Trend and Cycle in Bond Premia[*]

Monika Piazzesi Martin Schneider

Stanford & NBER Stanford & NBER

October 2008

Abstract

Common statistical measures of bond risk premia are volatile and countercyclical. This paper uses survey data on interest rate forecasts to construct subjective bond risk premia. Subjective premia are less volatile and not very cyclical; instead they are high only around the early 1980s. The reason for the discrepancy is that survey forecasts of interest rates are made as if both the level and the slope of the yield curve are more persistent than under common statistical models. The paper then proposes a consumption based asset pricing model with learning to explain jointly the difference between survey and statistical forecasts, and the evolution of subjective premia. Adaptive learning gives rise to inertia in forecasts, as well as changes in conditional volatility that help understand both features.

I Introduction

Many studies have documented that excess returns on long term bonds are predictable.[1] Indeed, postwar interest rate data reveal two striking patterns. On the one hand, high excess returns on long term bonds were typically preceded by a high spread between the long and short Treasury interest rates (a high "slope" of the yield curve). On the other hand, higher excess returns were typically preceded by a higher overall level of the yield curve. An investor who understood these two patterns would have predicted high excess returns on long bonds in times of high slope (for example, right after recessions) and in times of high level (especially in the early 1980s). Over the postwar period, the investor could have made a fortune trading on these predictions (borrowing short term and investing in long bonds in times of high slope or high level).

Why did investors not exploit these predictability patterns and thus make them disappear? There are two candidate reasons. The first is simply that investors' actual historical predictions of excess returns were different from the in-sample predictions found in today's statistical analysis. Investors may not have recognized the same patterns that we see today with the benefit of hindsight, at least not to the same extent. The second candidate reason is that changes in investors' assessment of risk were correlated with their excess return predictions. Even if investors predicted high excess returns on a strategy at some point in time, they may have chosen not to act on the prediction because they had simultaneously perceived an increase in the risk of the strategy, or because they had become more risk averse.

The standard approach in quantitative asset pricing studies is to focus almost exclusively on this second reason for predictability patterns. These studies assume that investors' historical predictions were identical to in-sample predictions derived today from statistical models, thus ruling out the first reason. Instead, most research effort bas been directed at documenting and explaining changes in risk assessment (e.g., Campbell and Cochrane 1995, Bansal and Yaron 2004, Wachter 2006, Hansen, Heaton, Lee, and Roussanov 2007.)

The goal of this paper is to document and explain the role of *both* candidate reasons for pre-

---

[1] As a concrete example, consider the excess return on a 10 year Treasury bond held over one year. It is defined as the return earned by holding the bond over the course of a year minus the one year interest rate available at the beginning of the year.

dictability – differences between actual historical and today's statistical predictions on the one hand, and changes in risk assessment on the other. The first part of the paper uses survey data on interest rate and inflation forecasts to document historical predictions of excess returns and to compare those to predictions from statistical models. There are two main results. First, historical predicted excess returns vary less than statistical predictions (by roughly one half). Second, historical predicted excess returns move less with the business cycle, but have a larger low frequency component that is correlated with inflation.

These findings imply that both candidate reasons for predictability patterns are important and that structural models should account for both. Indeed, a consumption based asset pricing model says that excess returns R_{t+1}^e should satisfy $E_t\left[R_{t+1}^e M_{t+1}\right] = 0$, where M_{t+1} is marginal utility and where the expectation is taken *under the subjective belief of the investor.* The expected excess return measured today with hindsight by an econometrician can thus be decomposed into

$$(1) \qquad \widehat{E}_t\left[R_{t+1}^e\right] \;\; = \;\; \underbrace{\widehat{E}_t\left[R_{t+1}^e\right] - E_t\left[R_{t+1}^e\right]}_{\text{difference in predictions}} - \underbrace{\text{cov}_t\left(M_{t+1}, R_{t+1}^e\right)/E_t M_{t+1}}_{\text{risk premium}}.$$

The standard approach to explaining predictability patterns (that is, changes in $\widehat{E}_t R_{t+1}^e$) assumes that the first component is zero and that changes in risk assessment move around risk premia. Our findings call for a theory that can jointly account for both components.

To this end, the second part of the paper studies a representative agent asset pricing model with learning and recursive utility. An adaptive learning algorithm gives rise to a sequence of subjective beliefs about consumption growth, inflation and interest rates. We compute predictions made by learning agents under their subjective belief, and thus obtain an explanation for the first term in (1). We check the relevance of this explanation by comparing the predictions made by learning agents to actual survey forecasts. Learning not only changes the forecasting rules used to make predictions, but also the risk perceived by investors. Our recursive utility model with subjective beliefs thus implies time varying subjective risk premia, providing an explanation for the second term in (1). We check the relevance of this explanation by examining the model's Euler equation errors.

We find that our model can help understand the movements in both components of statistical risk premia (1). On the one hand, adaptive learning provides a reason for systematic differences between statistical forecasts and survey forecasts. Adaptive learners react slowly to new information. They forecast relatively low interest rates as rates are rising, such as before 1980, and relatively high interest rates when rates are falling, such as after 1980. Moreover, they do not change their forecasts in response to sharp changes in rates, be they upward changes, as after the oil price shocks, or downward changes, as in the recessions after 1980. Overall, adaptive learning gives rise to a difference in predictions that is low before 1980, high after 1980 and countercyclical, much like the difference in predictions from surveys in (1), and like the statistical expected excess returns themselves.

