, which is the intertemporal budget constraint of the government, is zero. It is the intertemporal optimization of the households which ensures the budget constraint of the government holds. This holds regardless of the specification of fiscal policy. In this case, any increase in the net issuance of

[27] For an analysis of active and passive formulations of government policy with adaptive learning, the reader is referred to Eusepi and Preston (2008a).

[28] Please refer to Appendix A.6. for details of the derivation.

the one-period debt will not lot affect the consumption decision of the households.

When expectations are near-rational, household i is required to form forecasts of its future tax obligations, and the evolution of the bond prices, conditional on its information set. Out of the rational expectations equilibrium, it is no longer required for each household to understand that Ricardian equivalence must hold - an individual household may not recognize that its tax obligations are the same as all other households, and that the present value of government debt must equal the discounted sum of tax obligations.

Therefore, under adaptive learning, Ricardian equivalence will not hold in the transition to the rational expectations equilibrium. In the asymptotic limit, when the constant-gain learning algorithm converges to an ergodic distribution, households will continue to make small expectational errors in their conditional forecasts of the state variables. Thus, under the benchmark model, government bonds are *perceived* as net wealth, in the context of Barro (1974). The deviation from Ricardian Equivalence in a framework with adaptive learning has also been analyzed by Evans, Honkapohja and Mitra (2009).

Effects of Deficit Shocks Blinder and Yellen (2001) highlight two U.S. government budget-related episodes that affected the yield curve: in February 1993, Clinton introduced a budget reduction package. A second Clinton budget agreement in November 1999 declared that the Social Security surplus would be "off-budget" raised the fiscal bar. For the U.S. nominal yield curve data used here, the ten year yield fell by 1.3% between October 1992 and 1993, while the one year yield remained mostly constant. For the second period, the slope fell by approximately 1%. Bikbov and Chernov (2006) find that the long (ten-year) yield dropped by 1.5% between late 1992 and 1993, and the short yield remained constant.[29] For the second period, they show that the slope of the curve dropped by 2%. Using a factor model, the authors generate this effect on the yield curve using a latent factor that is found to be significantly positively correlated to the annual growth in public debt. However, these effects cannot be generated in an optimizing framework such as the one used above, with rational expectations.

In the adaptive learning framework, debt holdings have a non-trivial effect on consumption decisions, the effect of a deficit shock can be examined. The mechanism through which the shock would have a non-zero effect is by distorting the intertemporal substitution and income effects. When there is an increase in the net issuances of the one-period debt (a positive deficit shock), the households perceive this to be an increase in their net wealth holdings, as the conditional forecasts of future taxes are still constructed using beliefs from the past period, as are the forecasts of future yields. Thus, as the optimizing households choose to consume more, actual prices of the one-period riskless bonds fall, and the corresponding yields rise relative to other yields. As the short yield

[29] The authors use an unsmoothed Fama-Bliss approximation of the zero coupon bond prices of maturities three months to ten years.

rises, and the rise in long yields is smaller (by the Expectations Hypothesis), the yield spread falls. This effect would not be present in a rational expectations framework, as Ricardian Equivalence would hold exactly.

In this framework, it is also possible to explore the effects of a shift in issuances of the composition of debt between long and short term debt. This is left to future research.

7 Conclusion

This paper has attempted to construct a micro-founded optimization model with constant gain adaptive learning as a theory of expectations formation, to address empirical anomalies in the real and nominal yield curves.

The Expectations Hypothesis implies that when the yield spread rises, the expected future long yields must rise as well. This implication is found to be rejected for both the nominal U.S. yield curve, and the real yield curve for Index-Linked bonds in the U.K. That is, when the spread rises, expected future long rates fall. In the benchmark model used here, the subjective Expectations Hypothesis holds for the optimizing agents. However, the slope coefficient of the Campbell-Shiller (1991) regression is found negatively biased with respect to one, as in the empirical data. This is due to the amplification of intertemporal substitution effects - increases in the current short yields are misperceived by the adaptive learners as an increase in expected returns over their infinite horizon decision problem. As they demand greater savings in the framework with net zero supply, future expected yields rise less than the rise in current long rate.

The learning model also resolves a part of the excess volatility puzzle. Long yields generated by the model are more volatile than the corresponding yields in the rational expectations model. Additionally, the level of volatilities across the maturity structure are higher than the rational expectations model. These results hold for both the nominal and real yield curves. In case of the U.S. nominal yield curve, the long yield is slightly more volatile than its rational expectations counterpart of the excess volatility in the ten-year yield with respect to the one-year yield can be explained by the learning model. Also, the level of volatility at the long end (the ten-year yield) is approximately 10% higher than the rational expectations model.

Given that the constant gain learning model generates systematic forecast errors, I compare the properties of the errors implied by the model, with those of survey data, for different gain parameters. The model generates autocorrelation in forecast errors that are smaller than survey data for the benchmark gain, but for higher gain parameters, the fit with survey data improves. This may be considered a success of the model, as survey data imply limited rationality of agents. This cannot be obtained in a rational expectations analysis. Also, the inflation and one-period yield generated by the model are broadly consistent with U.S. data.

I use the model to explore different monetary policy regimes. The last two decades of U.S.

monetary policy have been characterized by a more aggressive response to inflation in the Taylor rule. I find that a higher Taylor rule parameter increases the volatilities of yields for the nominal and real yield curves.

Finally, I explore the effects of fiscal policy on yields in this framework - an increase in the supply of riskless one-period debt is found to increase the short yield. A smaller increase in longer yields implies that the yield spread falls. The result obtains through the violation of Ricardian equivalence in this framework.

The analysis here leads to several natural extensions. Under the benchmark model with rational expectations, the slope of the yield curve (real and nominal) is flat since I consider a first order approximation around the deterministic steady state. Under the learning model, this will be true as well. Although the self-referentiality in beliefs generates persistence in the path of yields, and increases their variance, the time varying beliefs (a_t, b_t) remain distributed normally around the rational expectations beliefs. Thus, the average of the yields remains the same, across the maturity structure. The framework used here can be extended to match the slope of the curve as well in atleast two different ways. First is to introduce an unobservable time trend in the one-period interest rate, which is extracted using a Kalman filter. Another avenue that can be explored is alternative utility specifications. Rudebusch and Swanson (2009) introduce Epstein-Zin preferences in a DSGE model, and find that the framework is successful at generating a sufficient term premium, although a long-run inflation risk must be introduced in to lower the risk aversion parameter. In the present framework, the time variation in risk premia generated by a recursive utility specification may no longer require high risk aversion to yield enough time-variation in the term premia due to the uncertainty generated by the adaptive learning process. Then, the empirically observed deviations in term structure can be decomposed into effects of adaptive learning by agents, and the risk premia term. It is also conjectured that such a framework will use empirically plausible values of parameter values, such as risk aversion, in contrast to those used in the literature that focusses only on the risk premia based explanation. Another extension of the analysis is a detailed empirical investigation, where the learning model would be estimated. Finally, the implications of the different gain parameters can be explored using a utility based welfare criterion - what is the loss in utility when the optimizing agents "forget" a larger history of data.

References

[1] Adam, K., A. Marcet and J.P. Nicolini, (2007), "Stock Market Volatility and Learning", *Working paper.*

[2] Ang, A. and M. Piazzesi, "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables", *Journal of Monetary Economics, 50, pp. 745-787.*

[3] Bacchetta, P., E. Mertens and E. van Wincoop, (2008), "Predictability in Financial Markets: What do Survey Expectations Tell Us?", *Journal of International Money and Finance, 28, pp. 406-426.*

[4] Barro, R.J., (1974), "Are Government Bonds Net Wealth?", *Journal of Political Economy, 82, pp. 1095-1117.*

[5] Barro, R.J., (1981), "Output Effects of Government Purchases", *Journal of Political Economy, 89, pp. 1086-1121.*

[6] Bikbov, R. and M. Chernov, (2006), "No-Arbitrage Macroeconomic Determinants of the Yield Curve", *Working paper.*

[7] Blinder, A.S. and J.L. Yellen, (2001), The Fabulous Decade - Macroeconomic Lessons from the 1990s*, Century Foundation Press, New York, New York*

[8] Branch, W.A. and G.W. Evans, (2006), "A Simple Recursive Forecasting Model", *Economics Letters, 91, pp. 158-166.*

[9] Calvo, G., (1983), "Staggered Prices in a Utility-Maximizing Framework", *Journal of Monetary Economics, 12, pp. 383-398.*

[10] Campbell, J. and J. Cochrane, (1998), "By Force of Habit: A Consumption Based Explanation of Aggregate Stock Market Behavior", *Journal of Political Economy, 107, pp. 205-251.*

[11] Campbell, J.Y. and R.J. Shiller, (1991), "Yield Spreads and Interest Rate Movements: A Bird's Eye View", *Review of Economic Studies, 58, 495-514.*

[12] Carceles-Poveda, E. and C. Giannitsarou, (2007), "Adaptive Learning in Practice", *Journal of Economic Dynamics and Control, 31 (8), 2659-2697.*

[13] Carceles-Poveda, E. and C. Giannitsarou, (2008), "Asset Pricing with Adaptive Learning", *Review of Economic Dynamics, 11(3), pp. 629-651.*

[14] Clarida, R., J. Gali and M. Gertler, (1999), "The Science of Monetary Policy: A New Keynesian Perspective", *Journal of Economic Literature, 37, pp. 1661-1707.*

[15] Cogley, T. and T.J. Sargent, (2005), "Anticipated Utility and Rational Expectations as Approximations of Bayesian Decision Making", *Working paper.*

[16] Dai, Q. and T. Phillippon, (2006), "Fiscal Policy and the Term Structure of Interest Rates", Working paper.

[17] Dai, Q. and K.J. Singleton, (2002), "Expectation puzzles, time varying risk premia and affine models of the term structure", *Journal of Finance, 63, pp. 415-441.*

[18] Dewachter, H. and M. Lyrio (2006), "Learning, Macroeconomic Dynamics and the Term Structure of Interest Rates", *Working paper.*

[19] Engen, E.M. and R.G. Hubbard, (2004), "Federal Government Debt and Interest Rates", *NBER Macroeconomics Annual, 19.*

[20] Evans, C.L. and D. A. Marshall, (2007), "Economic Determinants of the Nominal Treasury Yield Curve", *Journal of Monetary Economics, 54(7), pp. 1986-2003.*

[21] Eusepi, S. and B. Preston, (2008a), "Stabilizing Expectations under Monetary and Fiscal Policy Coordination", *Working paper.*

[22] Eusepi, S. and B. Preston, (2008b), "Expectations, Learning and Business Cycle Fluctuations", *NBER working paper 14181.*

[23] Evans, G. and A. Chakraborty, (2008), "Can Perpetual Learning Explain the Forward Premium Puzzle", *Journal of Monetary Economics, 55, pp. 477-490.*

[24] Evans, G. and S. Honkapohja, (2001), Learning and Expectations in Macroeconomics, *Princeton University Press, Princeton, New Jersey.*

[25] Evans, G., S. Honkapohja and K. Mitra, (2009), "Anticipated Fiscal Policy and Adaptive Learning", *Journal of Monetary Economics, forthcoming.*

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

[27] Froot, K., (1989), "New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates", *Journal of Finance, 44, pp. 283-305.*

[28] Gürkaynak, R., B. Sack and J. Wright, (2007a), "The TIPS Yield Curve and Inflation Compensation", *forthcoming, American Economic Journal: Macroeconomics.*

[29] Gürkaynak, R., B. Sack and J. Wright, (2007b), "The U.S. Treasury Yield Curve: 1961 to the Present", *Journal of Monetary Economics, 54(8), pp. 2291-2304.*

[30] Hördahl, P., O. Tristani and D. Vestin, (2008), "The Yield Curve and Macroeconomic Dynamics", *Economic Journal, 118(533), pp. 1937-1970.*

[31] Kreps, D., (1998), "Anticipated Utility and Dynamic Choice", *1997 Schwartz Lecture. In: D.P. Jacobs, E. Kalai and M. Kamien, Editors, Frontiers of Research in Economic Theory, Cambridge Univ. Press, Cambridge, England.*

[32] Laubach, T., R. J. Tetlow and J.C. Williams, (2007), "Learning and the Role of Macroeconomic Factors in the Term Structure of Interest Rates", *Working paper.*

[33] Lucas, R.E., (1978), "Asset Prices in Exchange Economy", *Econometrica, 46, 1426-1445.*

[34] Macaulay, F. R., (1938), Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields and Stock Prices in the United States since 1856, *NBER.*

[35] Mankiw, N.G., R. Reis and J. Wolfers, (2004), "Disagreement about Inflation Expectations", *NBER Macroeconomics Annual 2003, 18, pp. 209-248.*

[36] Mankiw, N.G. and L.H. Summers, (1984), "Do Long-Term Interest Rates Overreact to Short-Term Interest Rates?", *Brookings Papers on Economics Activity, Economic Studies Program, The Brookings Institution, Vol. 15, pp. 223-248.*

[37] Marcet, A. and T. J. Sargent, (1989), "Convergence of Least-Squares Learning in Environments with Hidden State Variables and Private Information", *Journal of Political Economy, pp.1306-1322.*

[38] Milani, F., (2007), "Expectations, Learning and Macroeconomic Persistence", *Journal of Monetary Economics, 54(7), pp. 2065-2082.*

[39] Muth, J.F., (1961), "Rational Expectations and the Theory of Price Movements", *Econometrica, 29, pp. 315-335.*

[40] Orphanides, A. and J.C. Williams, (2005), "The Decline of Activist Stabilization Policy: Natural Rate Misperceptions, Learning and Expectations", *Journal of Economic Dynamics and Control, 29(11), pp. 1927-1950.*

[41] Pasaogullari, M. and S. Tsonev, (2009), "The Term Structure of Inflation Compensation in the Nominal Yield Curve", *Working paper.*

[42] Piazzesi, M. and M. Schneider, (2007), "Equilibrium Yield Curves", *in Daron Acemoglu, Kenneth Rogoff, and Michael Woodford, NBER Macroeconomics Annual 2006 , Cambridge MA: MIT press p. 389-442*

[43] Piazzesi, M. and M. Schneider, (2009), "Trend and Cycle in Bond Premia", *Working paper.*

[44] Preston, B., (2005), "Learning about Monetary Policy Rules When Long-Horizon Expectations Matter", *International Journal of Central Banking, 1.*

[45] Rabanal, P. and J.F. Rubio-Ramirez, (2005), "Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach", *Journal of Monetary Economics, 52(6), pp. 1151-1166.*

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

[47] Sargent, T., (1993), Bounded Rationality in Macroeconomics, *Oxford University Press.*

[48] Seppälä, J., (2003), "The Term Structure of Real Interest Rates: Theory and Evidence from U.K. Index-Linked Bonds", *Journal of Monetary Economics, 51, pp. 1509-1549.*

[49] Smith, J. and J.B. Taylor, (2009), "The Term Structure of Policy Rules", *Journal of Monetary Economics, 56, pp. 907-917.*

[50] Taylor, J.B, (1999), "A Historical Analysis of Monetary Policy Rules", *in Monetary Policy Rules, edited by John B. Taylor.*

[51] Wachter, J., (2006), "A Consumption-Based Model of the Term Structure of Interest Rates", *Journal of Financial Economics, 79, pp. 365-399.*

[52] Woodford, M., (2003), Interest and Prices: Foundations of a Theory of Monetary Policy, *Princeton University Press.* $_{RS}$

Appendix

Appendix A.1. Intertemporal Optimization of Agents

The first order approximation of the optimality conditions of the households and firms is described below. The linearization is constructed around the following steady state: $\xi = 0$, $Y_t = \bar{Y}$ (defined below) and $\bar{P}_1^B = \beta$ (or $\bar{\imath}_1 = (1-\beta)/\beta$) with $\bar{\pi} = 1$. The hat variables denote the log deviations of the respective variable from its steady state value. For the one-period interest rate, the log deviation is defined as $\hat{\imath}_{1,t} = \log[(1+i_{1,t})/(1+\bar{\imath}_1)]$.

