Learning and the Yield Curve

Arunima Sinha[1]

Department of Economics, Fordham University

New York, USA

January 2015

[1] I would like to thank Bruce Preston, Ricardo Reis, Michael Woodford and John Donaldson for their advice and support. The analysis has benefitted greatly from the comments of the editor, and two anonymous referees. Discussions with seminar participants at Columbia University, Federal Reserve Banks of Dallas, New York and San Francisco, Hamilton College, Indian School of Business, Kansas University, Santa Clara University, UC Santa Cruz and UC Irvine are deeply appreciated. All errors are mine. Email: asinha3@fordham.edu.

Abstract

Arunima Sinha

Department of Economics, Fordham University

New York, USA

Two central implications of Expectations Hypothesis under rational expectations are inconsistent with yield curve data: (i) future expected long yields fall, instead of rising, when yield spread rises; (ii) long yields are excessively volatile relative to short yields. I propose an optimization framework in which boundedly rational agents use adaptive learning to form expectations. The belief structure rationalizes pattern of yields observed in the data so that the first puzzle does not arise with subjective expectations: intertemporal income and substitution effects are amplified relative to rational expectations. The second puzzle is partly accounted for by extra volatility due to parameter uncertainty.

JEL Classifications: E32, D83, D84

"*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 has been the classic theory used by policy makers and market participants to understand the behavior of the term structure of interest rates. According to the Hypothesis, the long-term yield is an average of future expected short-term yields. A central implication of this Hypothesis is that the yield spread (difference between the long- and short-term yields) is the sum of a constant risk premium and an optimal forecast of changes in future yields.

This consequence of the Expectations Hypothesis has been extensively tested for yield curve data using the Campbell-Shiller (1991) regression. In this regression, the difference between the $(n-1)$-period yield expected next period, and the current $n-$ period yield is regressed on the spread between the $n-$ and one-period yields.[2] With rational expectations, if the one-period yield is going to rise over the life of the $n-$period bond, then the $n-$period yield must be higher than the current one-period yield. In this case, the slope coefficient in the Campbell-Shiller regression should not be statistically different from one. For the U.S. nominal yield curve data, however, the slope coefficient is less than one at short maturities, and becomes negative at longer maturities. This implies that when the yield spread in the regression is high, the yield on the long-term bond falls over the life of the short-term bond, instead of rising, as predicted by the Hypothesis. The robustness of these findings on the slope coefficient, across sample periods and combinations of yield maturities, has been interpreted as a rejection of the Expectations Hypothesis in the data.

This paper poses the following question: can the empirically observed pattern in the Campbell-Shiller slope coefficients be rationalized by a framework in which conditional expectations of the long- and short-term bonds are formed using adaptive learning, instead of rational expectations? By construction, the Campbell-Shiller regression jointly tests the Expectations Hypothesis and that market participants have rational expectations. The approach in this paper is to separate these hypotheses: first, the term structure is derived from a general equilibrium model in which

[2] Campbell and Shiller (1991) refer to the spread between the current $n-$ and one-period yields as the "perfect foresight" spread.

optimizing agents use adaptive learning, and these yields are used to construct the Campbell-Shiller regression. Then the deviations of the resulting slope coefficient from one are tested.

The use of subjective expectations to explain the empirical estimates of the Campbell-Shiller coefficients is motivated by findings in previous literature. Froot (1989) uses survey data on yields to argue that the failure of the Expectations Hypothesis may be due to the failure of rational expectations, not the Hypothesis itself. Piazzesi, Salomao and Schneider (2013) show that expected excess returns can be explained by subjective expectations, in addition to time-varying risk premia.

The adaptive learning model is successful at generating the pattern of Campbell-Shiller slope coefficients observed in the data. The intuition is the following: consider an agent who is learning about the one-period bond yield. She derives the longer-term yields as an average of the future expected short yields; this is an optimization condition, and holds for the specified set of beliefs, even as they deviate from rational expectations. In case of a contractionary monetary policy shock, the one-period yield rises, and the spread in the Campbell-Shiller regression falls. On observing the shock at time t, the agent who is only using past data (upto period $t-1$) to form conditional expectations about future yields, perceives a fraction of the shock to be permanent, and the average one-period yield to be higher. In this case, her expectations about the $(n-1)$-period yield rise in period $t+1$.[3] Thus, as the yield spread falls, the expectations of the long-term yield rise over the life of the short-term bond.

The theoretical framework uses a micro-founded dynamic stochastic general equilibrium (DSGE) model with adaptive learning agents. Households choose optimal consumption, and their only means of saving are riskless bonds issued by the fiscal authority in zero net supply. Firms maximize profits and face Calvo (1983) pricing; the monetary authority sets the nominal short interest rate using the Taylor rule. Under learning, optimizing agents run a vector-autoregression on past data to form their conditional expectations about future aggregate variables. While updating their beliefs, they assign exponentially decaying weight on past observations. This is the constant-gain adaptive learning process for formation of beliefs. The Expectations Hypothesis holds since it is derived from the optimization problem of the agents, given the specified set of beliefs.

The process for expectations formation and the general equilibrium structure are central to the

[3] These expectations are determined by the (subjective) Expectations Hypothesis, and are higher than they were in the period of the shock, t.

analysis. Constant-gain learning implies that optimizing agents make systematic forecasting errors while forming conditional expectations; forecasts reported by the Survey of Professional Forecasters (SPF) exhibit similar errors as well. Data on survey expectations is used to specify the weight placed by optimizing agents on past observations, which is the 'gain' parameter. The median forecasts of the three-month treasury yield from the SPF exhibit a systematic autocorrelation in the forecast errors at different horizons; the gain in the benchmark calibration of the model minimizes the distance between the autocorrelation of the one-quarter ahead forecast errors of the three-month yield from the survey data and the learning model.

The success of the learning model can be further understood using these systematic errors made in forecasting bond yields. The forecasting error enters the regression error in the Campbell-Shiller regression, violating the orthogonality condition between the error and yield spread, causing a bias in the slope coefficient. Thus, the regression is mis-specified since it does not account for systematic errors being made by the agents. When agents have rational expectations, the systematic forecasting error is zero, and the corresponding bias in the slope coefficient relative to one is zero as well. Mis-specification of beliefs as the source of bias is in contrast to much of the literature which identifies the omission of a time-varying risk premium from the Campbell-Shiller regression as rationale for why the slope coefficient is found to be statistically different from one.

The interaction of learning dynamics with the general equilibrium framework generates self-referential behavior in the determination of aggregate variables, such as output, which is key to generating a negative bias in the slope coefficient with respect to one. To make consumption decisions in the current period, agents must form conditional expectations of interest rates, inflation and output over the infinite horizon; to form these expectations, agents use past observable data on these variables. This explains the economic mechanism of the adaptive learning model: under rational expectations, in response to a transitory, contractionary monetary policy shock, the optimizing agent correctly forecasts that the rise is temporary, and that the average interest rate is unchanged. As she lowers consumption and increases savings, the short yield falls. In the period after the shock, the agent is using the correct probability distribution to make yield forecasts. In this case, the expected long yield is lower, relative to the short yield in the period of the shock. Under rational expectations, the agent correctly attributes the forecasting error to the transitory shock. When conditional expectations are instead formed using adaptive learning, the agent attributes a fraction of the forecast errors to a permanent change in the parameters, and since she perceives the average interest rate to have increased, in the period after the monetary policy shock, she lowers her consumption even more than in the period of the shock. The fraction of the errors attributed to a permanent shift in the parameters depends on the value of the constant gain. In equilibrium, the fall in consumption (or equivalently output, as the benchmark considers an economy with no production) leads to a fall in savings, pushing up the expected long yield. In this case, as the yield spread falls, the expected long yield rises.

The general equilibrium model also yields a more realistic information set which optimizing agents use to form their consumption decisions: instead of only learning about the evolution of an endowment process, agents are modeled as learning about the output, inflation and interest rate processes. Finally, the explicit monetary policy rule allows for an analysis of the effects of different regimes on the evolution of beliefs and their consequences for the test of the Expectations Hypothesis. 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 at the long end of the yield curve.

The paper also explores a second implication of the Expectations Hypothesis: long yields should be less volatile than shorter yields; however, in the data, variance of long yields is almost as large (and in some cases, larger) as the shorter yields. I explore the performance of the learning model with respect to this excess volatility puzzle. In its benchmark calibration, the model is able to explain approximately 25% more of the excess volatility at the five-year horizon, relative to rational expectations. The intuition is that since the beliefs of optimizing agents are time-varying and no longer constant, the volatility of the yields generated by the model are a function of uncertainty in beliefs along with the volatility of the underlying state variables. I also find that under the benchmark calibration, the learning model generates variance of inflation and yields that are close to corresponding moments in U.S. data.

The rest of the paper is organized as follows: in section one, I present the empirical performance of the nominal and real yield curves for the U.S. and U.K. with respect to the implications of the Expectations Hypothesis. The benchmark DSGE model with adaptive learning is discussed in section two. The intuition for the mechanics of the learning model are first described using the limiting case of the flexible-price model with an exogenous endowment process in section three.

Analytical results are also derived here. Following model calibration and the performance of the model with respect to macroeconomic moments, I discuss the quantitative performance of the learning model with respect to the Campbell-Shiller regression and variance of yields. Further economic intuition for the model is also discussed. I then examine implications of the benchmark model under different monetary policy regimes and alternative values of key model parameters. Section four concludes.

1 Empirical Performance of Real and Nominal Yield Curves with respect to the Expectations Hypothesis

The specification of the Campbell-Shiller (1991) regression used in this analysis is the following:

$$i_{n-1,t+1} - i_{n,t} = \bar{\alpha} + \frac{\gamma}{n-1}(i_{n,t} - i_{1,t}) + \varepsilon_t. \tag{1}$$

Here $i_{n,t}$ denotes the yield of an $n-$period bond at time t.

I present results on the Campbell-Shiller coefficients and variance for real and nominal yield curves for the U.S. and U.K. The data on real yield curves is obtained from the inflation-indexed bonds issued. 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). [4] The estimates of the TIPS yield curve are taken from the daily data set constructed by Gürkaynak, Sack and Wright (2010). 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. The real yield curve is

[4] This is done 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. When the ratio of CPIs is less than one, no adjustment is made. 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.

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, TIPS are issued in terms of 5, 10 and 30 years.

upward sloping for the period considered.