On the other hand, adaptive learning gives rise to changes in perceived risk that in turn generate low frequency movements in subjective risk premia on long bonds. With recursive utility, the risk premium on an asset is high if the asset payoff either covaries a lot either with consumption growth or with news about future consumption growth. Under adaptive learning, the covariance between bond payoffs and consumption changes over time: after the experience of the 1970s, high interest rates were viewed much more as a predictor of low growth than in later years. As a result, adaptive learners viewed bonds as particularly unattractive around 1980, and demanded high subjective premia.

We build on a small literature which has shown that measuring subjective beliefs via surveys can help understand asset pricing puzzles. Froot (1989) argued that evidence against the expectations hypothesis of the term structure might be due to the failure of the (auxiliary) rational expectations assumption imposed in the tests rather than to failures of the expectations hypothesis itself. He used the Goldsmith-Nagan survey to measure interest rate forecasts and found that the failure of the expectations hypothesis for long bonds can be attributed to expectational errors. The findings from our reduced form model confirm Froot's results while including the BlueChip data set that allows for a longer sample as well as more forecast horizons and maturities. Moreover, our estimation jointly uses all data and recovers and characterizes the kernel M.[2] Several authors have explored the role

---

[2] Kim and Orphanides (2007) estimate a reduced-form term structure model using data on both interest rates and interest rate forecasts. They show that incorporating survey forecasts into the estimation sharpens the estimates of risk premia in small samples. In our language, they obtain more precise estimates of "statistical premia"; they are not interested in the properties of subjective risk premia for structural modelling. Chernov and Mueller (2008) adopt

of expectational errors in foreign exchange markets. Frankel and Froot (1989) show that much of the forward discount can be attributed to expectational errors. Gourinchas and Tornell (2004) use survey data to show that deviations from rational expectations can rationalize the forward premium and delayed overshooting puzzles. Bacchetta, Mertens and van Wincoop (2008) study expectational errors across a large number of asset markets.

The rest of the paper is structured as follows. Section II documents properties of survey forecasts. Subsection II.A takes a first look at the raw survey data for selected maturities and forecasting horizons. Subsection II.B compresses the information from the surveys using a time series model. Section III introduces the modeling framework and reports results from the structural model.

II Stylized Facts

We consider "zero coupon" bonds that pay off only once at some specified maturity date. Let $P_t^{(n)}$ denote the date t price of a zero coupon bond with maturity n that is, the bond pays off one dollar n periods from date t. The yield to maturity on this bond is defined as $i_t^{(n)} = -\log P_t^{(n)}/n$. It represents the per period interest rate earned from holding the bond to maturity if gains are continuously compounded.

The log excess return from holding the n period bond from date t to date $t+1$ is denoted $rx_{t,t+1}^{(n)}$. It is defined as the log capital gain on the n-period bond less the one period interest rate:

$$rx_{t,t+1}^{(n)} = \log P_{t+1}^{(n-1)} - \log P_t^{(n)} - i_t^{(1)}.$$

Excess returns over longer holding periods are defined analogously. We use lower case letters for logarithms, e.g. $p_t^{(n)} = \log P_t^{(n)}$ and so on. The log excess return on an n-period bond held from t to $t+h$ is thus

$$(2) \qquad rx_{t,t+h}^{(n)} = p_{t+h}^{(n-h)} - p_t^{(n)} - hi_t^{(h)},$$

where $i_t^{(h)}$ is the date t yield to maturity on an h-period bond. A high excess return on a long

---

a similar approach for inflation forecasting.

bond is earned if the capital gain exceeds the short term interest rate. This is more likely if the rate $i_t^{(n-h)}$ over the remaining life of the bond is relatively low.

Excess returns and forward rates

Excess returns on long bonds can also be written in terms of forward rates. A forward contract fixes a price (or, equivalently, a yield to maturity) for purchase or sale of a specified bond at some specified date in the future. Let $F_t^{(n-h,h)}$ denote the price fixed at date t for an $n-h$-period bond to be purchased at date $t+h$. In a frictionless market, locking in the purchase of an $n-h$-period bond for date $t+h$ is the same as borrowing $P_t^{(n)}$ dollars at the h-period rate at date t and using the borrowed money to purchase one n-period bond. In both cases, one ends up paying some amount of money at date $t+h$ in exchange for an $(n-h)$-period bond. The absence of arbitrage thus requires that the payments are the same under the two strategies, that is, the forward price to be paid at date $t+h$ is the same as the repayment on the h-period loan at date $t+h$:

(3)
$$F_t^{(n-h,h)} = P_t^{(n)} \exp(h i_t^{(h)}).$$

Let $f_t^{(n-h,h)} = -\log F_t^{(n-h,h)}/(n-h)$ denote the yield to maturity on the forward contract, or the "forward rate". Condition (3) implies that the forward rate can be expressed as a linear function of the n- and h-period interest rates

$$
\begin{aligned}
f_t^{(n-h,h)} &= \frac{n}{n-h} i_t^{(n)} - \frac{h}{n-h} i_t^{(h)} \\
&= i_t^{(h)} + \frac{n}{n-h} \left(i_t^{(n)} - i_t^{(h)} \right).
\end{aligned}
$$

If the maturity of the long bond n and a small maturity h of the forward contract, the forward rate behaves is essentially equal to the long bond rate. In contrast, if n is large but h is close to n, the forward rate behaves like a scaled version of the spread between the two rates.