The first order approximation of the household's Euler equation for the one-period asset price in (14) yields:

$$\hat{C}_t^i = \tilde{E}_t \hat{C}_{t+1}^i - \sigma(\hat{\imath}_{1,t} - \hat{\pi}_{t+1}) + (g_t - \tilde{E}_t g_{t+1}). \tag{48}$$

The flow budget constraint in (6) is iterated forwards, and its approximation is:

$$\tilde{E}_t \sum\nolimits_{j=0}^{\infty} \beta^j \hat{C}_{t+j}^i = \hat{W}_t^i + \tilde{E}_t \sum\nolimits_{j=0}^{\infty} \beta^j \hat{Y}_{t+j}^i. \tag{49}$$

Substituting (48) recursively into (49) yields (22):

$$\hat{C}_t^i = (1-\beta)\hat{W}_t^i + (1-\beta)\tilde{E}_t \sum\nolimits_{j=0}^{\infty} \beta^j \left[\hat{Y}_{t+j}^i - \sigma\beta(\hat{\imath}_{1,t+j} - \hat{\pi}_{t+j+1}) + \beta(g_{t+j} - g_{t+j+1}) \right].$$

To obtain (23), apply the market clearing conditions in (21):

$$\hat{Y}_t = (1-\beta)\tilde{E}_t \sum\nolimits_{j=0}^{\infty} \beta^j \left[\hat{Y}_{t+j} - \sigma\beta(\hat{\imath}_{1,t+j} - \hat{\pi}_{t+j+1}) + \beta(g_{t+j} - g_{t+j+1}) \right] \tag{50}$$

The output gap is defined as $\hat{x}_t = \log(Y_t/Y_t^n)$, and is used to rewrite (50) as:

$$\hat{x}_t = \tilde{E}_t \sum\nolimits_{j=0}^{\infty} \beta^j \left[(1-\beta)\hat{x}_{t+j+1} - \sigma\beta(\hat{\imath}_{1,t+j} - \tilde{E}_t\hat{\pi}_{t+j+1}) + \hat{r}_{t+j+1}^n \right], \tag{51}$$

which is the expression in (23), and the natural rate of interest $\hat{r}_t^n = (\hat{Y}_{t+1}^n - g_{t+1}) - (\hat{Y}_t^n - g_t)$.

Before deriving the approximation to the firm's optimization problem, the real marginal cost function is defined using $s_{t,t+j}$ as firm $k's$l marginal cost in period $t+j$:

$$s(y, Y, \bar{\xi}) = \frac{v_h(f^{-1}(y/A; \xi)}{u_c(Y, \xi)A} \frac{1}{f'(f^{-1}(y))}, \tag{52}$$

where $\bar{\xi} \equiv (\xi, A)$ is a vector of preference and technology shocks.

When prices are fully flexible, the price of firm k is a markup over its real marginal cost:

$$\frac{p_t(k)}{P_t} = \mu s(y_t(k), Y_t, \bar{\xi}_t), \tag{53}$$

where $\mu = \theta/(\theta - 1)$.Then, in equilibrium, the firms will face the symmetric problem, so that the price set by each firm k is P_t and its output is Y_t. This implies that $s(Y_t^n, Y_t^n, \bar{\xi}_t) = \mu^{-1}$. The natural rate of output Y_t^n is thus defined. This relation is also used to define the steady state level of output $\bar{Y}$ such that $s(\bar{Y}, \bar{Y}, 0) = \mu^{-1}$.

The linearization of (52) gives:

$$\hat{s}_{t,t+j}(k) = \omega \hat{y}_{t+j}(k) + \sigma^{-1}\hat{Y}_{t+j} - (\omega + \sigma^{-1})\hat{Y}_{t+j}^n, \tag{54}$$

where $\omega > 0$ is the elasticity of the real marginal cost function $s(.)$ with respect to $y_t(k)$. Aggregating the above relation yields:

$$\hat{s}_{t+j} = (\omega + \sigma^{-1})(\hat{Y}_{t+j} - \hat{Y}_{t+j}^n). \tag{55}$$

This implies the following relation between the real marginal cost of producing $y_t(k)$ and the aggregate output Y_t :

$$\hat{s}_{t,t+j}(k) = \hat{s}_{t+j} - \omega\theta \left[\hat{p}_t(k) - \sum\nolimits_{m=t+1}^{t+j} \hat{\pi}_m\right]. \tag{56}$$

Finally, to derive (28), differentiate (13) with respect to $p_t(k)$:

$$\tilde{E}_t \sum\nolimits_{j=0}^{\infty} \alpha^j Q_{t,t+j} Y_{t+j} P_{t+j}^{\theta} \left[p_t^*(k) - \mu P_{t+j} s_{t,t+j}(k)\right] = 0. \tag{57}$$

The discount factor $Q_{t,t+j}$ is defined in (15) and using the relation in (56) gives:

$$\hat{p}_t^* = \tilde{E}_t \sum\nolimits_{j=0}^{\infty} (\alpha\beta)^j \left[\frac{1-\alpha\beta}{1+\omega\theta}(\omega + \sigma^{-1})\hat{x}_{t+j} + \hat{\pi}_{t+j}\right], \tag{58}$$

which is the relation in (28). The can be rewritten in terms of (29) using the approximation to the aggregate price index in (11): $\hat{\pi}_t = \hat{p}_t^*(1-\alpha)/\alpha$.

Appendix A.2. Proof of Proposition 2

In the flexible price limit of the benchmark model considered here, the one-period asset price is only a function of the exogenous endowment process:

$$\hat{P}_{1,t} = a_t + b_t \hat{y}_{t-1} + \eta_t.$$

With rational expectations, the fixed points of the beliefs are $\bar{a} = 0$ and $\bar{b} = \rho(1-\rho)/\sigma$, where ρ is the AR(1) parameter of the endowment process. Then, the only exogenous processes that will be forecasted by optimizing households are $\hat{P}_{1,t}$, and $\hat{y}_t$.

The actual evolution of the one period asset price will be determined as:

$$\hat{P}_{1,t} = T_0^p(a_t, b_t) + T_z^p(a_t, b_t)\hat{y}_{t-1} + \eta_t$$

The $T-$ mappings are constructed as:

$$T^{0,p}(a_t, b_t) = \frac{-\beta}{1-\beta} a_t^p \tag{59}$$

$$T^{z,p}(a_t, b_t) = \rho \left[\frac{(1-\beta)\rho}{\sigma(1-\beta\rho)} - \frac{\beta b_t^{p,y}}{1-\beta\rho} \right] + \frac{\rho}{\sigma} \tag{60}$$

The rational expectations equilibrium (REE) is defined as a fixed point of the mappings in (59) and (60):

$$T(\bar{a}_t, \bar{b}_t) = (\bar{a}_t, \bar{b}_t)$$

Evans and Honkapohja (2001) use stochastic approximation results to show that the learning algorithm in (33) converges to this REE if the following ordinary differential equation is locally stable:

$$\frac{\partial}{\partial \tau}(a_t, b_t) = T(a_t, b_t) - (a_t, b_t) \tag{61}$$

where τ is "notional time". The REE is Expectationally Stable if the differential equation in (61) is locally stable in the neighborhood of the REE. Standard results for differential equations imply that a fixed point will be locally asymptotically stable when all eigenvalues of the Jacobian $D[T(a_t, b_t) - (a_t, b_t)]$ have negative real parts. Here D is the differentiation operator and the Jacobian is evaluated at the REE of interest. The Jacobian, evaluated at $(\bar{a}_t, \bar{b}_t)$ is:

$$J(\bar{a}_t, \bar{b}_t) = \begin{bmatrix} -\frac{\beta}{1-\beta} & 0 \\ 0 & -\frac{\beta\rho}{1-\beta\rho} \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

As $J(\bar{a}_t, \bar{b}_t)$ has negative eigenvalues for $0 < \beta, \rho < 1,$ the conditions for E-stability are satisfied.

To obtain the asymptotic distribution of the parameters, I consider the system where the intercept term is not estimated.

$$\begin{aligned} b_t &= b_{t-1} + gR_t^{-1}\hat{y}_{t-1}[db_{t-1}\hat{y}_{t-1} + V_{b,t-1}\varepsilon_t] \\ R_t &= R_{t-1} + g(\hat{y}_{t-1}^2 - R_{t-1}) \end{aligned}$$

Set $S_{t-1} = R_t$, and the system is rewritten as:

$$\begin{aligned} b_t &= b_{t-1} + gS_{t-1}^{-1}\hat{y}_{t-1}[db_{t-1}\hat{y}_{t-1} + V_{b,t-1}\varepsilon_t] \\ S_t &= S_{t-1} + g(\hat{y}_t^2 - S_{t-1}) \end{aligned}$$

This is in the standard form:

$$\theta_t = \theta_{t-1} + g\mathcal{H}(\theta_{t-1}, X_t)$$

where

$$\theta_t = \begin{pmatrix} b_t \\ S_t \end{pmatrix}, \mathcal{H}(\theta_{t-1}, X_t) = \begin{pmatrix} \mathcal{H}_b(\theta_{t-1}, X_t) \\ \mathcal{H}_S(\theta_{t-1}, X_t) \end{pmatrix}, X_t = \begin{pmatrix} b_t \\ b_{t-1} \\ \varepsilon_t \end{pmatrix}$$

and

$$\begin{aligned}
\mathcal{H}_b(\theta_{t-1}, X_t) &= S_{t-1}^{-1}\hat{y}_{t-1}[db_{t-1}\hat{y}_{t-1} + V_{b,t-1}\varepsilon_t] \\
\mathcal{H}_S(\theta_{t-1}, X_t) &= (\hat{y}_t^2 - S_{t-1})
\end{aligned}$$

For infinite horizon asymptotic results, by Theorem 7.9 of Evans and Honkapohja (2001), the distribution of θ_t can be approximated for small g and large t as:

$$\theta_t \sim N(\theta^{RE}, gC)$$

where

$$\begin{aligned}
\theta^{RE} &= (\rho(1-\rho), E\hat{y}_t^2)' \\
C &= \int_0^\infty e^{sB}\mathcal{R}^* e^{sB'} ds
\end{aligned}$$

Then,

$$\begin{aligned}
h_b(b, S) &= \lim_{t\longrightarrow\infty} E\mathcal{H}_b(\theta_{t-1}, X_t) \\
&= S^{-1}E\hat{y}_t^2\left[\frac{\rho}{\sigma} + \frac{\rho^2(\beta-1)}{\sigma(1-\beta\rho)} + b\left(\frac{-\beta\rho}{1-\beta\rho} - 1\right)\right] \\
h_S(b, S) &= \lim_{t\longrightarrow\infty} E\mathcal{H}_S(\theta_{t-1}, X_t) \\
&= E\hat{y}_t^2 - S
\end{aligned}$$

Also,

$$\begin{aligned}
B &= D_\theta h(\theta^{RE}) \\
&= \begin{pmatrix} (S^{RE})^{-1}E\hat{y}_t^2\left(\frac{-1}{1-\beta\rho}\right) & \frac{-1}{(S^{RE})^2}E\hat{y}_t^2.0 \\ 0 & -1 \end{pmatrix} \\
&= \begin{pmatrix} \left(\frac{-1}{1-\beta\rho}\right) & 0 \\ 0 & -1 \end{pmatrix} \\
&= \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix}
\end{aligned}$$

$$\mathcal{R}^{ij}(\theta) = \sum_{k=-\infty}^{\infty} cov[\mathcal{H}^i(\theta, X_k^\theta), \mathcal{H}^j(\theta, X_0^\theta)]$$

Considering only b :

$$\mathcal{H}_b(\theta, X^\theta) = S_{t-1}^{-1}\hat{y}_{t-1}[db\hat{y}_{t-1} + V_b\varepsilon_t]$$

$$\mathcal{R}^{ij}(b) = \sum_{k=-\infty}^{\infty}[... + cov[\mathcal{H}^i(\theta, X_0^\theta), \mathcal{H}^j(\theta, X_0^\theta)] + cov[\mathcal{H}^i(\theta, X_1^\theta), \mathcal{H}^j(\theta, X_0^\theta)] + ...]$$