Since the TIPS data set is relatively short, I 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 1985 onwards, and I use yields of maturities 2.5 to 20 years. I also consider two subsample periods: from January 1985 to September 1992, and from October 1992 to 2007. 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. For the entire sample, January 1985 to December 2007, the real yield curve is positively sloped. For the first subsample the slope continues to be positive, but for the second sample the yield curve has a small negative slope.

The estimates of the U.S. nominal yield curve are from the yields constructed by Gürkaynak, Sack and Wright (2007); the sample period is January 1972 to December 2007. I also check the robustness of the empirical analyses below by considering two subsamples, January 1972 to December 1979; and January 1984 to 2007. 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. The yield curve is upward sloping for the whole sample, January 1972 to December 2007, as well as for subsamples, January 1972 $-$ September 1992, and October 1992 $-$ December 2007.

1.1 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.[5] 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 table 1(a), I find that for the full sample period, the point estimates are smaller

[5] 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. Pflueger and Viciera (2011) construct a short real rate and then test the real Expectations Hypothesis using TIPS data. They find that the Hypothesis is strongly rejected in the 1999 $-$ 2009 sample.

than one, and the difference between the estimates and one is statistically significant. While this is also the case for the two subsamples, the rejections of the Hypothesis at the short end of the term structure are larger in the inflation targeting period and smaller at the long end relative to the pre 1992 sample.

Variances The empirical yield curves for the U.S. show the same qualitative pattern (table 2(a)) as predicted by the Expectations Hypothesis: the short end of the curve is more volatile relative to the long end. For the U.K. data, as shown in table 2(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.

1.2 Performance of the Nominal Yield Curves

Campbell-Shiller Regression The slope coefficients of the Campbell-Shiller regression are reported in table 1(b) 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. Table 1(c) shows the regression coefficients for U.K. nominal yields. Similar to the U.S., the coefficients are negatively biased with respect to one, and the difference is statistically significant. One of the notable patterns in both the country datasets is that the Campbell-Shiller coefficients at the short end are more negative than at the long end of the term structure in subsamples associated with lower inflation uncertainty (the Great Moderation for the U.S. and the inflation targeting regime in the U.K.).

Variances The term structure of variances for the U.S. nominal yield curve is shown in table 2(c). 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 also shown. For U.S. data (January 1984−December 2007), the ratio of the variance of the five year yield to the one-year is 0.91. For the calibrated model, the ratio is 0.75.[6] Thus, excess volatility of long yields in the data, relative to the short end of the term

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

structure, is larger than the predictions of the DSGE model described 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 U.K., as table 2(d) shows, the variance of long yields larger than the variance at the short end of the yield curve for the entire sample period, and for October 1992 to December 2007.

2 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 also supply labor hours to k monopolistically competitive firms. Asset markets are incomplete and the households have access to n riskless bonds. Each household optimally chooses its consumption and its holdings of each $n-$ period bond. As households do not own capital, wealth can only be held in the form of these riskless bonds. Firms face Calvo (1983) pricing. The monetary authority is assumed to follow the Taylor rule for specifying the short interest rate, and responds 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). Although the primary objective of this analysis is to explain the Campbell-Shiller results, it is useful to note that Milani (2007) finds that embedding adaptive learning in a DSGE framework generates the persistence observed in macro variables such the output gap and inflation in U.S. data.[7]

2.0.1 Households

The optimization problem of household i is:

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

[7] The introduction of learning is found to replace other mechanical sources of persistence such as habit formation and inflation indexation.

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

$$C_t^i = \left(\int_0^1 c_t^i(k)^{\frac{\theta-1}{\theta}} dk \right)^{\frac{\theta}{\theta-1}}, \tag{3}$$

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. Here $B_{n,t}^i$ denotes the net holdings of a bond of maturity n - period at time t by household i. 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:

$$\begin{aligned} C_t^i + \sum\nolimits_{m=1}^{n} P_{m,t}^B B_{m,t}^i &\leq Y_t^i + \tilde{W}_t^i; \\ \tilde{W}_{t+1}^i &= B_{1,t}^i + \sum\nolimits_{m=2}^{n} P_{m-1,t+1}^B B_{m,t}^i. \end{aligned} \tag{4}$$

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

$$P_t = \left(\int_0^1 p_t(k)^{1-\theta} dk \right)^{\frac{1}{1-\theta}}; \; P_t Y_t^i = W_t h_t^i + \int_0^1 \Pi_t(k) dk, \tag{5}$$

where W_t is the competitive wage, and $\Pi_t(k)$ denotes the profits from k accruing to the household. Households receive income in the form of wages and own an equal part of each firm, and therefore receive a common share of the profits from the sale of each firm's good.

The No-Ponzi condition holds:

$$\lim_{j \longrightarrow \infty} \tilde{E}_t P_{1,t,t+j}^B \tilde{W}_{t+j+1}^i \geq 0, \tag{6}$$

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

The optimization problem of household i is to choose $\{c_t^i(k), h_t^i(k), B_{m,t}^i\}$ for all k, m to maximize the present discounted sum of utilities subject to the constraints in (4) and (6), taking as given $\{p_{t+j}(k), w_{t+j}(k), \Pi_{t+j}(k), P_{m,t+j}^B, \xi_{t+j}\}$ for all j, for the subjective expectations operator $\tilde{E}_t$. The only difference with the standard maximization problem is that here $\tilde{E}_t$ is used to denote subjective expectations, and the expectations formation process will be described in section 2.1 below. However, beliefs are assumed to be homogeneous across households, although the individual households do have knowledge about the beliefs of other households. Firms (discussed below) are assumed to value future streams of income at the marginal value of aggregate income in terms of the marginal value of an additional unit of aggregate income today. This implies that a unit of income in each state and date $t+k$ is valued by the kernel: $\beta^k \frac{P_t U_c(Y_{t+k},\xi_{t+k})}{P_{t+k} U_c(Y_t,\xi_t)}$. This simplifying assumption is valid in the context of the symmetric equilibrium of the model.

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

$$\hat{C}_t^i = (1-\beta)\hat{W}_t^i + (1-\beta)\tilde{E}_t \sum_{j=0}^{\infty} \beta^j \left[\hat{Y}_{t+j}^i - \sigma\beta(\hat{\imath}_{1,t+j} - \hat{\pi}_{t+j+1}) + \beta(a_{t+j} - a_{t+j+1})\right]. \qquad (7)$$

Here $\hat{W}_t^i = W_t^i/(P_t\bar{Y})$ 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 $a_t = -U_{c\xi}\xi_t/\bar{C}U_c$ is an exogenous disturbance term. The hat variables denote 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. 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 goods and asset market clearing conditions, the consumption decision rule in terms of the output gap[8] yields:

$$\hat{x}_t = \tilde{E}_t \sum\nolimits_{j=0}^{\infty} \beta^j \left[(1-\beta)\hat{x}_{t+j+1} - \sigma\beta(\hat{\imath}_{1,t+j} - \tilde{E}_t\hat{\pi}_{t+j+1}) + \hat{r}^n_{t+j+1} \right]. \tag{8}$$

Here, $\hat{x}_t = \log(Y_t/Y_t^n)$, Y_t^n is the natural rate of output and $\hat{r}_t^n = (\hat{Y}_{t+1}^n - a_{t+1}) - (\hat{Y}_t^n - a_t)$ 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.[9]

2.0.2 Term Structure

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

$$\hat{P}^B_{n,t} = \left[\hat{P}^B_{1,t} + \tilde{E}_t \hat{P}^B_{n-1,t+1} \right]. \tag{9}$$

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

$$\hat{P}^B_{n,t} = \left[\hat{P}^B_{1,t} + \tilde{E}_t \hat{P}^B_{1,t+1} + ... + \tilde{E}_t \hat{P}^B_{1,t+(n-1)} \right]. \tag{10}$$

The corresponding $n-$period interest rates are:

$$\hat{\imath}_{n,t} = \frac{1}{n} \left[\hat{\imath}_{1,t} + \tilde{E}_t \hat{\imath}_{1,t+1} + ... + \tilde{E}_t \hat{\imath}_{1,t+(n-1)} \right]. \tag{11}$$

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

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

[9] Additionally, as noted by Preston (2005), under rational expectations, the expectation that (8) 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.

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

$$\hat{\imath}^R_{n,t} = \hat{\imath}_{n,t} - \frac{1}{n}\left[\tilde{E}_t\hat{\pi}_{t+1} + \tilde{E}_t\hat{\pi}_{t+2} + ... + \tilde{E}_t\hat{\pi}_{t+n}\right]. \tag{12}$$

2.0.3 Firms

The full linearization of the optimization problem of the firm leads to the New-Keynesian Phillips curve (derived in Appendix A.1.):

$$\hat{\pi}_t = \kappa\hat{x}_t + \tilde{E}_t \sum\nolimits_{j=0}^{\infty} (\alpha\beta)^j \left[\kappa\alpha\beta\hat{x}_{t+j+1} + (1-\alpha)\beta\hat{\pi}_{t+j+1}\right], \tag{13}$$

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.).[10]

2.0.4 Monetary Authority

Finally, the one-period interest rate evolves according to the rule:

$$\hat{\imath}_{1,t} = \bar{\imath}_t + \phi_x\hat{x}_t + \phi_\pi\hat{\pi}_t, \tag{14}$$

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

2.1 Adaptive Learning

The complete description of the framework requires a forecasting model for the optimizing agents, which can be used to construct forecasts of the variables that are exogenous to the decision problems of households and firms: output gap, inflation, one-period interest rate, the longer interest rates, and the vector of exogenous disturbances $r_t = (\hat{r}^n_t, \bar{\imath}_t)'$.

Before the specification of the forecasting model, the state space of the model can be further simplified. The structural relation in (11) 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

[10] As before, under the assumption, (13) 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 $\tilde{E}_t$.

of the long bond. I assume that (11) will be used by the household to make conditional forecasts of longer interest rates. 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 (11) can be used to form forecasts of the 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.[11] 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.

Then, the set of variables that must be forecast by the optimizing agents is denoted by the vector $z_t = \{\hat{x}_t, \hat{\pi}_t, \hat{\imath}_{1,t}\}$, along with r_t.

2.1.1 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 variables to be forecasted on the history of the vector of variables that can be used as the basis for a forecast.

Under the rational expectations equilibrium, the variables in z_t are a function of the disturbances $r_t = (\hat{r}^n_t, \bar{\imath}_t)'$. The dynamics of the disturbances are assumed to follow the state-space representation:

$$r_t = H r_{t-1} + \varepsilon_{r,t}, \tag{15}$$

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. I assume that optimizing agents know the parameters of the H matrix with probability one. This is a standard assumption in the adaptive learning literature (see for instance, Chakraborty and Evans (2008), and reduces the degree of uncertainty generated in the model.