It now follows that the excess return (2) can be rewritten as the (scaled) difference between the forward and the future spot rate:

$$rx_{t,t+h}^{(n)} = (n-h) \, (f_t^{(n-h,h)} - i_{t+h}^{n-h})$$

Excess returns are thus high if realized spot rates are lower than the associated forward rates.

Statistical and historical expected excess returns

This paper is about conditional expectations of excess returns. In particular, we want to distinguish between expected excess returns implied by a statistical model and historical predictions of excess returns. We write $\widehat{E}_t rx_{t,t+h}^{(n)}$ to denote the date t prediction from a statistical model, for example, a regression of log excess returns on variables known at date t. This is the object that the literature has been interested in: the stylized fact "predictability of excess returns on long bonds" means that $\widehat{E}_t rx_{t,t+h}^{(n)}$ moves around over time. Of course, the behavior of the conditional expectation depends on the particular statistical model used to compute it, and we will discuss several alternative models. In the literature, predictability has been established for a wide range of models.

We reserve the notation $E_t rx_{t,t+h}^{(n)}$ for historical predictions, or "true" expectations, of excess returns made at date t. To relate the two types of predictions, we decompose the expected excess return measured by a statistician as

$$
(4) \qquad \underbrace{\widehat{E}_t\left[rx_{t,t+h}^{(n)}\right]}_{} = E_t\left[rx_{t+h}^{(n,h)}\right] + \widehat{E}_t\left[rx_{t+h}^{(n-h)}\right] - E_t\left[rx_{t+h}^{(n-h)}\right]
$$

$$
= E_t\left[rx_{t+h}^{(n,h)}\right] + \widehat{E}_t\left[p_{t+h}^{(n-h)}\right] - E_t\left[p_{t+h}^{(n-h)}\right]
$$

$$
= \underbrace{E_t\left[rx_{t+h}^{(n,h)}\right]}_{} + \underbrace{(n-h)\left(E_t\left[i_{t+h}^{(n-h)}\right] - \widehat{E}_t\left[i_{t+h}^{(n-h)}\right]\right)}_{}
$$

statistical premium $\;=\;$ subjective premium $\;+\;$ subj. $-$ stat. interest-rate expectation

Here the second line reflects the fact that the only uncertain part in the excess return is the future bond price, and the third line follows from the definition of the yield to maturity. The decomposition shows that statistical expected excess returns can move around either because actual historical expected excess returns move around, or because statistical interest rate forecasts differ from historical interest rate forecasts.

In the next two subsections, we measure the two terms in the decomposition (4) with data on actual interest rates as well as survey forecasts. In subsection II.A, we take the simplest possible approach. We run regressions of excess returns on a set of date t interest rates to construct measures

7

of $\widehat{E}_t rx_{t,t+h}^{(n)}$ and $\widehat{E}_t i_{t+h}^{(n-h)}$, and we use the median survey forecast of $i_{t+h}^{(n-h)}$ as a measure of $E_t i_{t+h}^{(n-h)}$. This approach delivers a decomposition for a given bond maturity and a given forecast horizon, and thus provides a first look at the data. Unfortunately, the approach is limited by the nature of survey forecast data, which provide long samples for only a few bond maturities and forecast horizons. It also does not simultaneously process the information contained in the surveys for the many maturities and horizons that are available.

To compress all the available information from surveys, subsection II.B estimates a statistical model that describes the joint dynamics of actual interest rates and survey forecasts. The model is a standard linear state space system – both interest rates and survey forecasts are represented as linear functions of a small number of factors. We show that such a system does a decent job in describing the joint dynamics of interest rates and forecasts. We then proceed to use the system to show decompositions (4) for bonds and holding periods for which raw survey data are not available.

A   A first look at the data

We measure subjective expectations of interest rates with survey data from two sources. Both sources conduct comparable surveys that ask approximately 40 financial market professionals for their interest-rate expectations at the end of each quarter and record the median survey response. Our first source are the Goldsmith-Nagan surveys that were started in mid-1969 and continued until the end of 1986. These surveys ask participants about their one-quarter ahead and two-quarter ahead expectations of various interest rates, including the 3-month Treasury bill, the 12-month Treasury bill rate, and a mortgage rate. Our second source are Bluechip Financial Forecasts, a survey that was started in 1983 and continues until today. This survey asks participants for a wider range of expectation horizons (from one to six quarters ahead) and about a larger set of interest rates. The most recent surveys always include 3-month, 6-month and 1-year Treasury bills, the 2-year, 5-year, 10-year and 30-year Treasury bonds, and a mortgage rate.[3]

To measure interest-rate expectations from a statistical model, we estimate unrestricted VAR

---

[3]The survey questions ask for constant-maturity Treasury yield expectations. To construct zero-coupon yield expectations implied by the surveys, we use the following approximation. We compute the expected change in the n-year constant-maturity yield. We then add the expected change to the current n-year zero-coupon yield.

dynamics for a vector of interest rates with quarterly data over the sample 1952:2-2007:4 and compute their implied forecasts. Later, in section II.B, we will impose more structure on the VAR by assuming the absence of arbitrage and using a lower number of variables in the VAR, and thereby check the robustness of the empirical findings we document here. The vector of interest rates Y includes the 1-year, 2-year, 3-year, 4-year, 5-year, 10-year and 20-year zero-coupon yields. We use data on nominal zero-coupon bond yields with longer maturities from McCulloch and Kwon (1993). The sample for these data is 1952:2 - 1990:4. We augment these data with the new Gurkaynak, Sack, and Wright (2006) data. We compute the forecasts by running OLS directly on the system $Y_{t+h} = \mu + \phi Y_t + \varepsilon_{t+h}$, so that we can compute the h-horizon forecast simply as $\mu + \phi Y_t$.