At θ^{RE} :

$$\begin{aligned}
& cov[\mathcal{H}^i(\theta, X_t^\theta), \mathcal{H}^j(\theta, X_t^\theta)] \\
= \; & E([(S^{-1})^{RE}\hat{y}_{t-1}V_b^{RE}\varepsilon_t]^2) \\
= \; & \frac{1}{(E\hat{y}_t^2)^2}\left[\frac{-\beta\rho(1-\rho)}{\sigma(1-\beta\rho)} + \frac{\rho(\beta-1)}{1-\beta\rho} + \frac{1}{\sigma}\right]^2 E(\hat{y}_{t-1}^2\varepsilon_t^2) \\
= \; & \frac{1}{(E\hat{y}_t^2)^2}\frac{(1-\rho)^2}{\sigma^2}E\hat{y}_{t-1}^2\sigma_\varepsilon^2 \\
= \; & \frac{(1-\rho)^2}{\sigma^2}\frac{\sigma_\varepsilon^2}{E\hat{y}_t^2}
\end{aligned}$$

The last step is valid since the unconditional expectation is taken as $t \longrightarrow \infty$. Also, $\hat{y}_{t-1}^2, \varepsilon_t^2$ are independent random variables, and therefore, $E(\hat{y}_{t-1}^2\varepsilon_t^2) = E\hat{y}_{t-1}^2\sigma_\varepsilon^2$. Also, the second equality is valid as

$E((S^{-1})^{RE}\hat{y}_{t-1}V_b^{RE}\varepsilon_t) = 0$. Other sums in the series are:

$$\begin{aligned}
& cov[\mathcal{H}^i(\theta, X_{t+1}^\theta), \mathcal{H}^j(\theta, X_t)] \\
= \; & E([(S^{-1})^{RE}\hat{y}_t V_b^{RE}\varepsilon_{t+1}][(S^{-1})^{RE}\hat{y}_{t-1}V_b^{RE}\varepsilon_t]) \\
= \; & \frac{1}{(E\hat{y}_t^2)^2}\frac{(1-\rho)^2}{\sigma^2}E(\hat{y}_t\varepsilon_{t+1}\hat{y}_{t-1}\varepsilon_t) = 0
\end{aligned}$$

as

$$\begin{aligned}
\hat{y}_t &= \rho\hat{y}_{t-1} + \varepsilon_t \\
\hat{y}_t\varepsilon_{t+1}\hat{y}_{t-1}\varepsilon_t &= \rho\hat{y}_{t-1}^2\varepsilon_{t+1}\varepsilon_t + \varepsilon_t^2\varepsilon_{t+1}\hat{y}_{t-1} \\
E(\hat{y}_t\varepsilon_{t+1}\hat{y}_{t-1}\varepsilon_t) &= \rho E(\hat{y}_{t-1}^2)E(\varepsilon_{t+1})E(\varepsilon_t) + E(\varepsilon_t^2)E(\varepsilon_{t+1})E(\hat{y}_{t-1}) \\
&= 0
\end{aligned}$$

Therefore, for b :

$$\mathcal{R}^*(b) = \frac{(1-\rho)^2}{\sigma^2}\frac{\sigma_\varepsilon^2}{E\hat{y}_t^2}$$

To find C :

$$\begin{aligned}
& \frac{(1-\rho)^2}{\sigma^2}e^{sb_{11}}\frac{\sigma_\varepsilon^2}{E\hat{y}_t^2}e^{sb_{11}} \\
= \; & \frac{(1-\rho)^2}{\sigma^2}e^{2sb_{11}}\frac{\sigma_\varepsilon^2}{E\hat{y}_t^2}
\end{aligned}$$

and

$$\begin{aligned}
C &= \frac{(1-\rho)^2}{\sigma^2}\int_0^\infty e^{2sb_{11}}\frac{\sigma_\varepsilon^2}{E\hat{y}_t^2}ds \\
&= \frac{(1-\rho)^2}{\sigma^2}\frac{(1-\beta\rho)(1-\rho^2)}{2}
\end{aligned}$$

Appendix A.3. Proof of Proposition 3

For simplicity, I first consider the Expectations Hypothesis regression for $n = 2$:

$$y_{1,t+1} - y_{2,t} = \alpha + \gamma(y_{2,t} - y_{1,t}) + e_t \tag{62}$$

where

$$y_{1,t} = -T_{b,t-1}\hat{y}_{t-1} - V_{b,t-1}\varepsilon_t$$

Longer yields:

$$\begin{aligned} y_{1,t+1} &= -T_{b,t}\hat{y}_t - V_{b,t}\varepsilon_{t+1} \\ y_{2,t} &= \frac{1}{2}(y_{1,t} + E_t y_{1,t+1}) \end{aligned}$$

In the regression in (62), define X and Y where

$$\begin{aligned} X &= y_{2,t} - y_{1,t} \\ &= \frac{1}{2}(y_{1,t} + E_t y_{1,t+1}) - y_{1,t} \\ &= -\frac{1}{2}y_{1,t} + \frac{1}{2}(-T_{b,t-1}\hat{y}_t) \\ &= -\frac{1}{2}\left[-T_{b,t-1}\hat{y}_{t-1} - V_{b,t-1}\varepsilon_t\right] - \frac{1}{2}T_{b,t-1}\hat{y}_t \\ &= \frac{1}{2}T_{b,t-1}\hat{y}_{t-1} - \frac{1}{2}T_{b,t-1}(\rho\hat{y}_{t-1} + \varepsilon_t) + \frac{1}{2}V_{b,t-1}\varepsilon_t \\ &= \hat{y}_{t-1}\left[\frac{1}{2}T_{b,t-1} - \frac{1}{2}T_{b,t-1}\rho\right] + \frac{\varepsilon_t}{2}(V_{b,t-1} - T_{b,t-1}) \\ &= \frac{\hat{y}_{t-1}}{2}T_{b,t-1}(1-\rho) - \frac{\varepsilon_t}{2}(T_{b,t-1} - V_{b,t-1}) \\ Y &= y_{1,t+1} - y_{2,t} \\ &= \hat{y}_{t-1}\left[-T_{b,t}\rho + \frac{1}{2}T_{b,t-1}(1+\rho)\right] + \varepsilon_t(-T_{b,t} + \frac{1}{2}(V_{b,t-1} + T_{b,t-1})) - V_{b,t}\varepsilon_{t+1} \end{aligned}$$

For computing the asymptotic bias in γ:

$$\begin{aligned} bias &= \frac{p\lim_{T\longrightarrow\infty} T^{-1}\sum_{t=1}^{T}(X_t e_t) - -p\lim_{T\longrightarrow\infty}\left(T^{-1}\sum_{t=1}^{T}X_t\right)p\lim_{T\longrightarrow\infty}\left(T^{-1}\sum_{t=1}^{T}e_t\right)}{p\lim_{T\longrightarrow\infty}(T^{-1}\sum_{t=1}^{T}X_t^2 - (T^{-1}\sum_{t=1}^{T}X_t)^2)} \\ &= \frac{\begin{array}{c} p\lim_{T\longrightarrow\infty} T^{-1}\sum_{t=1}^{T}(X_t[\varepsilon_t(-T_{b,t} + \frac{1}{2}(V_{b,t-1} + T_{b,t-1})) - V_{b,t}\varepsilon_{t+1}]) \\ -p\lim_{T\longrightarrow\infty}\left(T^{-1}\sum_{t=1}^{T}X_t\right)p\lim_{T\longrightarrow\infty}\left(T^{-1}\sum_{t=1}^{T}e_t\right) \end{array}}{p\lim_{T\longrightarrow\infty}(T^{-1}\sum_{t=1}^{T}X_t^2 - (T^{-1}\sum_{t=1}^{T}X_t)^2)} \end{aligned}$$

Consider the following term:

$$T^{-1}\sum_{t=1}^{T}X_t = T^{-1}\sum_{t=1}^{T}\left[\frac{1}{2}T_{b,t-1}(1-\rho)\hat{y}_{t-1} - \frac{1}{2}\varepsilon_t(T_{b,t-1} - V_{b,t-1})\right]$$

From Evans and Honkapohja (2001), when the E-stability conditions are satisfied (as they are for the present case), for the constant gain algorithm, the following results are obtained: (a) the estimates of the updated coefficients are unbiased asymptotically, i.e., $E(b_t) \longrightarrow b^{RE}$ as $t \longrightarrow \infty$; (b) b_t approaches a limiting normal distribution $N(b^{RE}, gC)$ where C is the variance-covariance matrix of the updated coefficients.

From adaptive learning:

$$b_t = b_{t-1} + gR_t^{-1}\hat{y}_{t-1}[(T_{b,t-1} - b_{t-1})\hat{y}_{t-1} + V_{b,t-1}\varepsilon_t]$$

I further assume that R_t is not updated, and is equal to the RE value: $\bar{R}$. The relevant T-mappings are:

$$\begin{aligned} T_{b,t} &= \frac{\rho}{\sigma} + \rho\left[\frac{-\beta b_t}{1-\beta\rho} + \frac{\rho(\beta-1)}{\sigma(1-\beta\rho)}\right] \\ V_{b,t} &= \left[\frac{-\beta b_t}{1-\beta\rho} + \frac{\rho(\beta-1)}{\sigma(1-\beta\rho)}\right] + \frac{1}{\sigma} \end{aligned}$$

Then,

$$\begin{aligned} & p \lim_{T\longrightarrow\infty} T^{-1}\sum_{t=1}^{T}(T_{b,t-1}\hat{y}_{t-1}) \\ =& \; p \lim_{T\longrightarrow\infty} T^{-1}\sum_{t=1}^{T}\left[\left(\frac{\rho}{\sigma} + \rho\left[\frac{-\beta b_{t-1}}{1-\beta\rho} + \frac{\rho(\beta-1)}{\sigma(1-\beta\rho)}\right]\right)\hat{y}_{t-1}\right] \\ =& \; -\frac{\beta\rho}{1-\beta\rho} p \lim_{T\longrightarrow\infty} T^{-1}\sum_{t=1}^{T}(b_{t-1}\hat{y}_{t-1}) \end{aligned}$$

Here, noting that the AR process can be rewritten as:

$$\hat{y}_t = \sum_{j=0}^{\infty}\rho^j\varepsilon_{t-j}, \text{ for } \hat{y}_0 = 0$$

Thus,

$$\begin{aligned} & p \lim_{T\longrightarrow\infty} T^{-1}\sum_{t=1}^{T}(T_{b,t-1}\hat{y}_{t-1}) \\ =& \; -\frac{\beta\rho}{1-\beta\rho} p \lim_{T\longrightarrow\infty} T^{-1}\sum_{t=1}^{T}\left(b_{t-1}\sum_{j=0}^{\infty}\rho^j\varepsilon_{t-1-j}\right) \\ =& \; -\frac{\beta\rho}{1-\beta\rho} p \lim_{T\longrightarrow\infty} T^{-1}\sum_{t=1}^{T}(b_{t-1}\varepsilon_{t-1} + b_{t-1}\varepsilon_{t-2} + b_{t-1}\varepsilon_{t-3} + ...) \end{aligned}$$

The first term is:

$$\begin{aligned} (b_{t-1}\varepsilon_{t-1}) &= (b_{t-2}\varepsilon_{t-1} + g\bar{R}^{-1}\hat{y}_{t-2}[(T_{b,t-2} - b_{t-2})\hat{y}_{t-2} + V_{b,t-2}\varepsilon_{t-1}]\varepsilon_{t-1}) \\ &= (b_{t-2}\varepsilon_{t-1}) + gR_t^{-1}(\hat{y}_{t-1}^2(T_{b,t-2} - b_{t-2})\varepsilon_{t-1}) \\ &\quad + gR_t^{-1}(\hat{y}_{t-2}V_{b,t-2}\varepsilon_{t-1}\varepsilon_{t-1}) \end{aligned}$$

Since further recursions of b_t yield

$$\begin{aligned} b_{t-1} &= b_{t-2} + gR_{t-1}^{-1}\hat{y}_{t-2}[(T_{b,t-2} - b_{t-2})\hat{y}_{t-2} + V_{b,t-2}\varepsilon_{t-1}] \\ b_{t-2} &= b_{t-3} + gR_{t-2}^{-1}\hat{y}_{t-3}[(T_{b,t-3} - b_{t-3})\hat{y}_{t-3} + V_{b,t-3}\varepsilon_{t-2}] \end{aligned}$$

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T}(b_{t-1}\varepsilon_{t-1}) &= T^{-1}\sum_{t=1}^{T}\left[\left(b_{t-2}+gR_{t-1}^{-1}\hat{y}_{t-2}[(T_{b,t-2}-b_{t-2})\hat{y}_{t-2}+V_{b,t-2}\varepsilon_{t-1}]\right)\varepsilon_{t-1}\right] \\
&= T^{-1}\sum_{t=1}^{T}\left[b_{t-2}\varepsilon_{t-1}+gR_{t-1}^{-1}\hat{y}_{t-2}\varepsilon_{t-1}[(T_{b,t-2}-b_{t-2})\hat{y}_{t-2}+V_{b,t-2}\varepsilon_{t-1}]\right]
\end{aligned}
$$

Notice that $(b_{t-2}, \varepsilon_{t-1})$ are independent r.vs, as are $(\hat{y}_{t-2}, \varepsilon_{t-1})$. The first term on the RHS can be rewritten as:

$$
T^{-1}\sum_{t=1}^{T} b_{t-2}\varepsilon_{t-1} = (AB)^{-1}\sum_{b=1}^{B}\sum_{a=1}^{A} b_{A(b-2)+a-1}\varepsilon_{A(b-2)}
$$

where $T = AB$. For large A and B, and small $\bar{g}$, $b_{A(b-2)+a-1} \approx b_{A(b-2)}$:

$$
T^{-1}\sum_{t=1}^{T} b_{t-2}\varepsilon_{t-1} = B^{-1}\sum_{b=1}^{B} b_{A(b-2)} A^{-1}\sum_{a=1}^{A}\varepsilon_{A(b-2)+a}
$$