I assume that the agents understand the Minimum state variable form of the perceived process

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

for z_t, but will be updating their estimates of the parameters of the process. In this case, the perceived data generating process is:

$$z_t = a_t + b_t r_{t-1} + \eta_t, \tag{16}$$

where $a_t = [a_t^{\hat{x}}, a_t^{\hat{\pi}}, a_t^{\hat{i}_1}]'$ 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.[12]

2.1.2 Updating of Beliefs

Given the perceived data generating process in (16), 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:

$$\begin{aligned} \Omega_t &= \Omega_{t-1} + g^{-1}\Upsilon_{t-1}^{-1} q_{t-1}[z_t - \Omega_{t-1}' q_{t-1}]'; \\ \Upsilon_t &= \Upsilon_{t-1} + g^{-1}[q_{t-1}q_{t-1}' - \Upsilon_{t-1}], \end{aligned} \tag{17}$$

where $q_{t-1} = [1, r_{t-1}]$, 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. This controls the rate at which new information affects beliefs.[13]

A constant gain parameter g implies that the household puts greater weight on the more recent observations in the updating procedure. As the constant gain algorithm has found empirical support (see 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, the $n-$period

[12] The full perceived data generating process is given by:

$$\begin{pmatrix} z_t \\ r_t \end{pmatrix} = a_t + b_t \begin{pmatrix} z_{t-1} \\ r_{t-1} \end{pmatrix} + \eta_t.$$

However, the agents are assumed to understand that the coefficients corresponding to z_{t-1} will be zero, and the parameters of the H matrix are known with probability one.

[13] If g is a decreasing function over time, such as $g_t = 1/t$, the system of equations in (17) would be recursive representations of an ordinary least squares technique.

ahead forecast is:

$$\tilde{E}_t z_{t+n} = a_{t-1} + b_{t-1} H^{n-1} r_t, \ \forall n \geq 1. \tag{18}$$

Here a_{t-1} and b_{t-1} are the previous period's belief parameters.[14] 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 by using an anticipated utility argument as in Kreps (1998), 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) are 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.

2.1.3 Actual Data Generating Process

Substituting forecasts of the vector z_t in (18) 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. This process is consistent with the optimizing decision process of households:

$$z_t = T^0(a_{t-1}) + T^z(b_{t-1}) r_{t-1} + T^\varepsilon(b_{t-1}) \varepsilon_{r,t}. \tag{19}$$

The T matrices are functions of the model parameters. Self-referentiality in the learning model is generated because the observed values of z_t are used to form estimates of (a_t, b_t) in (17), and these are used to derive the process for z_t in (19).

2.1.4 Expectational Stability

The fixed point of the $T-$mappings in (19) 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 (18)) belongs to this class. If the self consistent equilibrium is Expectationally Stable

[14] Formation of forecasts at time t uses the beliefs from the past period, which justifies the use of (a_{t-1}, b_{t-1}) in (18).

(E-stable), it 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.[15] In (11), long interest rates are linear functions of the one-period yield and its expected future realizations, and $\hat{\imath}_{1,t}$ is the only relevant yield that must be forecast. 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. The model of the economy can be summarized: equilibrium dynamics determined by (8) and (13), the determination of asset prices in (11) and (12), the interest rate rule in (14), along with the system of equations used for forecasting in (17) and (18).

3 Results

Before presenting the predictions of the adaptive learning framework for real and nominal yields with respect to the two term structure moments being considered, I discuss the intuition for the results using the flexible price model. The calibration of model parameters used in the numerical analysis follows.

3.1 Flexible Price Model

The intuition for the mechanics of the learning model is first illustrated in the limiting 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 (1978) 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,

[15] This result is obtained for decreasing gain adaptive learning, that is g is a decreasing function of time.

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 with variance σ^2. This can be interpreted as the technology shock in the benchmark model with nominal rigidities. The limiting case of the benchmark model satisfies the E-stability criterion.

For the flexible-price limit of the benchmark model, the distribution of the belief parameters can be characterized as follows:

Proposition 1 *With flexible prices, under constant gain learning for a small gain* $g > 0$, *and large enough* gt, *the* 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. ■

The distribution of the coefficient will be used to derive analytically the bias in Campbell-Shiller coefficients in Proposition 2 below.

3.1.1 Campbell-Shiller Coefficients in the Flexible-Price Model

The flexible-price limit of the benchmark model allows for an analytical characterization of the bias in the Campbell-Shiller slope coefficient. In this case, under rational expectations, the one-period asset yield is determined entirely by the realization of the exogenous endowment process:

$$\hat{\imath}_{1,t} = \bar{T}^0 + \bar{T}^z\hat{y}_{t-1} + \bar{T}^\varepsilon\varepsilon_t^y. \tag{20}$$

Then, the rational expectations model does not rejection 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. 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 1 above). Due to the nature of the $T-$mappings, the paths of the asset prices are more complex: the law of motion in (19) 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 flexible-price case of the benchmark model:

Proposition 2 *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 mis-specification 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 mis-specification 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 mis-specification 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.

The dynamic responses of yields in response to an endowment shock for the Lucas economy further illustrate the difference between rational expectations and learning.[16] In response to the positive technology shock, the one-period bond price (and expected one-year 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

[16] Analytical proof available upon request.

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. A fraction of this decline is perceived to be permanent, and the household 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.

3.1.2 Volatilities in the Flexible-Price Model

For the limiting case with flexible prices, the distribution of beliefs is derived analytically in proposition 1. 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 3 *With flexible prices, under constant gain learning for small $g > 0$, and large gt:*

$$Var(y^L) > Var(y^{RE}) \tag{21}$$

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.4. ■

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.

3.2 Model Parameters and Solution

Before analyzing the quantitative results from the full model, two sets of parameters need to be specified - 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 since agents "forget" more data. The difference is large - a gain of 0.01 places 74% of the weight relative to rational expectations on an observation thirty quarters ago; the corresponding weight placed by a gain of 0.001 is 97%. The learning literature is mixed on the value of the gain parameter: Orphanides and Williams (2005) use a gain of 0.02, while Eusepi and Preston (2011) estimate a value of 0.002 for a real business cycle model with adaptive learning.

In order to discipline the gain parameter, I use survey data on expectations, which have been found to be reasonable approximations of subjective expectations in the literature (Bacchetta, Mertens and van Wincoop (2008); Froot (1989); Piazzesi, Salomao and Schneider (2013)). Survey expectations exhibit systematic forecasting errors at different horizons, and this can be considered as evidence of the limited rationality of professional forecasters and market participants. The systematic forecast errors generated by the self referential nature of the learning model can be considered a success of the framework.

Two survey datasets used: the Survey of Professional Forecasters (SPF) and the Michigan Survey of Consumer Finances (MSCF). Details of the surveys are shown in table 3. The survey forecast error is the difference between the realization of the variable m at time t, and the forecast one quarter ago, $E_{t-1}m_t$. I compute the forecast errors from the model in the same way. The survey data is used from 1992:Q1 to 2006:Q4; this is the time period for which the full length of survey forecasts is available for all series, including the ten-year Treasury bond yield[17].

The benchmark gain minimizes the distance between the autocorrelation in one-quarter ahead forecast errors of the three-month nominal interest rate from mean SPF forecasts, and the corresponding autocorrelation in the three-month nominal yield forecast error of the learning model.

[17] The three-month Treasury bill yield is obtained from St. Louis FRED database, and the ten-year yield is from the Gürkaynak, Sack and Wright (2007) data.

The autocorrelation in the one-quarter ahead forecast errors of the three-month nominal interest rate in mean SPF forecasts is 0.24. Table 4 shows the autocorrelation in forecast errors from the SPF survey data, and the learning model, for the three-month and ten-year yields, for the benchmark gain of 0.009. The autocorrelation implied by the rational expectations model is also shown. Additionally, the autocorrelation in the one-year ahead forecast errors of the inflation rate with the MSCF forecasts is 0.37,[18] and the alternative gain corresponding to forecast errors in inflation from the learning model is 0.01[19].

The remaining parameters of the model are: the Calvo frequency of price adjustment (α), the discount factor on a quarterly basis (β), the intertemporal elasticity of substitution (σ), the standard deviations of the technology, monetary policy and preference shocks, the persistence in these shocks (denoted by the AR(1) parameters), and the weight of output gap and inflation in the Taylor rule.

The frequency of price adjustment, α, is 0.75, corresponding to an yearly price re-optimization; β, the discount factor is 0.99, implying a quarterly real interest rate of approximately 4% and σ, the intertemporal elasticity of substitution is 0.2. These are in the range of these parameters used by Hördahl, Tristani and Vestin (2007) and Rudebusch and Swanson (2008) for U.S. data. The parameters in the monetary policy rule satisfy the conditions for the Taylor principle; the weight on inflation is 1.5 and on output gap is 0.9.

The persistence and standard deviation for the preference shock are set at 0.95 and 0.06 respectively, using the values from Ravenna and Seppälä (2008) and Rabanal and Rubio-Ramirez (2005). For the monetary policy shock, the persistence and standard deviation are 0.78 and 0.004, and these are in the range of values used by Rudebusch and Swanson (2008). The standard deviation for the technology shock is 0.01; this is calibrated to match the variance of model-implied output series with that of detrended U.S. output data for the period 1972:Q1 - 2006:Q4[20]. The persistence parameter for the technology shock is 0.9, and is set using the value from Ravenna and Seppälä (2008). These parameters are shown in table 5.

It is useful to compare the moments of the macroeconomic variables, implied by the rational

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

[19] The implied autocorrelation in the one-year ahead forecast errors in inflation from the learning model is 0.32.

[20] The data for output is obtained using the Quarterly Real GDP series, in 2009 chained dollars from the St. Louis FRED.

expectations and learning models, with corresponding moments in U.S. data. Table 6 shows the standard deviation and autocorrelations of the series implied by the models, along with the moments of the U.S. data for the period 1972:Q1 - 2006:Q4. The inflation series is computed using the series on Consumer Price Index (CPI)[21], from the Bureau of Economic Analysis, and the FRED three-month Treasury bill rate is used for the short-term nominal interest rate. The ten-year yield is obatined from the Gürkaynak, Sack and Wright (2007) dataset.

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 mean levels. However, as the learning beliefs are dispersed around the rational expectations means, the variance of the processes is expected to be higher than the rational expectations analog. As table 6 shows, for the benchmark gain parameter, this is indeed the case. In case of output, the standard deviation is 2.04 for the learning model, and 1.16 for the rational expectations case. The same patterns are found for the one-year and ten-year yields. The autocorrelations in the learning model are somewhat higher than the data; for the one-year yield, the autocorrelation is found to be 0.97, compared to 0.93 for U.S. data.