We can evaluate equation (4) based on our survey measures of subjective interest-rate expectations $\hat{E}_t i_{t+h}^{(n-h)}$ and the VAR measures of expectations $\hat{E}_t i_{t+h}^{(n-h)}$ for different maturities n and different horizons h. Figure 1 plots the left-hand side of equation (4), expected excess returns under VAR beliefs as a black line, and the second term on the right-hand side of the equation, the difference between subjective and VAR interest-rate expectations, as a gray line. For the short post-1983 sample for which we have Bluechip data, we have data for many maturities n and many forecasting horizons h. The lower two panels of Figure 1 use maturities $n = 3$ years and 11 years and a horizon of $h = 1$ year, so that we deal with expectations of the $n - h = 2$ year and 10 year interest rate. These combinations of n and h are in the Bluechip survey, and the VAR includes these two maturities as well so that the computation of expectations is easy.

For the long post-1970 sample, we need to combine data from the Goldsmith-Nagan and Bluechip surveys. The upper left panel shows the $n = 1.5$ year bond and $h = 6$ month holding period from the estimated VAR (which includes the $n - h = 1$ year yield.) This works because both surveys include the $n - h = 1$ year interest rate and a $h = 6$-month horizon. The VAR delivers a 6-month ahead expectation of the 1-year interest rate. For long bonds, we do not have consistent survey data over this long sample. To get a rough idea of long-rate expectations during the Great Inflation, we take the Goldsmith-Nagan data on expected mortgage-rate changes and the Bluechip data on expected 30-year Treasury-yield over the next $h = 2$ quarters and add them to the current 20-year zero-coupon yield. The VAR produces a $h = 2$ quarter ahead forecast of the 20-year yield.

Figure 1 also shows NBER recessions as shaded areas. The plots indicate that expected excess
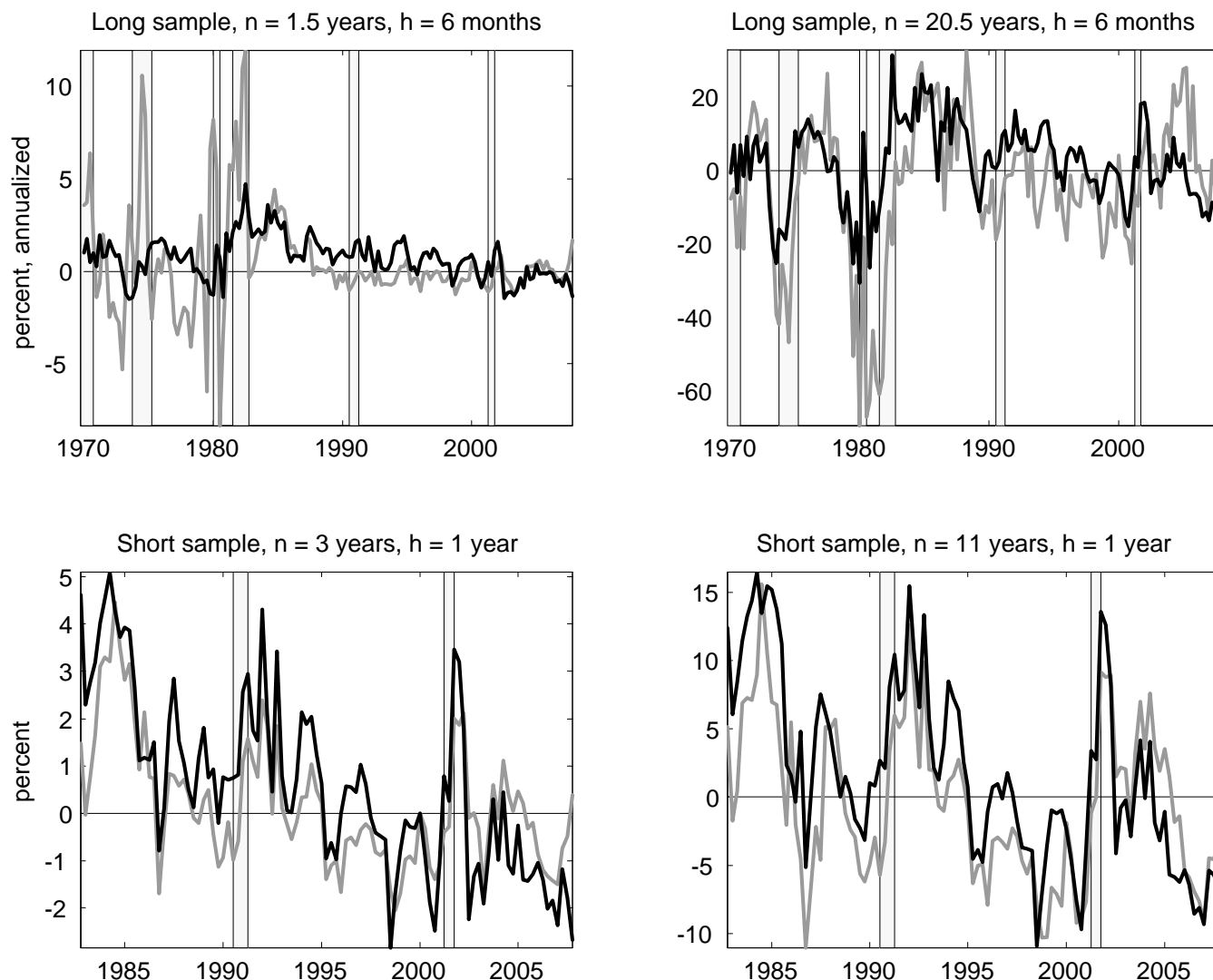
Figure 1: Each panel shows expectations of excess returns derived from a VAR in black (the left-hand side of equation (4)) and the difference between subjective and VAR interest-rate expectations in gray (the second term on the right-hand side of the equation) for the indicated bond maturity n and holding period/forecast horizon h. Shaded areas indicate NBER recessions. The numbers are annualized and in percent. The upper panels show data over a longer sample than the lower panels.