Using the weak law of large numbers:

$$
A^{-1}\sum_{a=1}^{A}\varepsilon_{A(b-2)+a} \xrightarrow{p} E(\varepsilon_t) = 0,\ B^{-1}\sum_{b=1}^{B} b_{A(b-2)} \xrightarrow{p} E(b_t) = b^{RE}
$$

I get

$$
T^{-1}\sum_{t=1}^{T} b_{t-2}\varepsilon_{t-1} = 0
$$

Similar reasoning for the other terms will imply:

$$
T^{-1}\sum_{t=1}^{T}(T_{b,t-1}\hat{y}_{t-1}) = 0
$$

For the second term in $T^{-1}\sum_{t=1}^{T}(X_t)$:

$$
\begin{aligned}
& T^{-1}\sum_{t=1}^{T}[\varepsilon_t(T_{b,t-1}-V_{b,t-1})] \\
=\ & T^{-1}\sum_{t=1}^{T}\left[\varepsilon_t\left(\left(\frac{\rho}{\sigma}-\frac{1}{\sigma}\right)+(\rho-1)\left[\frac{-\beta b_{t-1}}{1-\beta\rho}+\frac{\rho(\beta-1)}{1-\beta\rho}\right]\right)\right] \\
=\ & T^{-1}\sum_{t=1}^{T}(\varepsilon_t(1-\rho)\frac{\beta b_{t-1}}{1-\beta\rho}) \\
=\ & (1-\rho)\frac{\beta}{1-\beta\rho}T^{-1}\sum_{t=1}^{T}(\varepsilon_t b_{t-1})
\end{aligned}
$$

Now $T^{-1}\sum_{t=1}^{T}(\varepsilon_t b_{t-1})$ can be written as:

$$\begin{aligned} & T^{-1}\sum_{t=1}^{T}[\varepsilon_t(b_{t-2}+gR_t^{-1}\hat{y}_{t-2}[(T_{b,t-2}-b_{t-2})\hat{y}_{t-2}+V_{b,t-2}\varepsilon_{t-1}]) \\ = \; & T^{-1}\sum_{t=1}^{T}(\varepsilon_t b_{t-2})+gR_t^{-1}T^{-1}\sum_{t=1}^{T}(\varepsilon_t\hat{y}_{t-2}[(T_{b,t-2}-b_{t-2})\hat{y}_{t-2}+V_{b,t-2}\varepsilon_t]) \end{aligned}$$

And

$$p\lim_{T\longrightarrow\infty}T^{-1}\sum_{t=1}^{T}(\varepsilon_t b_{t-1})=0$$

Therefore,

$$p\lim_{T\longrightarrow\infty}T^{-1}\sum_{t=1}^{T}(X_t)=0$$

Similar reasoning as above implies:

$$p\lim_{T\longrightarrow\infty}T^{-1}\sum_{t=1}^{T}(e_t)=0$$

Now for $T^{-1}\sum_{t=1}^{T}(X_t e_t)$:

$$\begin{aligned} & T^{-1}\sum_{t=1}^{T}(X_t e_t) \\ = \; & T^{-1}\sum_{t=1}^{T}\left[\begin{array}{c} \frac{1}{2}T_{b,t-1}(1-\rho)\hat{y}_{t-1}\varepsilon_t(-T_{b,t}+\frac{1}{2}(V_{b,t-1}+T_{b,t-1})) \\ +\frac{1}{2}T_{b,t-1}(1-\rho)\hat{y}_{t-1}V_{b,t}\varepsilon_{t+1} \\ -\frac{1}{2}\varepsilon_t(T_{b,t-1}-V_{b,t-1})\varepsilon_t(-T_{b,t}+\frac{1}{2}(V_{b,t-1}+T_{b,t-1})) \\ -\frac{1}{2}\varepsilon_t(T_{b,t-1}-V_{b,t-1})V_{b,t}\varepsilon_{t+1} \end{array}\right] \end{aligned}$$

Consider the following:

$$\begin{aligned} b_t &= b_{t-1}+g\bar{R}^{-1}\hat{y}_{t-1}[(T_{b,t-1}-b_{t-1})\hat{y}_{t-1}+V_{b,t-1}\varepsilon_t] \\ &= b_{t-1}+g\bar{R}^{-1}\hat{y}_{t-1}[(T_{b,t-1}-b_{t-1})\hat{y}_{t-1}+V_{b,t-1}\varepsilon_t] \\ b_t^2 &= b_{t-1}^2+(g\bar{R}^{-1}\hat{y}_{t-1}[(T_{b,t-1}-b_{t-1})\hat{y}_{t-1}+V_{b,t-1}\varepsilon_t])^2 \\ & \quad +2b_{t-1}g\bar{R}^{-1}\hat{y}_{t-1}[(T_{b,t-1}-b_{t-1})\hat{y}_{t-1}+V_{b,t-1}\varepsilon_t] \end{aligned}$$

Now, b_t^2 and ε_t are independent rvs as b_t and ε_t are independent since the square function is Borel measurable.Therefore:

$$p\lim_{T\longrightarrow\infty}T^{-1}\sum_{t=1}^{T}(b_t^2\varepsilon_t)=p\lim_{T\longrightarrow\infty}\left[T^{-1}\sum_{t=1}^{T}(b_t^2)T^{-1}\sum_{t=1}^{T}(\varepsilon_t)\right]=0$$

For $b_{t+1}b_t\varepsilon_{t+1}\varepsilon_t$:

$$b_{t+1}b_t\varepsilon_{t+1}\varepsilon_t=b_t^2\varepsilon_{t+1}\varepsilon_t+gR_{t+1}^{-1}\hat{y}_t b_t\varepsilon_{t+1}\varepsilon_t[f(b_t)\hat{y}_t+V_{b,t}\varepsilon_{t+1}]$$

As ε_t are iid, $\varepsilon_{t+1}\varepsilon_t$ can be treated as a new rv. Here, as b_t^2 is independent of ε_{t+1}, and of ε_t, and $\varepsilon_{t+1},\varepsilon_t$

are independent of each other:

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}(b_t^2\varepsilon_{t+1}\varepsilon_t) &= T^{-1}\sum_{t=1}^{T}(b_t^2)T^{-1}\sum_{t=1}^{T}(\varepsilon_{t+1})T^{-1}\sum_{t=1}^{T}(\varepsilon_t) = 0 \\ & \quad and, \\ T^{-1}\sum_{t=1}^{T}(b_t^2\varepsilon_t^2) &= T^{-1}\sum_{t=1}^{T}(b_t^2)T^{-1}\sum_{t=1}^{T}(\varepsilon_t^2) = \sigma_\varepsilon^2 T^{-1}\sum_{t=1}^{T}(b_t^2) \end{aligned}$$

Similiar arguments apply to other terms. Then, $T^{-1}\sum_{t=1}^{T}(X_t e_t)$ reduces to:

$$T^{-1}\sum_{t=1}^{T}(X_t e_t) = -\frac{1}{2}T^{-1}\sum_{t=1}^{T}[\varepsilon_t^2(T_{b,t-1}-V_{b,t-1})(-T_{b,t}+\frac{1}{2}(V_{b,t-1}+T_{b,t-1}))]$$

Now,

$$T_{b,t-1}-V_{b,t-1} = \left(\frac{\rho}{\sigma}-\frac{1}{\sigma}\right)+(\rho-1)\left[\frac{-\beta b_{t-1}}{1-\beta\rho}+\frac{\rho(\beta-1)}{\sigma(1-\beta\rho)}\right]$$

Then,

$$\begin{aligned} & (T_{b,t-1}-V_{b,t-1})T_{b,t} \\ = & \left[\left(\frac{\rho}{\sigma}-\frac{1}{\sigma}\right)+(1-\rho)\left[\frac{\beta b_{t-1}}{1-\beta\rho}+\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\right]\right]\left[\frac{\rho}{\sigma}-\rho\left[\frac{\beta b_t}{1-\beta\rho}+\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\right]\right] \\ = & \left(\frac{\rho}{\sigma}-\frac{1}{\sigma}\right)\frac{\rho}{\sigma}-\left(\frac{\rho}{\sigma}-\frac{1}{\sigma}\right)\rho\left[\frac{\beta b_t}{1-\beta\rho}+\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\right] \\ & +(1-\rho)\frac{\rho}{\sigma}\left[\frac{\beta b_{t-1}}{1-\beta\rho}+\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\right]-(1-\rho)\left[\frac{\beta b_{t-1}}{1-\beta\rho}+\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\right]\rho\left[\frac{\beta b_t}{1-\beta\rho}+\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\right] \end{aligned}$$

Therefore,

$$
\begin{aligned}
& T^{-1}\sum_{t=1}^{T}[(T_{b,t-1}-V_{b,t-1})T_{b,t}] \\
= \quad & T^{-1}\sum_{t=1}^{T}\left[\begin{array}{c} \left(\frac{\rho}{\sigma}-\frac{1}{\sigma}\right)\frac{\rho}{\sigma}-\left(\frac{\rho}{\sigma}-\frac{1}{\sigma}\right)\rho\left[\frac{\beta b_t}{1-\beta\rho}+\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\right] \\ +(1-\rho)\frac{\rho}{\sigma}\left[\frac{\beta b_{t-1}}{1-\beta\rho}+\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\right]-(1-\rho)\rho\left[\frac{\beta b_{t-1}}{1-\beta\rho}+\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\right]\left[\frac{\beta b_t}{1-\beta\rho}+\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\right] \end{array}\right] \\
= \quad & \left(\frac{\rho}{\sigma}-\frac{1}{\sigma}\right)\frac{\rho}{\sigma}-\left(\frac{\rho}{\sigma}-\frac{1}{\sigma}\right)\rho\left[\frac{\beta\rho(1-\rho)}{\sigma(1-\beta\rho)}+\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\right] \\
& +(1-\rho)\frac{\rho}{\sigma}\left[\frac{\beta\rho(1-\rho)}{\sigma(1-\beta\rho)}+\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\right] \\
& -(1-\rho)\rho T^{-1}\sum_{t=1}^{T}\left(\frac{\beta b_{t-1}}{1-\beta\rho}\frac{\beta b_t}{1-\beta\rho}\right)-(1-\rho)\rho T^{-1}\sum_{t=1}^{T}\left(\frac{\beta b_{t-1}}{1-\beta\rho}\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\right) \\
& -(1-\rho)\rho T^{-1}\sum_{t=1}^{T}\left(\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\frac{\beta b_t}{1-\beta\rho}\right)-(1-\rho)\rho\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)} \\
= \quad & \left(\frac{\rho}{\sigma}-\frac{1}{\sigma}\right)\frac{\rho}{\sigma}-\left(\frac{\rho}{\sigma}-\frac{1}{\sigma}\right)\frac{\rho}{\sigma}+(1-\rho)\left(\frac{\rho}{\sigma}\right)^2 \\
& -(1-\rho)\rho T^{-1}\sum_{t=1}^{T}\left(\frac{\beta b_{t-1}}{1-\beta\rho}\frac{\beta b_t}{1-\beta\rho}\right)-(1-\rho)\rho\frac{\beta}{1-\beta\rho}\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\frac{\rho(1-\rho)}{\sigma} \\
& -(1-\rho)\rho\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\frac{\beta}{1-\beta\rho}\frac{\rho(1-\rho)}{\sigma}-(1-\rho)\rho\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)}\frac{\rho(1-\beta)}{\sigma(1-\beta\rho)} \\
= \quad & (1-\rho)\left(\frac{\rho}{\sigma}\right)^2-(1-\rho)\rho T^{-1}\sum_{t=1}^{T}\left(\frac{\beta b_{t-1}}{1-\beta\rho}\frac{\beta b_t}{1-\beta\rho}\right) \\
& -2\frac{\rho^2(1-\rho)^2\beta\rho(1-\beta)}{\sigma^2(1-\beta\rho)^2}-\frac{\rho^3(1-\rho)(1-\beta)^2}{\sigma^2(1-\beta\rho)^2}
\end{aligned}
$$

Also,

$$
\begin{aligned}
& (T_{b,t-1}^2 - V_{b,t-1}^2) \\
= \; & \left(\frac{\rho}{\sigma} + \rho\left[\frac{-\beta b_{t-1}}{1-\beta\rho} + \frac{\rho(\beta-1)}{\sigma(1-\beta\rho)}\right]\right)^2 \\
& -\left(\frac{1}{\sigma} + \left[\frac{-\beta b_{t-1}}{1-\beta\rho} + \frac{\rho(\beta-1)}{\sigma(1-\beta\rho)}\right]\right)^2 \\
= \; & b_{t-1}^2\left[\left(\frac{-\beta\rho}{1-\beta\rho}\right)^2 - \left(\frac{-\beta}{1-\beta\rho}\right)^2\right] + \\
& -2b_t\left[\left(\frac{\beta\rho}{1-\beta\rho}\left(\frac{\rho^2(\beta-1)}{\sigma(1-\beta\rho)} + \frac{\rho}{\sigma}\right)\right) + \left(\left(\frac{-\beta}{1-\beta\rho}\right)\left(\frac{\rho(\beta-1)}{\sigma(1-\beta\rho)} + \frac{1}{\sigma}\right)\right)\right] \\
& +\left[\left(\frac{\rho^2(\beta-1)}{\sigma(1-\beta\rho)} + \frac{\rho}{\sigma}\right)^2 - \left(\frac{\rho(\beta-1)}{1-\beta\rho} + \frac{1}{\sigma}\right)^2\right] \\
= \; & -b_t^2\left[\frac{\beta^2(1-\rho^2)}{(1-\beta\rho)^2}\right] + 2b_t\left[\frac{\beta(1-\rho)}{\sigma(1-\beta\rho)^2}(1-\rho^2)\right] \\
& -\frac{(1-\rho)^2}{\sigma^2(1-\beta\rho)^2}\left[1-\rho^2\right]
\end{aligned}
$$

Note that $E(b_t^2)$ as $t \longrightarrow \infty$,:

$$
\begin{aligned}
& E(b_t^2) = [V(b_t) + (E(b_t))^2] \\
= \; & gC + \frac{\rho^2(1-\rho)^2}{\sigma^2}
\end{aligned}
$$