It is also useful here to discuss the implications of the framework used here for the steady state slope of the term structure. The U.S. nominal and real yield curves over the period of the analysis have a positive slope. Matching this term structure moment, while simultaneously explaining the Campbell-Shiller results and excess volatilities of yields has proven extremely challenging. Wachter (2006) successfully uses Campbell-Cochrane (1998) habit formation preferences in an endowment economy model to generate the Campbell-Shiller coefficients and term premium. However, Rudebusch and Swanson (2008) analyze several variants of general equilibrium models, and find that these are unable to satisfactorily match the above moments. [22] Kozicki and Tinsley (2001) and Fuhrer (1996) both find that incorporating learning in forming expectations help to explain the deviations from the Expectations Hypothesis. The shifting endpoints in former, and varying monetary policy rule coefficients in the latter generate similar effects to the time-varying risk premium.

[21] Inflation is computed as $\pi_t = 400 \ln (P_t / P_{t-1})$.

[22] Along with habit formation preferences, quadratic labor adjustment costs and real wage rigidities are introduced by the authors. However, both the term premium and Campbell-Shiller coefficients are significantly different from the data. In a subsequent paper (Rudebusch and Swanson, 2010), the authors introduce long-run nominal risk in a model with Epstein-Zin preferences preferences, which is able to generate a positive slope for the nominal term structure, but not for the real yield curve.

In the present exercise, the first-order linearization of optimal decision rules implies that the impact of learning beliefs on the risk-premium are not considered. Therefore, the steady state yield curve under constant-gain learning will be approximately the same as the rational expectations case, since beliefs are centered around the rational expectations means. However, this will not affect the other two moments being considered as the main channel to explain their evolution is the self-referential learning process by the optimizing households[23].

To numerically analyze the model, I initialize beliefs at their rational expectations values, and simulate 500 draws of the model for 1000 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. The dispersion of beliefs under the learning model around the time-invariant beliefs is a central feature of the learning model.

3.3 Campbell-Shiller Regression

I first present numerical estimates of the Campbell-Shiller regression coefficients for the real and nominal yield curves. I then discuss the results obtained in the adaptive learning framework in context of the existing literature.

3.3.1 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-quarter interest rate, and the results are reported for the same forecasting horizon as in the data, between two and ten years.

The second column in table 7(a) reports the slope coefficients for the benchmark gain. 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 can be rejected at the 95% confidence level.

[23] Allowing for a risk premium will alter this result on the steady state term structure, as the belief formation process will affect the covariance between consumption and asset prices.

The fourth column in table 7(a) reports the Campbell-Shiller coefficients under rational expectations, in case of the nominal yield curve. They are statistically not different from one at the 95% confidence interval. The rational expectations case is obtained when the gain coefficient 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 mis-specification in (1), other than sampling error. Column six shows the slope coefficients for nominal yields when the alternative gain parameter is used. Table 7(b) 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 understood 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 900. The pointwise median response in the difference of the trajectory of the relevant variables (with and without the shock) is then considered.[24]

Figure 1 shows the response of the output gap, inflation and the regressor and regressand of the Campbell-Shiller regression.[25] 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 1(a). 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 becomes important.

In the period following the shock, 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 with

[24] This experiment is repeated for 500 draws. The construction of the dynamic responses follows Eusepi and Preston (2011). A monetary policy shock is considered as it has been shown to have persistent, significant effects on the slope of the yield curve in structural VARs. See for instance, Evans and Marshall (2002).

[25] 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 results will be the same. The short yield denotes the one-period yield.

respect to the current long yield will be consistent with the Expectations Hypothesis. In this case, the agents correctly attribute all of their forecasting error to the transitory monetary policy shock. In terms of the difference between the expected long yield 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 long rate 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, households do not recognize that the entire rise in the yield is due to the transitory monetary policy shock. The agents attribute a fraction of their forecasting error to a permanent rise in the one-period yield at time t, and subsequently, an increase in the expected average returns, across the infinite horizon decision problem of the optimizing households. The fraction of the forecasting error attributed to a permanent change in the one-period yield is determined by the magnitude of the constant gain parameter. Subsequently, yields are also 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 savings for the infinite horizon problem. This is seen from the responses of the output gap and inflation in figures 1(c) and 1(d) - 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, as shown in figure 1(b).

3.3.2 Connections to the Literature

The analyses of Kozicki and Tinsley (2001) and Fuhrer (1996) use shifts in agents' expectations about monetary policy to explain the rejections of the Expectations Hypothesis in the data. In the first paper, the authors link changes in long-run forecasts of short yields to shifts in perceptions about the inflation target. Adaptive learning is one of the methods used to model the agents' behavior as they update their estimates of the long-run inflation target. These shifting endpoints in the short rates are incorporated into the determination of longer yields, and the Expectations Hypothesis is no longer rejected. The present analysis has a broader scope as it models the agents as facing uncertainty about the aggregate economy: they forecast future output gap, inflation and interest rates, and monetary policy is only one of the sources of uncertainty in beliefs. The rejections of the Expectations Hypothesis in the nominal term structures, during the Great Moderation and inflation targeting regimes in the U.S. and U.K. respectively, further illustrates the importance of these sources of uncertainty and of the general equilibrium framework which provides a tight link between different macroeconomic shocks and the yield curve.

Fuhrer (1996) models the short rate as being determined by the Federal Reserve, in response to output gap and inflation. He finds that the changes in the Federal Reserve's inflation target and response coefficients (to output gap and inflation) lead to variations in the long nominal rates of the magnitude that are observed in the data.[26] Fuhrer (1996) concludes that if shifts in the expectations formations process of future short rates is accounted for, then the Expectations Hypothesis fares well relative to the data.

Since a rich literature has attempted to explain the findings on the Campbell-Shiller coefficients by allowing for a time-varying term premia and subjective expectations, it is also useful to interpret the negative bias with respect to one in γ as the under-reaction of expected future yields of maturity $(n-1)$ to changes in the current short yield. In Froot's (1989) 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. The present paper expounds on Froot's (1989) analysis because expectational errors entirely account for the rejections of the rational Expectations Hypothesis since the first-order approximation of the model eliminates the time-varying risk premia. In addition, the adaptive learning approach pursued here provides a theory of expectations that generates systematic expectational errors.

[26] The long rates are derived using the Expectations Hypothesis.

Mankiw and Summers (1984) also 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 fundamentals than to future fundamentals.[27] In this context, the benchmark model used here shows that adaptive learning generates persistence in longer yields relative to shorter yields, supporting the finding of Mankiw and Summers that expected future yields do not overreact to short yield changes.

Recent work by Piazzesi, Salomao and Schneider (2013) highlights the importance of subjective expectations. The authors show 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 the authors, in an endowment economy 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 also 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.

It is useful 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

[27] The authors use the term premia to explain the rejections of the Expectations Hypothesis.

Seppälä (2008) 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. However, in their model, the authors 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.[28] Rudebusch and Swanson (2012) 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. The model successfully fits the nominal yield curve, but would potentially have difficulty in generating the right slope of the real term structure observed in U.S. TIPS data. [29]

Finally, Nimark (2012) uses a model of trading to show that when all traders do not have access to the same information, their non-nested information sets imply that individual traders can systematically exploit excess returns. They are able to take advantage of the forecasting errors of the other traders in the model, even when no trader is better informed than the other. In Nimark's analysis, the traders are rational, and dispersion in their expectations about bond returns are caused by observing different signals. Under perfect information, the Expectations Hypothesis holds. However, when information sets are non-nested, and long bonds are traded frequently (and not only held until maturity), the Hypothesis no longer holds, and excess returns are predictable. This result contrasts with the analysis of the current paper: here, although each agent does not know the beliefs of others, all subjective beliefs are assumed to be identical. Therefore, the Expectations Hypothesis holds here, and is a direct implication of the optimization problem of the agents.

3.4 Volatility of Yields

As for the Campbell-Shiller regression coefficients, I discuss the numerical result, followed by their context in the existing literature.

---

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

[29] In this 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.

3.4.1 Numerical Analysis

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 table 8(a), the variances of yields are larger than their rational expectations counterparts, for different values of the gain considered. For higher gains, the variances of the corresponding yields are larger under learning. This is expected - as the mis-specification 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, the volatility of the ten-year yield under learning is approximately twice the volatility of the ten-year yield in 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 five-year volatility to the one-year volatility is approximately 25% larger for the learning model, relative to rational expectations .

The variances of real yields (table 8(b)) 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.

It is also useful to analyze the role of price rigidities on yield variances in the learning model. This can be understood in the context of the rational expectations model. As can be seen from table 8(a), when the rational expectations model is calibrated to U.S. data, the level of nominal 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. 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 yields increases across the spectrum. These are much larger than the variances for U.S. data for the $1972-2007$ period. I discuss this result further in the context of the learning model below.

3.4.2 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 (2008) 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). 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.[30] Thus, price stickiness is key to lowering the volatility of yields at the short end of the curve and preventing a very steep decline in variances across the maturity structure, as would be predicted by the Expectations Hypothesis. 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. In the framework used here, price stickiness lowers the variance at the short end of the yield curve as well.

That adaptive learning can generate larger yield volatilities as shown here, has been discussed elsewhere in the literature as well. 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.

3.5 Effect of Different Monetary Policy Regimes in the Benchmark Model

Here I explore the implications of the model 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?

Taylor (1999) and Smith and Taylor (2009) discuss the change in the response coefficients in the monetary policy rule response to inflation and output gap. The 1980s and 1990s have been

[30] In the Hördahl et. al. (2007) model, the variance of the one-quarter interest rate increases by five times when $\alpha \longrightarrow 0$.

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.[31]

In terms of the Campbell-Shiller slope coefficients, as can be seen from table 1(b), the slope coefficients for U.S. nominal yield curve data are less negative for the period between $1984 - 2007$ at the long end of the yield curve relative to the entire period. 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. The model estimates are presented in the second column of table 9(a)

As can be seen from comparing the last two columns in table 9(b), the volatility of yields increases when the central bank's response to inflation becomes more aggressive. This effect is transmitted throughout the term structure. It is interesting to note that the differences for the two term structure moments at different values of the Taylor parameter for inflation are larger at the longer end of the yield curve.

3.6 Robustness with respect to Other Parameter Values

It is useful to analyze the implications of different parameter values on the term structure moments considered. Two parameters are of particular interest: the degree of price stickiness and the intertemporal elasticity of substitution. Table 10 shows the results for the Campbell-Shiller regression coefficients and the volatilities for various values of these parameters.