returns computed from the VAR and the difference between subjective and VAR interest-rate expectations have common business-cycle movements. The patterns appear more clearly in the lower panels which use longer (1 year) horizons. This is not surprising in light of the existing predictability literature which documents that expected excess returns on bonds and other assets are countercyclical when we look at longer holding periods, such as one year (e.g., Cochrane and

Piazzesi 2005.) In particular, expected excess returns are high right after recession troughs. The lower panels show indeed high values for both series around and after the 1991 and 2001 recessions. The series are also high in 1984 and 1996, which are years of slower growth (as indicated, for example, by employment numbers) although they were not classified as recessions.

For shorter holding periods, the patterns are also there in the data but they are much weaker. However, the upper panels show additional recessions where similar patterns appear. For example, the two series in both panels are high in the 1970, 1974, 1980 and 1982 recessions or shortly afterwards. (As we can see in the upper panels, expected excess returns for short holding periods are large when annualized. Of course, the risk involved in these investment strategies is high, and so they are not necessarily attractive.)

<div align="center">

TABLE 1: FORECAST ERRORS REGRESSIONS FROM SURVEYS

</div>

			maturity				
	1 year	2 year	3 year	5 year	7 year	10 year	30 year
YS	0.22	0.53	0.89	1.72	2.70	4.08	12.98
t-stat	0.60	0.84	1.05	1.46	1.85	2.09	2.18
R^2	0.02	0.03	0.04	0.06	0.08	0.10	0.14
CP R^2	0.14	0.10	0.08	0.09	0.11	0.13	0.19

Note: This table reports results from regressing $(n-h)\left(E_t i_{t+h}^{(n-h)} - i_{t+h}^{(n-h)}\right)$ on time t variables over the Bluechip sample. The forecast $E_t i_{t+h}^{(n-h)}$ is measured as median forecast in the Bluechip survey. The "YS" row reports the slope coefficient from a regression on a constant and the spread between the 5-year yield and the 3-month short rate, and its t-statistic computed with Hansen-Hodrick standard errors with 4 quarterly lags. The "CP R^2" uses five forward rates as right-hand side variables.

The above findings depend on a particular choice of a statistical model, or VAR. We can see the business-cycle patterns more directly by looking at realized survey forecast errors $E_t i_{t+h}^{(n-h)} - i_{t+h}^{(n-h)}$ instead of forecast differences relative to the VAR. Table 1 reports results from regressions of these realized survey forecast errors, scaled by $(n-h)$ as on the right-hand side of equation (4), on time t variables. The row labeled "YS" reports the slope coefficients of a regression on the time t yield spread between the 5-year bond and the 1-quarter bond. The point estimates in these regression are positive, so that survey forecast errors $E_t i_{t+h}^{(n-h)} - i_{t+h}^{(n-h)}$ are countercyclical and thus

systematic over the business cycle. In particular, survey interest rate forecasts tend to be above the subsequent realized values in periods of high spreads, and thus recessions. The slope coefficient estimates increase with the maturity of the bond whose interest rate we are forecasting and are significant for long bonds (10 and 30 years.) The R^2 in these regression range from 2% to 14%. We also report the R^2 from a regression on five forward rates, as in Cochrane and Piazzesi (2005), indicated "CP R^2". Naturally, these R^2 are higher, and range from 8% - 19%.

A potential concern with Bluechip forecast data is that the survey is not anonymous, and so career concerns of survey respondents may matter. To address this concern, we also measure subjective interest-rate expectations using the Survey of Professional Forecasters. Starting in 1992, the SPF reports median interest-rate forecasts for the 10-year Treasury bond over various forecast horizons. We find that median forecasts from the SPF are similar to those from the Bluechip survey. Importantly, the differences between SPF forecasts and VAR expectations show the same patterns as those documented in Figure 1.

To sum up, the evidence presented in this section suggests that subjective interest-rate expectations deviate from the expectations that we commonly measure from statistical models. Figure 1 suggest that these deviations may also be responsible for the time-variation in statistical bond premia.

B   Joint dynamics of interest rates and survey forecasts

We describe the dynamics of interest rates, macro variables, and median survey forecasts by a state space system with a small number of factors. This approach is known to work well, because interest rates of different maturities are highly correlated. For example, in a typical cross section of interest rates, the first principal component typically explains more 90% of the variation. We assume that there is a vector of four factors that can be represented by a vector AR(1) process with mean zero:

$$(5) \qquad\qquad\qquad x_t = \phi_x x_{t-1} + \sigma_x e_t,$$

where e_t is a sequence of iid normal shocks with variance Ω.