Then,

$$
\begin{aligned}
& T^{-1}\sum_{t=1}^{T}(T_{b,t}^2 - V_{b,t}^2) \\
= \; & \left[\frac{\beta^2(\rho^2-1)}{(1-\beta\rho)^2}\right]\left(gC + \rho^2(1-\rho^2)\right) \\
& +2\frac{\rho(1-\rho)}{\sigma}\left[\frac{\beta(1-\rho)}{\sigma(1-\beta\rho)^2}(1-\rho^2)\right] \\
& -\frac{(1-\rho)^2}{\sigma^2(1-\beta\rho)^2}\left[1-\rho^2\right]
\end{aligned}
$$

Finally, we have:

$$
\begin{aligned}
& T^{-1}\sum_{t=1}^{T}(X_t e_t) \\
= & -\frac{1}{2}T^{-1}\sum_{t=1}^{T}[-\varepsilon_t^2(T_{b,t-1}-V_{b,t-1})T_{b,t}+\frac{1}{2}\varepsilon_t^2(T_{b,t-1}^2-V_{b,t-1}^2)] \\
= & \frac{1}{2}E(\varepsilon_t^2)E((T_{b,t-1}-V_{b,t-1})T_{b,t}) \\
& +\frac{1}{2}E(\varepsilon_t^2)E(T_{b,t-1}^2-V_{b,t-1}^2) \\
= & \frac{\sigma_\varepsilon^2}{2}\left[\begin{array}{c} (1-\rho)\left(\frac{\rho}{\sigma}\right)^2-(1-\rho)\rho T^{-1}\sum_{t=1}^{T}\left(\frac{\beta b_{t-1}}{1-\beta\rho}\frac{\beta b_t}{1-\beta\rho}\right) \\ -2\frac{\rho^2(1-\rho)^2\beta\rho(1-\beta)}{\sigma^2(1-\beta\rho)^2}-\frac{\rho^3(1-\rho)(1-\beta)^2}{\sigma^2(1-\beta\rho)^2} \end{array}\right] \\
& +\frac{\sigma_\varepsilon^2}{2}\left[\begin{array}{c} \left[\frac{\beta^2(\rho^2-1)}{(1-\beta\rho)^2}\right]\left(gC+\rho^2(1-\rho^2)\right) \\ +2\frac{\rho(1-\rho)}{\sigma}\left[\frac{\beta(1-\rho)}{\sigma(1-\beta\rho)^2}(1-\rho^2)\right] \\ -\frac{(1-\rho)^2}{\sigma^2(1-\beta\rho)^2}\left[1-\rho^2\right] \end{array}\right]
\end{aligned}
$$

To determine $E(b_t b_{t+1})$:

$$
\begin{aligned}
b_{t+1} &= b_t+gR_{t+1}^{-1}\hat{y}_t[(T_{b,t}-b_t)\hat{y}_t+V_{b,t}\varepsilon_{t+1}] \\
b_{t+1}b_t &= b_t^2+gR_{t+1}^{-1}\hat{y}_t[(T_{b,t}-b_t)\hat{y}_t b_t+V_{b,t}\varepsilon_{t+1}b_t] \\
E(b_{t+1}b_t) &= E(b_t^2)+gR_{t+1}^{-1}E((T_{b,t}-b_t)b_t\hat{y}_t^2) \\
& \quad +gR_{t+1}^{-1}E(\hat{y}_t V_{b,t}\varepsilon_{t+1}b_t)
\end{aligned}
$$

The remaining terms are:

$$
\begin{aligned}
& E((T_{b,t}-b_t)b_t\hat{y}_t^2) \\
= & E\left(\left[\rho+\frac{\rho^2(\beta-1)}{1-\beta\rho}+b_t\left(\frac{-\beta\rho}{1-\beta\rho}\right)\right]b_t\hat{y}_t^2\right)
\end{aligned}
$$

As $\hat{y}_t^2$ can be written in terms of the errors, similar arguments as above imply that $E(b_t\hat{y}_t^2)$ and $E(b_t^2\hat{y}_t^2)=0$. Thus,

$$E(b_t b_{t+1})=E(b_t^2)$$

Therefore,

$$\frac{p\lim_{T\longrightarrow\infty}T^{-1}\sum_{t=1}^{T}(X_t e_t)}{p\lim_{T\longrightarrow\infty}(T^{-1}\sum_{t=1}^{T}X_t^2}-1=\frac{\sigma_\varepsilon^2}{2}\left[A-\frac{\beta^2(1-\rho)}{(1-\beta\rho)^2}E(b_t^2)\right]-1$$

where

$$
\begin{aligned}
A = & (1-\rho)\left(\frac{\rho}{\sigma}\right)^2-2\frac{\rho^2(1-\rho)^2\beta\rho(1-\beta)}{\sigma^2(1-\beta\rho)^2}-\frac{\rho^3(1-\rho)(1-\beta)^2}{\sigma^2(1-\beta\rho)^2} \\
& +2\frac{\rho(1-\rho)}{\sigma}\left[\frac{\beta(1-\rho)}{\sigma(1-\beta\rho)^2}(1-\rho^2)\right]-\frac{(1-\rho)^2}{\sigma^2(1-\beta\rho)^2}\left[1-\rho^2\right]
\end{aligned}
$$

is negative. Thus, the bias is negative, and the slope estimator γ is less than one.

For longer yields, the same reasoning applies, as the $T-$mappings are monotonic transformations of the mappings considered here.

Appendix A.4.: Dynamic response of yields for flexible price economy

Here I show the response of long and short yields, in response to a positive endowment shock for the flexible price economy ($\kappa \longrightarrow \infty$).

Under Lucas economy, with only the slope coefficient being updated, the one-period price process is:

$$P_{1,t} = T(b_{t-1})\hat{y}_{t-1} + V(b_{t-1})\varepsilon_t$$

where

$$\hat{y}_t = \rho\hat{y}_{t-1} + \varepsilon_t$$

In the RE case, $\bar{T} = \frac{\rho(1-\rho)}{\sigma}$, and $\bar{V} = \frac{(1-\rho)}{\sigma}$

Assume that the ρ parameter is known. The T mapping is

$$\rho\left[-\frac{(1-\beta)\rho}{\sigma(1-\beta\rho)} - \frac{\beta b_t}{1-\beta\rho}\right] + \frac{\rho}{\sigma}$$

For $\sigma = 1$:

$$\begin{aligned} T &= \rho\left[-\frac{(1-\beta)\rho}{(1-\beta\rho)} - \frac{\beta b_t}{1-\beta\rho}\right] + \rho \\ &= \frac{\rho(1-\rho)}{(1-\beta\rho)} - \frac{\beta\rho b_t}{1-\beta\rho} \end{aligned}$$

When there is a positive endowment shock, given that this is a endowment economy, with no production, the price of the one period bond must rise. At time t, when the shock hits, and if beliefs were at the RE mean. Yields in this framework are:

$$\begin{aligned} y_{1,t} &= -T(b_{t-1})\hat{y}_{t-1} - V(b_{t-1})\varepsilon_t \\ \frac{\partial y_{1,t}}{\partial \varepsilon_t} &= -V(b_{t-1}) \end{aligned}$$

and the two period yield is

$$\begin{aligned} y_{2,t} &= \frac{1}{2}\left[y_{1,t} + E_t y_{1,t+1}\right] \\ &= \frac{1}{2}\left[y_{1,t} - T(b_{t-1})\hat{y}_t\right] \\ &= \frac{1}{2}\left[-T(b_{t-1})\hat{y}_{t-1} - V(b_{t-1})\varepsilon_t - T(b_{t-1})\hat{y}_t\right] \\ \frac{\partial y_{2,t}}{\partial \varepsilon_t} &= \frac{1}{2}\left[-V(b_{t-1}) - T(b_{t-1})\right] \end{aligned}$$

In response to a positive endowment shock under RE, prices of bonds will rise, or correspondingly the yields will fall.

The RHS of the CS regression is:

$$\begin{aligned} RHS &= y_{2,t} - y_{1,t} \\ &= \frac{1}{2}\left[y_{1,t} + E_t y_{1,t+1}\right] - y_{1,t} \\ &= \frac{1}{2}\left[E_t y_{1,t+1} - y_{1,t}\right] \\ &= \frac{1}{2}\left[-T(b_{t-1})\hat{y}_t + T(b_{t-1})\hat{y}_{t-1} + V(b_{t-1})\varepsilon_t\right] \end{aligned}$$

Response to a shock at time t :

$$\frac{\partial RHS}{\partial \varepsilon_t} = -\frac{1}{2}T(b_{t-1}) + \frac{1}{2}V(b_{t-1})$$

Assuming that beliefs were at their RE means at time of impact, the change in the spread is the same for learning and RE:

$$\begin{aligned} \frac{\partial RHS}{\partial \varepsilon_t} &= -\frac{1}{2}\rho(1-\rho) + \frac{1}{2}(1-\rho) \\ &= \frac{1}{2}(1-\rho)^2 > 0 \end{aligned}$$

The LHS of the CS regression is

$$LHS = y_{1,t+1} - y_{2,t}$$

At time of impact, under RE

$$\begin{aligned} & E_t(y_{1,t+1} - y_{2,t}) \\ &= \left(E_t y_{1,t+1} - \frac{1}{2}\left[y_{1,t} + E_t y_{1,t+1}\right]\right) \\ &= \left(-\frac{1}{2}y_{1,t} + \frac{1}{2}E_t y_{1,t+1}\right) \\ &= \frac{1}{2}\left(-y_{1,t} + E_t y_{1,t+1}\right) \\ &= \frac{1}{2}\left(T(b_{t-1})\hat{y}_{t-1} + V(b_{t-1})\varepsilon_t - T(b_{t-1})\hat{y}_t\right) \\ \frac{\partial LHS}{\partial \varepsilon_t} &= \frac{1}{2}\left[V(b_{t-1}) - T(b_{t-1})\right] \\ &= \frac{1}{2}\left[(1-\rho) - \rho(1-\rho)\right] = \frac{1}{2}(1-\rho)^2 > 0 \end{aligned}$$

The impact under learning will be the same in the period of impact also, that is positive

Next period under RE:

$$\begin{aligned}
& E_t(y_{1,t+2} - y_{2,t+1}) \\
= \; & E_t\left[-\rho(1-\rho)\hat{y}_{t+1} - (1-\rho)\varepsilon_{t+2}\right] \\
& -E_t\frac{1}{2}\left[y_{1,t+1} - \rho(1-\rho)\hat{y}_{t+1}\right] \\
= \; & -\rho(1-\rho)\hat{y}_{t+1} - \frac{1}{2}y_{1,t+1} + \frac{1}{2}\rho(1-\rho)\hat{y}_{t+1} \\
= \; & -\frac{1}{2}y_{1,t+1} - \frac{1}{2}\rho(1-\rho)\hat{y}_{t+1}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial E_t(y_{1,t+2} - y_{2,t+1})}{\partial \varepsilon_t} &= \frac{1}{2}\rho(1-\rho) - \frac{1}{2}\rho^2(1-\rho) \\
&= \frac{1}{2}\rho(1-\rho)^2 > 0
\end{aligned}$$

Under learning, in the next period:

$$\begin{aligned}
& E_t(y_{1,t+2} - y_{2,t+1}) \\
= \; & \frac{-1}{2}T(b_t)\left(\rho\hat{y}_t + \varepsilon_{t+1}\right) + \frac{1}{2}T(b_t)\left(\rho\hat{y}_{t-1} + \varepsilon_t\right) + \frac{1}{2}V(b_t)\varepsilon_{t+1}
\end{aligned}$$

$$\frac{\partial E_t(y_{1,t+2} - y_{2,t+1})}{\partial \varepsilon_t} = \frac{-1}{2}\rho\frac{\partial}{\partial \varepsilon_t}T(b_t)\varepsilon_t + \frac{1}{2}\frac{\partial}{\partial \varepsilon_t}T(b_t)\varepsilon_t$$

Using the recursive process for b_t :

$$b_t = b_{t-1} + g\frac{\hat{y}_{t-1}}{E\hat{y}_t^2}\left[(T(b_{t-1}) - b_{t-1})\hat{y}_{t-1} + V(b_{t-1})\varepsilon_t\right]$$

$$\begin{aligned}
& T(b_t)\varepsilon_t \\
= \; & \left[\frac{\rho(1-\rho)}{(1-\beta\rho)} - \frac{\beta\rho}{1-\beta\rho}\left(b_{t-1} + g\frac{1}{E\hat{y}_t^2}\left[\underbrace{(T(b_{t-1}) - b_{t-1})}_{0}\hat{y}_{t-1} + V(b_{t-1})\varepsilon_t\right]\right)\right]\varepsilon_t \\
= \; & \rho(1-\rho) - \frac{\beta\rho}{1-\beta\rho}g\frac{1}{E\hat{y}_t^2}(1-\rho)
\end{aligned}$$

(If $\hat{y}_{t-1} = 1$) For the entire LHS term, the response is:

$$\begin{aligned}
& \frac{1}{2}(1-\rho)\left[\rho(1-\rho) - \frac{\beta\rho}{1-\beta\rho}g\frac{1}{E\hat{y}_t^2}(1-\rho)\right] \\
= \; & \frac{1}{2}\rho(1-\rho)^2 - \frac{1}{2}\rho(1-\rho)^2\frac{\beta}{1-\beta\rho}\frac{g}{E\hat{y}_t^2}
\end{aligned}$$

This will reduce to the RE response when $g = 0$. Otherwise it is strictly smaller than the RE response.

The reponse under learning will be negative if:

$$1 < \frac{\beta}{1-\beta\rho}\frac{g}{E\hat{y}_t^2}$$

To analyze the properties with respect to the model parameters:

$$\begin{aligned}
A &= \frac{\beta}{1-\beta\rho}\frac{g}{E\hat{y}_t^2} - 1 \\
\frac{\partial A}{\partial \rho} &= -\frac{\beta}{(1-\beta\rho)^2}\frac{g}{E\hat{y}_t^2}(-\beta) > 0 \\
\frac{\partial A}{\partial E\hat{y}_t^2} &= -\frac{\beta}{1-\beta\rho}\frac{g}{(E\hat{y}_t^2)^2} < 0 \\
\frac{\partial A}{\partial g} &= \frac{\beta}{1-\beta\rho}\frac{1}{E\hat{y}_t^2} > 0
\end{aligned}$$

That is, for higher ρ and higher g, the deviations from RE will be larger.