Greater flexibility in prices increases the level of term structure volatilities in both the rational expectations and learning frameworks. As noted in table 10, when the Calvo parameter is set to 0.60, the volatility of the short-term yield in the rational expectations case rises to 5.78. At the long end of the yield curve, a similar increase in volatility is observed. For the learning framework, the results are similar and both the short and long ends of the term structure experience higher volatilities. A change in price flexibility, however, does not have any quantitatively significant effects on the Campbell-Shiller coefficients. In the flexible-price model, the negative bias in the

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

regression coefficients is derived for the limiting case of $\alpha \to 0$.[32]

When the intertemporal elasticity of substitution is increased[33], the Campbell-Shiller coefficients are found to be closer to one; that is, the magnitude of the deviation from the Expectations Hypothesis in the learning model decreases. In response to the transitory monetary policy shock, for a larger intertemporal elasticity of substitution, the households are more willing to substitute intertemporally under the rational expectations and the learning frameworks. Under the learning model, while the optimizing agents still misperceive the transitory change to be permanent, the effect on consumption and consequent expected short-term yields is more muted. In terms of the yield volatility, the levels are smaller at both the short and long ends of the yield curve.

4 Conclusion

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

The Expectations Hypothesis implies that when the yield spread is high, long yields must rise over the life of the short yields. This implication does not hold, when tested using the Campbell-Shiller regression, for both the nominal and real yield curves. This is manifested in the Campbell-Shiller regression as slope coefficients which are smaller than one, and negative at long maturities. The regression, however, is constructed to jointly test the hypotheses of rational expectations and the Expectations Hypothesis. The analysis in this paper separates these by using adaptive learning. In this framework, the subjective Expectations Hypothesis holds since it is derived using the optimization problem of agents, but rational expectations do not.

When the yields from the learning model are used to construct the Campbell-Shiller regression, the slope coefficients match the pattern found in the empirical data. Learning introduces a systematic forecasting error, and consequently, the regression error is correlated with the yield spread. The orthogonality condition for the regression is violated, and a bias is introduced. The negative bias with respect to one is further explained due to the amplification of intertemporal substitution and

[32] In the rational expectations model, the response of output gap and inflation to the monetary policy shock in case of different values of price stickiness are only found to be marginally different.

[33] For the present utility specification, this implies a decrease in the curvature of the utility function.

income 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. The analysis therefore suggests that rational expectations based econometric tests of the Expectations Hypothesis, such as the Campbell-Shiller regression, are misspecified.

The learning model also resolves a part of the excess volatility puzzle. Long yields generated by the model are more volatile relative to the short yields than the corresponding volatilities in the rational expectations model. Additionally, the level of volatilities across the maturity structure are higher than the rational expectations model due to increased parameter uncertainty. These results hold for both the nominal and real yield curves.

The present analysis endogenously generates time variation in second moments via the expectations formation process. Through the lens of rational expectations models, this would appear as time variation in the risk premium. Therefore, the framework may be considered as nesting the risk-premia based explanations. In addition, the bias in Campbell-Shiller coefficients is characterized analytically. Finally, by using a micro-founded framework with the endogenous determination of output, inflation and interest rates, the model provides a theoretical link between the economy and the yield curve, in the spirit of the macro-finance literature.

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 (2012) 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 to lower the risk aversion parameter. In the present framework, time variation in risk premia generated by a recursive utility specification may no longer require high risk aversion 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. 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] Bacchetta, Philippe, Elmar Mertens and Eric van Wincoop. (2008). "Predictability in Financial Markets: What do Survey Expectations Tell Us?". Journal of International Money and Finance, 28: 406-426.

[2] Branch, William A. and George.W. Evans. (2006). "A Simple Recursive Forecasting Model". Economics Letters, 91: 158-166.

[3] Calvo, Guillermo. (1983). "Staggered Prices in a Utility-Maximizing Framework". Journal of Monetary Economics, 12: 383-398.

[4] Campbell, John Y. and John H. Cochrane. (1998). "By Force of Habit: A Consumption Based Explanation of Aggregate Stock Market Behavior". Journal of Political Economy, 107: 205-251.

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

[6] Chakraborty, Avik and George Evans. (2008). "Can Perpetual Learning Explain the Forward Premium Puzzle". Journal of Monetary Economics, 55: 477-490.

[7] Clarida, Richard, Jordi Gali and Mark Gertler. (1999). "The Science of Monetary Policy: A New Keynesian Perspective". Journal of Economic Literature, 37: 1661-1707.

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

[9] Dick, Christian D., Malik Schmeling and Andreas Schrimpf. (2013). "Macro-expectations, Aggregate Uncertainty, and Expected Term Premia". European Economic Review, 58: 58-80.

[10] Eusepi, Stefano, and Bruce Preston. (2011). "Expectations, Learning and Business Cycle Fluctuations". American Economic Review, forthcoming.

[11] Evans, George and Seppo Honkapohja. (2001). Learning and Expectations in Macroeconomics. Princeton University Press, Princeton, New Jersey.

[12] Evans, Charles L. and David Marshall. (2002). "Economic Determinants of the Nominal Treasury Yield Curve". Working paper 2001-16, Federal Reserve Bank of Chicago.

[13] Froot, Kenneth. (1989). "New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates". Journal of Finance, 44: 283-305.

[14] Fuhrer, Jeffery. (1996). "Monetary Policy Shifts and Long-Term Interest Rates". The Quarterly Journal of Economics, 111(4): 1183-1209.

[15] Gürkaynak, Refet, Brian Sack and Jonathan Wright. (2010). "The TIPS Yield Curve and Inflation Compensation". American Economic Journal: Macroeconomics, 2(1): 70-92.

[16] Gürkaynak, Refet, Brian Sack and Jonathan Wright. (2007). "The U.S. Treasury Yield Curve: 1961 to the Present". Journal of Monetary Economics, 54(8): 2291-2304.

[17] Hördahl, Peter, Oreste Tristani and David Vestin. (2008). "The Yield Curve and Macroeconomic Dynamics". Economic Journal, 118(533): 1937-1970.

[18] Kreps, David. (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.

[19] Kozicki, Sharon and Peter A. Tinsley. (2001). "Shifting Endpoints in the Term Structure of Interest Rates". Journal of Monetary Economics, 47(3): 613-652.

[20] Laubach, Thomas, Robert J. Tetlow and John C. Williams. (2007). "Learning and the Role of Macroeconomic Factors in the Term Structure of Interest Rates". Working paper.

[21] Lucas, Robert E. (1978). "Asset Prices in Exchange Economy", Econometrica, 46: 1426-1445.

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

[23] Mankiw, N. Gregory, Ricardo Reis and Justin Wolfers. (2004). "Disagreement about Inflation Expectations". NBER Macroeconomics Annual 2003, Cambridge MA, MIT press: 209-248.

[24] Mankiw, N. Gregory and Lawrence 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: 223-248.

[25] Marcet, Adam and Thomas J. Sargent. (1989). "Convergence of Least-Squares Learning in Environments with Hidden State Variables and Private Information". Journal of Political Economy: 1306-1322.

[26] Milani, Fabio. (2007). "Expectations, Learning and Macroeconomic Persistence". Journal of Monetary Economics, 54(7): 2065-2082.

[27] Muth, John F. (1961). "Rational Expectations and the Theory of Price Movements". Econometrica, 29: 315-335.

[28] Nimark, Kristoffer P. (2012). "Speculative Dynamics in the Term Structure of Interest Rates". Working paper.

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

[30] Pflueger, Carolin and Luis M. Viceira. (2011). "Inflation-Indexed Bonds and the Expectation Hypothesis". Annual Review of Financial Economics, 3: 139-158.

[31] Piazzesi, Monika and Martin Schneider. (2007). "Equilibrium Yield Curves". NBER Macroeconomics Annual 2006, Cambridge MA, MIT press: 389-442.

[32] Piazzesi, Monika, Juliana Salomao and Martin Schneider. (2013). "Trend and Cycle in Bond Premia". Working paper.

[33] Preston, Bruce. (2005). "Learning about Monetary Policy Rules When Long-Horizon Expectations Matter". International Journal of Central Banking, 1(2): 81-126.

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

[35] Ravenna, Federico and Juha Seppälä. (2008). "Monetary Policy and Rejections of the Expectations Hypothesis". Working paper.

[36] Rudebusch, Glenn D. and Eric T. Swanson. (2008). "Explaining the Bond Premium Puzzle with a DSGE Model". Journal of Monetary Economics, 55: S111-S126.

[37] Rudebusch, Glenn and Eric T. Swanson. (2012). "The Bond Premium in a DSGE Model with Long-Run Real and Nominal Risks". American Economic Journal: Macroeconomics, 4(1): 105-43.

[38] Sargent, Thomas. (1993). Bounded Rationality in Macroeconomics, Oxford University Press.

[39] Smith, Josephine and John B. Taylor. (2009). "The Term Structure of Policy Rules". Journal of Monetary Economics, 56: 907-917.

[40] Taylor, John B. (1999). "A Historical Analysis of Monetary Policy Rules". In Monetary Policy Rules, edited by John B. Taylor.

[41] Wachter, Jessica. (2006). "A Consumption-Based Model of the Term Structure of Interest Rates". Journal of Financial Economics, 79: 365-399.

[42] Woodford, Michael. (2003). Interest and Prices: Foundations of a Theory of Monetary Policy. Princeton University Press.