The yield to maturity on an n-period zero-coupon bond can be represented as a linear function of the factors:

$$(6) \qquad i_t^{(n)} = a_n + b_n^\top x_t.$$

Under our model, interest rates of all maturities are thus deterministic functions of the same small number of factors. If the coefficients a_n and b_n were unrestricted in the estimation, the resulting model would generally imply that there are opportunities for riskless arbitrage in the bond market.

Since riskless arbitrage opportunities would be quickly eliminated by bond market participants, it makes sense to rule them out from the beginning. We thus restrict the coefficients (a_n, b_n), which also makes the model more parsimonious. In particular, we assume that there exist a 4×1 vector μ_Q and a 4×4 matrix ϕ_Q such that

$$
\begin{aligned}
(7) \qquad a_{n+1} &= \frac{n}{n+1} a_n - \frac{1}{n+1} B_n^\top \mu_Q - \frac{1}{2} B_n^\top \sigma_x \Omega \sigma_x^\top B_n + \frac{1}{n+1} a_1, \\
b_{n+1}^\top &= \frac{n}{n+1} b_n^\top \phi_Q + \frac{1}{n+1} b_1^\top.
\end{aligned}
$$

Standard arguments (explained in detail in the appendix) imply that these restrictions rule out riskless arbitrage.

In addition to interest rates, we also want to describe the dynamics of macro variables such as inflation and consumption growth. We collect these variables in a vector h_t and assume that they are linear functions of the factor plus noise:

$$(8) \qquad h_t = \mu_h + \phi_h x_{t-1} + \sigma_h e_t.$$

The complete model of actual yields and macro variables model is thus summarized by (5)-(8). It is parametrized by the factor dynamics parameters (ϕ_x, σ_x), the yield coefficient parameters $(a_1, b_1, \mu_Q, \phi_Q)$, the macro variable parameters $(\mu_h, \phi_h, \sigma_h)$ and the variance of the shocks Ω.

In this subsection, the statistical forecasts $\widehat{E}_t$ of interest rates and macro variables are computed directly from the estimated state space system. The k-period ahead forecasts on the macro variables

and the n-period interest rate at date t are given by

$$\widehat{E}_t h_{t+k} = \mu_h + \phi_h \left(\phi_x\right)^{k-1} x_t,$$

(9)
$$\widehat{E}_t i_{t+k}^{(n)} = a_n + b_n^\top \left(\phi_x\right)^k x_t.$$

We assume that survey forecasts follow the same functional form as the statistical forecasts (9). We also assume that they are correct *on average* and that they respect the deterministic arbitrage-free relationship (6) between factors and interest rates. However, survey forecasts may differ from statistical forecasts in the sensitivity to the different factors. Our model of the k-period ahead survey forecasts on the macro variables and the n-period interest rate at date t is thus

$$E_t h_{t+k} = \mu_h + \phi_h^* \left(\phi_x^*\right)^{k-1} x_t,$$

$$E_t i_{t+k}^{(n)} = a_n + b_n^\top \left(\phi_x^*\right)^k x_t.$$

The survey forecasts are therefore parametrized by matrices ϕ_h^* and ϕ_x^*.

C Estimation

The estimation proceeds in three steps. We first estimate the joint distribution of the factors (equations (5)), the macro variables (equation (8)) and two interest rates. In a second step, we use data on many other interest rates to estimate their coefficients in (6). Finally, we use data on survey forecasts for many forecast horizons and maturities to estimate the parameters in (9).

Data

The data consists of quarterly observations over the sample 1952:2–2007:3. The data on zero-coupon interest rates and survey forecasts are the same as in section A. Moreover, we measure per capita consumption growth Δc_t and its corresponding inflation series π_t from data on nondurables and services in the NIPA tables. (To construct appropriate quantity and price indices, we use the corresponding lines in NIPA tables 2.3.3, 2.3.4 and 2.3.5.) The population series is also from NIPA (line 38 from table 2.1). We also use measures of subjective inflation expectations from the Survey

of Professional Forecasters. This survey is conducted at a quarterly frequency during the years 1968:4-2007:3.

Step 1: Factor dynamics and macro variables

We consider the joint dynamics of inflation, consumption growth, and two interest rates, the short (q-quarter) rate and the spread between the 5-year and 1-quarter interest rates. From equations (5)-(8), these variables can be represented by a four variable state space system with four factors

$$
\begin{aligned}
(10) \qquad z_t &= \mu_z + \phi_z x_t + \sigma_z e_t, \\
x_t &= \phi_x x_{t-1} + \sigma_x e_t.
\end{aligned}
$$

The observation equation for the vector $z_t = (i_t^{(1)}, i_t^{(20)} - i_t^{(1)}, h_t)$ is

$$
\begin{pmatrix} i_t^{(1)} \\ i_t^{(20)} - i_t^{(1)} \\ h_t \end{pmatrix} = \begin{pmatrix} a_1 \\ a_{20} - a_1 \\ \mu_h \end{pmatrix} + \begin{pmatrix} b_1 \phi_x \\ (b_{20} - b_1) \phi_x \\ \phi_h \end{pmatrix} x_{t-1} + \begin{pmatrix} b_1 \sigma_x \\ (b_{20} - b_1) \sigma_x \\ \sigma_h \end{pmatrix} e_t.
$$

We impose three sets of restrictions on this system. First, we identify the first two factors with the demeaned short rate and the demeaned spread. Formally, we set b_1 equal to the first unit vector, and we let $b_{20}^\top = (1, 1, 0, 0)$. As a result, the first two observation equations become simply copies of the first two state equations. Second, we set the first two rows of σ_x equal to $\begin{pmatrix} I_{(2 \times 2)} & 0 \end{pmatrix}$. Third, we identify the third and fourth factor with expected inflation and expected consumption growth. Formally, we set $\phi_h = \begin{pmatrix} 0_{(2x2)} & I_2 \end{pmatrix}$ and $\sigma_h = \begin{pmatrix} 0_{(2 \times 2)} & I_2 \end{pmatrix}$. In contrast to the first two factors, the third and fourth factor are latent. Even after the restrictions, the system is more flexible than a first order VAR in the four observables. This is because the presence of the latent factors allows for MA(1) style dynamics in the macro variables (which are important for capturing the dynamics of inflation.)