A. 5: Proof of Proposition 4

Under RE, the one period asset price is:

$$\begin{aligned}
\hat{P}_{1,t}^{RE} &= \frac{\rho(1-\rho)}{\sigma}\hat{y}_{t-1} + \frac{(1-\rho)}{\sigma}\varepsilon_t \\
V(P_{1,t}^{RE}) &= E\left[\frac{\rho(1-\rho)}{\sigma}\hat{y}_{t-1} + \frac{(1-\rho)}{\sigma}\varepsilon_t\right]^2 \\
&= \frac{\rho^2(1-\rho)^2}{\sigma^2}E(\hat{y}_{t-1}^2) + \frac{(1-\rho)^2}{\sigma^2}E(\varepsilon_t^2)
\end{aligned}$$

Under learning:

$$\begin{aligned}
\hat{P}_{1,t}^{L} &= T_a + T_b\hat{y}_{t-1} + V_b\varepsilon_t \\
V(\hat{P}_{1,t}^{L}) &= E\left[\begin{array}{c} T_a + T_b\hat{y}_{t-1} + V_b\varepsilon_t \\ -E(T_a + T_b\hat{y}_{t-1} + V_b\varepsilon_t) \end{array}\right]^2
\end{aligned}$$

For evaluating the unconditional expectations, as $g \longrightarrow \bar{g}$, gt becomes large:

$$\begin{aligned}
E(T_a) &= 0 \\
E(T_b\hat{y}_{t-1}) &= E\left[\left(\rho\left[\frac{-\beta b_{t-1}}{1-\beta\rho} - \frac{(1-\beta)\rho}{\sigma(1-\beta\rho)}\right] + \frac{\rho}{\sigma}\right)\hat{y}_{t-1}\right] \\
&= E\left[\rho\frac{-\beta b_{t-1}\hat{y}_{t-1}}{1-\beta\rho} - \frac{(1-\beta)\rho^2}{\sigma(1-\beta\rho)}\hat{y}_{t-1} + \frac{\rho}{\sigma}\hat{y}_{t-1}\right] \\
&= 0
\end{aligned}$$

Similarly,

$$E(V_b\varepsilon_t) = 0$$

The mappings are:

$$\begin{aligned}
T_a &= \frac{-\beta}{1-\beta}a_t \\
T_b &= \rho\left[\frac{-\beta b_t}{1-\beta\rho} - \frac{(1-\beta)\rho}{\sigma(1-\beta\rho)}\right] + \frac{\rho}{\sigma} \\
V_b &= \left[\frac{-\beta b_t}{1-\beta\rho} - \frac{(1-\beta)\rho}{\sigma(1-\beta\rho)}\right] + \frac{1}{\sigma}
\end{aligned}$$

To evaluate the variances:

$$E(T_a^2) = \frac{\beta^2}{(1-\beta)^2} E(a_t^2)$$

Other terms are:

$$\begin{aligned}
& T_b^2 \hat{y}_{t-1}^2 \\
= & \left(\rho\left[\frac{-\beta b_t}{1-\beta\rho} - \frac{(1-\beta)\rho}{\sigma(1-\beta\rho)}\right] + \frac{\rho}{\sigma}\right)^2 \hat{y}_{t-1}^2 \\
= & \left(\frac{\beta\rho}{1-\beta\rho}\right)^2 b_t^2 \hat{y}_{t-1}^2 + \left(-\frac{(1-\beta)\rho^2}{\sigma(1-\beta\rho)} + \frac{\rho}{\sigma}\right)^2 \hat{y}_{t-1}^2 \\
& +2\left(\frac{-\beta\rho}{1-\beta\rho}\right)\left(-\frac{(1-\beta)\rho^2}{\sigma(1-\beta\rho)} + \frac{\rho}{\sigma}\right) b_t \hat{y}_{t-1}^2
\end{aligned}$$

$$\begin{aligned}
& E(T_b^2 \hat{y}_{t-1}^2) \\
= & \left(\frac{\beta\rho}{1-\beta\rho}\right)^2 E(b_t^2 \hat{y}_{t-1}^2) \\
& + \left(-\frac{(1-\beta)\rho^2}{\sigma(1-\beta\rho)} + \frac{\rho}{\sigma}\right)^2 E(\hat{y}_{t-1}^2) \\
& +2\left(\frac{-\beta\rho}{1-\beta\rho}\right)\left(-\frac{(1-\beta)\rho^2}{\sigma(1-\beta\rho)} + \frac{\rho}{\sigma}\right) E(b_t \hat{y}_{t-1}^2) \\
= & \left(-\frac{(1-\beta)\rho^2}{\sigma(1-\beta\rho)} + \frac{\rho}{\sigma}\right)^2 E(\hat{y}_{t-1}^2)
\end{aligned}$$

$$\begin{aligned}
& V_b^2 \varepsilon_t^2 \\
= & \left(\left[\frac{-\beta b_{t-1}}{1-\beta\rho} - \frac{(1-\beta)\rho}{\sigma(1-\beta\rho)}\right] + \frac{1}{\sigma}\right)^2 \varepsilon_t^2 \\
= & \left(\frac{-\beta}{1-\beta\rho}\right)^2 b_{t-1}^2 \varepsilon_t^2 + \left(-\frac{(1-\beta)\rho}{\sigma(1-\beta\rho)} + \frac{1}{\sigma}\right)^2 \varepsilon_t^2 \\
& +2\left(\frac{-\beta}{1-\beta\rho}\right)\left(-\frac{(1-\beta)\rho}{\sigma(1-\beta\rho)} + \frac{1}{\sigma}\right) b_{t-1} \varepsilon_t^2 \\
& E(V_b^2 \varepsilon_t^2) \\
= & \left(-\frac{(1-\beta)\rho}{\sigma(1-\beta\rho)} + \frac{1}{\sigma}\right)^2 E(\varepsilon_t^2)
\end{aligned}$$

The covariance terms are also zero: $E(T_a T_b \hat{y}_{t-1}) = E(T_b V_b \hat{y}_{t-1} \varepsilon_t) = E(T_a V_b \varepsilon_t) = 0$. Then, the variance of the one-period price under learning is:

$$\begin{aligned}
& V(\hat{P}_{1,t}^L) \\
= & \frac{\beta^2}{(1-\beta)^2} E(a_t^2) + \left(-\frac{(1-\beta)\rho^2}{\sigma(1-\beta\rho)} + \frac{\rho}{\sigma}\right)^2 E(\hat{y}_{t-1}^2) \\
& + \left(-\frac{(1-\beta)\rho}{\sigma(1-\beta\rho)} + \frac{1}{\sigma}\right)^2 E(\varepsilon_t^2)
\end{aligned}$$

The difference between learning and RE variance is given by:

$$
\begin{aligned}
& V(\hat{P}^{L}_{1,t}) - V(\hat{P}^{RE}_{1,t}) \\
= & \frac{\beta^2}{(1-\beta)^2}E(a_t^2) + \left(-\frac{(1-\beta)\rho^2}{\sigma(1-\beta\rho)} + \frac{\rho}{\sigma}\right)^2 E(\hat{y}^2_{t-1}) \\
& + \left(-\frac{(1-\beta)\rho}{\sigma(1-\beta\rho)} + \frac{1}{\sigma}\right)^2 E(\varepsilon_t^2) \\
& - \frac{\rho^2(1-\rho)^2}{\sigma^2}E(\hat{y}^2_{t-1}) - \frac{(1-\rho)^2}{\sigma^2}E(\varepsilon_t^2) \\
= & \frac{\beta^2}{(1-\beta)^2}E(a_t^2) \\
& + E(\hat{y}^2_{t-1})\left[\left(-\frac{(1-\beta)\rho^2}{\sigma(1-\beta\rho)} + \frac{\rho}{\sigma}\right)^2 - \frac{\rho^2(1-\rho)^2}{\sigma^2}\right] \\
& + E(\varepsilon_t^2)\left[\left(-\frac{(1-\beta)\rho}{\sigma(1-\beta\rho)} + \frac{1}{\sigma}\right)^2 - \frac{(1-\rho)^2}{\sigma^2}\right]
\end{aligned}
$$

The constant terms in second and third terms are positive for all values of σ and $\beta, \rho \in (0,1)$. Therefore:

$$V(\hat{P}^{L}_{1,t}) = V(\hat{P}^{RE}_{1,t}) + f(V(a_t), \text{positive constants})$$

As the one-period yield is a linear transformation of the one-period price, the result follows.

A.6: Intertemporal optimization of households with non-zero government debt

Government expenditures are introduced directly into the utility function of households, and the optimization problem of the household is:

$$\max_{\{C_t^i, B^i_{1,t}, B^i_{2,t}, \dots, B^i_{n,t}\}} \tilde{E}_t \sum\nolimits_{j=0}^{\infty} \beta^j \left(U(C^i_{t+j} + \mu g; \xi_{t+j}) - \int_0^1 v(h^i_{t+j}(k); \xi_{t+j}) dk \right). \tag{63}$$

Here g is the quantity of government expenditures. The utility specification follows Barro (1981): μ is a parameter governing the magnitude of the derivative of the marginal utility of the private consumption C with respect to g. As g is assumed to be constant, it does not affect any results in the analysis. This assumption is made for analytical tractability.

The flow budget constraint of household i is:

$$
\begin{aligned}
C_t^i + \sum\nolimits_{m=1}^{n} P^B_{m,t} B^i_{m,t} &\leq Y_t^i + \tilde{W}^i_t - T^i_t; \\
\tilde{W}^i_{t+1} &= B^i_{1,t} + \sum\nolimits_{m=2}^{n} P^B_{m-1,t+1} B^i_{m,t},
\end{aligned} \tag{64}
$$

where T^i_t is the tax obligation of each household. Other conditions of the household's maximization, as well as the specification of the firm's and monetary authority's problem are unchanged.

The goods and asset market clearing conditions are:

$$
\begin{aligned}
\int_0^1 C_t^i di + g &= Y_t; \\
\int_0^1 B^i_{m,t} di &= B^s_{m,t}.
\end{aligned} \tag{65}
$$

I consider a first order log-linear approximation around the steady state level of wealth $\bar{W}$, taxes $\bar{T}$, output level $\bar{Y}$, and the one-period bond price β. Then, the optimal consumption decision rule for household i, derived as in Appendix A.1., is:

$$\begin{aligned}\hat{C}_t^i &= s_C^{-1} s_W (1-\beta)\hat{W}_t^i + s_C^{-1}(1-\beta)\tilde{E}_t\left[\sum\nolimits_{j=0}^{\infty}\beta^j \hat{Y}_{t+j} - s_T \sum\nolimits_{j=0}^{\infty}\beta^j \hat{T}_{t+j}^i\right] \\ &\quad + s_C^{-1}\beta(1-s_T)\hat{E}_t \sum\nolimits_{j=0}^{\infty}\beta^j \hat{P}_{1,t+j}^B,\end{aligned} \tag{66}$$

where $s_C = \frac{\bar{C}}{\bar{Y}}$; $s_W = \frac{\bar{B}_1}{\bar{Y}}$ and $s_T = \frac{\bar{T}}{\bar{Y}}$. In the model described in section 6.2 above, g is assumed to be zero, so that $s_C = 1$. Summing consumption and wealth holdings over the the i households, and imposing the market clearing conditions in (65), and rewriting the consumption decision rule in terms of the output gap as in Appendix A.1. above:

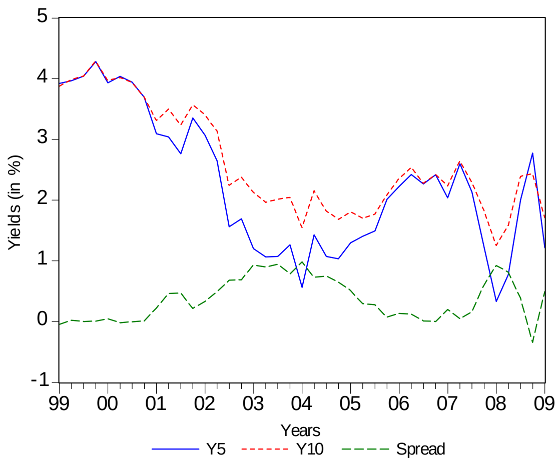
$$\begin{aligned}\hat{x}_t &= \tilde{E}_t \sum\nolimits_{j=0}^{\infty}\beta^j \left[\begin{array}{c}(1-\beta)\hat{x}_{t+j+1} - \sigma\beta(\hat{\imath}_{1,t+j} - \tilde{E}_t\hat{\pi}_{t+j+1}) \\ +\hat{r}_{t+j+1}^n\end{array}\right] \\ &\quad + s_T \left[\begin{array}{c}\frac{(\hat{b}_{1,t} - \hat{\pi}_t)}{\beta} - \frac{\hat{\tau}_t}{\beta} \\ +\tilde{E}_t \sum_{j=0}^{\infty}\left[(\hat{\imath}_{1,t+j} - \tilde{E}_t\hat{\pi}_{t+j+1}) - (1-\beta)\hat{\tau}_{t+j+1}\right]\end{array}\right].\end{aligned} \tag{67}$$

Here $\tau_t = T_t/P_t$, $b_{1,t} = B_{1,t}/P_{t-1}$. Other variables are as defined above in Appendix A.1.