A Mathematical Appendix

Appendix A.1. Intertemporal Optimization of Agents:

The first order approximation of the optimality conditions of the households and firms 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 yields:

$$\hat{C}_t^i = \tilde{E}_t \hat{C}_{t+1}^i - \sigma(\hat{\imath}_{1,t} - \hat{\pi}_{t+1}) + (g_t - \tilde{E}_t g_{t+1}). \tag{22}$$

The flow budget constraint in (4) is iterated forwards, and its approximation is:

$$\tilde{E}_t \sum_{j=0}^{\infty} \beta^j \hat{C}_{t+j}^i = \hat{W}_t^i + \tilde{E}_t \sum_{j=0}^{\infty} \beta^j \hat{Y}_{t+j}^i. \tag{23}$$

Substituting (22) recursively into (23) yields:

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

The output gap is defined as $\hat{x}_t = \log(Y_t/Y_t^n)$. Aggregating across i households, and applying market clearing conditions yields:

$$\hat{x}_t = \tilde{E}_t \sum_{j=0}^{\infty} \beta^j \left[(1-\beta)\hat{x}_{t+j+1} - \sigma\beta(\hat{\imath}_{1,t+j} - \tilde{E}_t \hat{\pi}_{t+j+1}) + \hat{r}_{t+j+1}^n\right], \tag{24}$$

with 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 $k's$marginal cost in period $t+j$:

$$s(y,Y,\bar{\xi}) = \frac{v_h(f^{-1}(y/A;\xi)}{u_c(Y,\xi)A}\frac{1}{f'(f^{-1}(y))}, \tag{25}$$

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, $\frac{p_t(k)}{P_t} = \mu s(y_t(k), Y_t, \bar{\xi}_t)$,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 (25) gives $\hat{s}_{t,t+j}(k) = \omega\hat{y}_{t+j}(k) + \sigma^{-1}\hat{Y}_{t+j} - (\omega+\sigma^{-1})\hat{Y}^n_{t+j}$,where $\omega > 0$ is the elasticity of the real marginal cost function $s(.)$ with respect to $y_t(k)$. Aggregating this relation yields $\hat{s}_{t+j} = (\omega+\sigma^{-1})(\hat{Y}_{t+j} - \hat{Y}^n_{t+j})$.This implies the following relation between the real marginal cost of producing $y_t(k)$ and the aggregate output Y_t : $\hat{s}_{t,t+j}(k) = \hat{s}_{t+j} - \omega\theta\left[\hat{p}_t(k) - \sum_{m=t+1}^{t+j}\hat{\pi}_m\right]$. Finally, to derive the Phillips curve, differentiate the firm's optimization problem with respect to $p_t(k)$, and use this relation to get:

$$\hat{p}_t^* = \tilde{E}_t\sum\nolimits_{j=0}^{\infty}(\alpha\beta)^j\left[\frac{1-\alpha\beta}{1+\omega\theta}(\omega+\sigma^{-1})\hat{x}_{t+j} + \hat{\pi}_{t+j}\right]. \tag{26}$$

The can be rewritten using the approximation to the aggregate price index: $\hat{\pi}_t = \hat{p}_t^*(1-\alpha)/\alpha$.

Appendix A.2. Proof of Proposition 1:

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. The only process that will be forecasted by optimizing households is $\hat{P}_{1,t}$, since $\hat{y}_t$ is assumed to be known with probability one. The actual evolution of the one period asset price will be determined as $\hat{P}_{1,t} = T_0^p(a_t,b_t) + T_z^p(a_t,b_t)\hat{y}_{t-1} + \eta_t$. Now, the $T-$ mappings are constructed as:

$$T^{0,p}(a_t,b_t) = \frac{-\beta}{1-\beta}a_t^p \tag{27}$$

$$T^{z,p}(a_t,b_t) = \rho\left[\frac{(1-\beta)\rho}{\sigma(1-\beta\rho)} - \frac{\beta b_t^{p,y}}{1-\beta\rho}\right] + \frac{\rho}{\sigma} \tag{28}$$

The rational expectations equilibrium (REE) is defined as a fixed point of the mappings in (27) and (28): $T(\bar{a}_t, \bar{b}_t) = (\bar{a}_t, \bar{b}_t)$. Evans and Honkapohja (2001) use stochastic approximation results to show that the learning algorithm in (17)[34] converges to this REE if the following ordinary differential equation is locally stable: $\frac{\partial}{\partial\tau}(a_t,b_t) = T(a_t,b_t) - (a_t,b_t)$.The required 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

[34] Equation reference in original paper.

term is not estimated, and set $S_{t-1} = R_t$ to get:

$$\begin{aligned} b_t &= b_{t-1} + gS_{t-1}^{-1}\hat{y}_{t-1}[db_{t-1}\hat{y}_{t-1} + V_{b,t-1}\varepsilon_t] \\ S_t &= S_{t-1} + g(\hat{y}_t^2 - S_{t-1}) \end{aligned} \tag{29}$$

This is in the standard form: $\theta_t = \theta_{t-1} + g\mathcal{H}(\theta_{t-1}, X_t)$ where

$$\theta_t = \begin{pmatrix} b_t \\ S_t \end{pmatrix}, \mathcal{H}(\theta_{t-1}, X_t) = \begin{pmatrix} \mathcal{H}_b(\theta_{t-1}, X_t) \\ \mathcal{H}_S(\theta_{t-1}, X_t) \end{pmatrix}, X_t = \begin{pmatrix} b_t \\ b_{t-1} \\ \varepsilon_t \end{pmatrix} \tag{30}$$

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

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

$$\begin{aligned} h_b(b, S) &= \lim_{t \longrightarrow \infty} E\mathcal{H}_b(\theta_{t-1}, X_t) = S^{-1}E\hat{y}_t^2 \left[\frac{\rho}{\sigma} + \frac{\rho^2(\beta-1)}{\sigma(1-\beta\rho)} + b\left(\frac{-\beta\rho}{1-\beta\rho} - 1\right)\right] \\ h_S(b, S) &= \lim_{t \longrightarrow \infty} E\mathcal{H}_S(\theta_{t-1}, X_t) = E\hat{y}_t^2 - S \end{aligned} \tag{31}$$

Also,

$$B = D_\theta h(\theta^{RE}) = \begin{pmatrix} \left(\frac{-1}{1-\beta\rho}\right) & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix} \tag{32}$$

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

Considering only b, $\mathcal{H}_b(\theta, X^\theta) = S_{t-1}^{-1}\hat{y}_{t-1}[db\hat{y}_{t-1} + V_b\varepsilon_t]$ and $\mathcal{R}^{ij}(b) = \sum_{k=-\infty}^{\infty}[... + cov[\mathcal{H}^i(\theta, X_0^\theta), \mathcal{H}^j(\theta, X_0^\theta)] + cov[\mathcal{H}^i(\theta, X_1^\theta), \mathcal{H}^j(\theta, X_0^\theta)] + ...]$. At θ^{RE}, $cov[\mathcal{H}^i(\theta, X_t^\theta), \mathcal{H}^j(\theta, X_t^\theta)] = \frac{(1-\rho)^2}{\sigma^2}\frac{\sigma_\varepsilon^2}{E\hat{y}_t^2}$. This is valid since the unconditional expectation is taken as $t \longrightarrow \infty$. Since $\hat{y}_{t-1}^2, \varepsilon_t^2$ are independent random variables, $E(\hat{y}_{t-1}^2\varepsilon_t^2) = E\hat{y}_{t-1}^2\sigma_\varepsilon^2$. Other sums in the series are $cov[\mathcal{H}^i(\theta, X_{t+1}^\theta), \mathcal{H}^j(\theta, X_t)] = \frac{1}{(E\hat{y}_t^2)^2}\frac{(1-\rho)^2}{\sigma^2}E(\hat{y}_t\varepsilon_{t+1}\hat{y}_{t-1}\varepsilon_t) = 0$. Therefore, for b, $\mathcal{R}^*(b) = \frac{(1-\rho)^2}{\sigma^2}\frac{\sigma_\varepsilon^2}{E\hat{y}_t^2}$ and $C = \frac{(1-\rho)^2}{\sigma^2}\frac{(1-\beta\rho)(1-\rho^2)}{2}$.

Appendix A.3. Proof of Proposition 2:

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

$$\hat{\imath}_{1,t+1} - \hat{\imath}_{2,t} = \alpha + \gamma(\hat{\imath}_{2,t} - \hat{\imath}_{1,t}) + e_t \tag{33}$$

where $\hat{\imath}_{1,t} = -T_{b,t-1}\hat{y}_{t-1} - V_{b,t-1}\varepsilon_t$ and $\hat{\imath}_{1,t+1} = -T_{b,t}\hat{y}_t - V_{b,t}\varepsilon_{t+1}$, $\hat{\imath}_{2,t} = \frac{1}{2}(y_{1,t} + E_t y_{1,t+1})$.

In the regression in (33), define X and Y where

$$\begin{aligned} X &= \hat{\imath}_{2,t} - \hat{\imath}_{1,t} = \frac{\hat{y}_{t-1}}{2}T_{b,t-1}(1-\rho) - \frac{\varepsilon_t}{2}(T_{b,t-1} - V_{b,t-1}) \\ Y &= \hat{\imath}_{1,t+1} - \hat{\imath}_{2,t} = \hat{y}_{t-1}\left[-T_{b,t}\rho + \frac{1}{2}T_{b,t-1}(1+\rho)\right] + \varepsilon_t(-T_{b,t} + \frac{1}{2}(V_{b,t-1} + T_{b,t-1})) - V_{b,t}\varepsilon_{t+1} \end{aligned} \tag{34}$$

For computing the asymptotic bias in γ:

$$bias = \frac{\begin{array}{c} p\lim_{T\longrightarrow\infty} T^{-1}\sum_{t=1}^{T}(X_t[\varepsilon_t(-T_{b,t}+\frac{1}{2}(V_{b,t-1}+T_{b,t-1}))-V_{b,t}\varepsilon_{t+1}]) \\ -p\lim_{T\longrightarrow\infty}\left(T^{-1}\sum_{t=1}^{T}X_t\right)p\lim_{T\longrightarrow\infty}\left(T^{-1}\sum_{t=1}^{T}e_t\right)\end{array}}{p\lim_{T\longrightarrow\infty}(T^{-1}\sum_{t=1}^{T}X_t^2-(T^{-1}\sum_{t=1}^{T}X_t)^2)} \tag{35}$$

Consider the following term:

$$T^{-1}\sum_{t=1}^{T}X_t = T^{-1}\sum_{t=1}^{T}\left[\frac{1}{2}T_{b,t-1}(1-\rho)\hat{y}_{t-1}-\frac{1}{2}\varepsilon_t(T_{b,t-1}-V_{b,t-1})\right] \tag{36}$$

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:

$$\begin{aligned} T_{b,t} &= \frac{\rho}{\sigma}+\rho\left[\frac{-\beta b_t}{1-\beta\rho}+\frac{\rho(\beta-1)}{\sigma(1-\beta\rho)}\right] \\ V_{b,t} &= \left[\frac{-\beta b_t}{1-\beta\rho}+\frac{\rho(\beta-1)}{\sigma(1-\beta\rho)}\right]+\frac{1}{\sigma}\end{aligned} \tag{37}$$

Then,

$$p\lim_{T\longrightarrow\infty}T^{-1}\sum_{t=1}^{T}(T_{b,t-1}\hat{y}_{t-1}) = -\frac{\beta\rho}{1-\beta\rho}p\lim_{T\longrightarrow\infty}T^{-1}\sum_{t=1}^{T}\left(b_{t-1}\sum_{j=0}^{\infty}\rho^j\varepsilon_{t-1-j}\right) \tag{38}$$