We estimate the unconditional means $\mu_z = (a_1, a_{20} - a_1, \mu_h)$ from the sample means of the short rate, the spread, inflation and consumption growth. In the table, we report these means in percent,

so that $a_1 \times 100 = 1.28$ means a 5.12% average short rate in annualized terms. The remaining 34 parameters $(\sigma_z, \phi_x, \sigma_x)$ are estimated by maximum likelihood. The parameter estimates also deliver a sequence of estimates $\hat{x}_t$ for the realizations of the state variables, starting from $x_0 = 0$. Of course, the first two components of $\hat{x}_t$ are equal to the demeaned short rate and spread. However, the third and fourth factors are latent, and thus their respective entries of $\hat{x}_t$ represent their conditional expectations given the data.

<div align="center">

TABLE 2: ESTIMATED DYNAMICS

PANEL A: MAXIMUM LIKELIHOOD OF STATE SPACE SYSTEM

</div>

	$\mu_z \times 100$	chol(Ω)			
$i_t^{(1)}$	1.28	0.237	0	0	0
	–	(0.012)	–	–	–
$i_t^{(20)} - i_t^{(1)}$	0.25	−0.110	0.118	0	0
	–	(0.007)	(0.015)	–	–
π_t	0.86	0.030	0.0211	0.253	0
	–	(0.002)	(0.003)	(0.022)	–
Δc_t	0.51	−0.047	0.022	−0.083	0.411
	–	(0.006)	(0.002)	(0.004)	(0.022)

	ϕ_x				σ_x			
$i_t^{(1)}$	0.894	0.035	−0.016	0.31	1	0	0	0
	(0.029)	(0.002)	(0.001)	(0.004)				
$i_t^{(20)} - i_t$	0.038	0.827	−0.069	−0.213	0	1	0	0
	(0.004)	(0.049)	(0.006)	(0.040)				
π_t	−0.016	−0.079	0.980	0.118	0.079	−0.286	−0.137	0.252
	(0.019)	(0.008)	(0.036)	(0.008)	(0.004)	(0.015)	(0.012)	(0.021)
Δc_t	0.023	0.083	−0.068	0.584	0.532	0.101	0.047	0.283
	(0.028)	(0.026)	(0.009)	(0.031)	(0.014)	(0.019)	(0.004)	(0.012)

Note: This table contains the maximum likelihood estimates for the state space system

$$z_t = \mu_z + \eta_z x_t + e_t$$
$$x_t = \phi_x x_{t-1} + \sigma_x e_t$$

where e_t is an iid normal shock with variance Ω. The observables are $z_t = (i_t^{(1)}, i_t^{(20)} - i_t^{(1)}, \pi_t, \Delta c_t)$, and the state vector x_t consists of $i_t^{(1)}, i_t^{(20)} - i_t^{(1)}$, expected inflation, and expected consumption growth, starting from $x_0 = 0$. Standard errors are in brackets. The sample is quarterly, 1952:2-2007:3.

Step 2: Yield coefficients

The second step is to estimate the interest rate parameters μ_Q and ϕ_Q . Here we take as given the estimate of the mean short rate a_1 and the factors $\hat{x}_t$ derived in step 1. For every pair (μ_Q, ϕ_Q), we can form samples of predicted zero coupon yields

$$(11) \qquad\qquad i_t^{(n)} = a_n + b_n^\top \hat{x}_t,$$

where the coefficients (a_n, b_n) are computed from (μ_Q, ϕ_Q) and a_1 via the recursion (7). We estimate μ_Q and ϕ_Q by minimizing, for a set of maturities, the sum of squared fitting errors, that is, differences between actual yields and predicted yields computed from (11). We use yields of maturities 1, 2, 5, 10, 15, 20 and 30 years. We also impose the constraint that the spread between the 5-year and 1-quarter rate that serves as a factor be matched exactly.

TABLE 2: ESTIMATED DYNAMICS

PANEL B: YIELD COEFFICIENTS ON FACTORS

	l_0 from $\mu_Q = -\sigma_x \Omega l_0$	l_1 from $\phi_Q = \phi_x - \sigma_x \Omega l_1$			
$i_t^{(1)}$	−33.84	−56.23	9.23	−188.41	81.22
	(1.73)	(7.13)	(0.72)	(3.11)	(1.26)
$i_t^{(20)} - i_t^{(1)}$	20.44	−94.11	−71.78	−73.24	295.41
	(0.73)	(3.29)	(11.52)	(9.37)	(37.72)
π_t	48.96	65.72	2.46	−59.67	−23.48
	(17.77)	(1.17)	(0.23)	(8.44)	(5.78)
Δc_t	146.99	115.03	−34.10	−143.15	−121.83
	(3.10)	(5.12)	(2.22)	(35.45)	(4.56)

NOTE: We estimate μ_Q and ϕ_Q by minimizing the squared fitting errors of the model. We parametrize them as follows: $\mu_Q = -\sigma_x \Omega l_0$ and $\phi_Q = \phi_x - \sigma_x \Omega l_1$. Standard errors are computed by GMM, taking into account the two-step nature of the estimation.