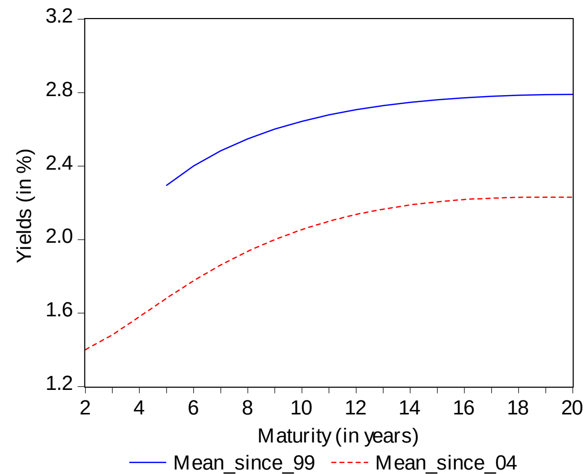
Tables and Figures

FIGURE 1: REAL YIELD CURVES

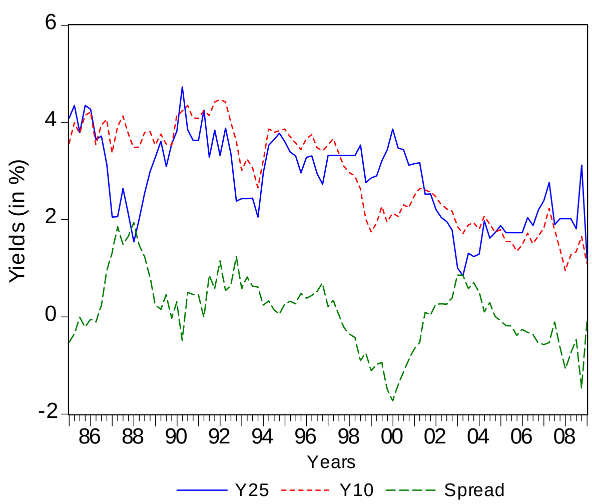
1A: U.S. TIPS YIELDS



1B: U.S. TIPS SLOPE



1C: U.K. INDEX-LINKED YIELDS



1D: U.K. INDEX-LINKED SLOPE

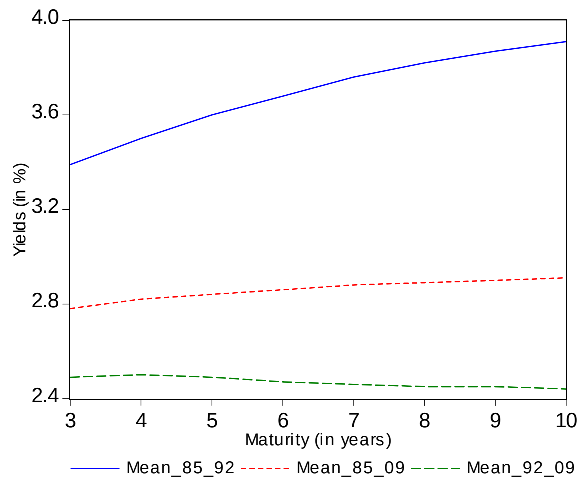
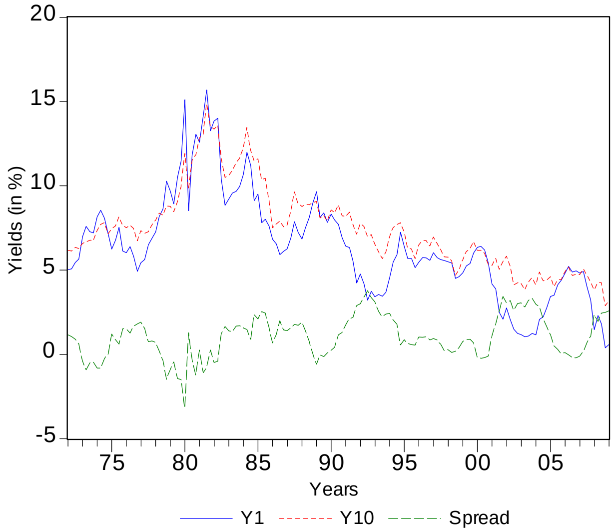
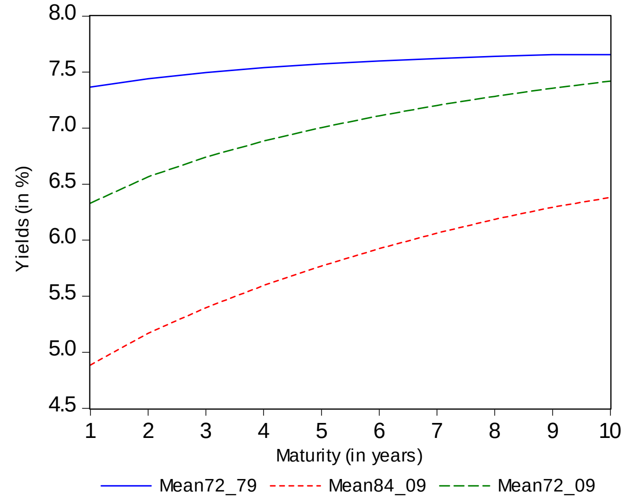


Figure 2: Nominal Yield Curves

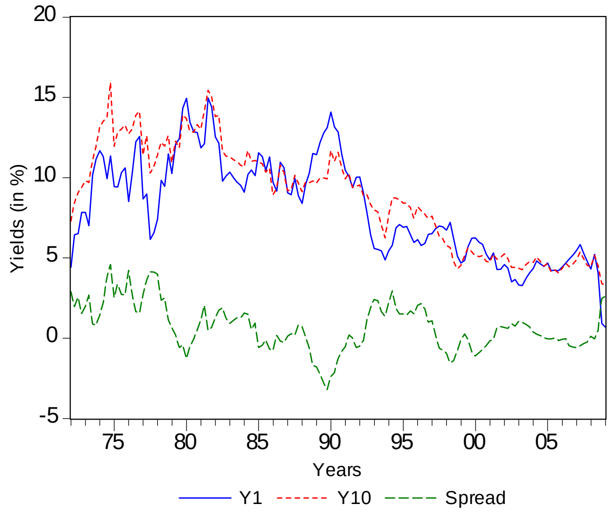
2a: U.S. Nominal Yields



2b: U.S. Nominal Slope



2c: U.K. Nominal Yields



2d: U.K. Nominal Slope

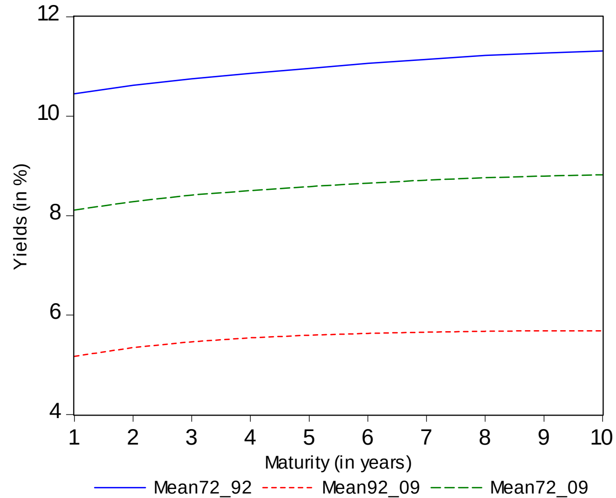
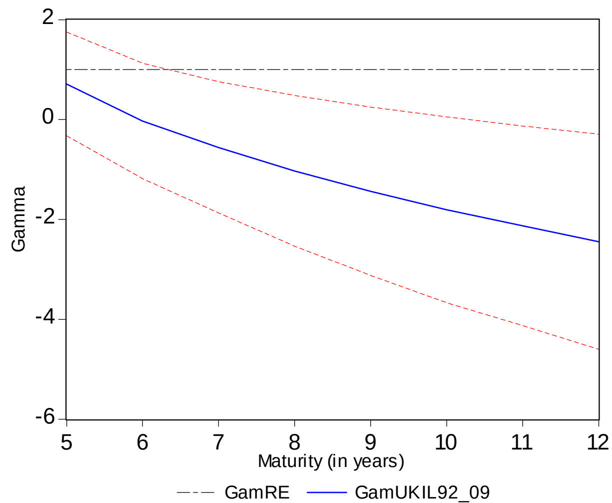


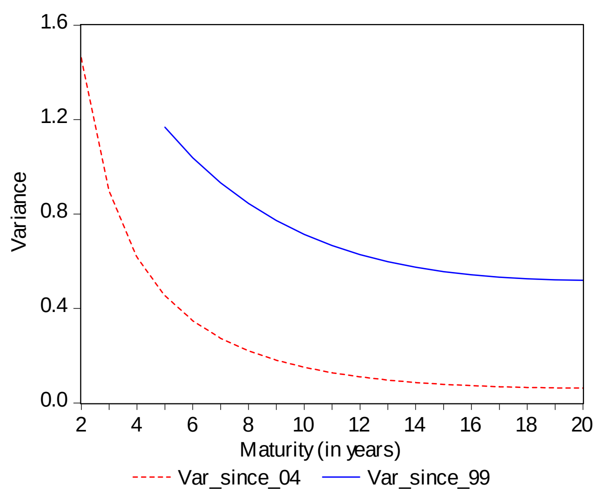
Figure 3: Campbell-Shiller Slope Coefficients
U.K. Index-Linked Yields



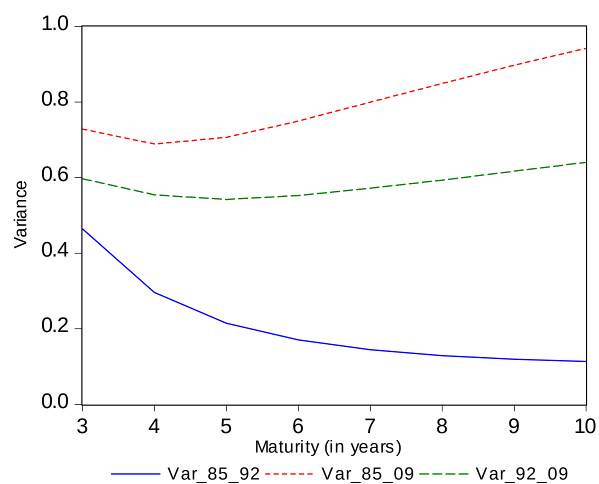
Note: The dotted red lines show the 95% confidence intervals. The Rational Expectations slope coefficient is one, and is shown as the dashed black line.

Figure 4: Real Yield Variances

4a: U.S. Tips Yields



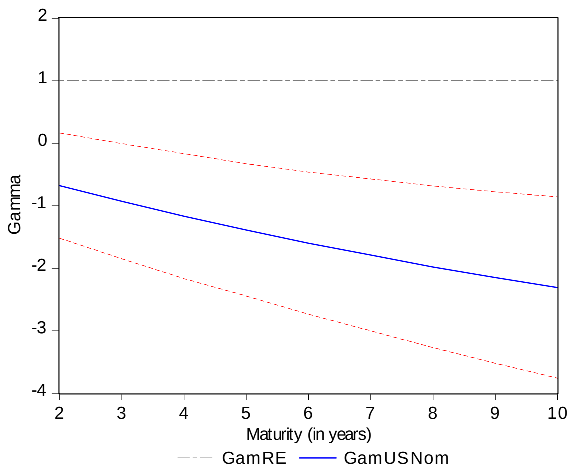
4b: U.K. Index-Linked Yields



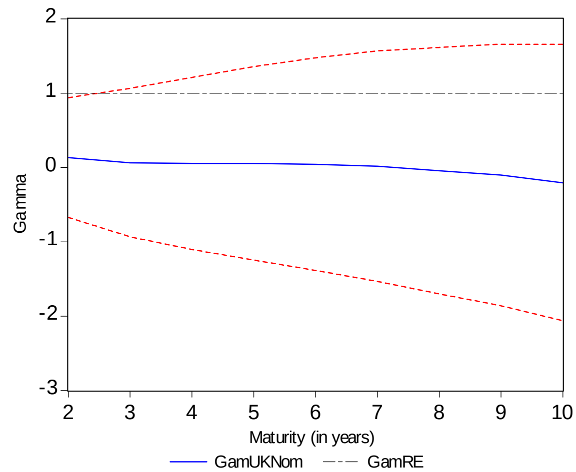
Performance of Nominal Yields with respect to Implications of the Expectations Hypothesis

Figure 5: Campbell-Shiller Slope Coefficients

5A: U.S. Nominal Yields



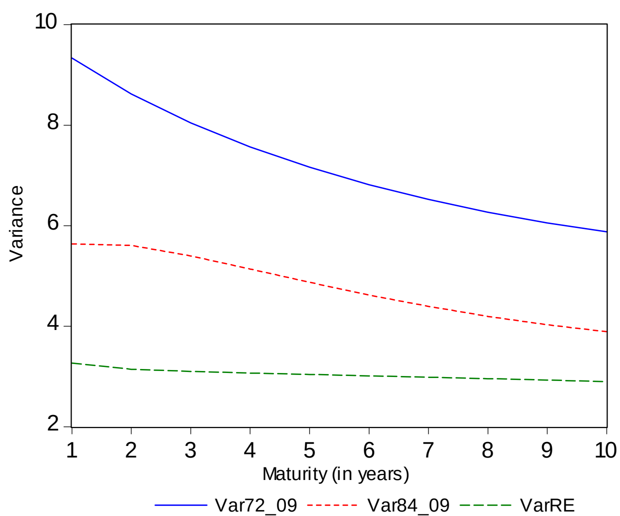
5A: U.K. Nominal Yields



Note: The dotted red lines show the 95% confidence intervals. The Rational Expectations slope coefficient is one, and is shown as the dashed black line. For the U.K., the regression coefficients are for the sub-sample 1992:4-2009:1.