Using the independence of random variables, and the weak law of large numbers, $T^{-1}\sum_{t=1}^{T}(T_{b,t-1}\hat{y}_{t-1}) = 0$. For the second term in $T^{-1}\sum_{t=1}^{T}(X_t)$:

$$T^{-1}\sum_{t=1}^{T}[\varepsilon_t(T_{b,t-1}-V_{b,t-1})] = (1-\rho)\frac{\beta}{1-\beta\rho}T^{-1}\sum_{t=1}^{T}(\varepsilon_t b_{t-1}) \tag{39}$$

Rewriting $T^{-1}\sum_{t=1}^{T}(\varepsilon_t b_{t-1})$, and applying $p\lim_{T\longrightarrow\infty}T^{-1}\sum_{t=1}^{T}(\varepsilon_t b_{t-1}) = p\lim_{T\longrightarrow\infty}T^{-1}\sum_{t=1}^{T}(X_t) = 0$, we get:

$$\begin{aligned} & T^{-1}\sum_{t=1}^{T}[\varepsilon_t(b_{t-2}+gR_t^{-1}\hat{y}_{t-2}[(T_{b,t-2}-b_{t-2})\hat{y}_{t-2}+V_{b,t-2}\varepsilon_{t-1}]) \\ = \; & T^{-1}\sum_{t=1}^{T}(\varepsilon_t b_{t-2})+gR_t^{-1}T^{-1}\sum_{t=1}^{T}(\varepsilon_t\hat{y}_{t-2}[(T_{b,t-2}-b_{t-2})\hat{y}_{t-2}+V_{b,t-2}\varepsilon_t])\end{aligned} \tag{40}$$

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

$$T^{-1}\sum_{t=1}^{T}(X_t e_t) = T^{-1}\sum_{t=1}^{T}\left[\begin{array}{c} \frac{1}{2}T_{b,t-1}(1-\rho)\hat{y}_{t-1}\varepsilon_t(-T_{b,t}+\frac{1}{2}(V_{b,t-1}+T_{b,t-1})) \\ +\frac{1}{2}T_{b,t-1}(1-\rho)\hat{y}_{t-1}V_{b,t}\varepsilon_{t+1} \\ -\frac{1}{2}\varepsilon_t(T_{b,t-1}-V_{b,t-1})\varepsilon_t(-T_{b,t}+\frac{1}{2}(V_{b,t-1}+T_{b,t-1})) \\ -\frac{1}{2}\varepsilon_t(T_{b,t-1}-V_{b,t-1})V_{b,t}\varepsilon_{t+1} \end{array}\right] \tag{41}$$

Again, using the independence of b_t^2 and ε_t, $T^{-1}\sum_{t=1}^{T}(X_t e_t)$ reduces to:

$$T^{-1}\sum_{t=1}^{T}(X_t e_t) = -\frac{1}{2}T^{-1}\sum_{t=1}^{T}[\varepsilon_t^2(T_{b,t-1}-V_{b,t-1})(-T_{b,t}+\frac{1}{2}(V_{b,t-1}+T_{b,t-1}))] \tag{42}$$

Also,

$$\begin{aligned} & T^{-1}\sum_{t=1}^{T}[(T_{b,t-1}-V_{b,t-1})T_{b,t}] \\ = \; & (1-\rho)\left(\frac{\rho}{\sigma}\right)^2 - (1-\rho)\rho T^{-1}\sum_{t=1}^{T}\left(\frac{\beta b_{t-1}}{1-\beta\rho}\frac{\beta b_t}{1-\beta\rho}\right) \\ & -2\frac{\rho^2(1-\rho)^2\beta\rho(1-\beta)}{\sigma^2(1-\beta\rho)^2} - \frac{\rho^3(1-\rho)(1-\beta)^2}{\sigma^2(1-\beta\rho)^2} \end{aligned} \tag{43}$$

and as $t \longrightarrow \infty$ $E(b_t^2) = gC + \frac{\rho^2(1-\rho)^2}{\sigma^2}$,

$$\begin{aligned} & T^{-1}\sum_{t=1}^{T}(T_{b,t}^2 - V_{b,t}^2) = \left[\frac{\beta^2(\rho^2-1)}{(1-\beta\rho)^2}\right]\left(gC+\rho^2(1-\rho^2)\right) \\ & +2\frac{\rho(1-\rho)}{\sigma}\left[\frac{\beta(1-\rho)}{\sigma(1-\beta\rho)^2}(1-\rho^2)\right] - \frac{(1-\rho)^2}{\sigma^2(1-\beta\rho)^2}\left[1-\rho^2\right] \end{aligned} \tag{44}$$

Finally, I have:

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}(X_t e_t) = \frac{\sigma_\varepsilon^2}{2} & \left[\begin{array}{c} (1-\rho)\left(\frac{\rho}{\sigma}\right)^2 - (1-\rho)\rho T^{-1}\sum_{t=1}^{T}\left(\frac{\beta b_{t-1}}{1-\beta\rho}\frac{\beta b_t}{1-\beta\rho}\right) \\ -2\frac{\rho^2(1-\rho)^2\beta\rho(1-\beta)}{\sigma^2(1-\beta\rho)^2} - \frac{\rho^3(1-\rho)(1-\beta)^2}{\sigma^2(1-\beta\rho)^2} \end{array}\right] \\ +\frac{\sigma_\varepsilon^2}{2} & \left[\begin{array}{c} \left[\frac{\beta^2(\rho^2-1)}{(1-\beta\rho)^2}\right]\left(gC+\rho^2(1-\rho^2)\right) \\ +2\frac{\rho(1-\rho)}{\sigma}\left[\frac{\beta(1-\rho)}{\sigma(1-\beta\rho)^2}(1-\rho^2)\right] \\ -\frac{(1-\rho)^2}{\sigma^2(1-\beta\rho)^2}\left[1-\rho^2\right] \end{array}\right] \end{aligned} \tag{45}$$

Using $E(b_t\hat{y}_t^2)$ and $E(b_t^2\hat{y}_t^2) = 0$, and $E(b_t b_{t+1}) = E(b_t^2)$, we get:

$$\frac{p\lim_{T\longrightarrow\infty} T^{-1}\sum_{t=1}^{T}(X_t e_t)}{p\lim_{T\longrightarrow\infty}(T^{-1}\sum_{t=1}^{T} X_t^2} - 1 = \frac{\sigma_\varepsilon^2}{2}\left[A - \frac{\beta^2(1-\rho)}{(1-\beta\rho)^2}E(b_t^2)\right] - 1 \tag{46}$$

where

$$A = (1-\rho)\left(\frac{\rho}{\sigma}\right)^2 - 2\frac{\rho^2(1-\rho)^2\beta\rho(1-\beta)}{\sigma^2(1-\beta\rho)^2} - \frac{\rho^3(1-\rho)(1-\beta)^2}{\sigma^2(1-\beta\rho)^2} + 2\frac{\rho(1-\rho)}{\sigma}\left[\frac{\beta(1-\rho)}{\sigma(1-\beta\rho)^2}(1-\rho^2)\right] - \frac{(1-\rho)^2}{\sigma^2(1-\beta\rho)^2}\left[1-\rho^2\right] \tag{47}$$

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.

A. 4: Proof of Proposition 3:

Under RE, the one-period asset price is $\hat{P}_{1,t}^{RE} = \frac{\rho(1-\rho)}{\sigma}\hat{y}_{t-1} + \frac{(1-\rho)}{\sigma}\varepsilon_t$ and its variance is $V(P_{1,t}^{RE}) = \frac{\rho^2(1-\rho)^2}{\sigma^2}E(\hat{y}_{t-1}^2) + \frac{(1-\rho)^2}{\sigma^2}E(\varepsilon_t^2)$. Under learning, the corresponding variables are $\hat{P}_{1,t}^{L} = T_a + T_b\hat{y}_{t-1} + V_b\varepsilon_t$, and $V(\hat{P}_{1,t}^{L}) = E\left[T_a + T_b\hat{y}_{t-1} + V_b\varepsilon_t - E(T_a + T_b\hat{y}_{t-1} + V_b\varepsilon_t)\right]^2$.

For evaluating the unconditional expectations, as $g \longrightarrow \bar{g}$, gt becomes large, $E(T_a) = 0$ and $E(T_b\hat{y}_{t-1}) = E\left[\left(\rho\left[\frac{-\beta b_{t-1}}{1-\beta\rho} - \frac{(1-\beta)\rho}{\sigma(1-\beta\rho)}\right] + \frac{\rho}{\sigma}\right)\hat{y}_{t-1}\right] = 0$. Similarly, $E(V_b\varepsilon_t) = 0$.

Now, the mappings are:

$$T_a = \frac{-\beta}{1-\beta}a_t;\ T_b = \rho\left[\frac{-\beta b_t}{1-\beta\rho} - \frac{(1-\beta)\rho}{\sigma(1-\beta\rho)}\right] + \frac{\rho}{\sigma} \qquad V_b = \left[\frac{-\beta b_t}{1-\beta\rho} - \frac{(1-\beta)\rho}{\sigma(1-\beta\rho)}\right] + \frac{1}{\sigma} \tag{48}$$

Then the difference between learning and RE variance is given by:

$$V(\hat{P}_{1,t}^{L}) - V(\hat{P}_{1,t}^{RE}) = \frac{\beta^2}{(1-\beta)^2}E(a_t^2) + E(\hat{y}_{t-1}^2)\left[\left(-\frac{(1-\beta)\rho^2}{\sigma(1-\beta\rho)} + \frac{\rho}{\sigma}\right)^2 - \frac{\rho^2(1-\rho)^2}{\sigma^2}\right] + E(\varepsilon_t^2)\left[\left(-\frac{(1-\beta)\rho}{\sigma(1-\beta\rho)} + \frac{1}{\sigma}\right)^2 - \frac{(1-\rho)^2}{\sigma^2}\right] \tag{49}$$

The constant terms in second and third terms are positive for all values of σ and $\beta, \rho \in (0,1)$. Therefore, $V(\hat{P}_{1,t}^{L}) = V(\hat{P}_{1,t}^{RE}) + f(V(a_t)$,positive constants). As the one-period yield is a linear transformation of the one-period price, the result follows.