Figure 6: Nominal Yield Variances

6A: U.S. Nominal Yields



6B: U.K. Nominal Yields

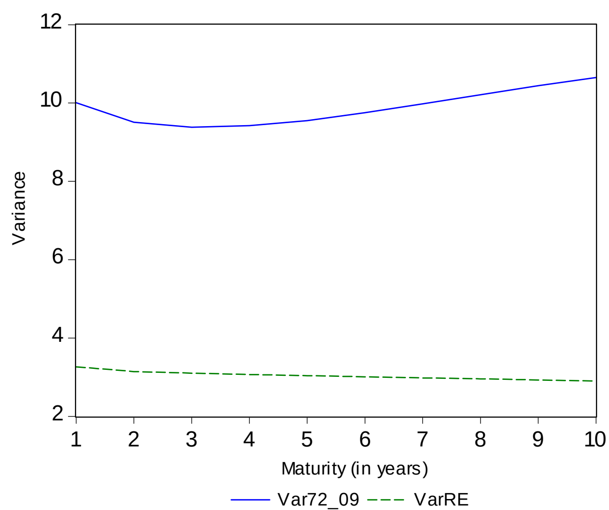
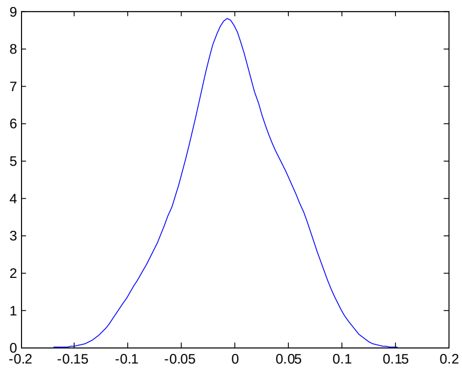
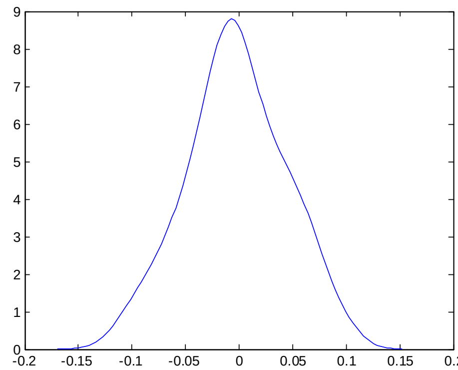


Figure 7: Distribution of Beliefs in Benchmark Model

7a: Distribution of $a^{\hat{x}}$



7b: Distribution of a^{π}



7c: Distribution of $a^{\hat{\imath}}$

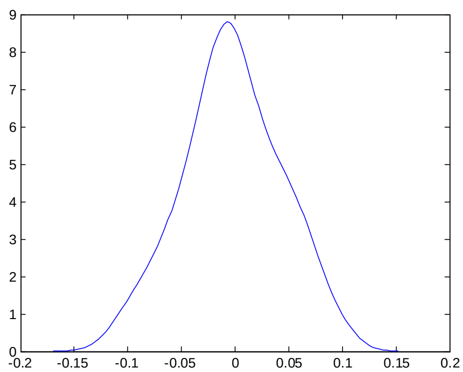
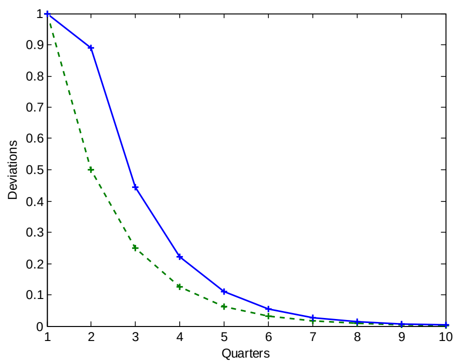
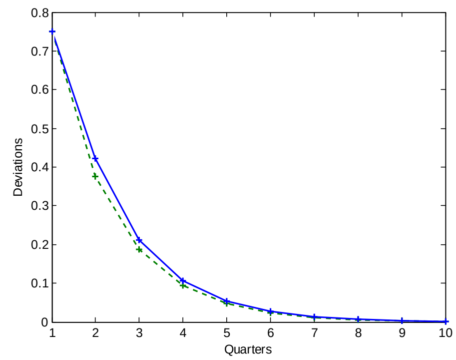


Figure 8: Dynamic Responses under Rational Expectations and Learning

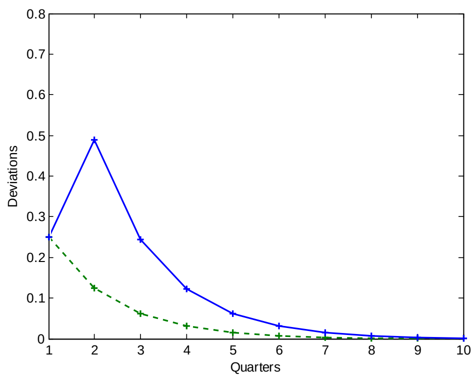
8a: Response of $\hat{\imath}_{1,t}$



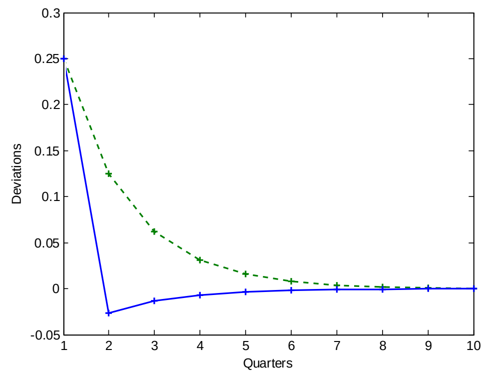
8b: Response of $\hat{\imath}_{2,t}$



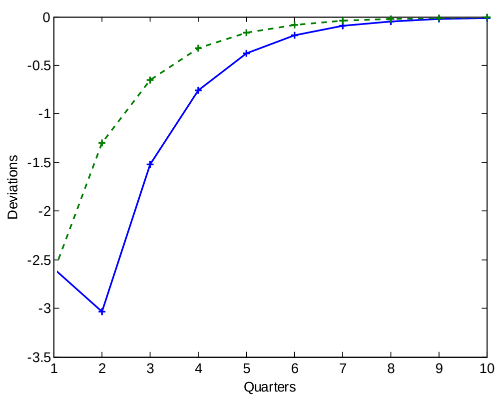
8c: Response of $-(\hat{\imath}_{2,t} - \hat{\imath}_{1,t})$



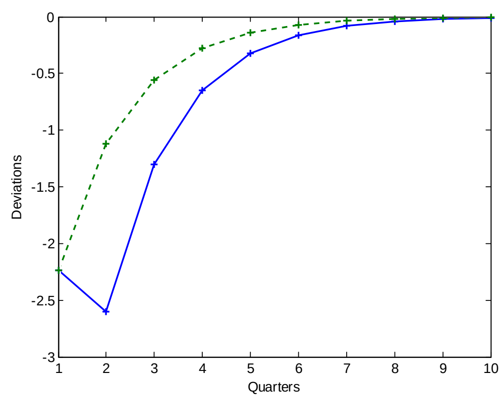
8d: Response of $-(E_t\hat{\imath}_{1,t+1} - \hat{\imath}_{2,t})$



8e: Response of Output Gap



8f: Response of Inflation



Note: The deviations are in percentages. The dotted green line is the response under Rational Expectations, and the solid blue line is the Learning response. For the purpose of the impulse response, I lower the persistence of the monetary policy shock to 0.5 from 0.98. These are the responses for the benchmark gain parameter $g = 0.05$. The expectations under learning are those implied by the actual data generating process, as shown in (35).

TABLE 1: WEIGHTS UNDER CONSTANT GAIN

g	$10Q$	$20Q$	$30Q$
0.002	0.98	0.96	0.94
0.02	0.83	0.68	0.55
0.2	0.13	0.01	0.00

Note: The entries are interpreted as the weight $(1-g)^{i-1}$ placed by constant gain g to an observation i periods back.

Performance of Benchmark Model for Campbell-Shiller Coefficients

Table 2: Campbell-Shiller Slope Coefficients for Nominal Yields

$n(Years)$	$U.S.Data$	$g_2 = 0.05$		$g_5 = 0$	
	γ	γ	$Rej.$	γ	$Rej.$
2	-0.68	0.02	64%	1.00	95%
3	-0.93	-0.49	64%	1.01	96%
4	-1.17	-0.93	65%	1.01	96%
5	-1.39	-1.30	67%	1.01	95%
6	-1.6	-1.61	68%	1.00	95%
7	-1.79	-1.87	69%	1.02	96%
8	-1.98	-2.08	70%	1.02	96%
9	-2.15	-2.26	72%	1.00	96%
10	-2.31	-2.39	73%	1.00	95%

Note: The 'Rejections' column, denoted 'Rej.', refers to the percentage of times that the Expectations Hypothesis cannot be be rejected.

Table 3: Campbell-Shiller Slope Coefficients for Real Yields

$n(Years)$	$U.K.Data$	$g_1 = 0.01$		$g_2 = 0.02$		$g_5 = 0$	
	γ	γ	$Rej.$	γ	$Rej.$	γ	$Rej.$
2	na	-0.40	66%	-1.01	57%	1.00	95%
3	na	-1.01	67%	-1.93	57%	1.00	96%
4	na	-1.49	68%	-2.70	58%	1.01	96%
5	0.71	-1.85	69%	-3.33	60%	1.01	95%
6	-0.03	-2.12	71%	-3.84	62%	1.00	95%
7	-0.56	-2.30	72%	-4.24	62%	0.99	95%
8	-1.03	-2.41	73%	-4.54	63%	1.00	96%
9	-1.44	-2.47	75%	-4.77	64%	1.01	96%
10	-1.81	-2.48	76%	-4.93	66%	1.01	95%

PERFORMANCE OF BENCHMARK MODEL FOR YIELD VOLATILITIES

TABLE 4: VOLATILITIES FOR NOMINAL YIELDS

$n(Years)$	$U.S.Data$	$g_3 = 0.05$	$g_5 = 0$
1	2.99	1.74	1.65
2	2.97	1.72	1.62
3	2.92	1.69	1.57
4	2.80	1.66	1.56
5	2.73	1.63	1.53
6	2.66	1.60	1.50
7	2.60	1.57	1.47
8	2.55	1.54	1.44
9	2.51	1.51	1.42
10	2.47	1.48	1.39

TABLE 5: VOLATILITIES FOR REAL YIELDS

$n(Years)$	$U.S.Data$		$g_1 = 0.02$	$g_5 = 0$
	$(1999 - 2009)$	$(2004 - 2009)$		
1	na	1.20	1.85	1.80
2	na	0.96	1.82	1.78
3	na	0.81	1.79	1.72
4	na	0.69	1.75	1.67
5	1.12	0.60	1.72	1.61
6	1.05	0.53	1.69	1.54
7	0.99	0.47	1.66	1.49
8	0.94	0.42	1.63	1.44
9	0.90	0.38	1.59	1.37
10	0.87	0.34	1.57	1.31

Table 6: Survey Details

Survey	**SPF**	**MSCF**
Dates	Since 1990Q1 by FRB Philadelphia	Since 1964 by Survey Research Center at University of Michigan
Population	Industry economists	Households
Sample Variables	3 month Tbill, CPI Inflation	Changes in prices
Sample question	"Fill in your response, in level or growth, for the following variables"	"During the next 12 months, do you think that prices in general will go up, or go down, or stay where they are now?"

Table 7: Testable Implications

Series $(1972:1-2009:2)$	*Moment*	*Data*	$g=0$	$g=0.05$
Inflation	*Stdev*	1.74	1.28	1.39
1 year yield	*Autocorr*	0.94	0.94	0.98
	Stdev	2.99	1.65	1.74

Note: The data shown here is for U.S. for the period between 1972:1-2009:2. Inflation is computed from the CPI measure from the Bureau of Economic Analysis, where $\pi_t = 400\ln(P_t/P_{t-1})$.

Performance of Benchmark Model for Different Gain Parameters

Table 8: Campbell-Shiller Slope Coefficients for Nominal Yields

$n(Years)$	$U.S.Data$	$g_1 = 0.02$		$g_4 = 0.06$	
	γ	γ	$Rej.$	γ	$Rej.$
2	-0.68	0.21	70%	-0.23	57%
3	-0.93	-0.18	71%	-0.88	58%
4	-1.17	-0.50	72%	-1.45	58%
5	-1.39	-0.76	74%	-1.95	58%
6	-1.6	-0.97	75%	-2.39	59%
7	-1.79	-1.13	77%	-2.76	60%
8	-1.98	-1.24	77%	-3.08	61%
9	-2.15	-1.33	79%	-3.35	62%
10	-2.31	-1.38	80%	-3.57	63%

Table 9: Volatilities for Nominal Yields

$n(Years)$	$g_1 = 0.02$	$g_4 = 0.06$
1	1.72	1.78
2	1.69	1.77
3	1.66	1.74
4	1.63	1.71
5	1.60	1.68
6	1.57	1.65
7	1.54	1.62
8	1.51	1.59
9	1.48	1.56
10	1.45	1.53

Performance of Benchmark Model for Different Taylor Rule Parameters

Table 10: Campbell-Shiller Slope Coefficients for Nominal Yields for $g = 0.05$

$n(Years)$	$U.S.Data$	$\phi_\pi = 1.5$		$\phi_\pi = 4$		$\phi_\pi = 15$	
	γ	γ	$Rej.$	γ	$Rej.$	γ	$Rej.$
2	-0.01	0.02	64%	0.15	65%	0.39	68%
3	-0.25	-0.49	64%	-0.29	66%	0.06	68%
4	-0.45	-0.93	65%	-0.68	66%	-0.22	68%
5	-0.63	-1.30	67%	-1.03	67%	-0.47	69%
6	-0.79	-1.61	68%	-1.33	68%	-0.69	69%
7	-0.91	-1.87	69%	-1.58	69%	-0.87	70%
8	-1.01	-2.08	70%	-1.79	69%	-1.02	70%
9	-1.09	-2.26	72%	-1.97	70%	-1.15	70%
10	-1.09	-2.39	73%	-2.11	71%	-1.24	70%

Table 11: Volatilities for Nominal Yields for $g = 0.05$

$n(Years)$	$U.S.Data$	$\phi_\pi = 1.5$	$\phi_\pi = 4$	$\phi_\pi = 1.5$	$\phi_\pi = 4$
1	3.32	1.74	2.98	1.65	2.83
2	3.26	1.72	2.95	1.62	2.78
3	3.18	1.69	2.90	1.57	2.73
4	3.10	1.66	2.85	1.56	2.67
5	3.03	1.63	2.80	1.53	2.62
6	2.96	1.60	2.74	1.50	2.57
7	2.90	1.62	2.69	1.47	2.53
8	2.84	1.59	2.64	1.44	2.48
9	2.79	1.56	2.59	1.42	2.43
10	2.75	1.53	2.55	1.39	2.39