Table 1a: Campbell-Shiller Slope Coefficients for U.K. Real Yields

n (Years)	1985:1-2007:4 γ	1985:1-1992:3 γ	1992:4-2007:4 γ
3.5	-0.80	-0.45	-0.98
10.5	-0.55	-0.94	-0.38
15.5	-0.42	-0.97	-0.28

Table 1b: Campbell-Shiller Slope Coefficients for U.S. Nominal Yields

n (Years)	1972:1-2007:4 γ	1972:1-1979:4 γ	1984:1-2007:4 γ
1	-1.55	-1.53	-1.61
5	-2.71	-1.66	-3.23
10	-3.44	-1.67	-3.25

Table 1c: Campbell-Shiller Slope Coefficients for U.K. Nominal Yields

n (Years)	1972:1-2007:4 γ	1972:1-1992:3 γ	1992:4-2007:4 γ
2	-0.55	-0.53	-0.64
5	-0.11	-0.10	-0.23
10	-0.37	-0.49	-0.16

Notes The regression in (1) is constructed using appropriate yields in table 1. In table (1a), the short yield is of maturity 2.5 years, in table (1b) this is 3 months and in table (1c), it is 1 year. These coefficients are statistically different from one at all conventional levels of significance.

Table 2a: Variances of U.S. Real Yields

n (Years)	2004:1-2007:4 σ^2	1999:1-2007:4 σ^2
2	1.09	na
5	0.36	1.22
10	0.11	0.74
20	0.05	0.57

Table 2b: Variances of U.K. Real Yields

n (Years)	1985:1-2007:4 σ^2	1985:1-1992:3 σ^2	1992:4-2007:4 σ^2
2.5	0.74	0.63	0.60
5.5	0.68	0.17	0.52
10.5	0.90	0.09	0.62
15.5	1.07	0.10	0.74

Table 2c: Variances of U.S. Nominal Yields

n (Years)	1972:1-2007:4 σ^2	1972:1-1979:4 σ^2	1984:1-2007:4 σ^2	Calibrated Model σ^2
1	8.86	2.80	5.35	1.81
5	6.93	1.00	4.91	1.37
10	5.82	0.80	4.21	0.95

.

Table 2d: Variances of U.K. Nominal Yields

n (Years)	1972:1-2007:4 σ^2	1972:1-1992:3 σ^2	1992:4-2007:4 σ^2
1	9.30	4.21	1.17
5	8.98	2.62	1.96
10	10.22	3.13	2.39

.

Table 3: 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 4: Forecast Errors

Autocorr in FE	1-Qtr Lag FE SPF-Mean FE	Learning model-implied FE	RE model-implied FE
3-month Bill	0.24	0.21	0.01
10-year Bond	0.19	0.11	0.00

Notes This table shows the autocorrelation in forecast errors (FE), computed as the difference between the realization of the yields m_t, and the expected values, $E_{t-1}m_t$, for the SPF survey data (for 1992:Q1 to 2006:Q4), and the learning and rational expectations models. The mean expectations reported in the SPF are used here.

Table 5: Model Parameters

Coeff	Value	Coeff	Value	Coeff	Value
α	0.75	ρ_{tech}	0.90	std($\varepsilon_{tech,t}$)	0.01
β	0.99	ρ_{mp}	0.78	std($\varepsilon_{mp,t}$)	0.004
σ	0.20	ρ_{pref}	0.95	std($\varepsilon_{pref,t}$)	0.06
		ϕ_π	1.50	ϕ_x	0.90

Notes This table shows the values of the model parameters used in the analysis. The AR(1) parameters of the technology, monetary policy and preference shocks are ρ_{tech}, ρ_{mp} and ρ_{pref}. The corresponding standard deviations are std($\varepsilon_{tech,t}$), std($\varepsilon_{mp,t}$) and std($\varepsilon_{pref,t}$). The Calvo parameter is α, the quarterly discount factor is β, and the intertemporal elasticity of substitution is σ. The parameters in the Taylor rule for inflation and output gap are given by ϕ_π and ϕ_x.

Table 6: Implied Macroeconomic Moments

Moments	U.S. Data Stdev	RE Model Stdev	Learning Model Stdev
Output	2.13	1.16	2.04
Inflation	1.06	0.69	0.85
One-year yield	2.97	2.01	2.36
Ten-year yield	2.41	0.75	1.27

Notes Inflation is computed from the CPI measure from the Bureau of Economic Analysis, where $\pi_t = 400ln(P_t/P_{t-1})$. The moments for the U.S. data are computed for the 1972:Q1 - 2006:Q4 period.

Table 7a: Campbell-Shiller Slope Coefficients for Nominal Yields

n (Years)	Model γ $g = 0.009$	% Rej. of EH	Model γ $g = 0$ (RE)	% Rej. of EH	Model γ $g = 0.01$	% Rej. of EH
1	-0.94	98.5%	0.99	4.5%	-1.00	97%
5	-2.00	69%	1.03	4%	-2.51	73%
10	-2.56	60%	1.04	4.5%	-3.38	67%

Table 7b: Campbell-Shiller Slope Coefficients for Real Yields

n (Years)	Model γ $g = 0.009$	% Rej. of EH	Model γ $g = 0$ (RE)	% Rej. of EH	Model γ $g = 0.01$	% Rej. of EH
1	-0.85	97%	1.06	5.5%	-0.91	95%
5	-3.74	93%	1.03	3.5%	-4.00	97%
10	-4.00	87%	1.05	3.5%	-4.96	93%

Notes The regression in (1) is constructed using appropriate yields in table 5. In all sub-tables, the short yield is of maturity one quarter. The percentage of times the Expectations Hypothesis is rejected is reported in column '% Rej. of EH'.

Table 8a: Variances of Nominal Yields

n (Years)	Model σ^2 $g = 0.009$	Model σ^2 $g = 0$ (RE)	Model σ^2 $g = 0.01$
1	5.44	3.93	6.00
5	2.80	1.92	3.42
10	1.69	0.58	2.31

Table 8b: Variances of Real Yields

n (Years)	Model σ^2 $g = 0.009$	Model σ^2 $g = 0$ (RE)	Model σ^2 $g = 0.01$
1	2.98	1.99	3.10
5	1.34	0.94	1.99
10	0.88	0.66	1.27

Table 9a: Campbell-Shiller Slope Coefficients for Nominal Yields for Different Monetary Policy Regimes

n (Years)	Model γ $\phi_\pi = 4$	% Rej. of EH	Model γ $\phi_\pi = 15$	% Rej. of EH
1	-0.84	92%	-0.81	83%
5	-1.77	68%	-1.42	70%
10	-2.00	65%	-1.94	63%

Table 9b: Variances for Nominal Yields for Different Monetary Policy Regimes

n (Years)	Model $\phi_\pi = 4$ σ^2	Model $\phi_\pi = 15$ σ^2
1	5.87	7.01
5	3.76	4.80
10	1.96	3.02

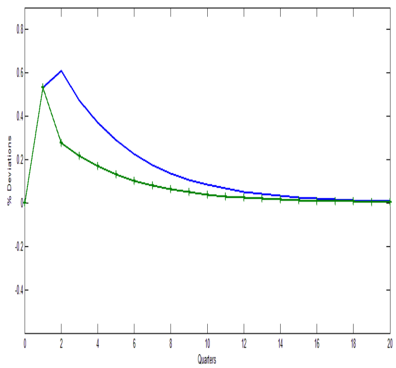
Table 10: Nominal Yields Moments for Different Parameter Values

Parameters	CS coeff		Variance	
	γ	γ	σ^2	σ^2
Years→	$n=1$	$n=10$	$n=1$	$n=10$
Benchmark *RE*	0.99 (3.5%)	1.04 (4%)	3.93	0.58
Benchmark *Learning*	−0.94 (96%)	−2.56 (59%)	5.44	1.69
$\alpha = 0.60$ *RE*	1.00 (5%)	1.06 (4.5%)	5.78	1.26
$\alpha = 0.60$ *Learning*	−0.85 (97%)	−2.70 (67%)	7.22	3.21
$IES = 0.4$ *RE*	1.01 (5%)	1.05 (3.5%)	3.02	0.33
$IES = 0.4$ *Learning*	−0.10 (84%)	−0.97 (56%)	4.57	1.40

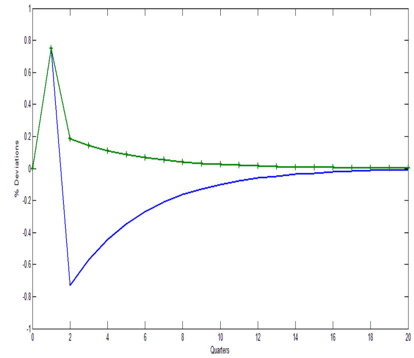
Notes This table reports the Campbell-Shiller slope coefficients and yield variances for the one- and ten-year maturity, for different values of parameters. Here, α is the Calvo parameter, and IES denotes the Intertemporal Elasticity of Substitution. The numbers in brackets in first two columns show the percentage of rejections of the Expectations Hypothesis. In these exercises, the gain is held fixed at the benchmark value.

Figure 1: Dynamic Responses under Rational Expectations and Learning

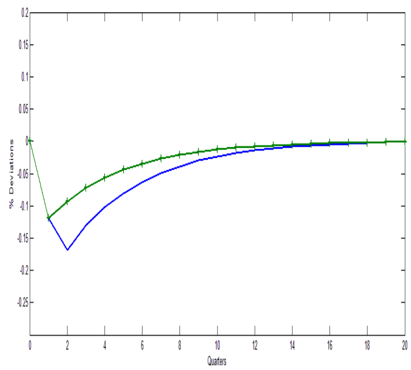
1a: Response of $-(\hat{\imath}_{2,t} - \hat{\imath}_{1,t})$



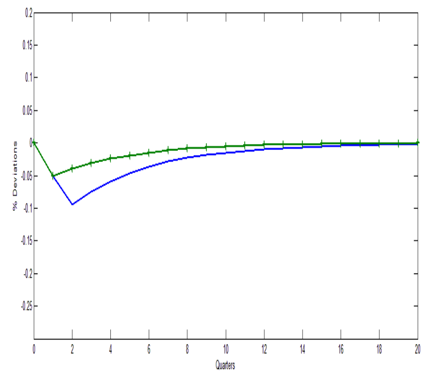
1b: Response of $-(E_t\hat{\imath}_{1,t+1} - \hat{\imath}_{2,t})$



1c: Response of Output Gap



1d: Response of Inflation



Note: The contractionary monetary policy shock occurs in period 1. The deviations from steady state are in percentages. The hatched green line denotes the response under Rational Expectations, and the solid blue line is the Learning response. These are the responses for the benchmark gain parameter $g = 0.009$.