For the estimation, we parametrize (μ_Q, ϕ_Q) as follows: $\mu_Q = -\sigma_x \Omega l_0$ and $\phi_Q = \phi_x - \sigma_x \Omega l_1$ and estimate the 4×1 vector l_0 and the 4×4 matrix l_1. The advantage of this parametrization is that (μ_Q, ϕ_Q) is equal to (μ_x, ϕ_x) if $l_0 = 0$ and $l_1 = 0$. Panel B of Table 2 reports the estimates of l_0 and l_1. The standard errors (in brackets) take into account the two-step nature of the estimation.

This is done by GMM; we stack the moment conditions from the first step MLE (which are the scores of the likelihood function) and the moment conditions from the second step NLS estimation (which are the first-order conditions for the minimization.) More details are in the Appendix.

Step 3: Subjective state space system

The third step is to estimate the parameters ϕ_x^* and ϕ_h^* that determine how survey forecasts depend on the factors. Here we take as given the factor estimates $\hat{x}_t$ and the means of the macro variables $\hat{\mu}_h$ from step 1 as well as the interest rate coefficient estimates $\left(\hat{a}_n, \hat{b}_n\right)$ from step 2. For every (ϕ_x^*, ϕ_h^*), we can form samples of predicted survey forecasts of macro variables and interest rates

$$
\begin{aligned}
E_t h_{t+h} &= \widehat{\mu}_h + \phi_h^* \left(\phi_x^*\right)^{k-1} \hat{x}_t, \\
E_t i_{t+h}^{(n)} &= \hat{a}_n + \hat{b}_n^\top \left(\phi_x^*\right)^k \hat{x}_t.
\end{aligned}
$$

(12)

We estimate ϕ_x^* and ϕ_h^* by minimizing a sum of squared fitting errors, that is, differences between median survey forecasts and predicted forecasts computed as in (12). The maturities and horizons for the interest rate forecasts differ by sample period. For 1970:1-1982:2, we use Goldsmith-Nagan data and consider a horizon of 2 quarters and maturities 1 year and 20 years. For 1982:3-2007:3, we Bluechip data and consider horizons of 2 and 4 quarters, and maturities of 1 quarter, as well as 1, 2, 3, 5, 7, 10 and 30 years. For inflation forecasts, we use a horizon of 4 quarters over the sample 1968:2-2007:3. The only variable for which we do not have survey data is per capita consumption of nondurables and services. Here, we impose that the forecasts from the subjective state space system are equal to the forecasts from the state-space system.

For the estimation, we parametrize the coefficients using a 4×4 matrix k which satisfies $\phi_x^* = \phi_x - \sigma_x \Omega k$ and $\phi_h^* = \phi_h - \sigma_h \Omega k$. Panel C in Table 2 reports the estimated k. Again, this parametrization highlights that $\phi_x^* = \phi_x$ and $\phi_h^* = \phi_h$ if $k = 0$. Panel C also reports GMM standard errors that take into account the three-step nature of this estimation. (We plan on investigating the small sample properties of these standard errors, since they seem low.) The first two steps are the MLE of the state space and the NLS for the yield coefficients. The last, third, step is also a NLS estimation, and we can stack its first-order conditions together with the others. Details

are in the Appendix.

TABLE 2: ESTIMATED DYNAMICS

PANEL C: SUBJECTIVE STATE SPACE SYSTEM

$$k \text{ from } \phi_x^* = \phi_x - \sigma_x \Omega k, \ \phi_h^* = \phi_h - \sigma_h \Omega k$$

$i_t^{(1)}$	-46.93	24.99	-65.31	-19.51
	(0.01)	(0.01)	(0.02)	(0.01)
$i_t^{(20)} - i_t^{(1)}$	-42.93	-86.90	-128.37	-4.32
	(0.73)	(0.01)	(0.02)	(0.001)
π_t	66.80	14.31	66.80	14.61
	(0.02)	(0.005)	(0.02)	(0.003)
Δc_t	152.24	-74.11	65.01	21.69
	(0.04)	(0.02)	(0.02)	(0.001)

NOTE: We estimate ϕ_x^* and ϕ_h^* by minimizing the squared difference between the subjective forecasts and survey forecasts. The table reports the estimated k in $\phi_x^* = \phi_x - \sigma_x \Omega k$, and $\phi_h^* = \phi_h - \sigma_h \Omega k$. Standard errors are computed by GMM, taking into account the three-step nature of the estimation.

D Results

The appendix reports our estimates for subjective beliefs and beliefs derived from the statistical model. To understand the estimated statistical dynamics, we report covariance functions which completely characterize the Gaussian state space system. Figure 2 plots covariance functions computed from the state space system and the raw data. At 0 quarters, these represent variances and contemporaneous covariances. The black lines from the system match the gray lines in the data quite well. To interpret the units, consider the upper left panel. The quarterly variance of the short rate is 0.51 in the data which amounts to $\sqrt{0.51 \times 4} = 1.43$ percent annualized volatility. Figure 2 shows that all four variables are positively autocorrelated. For example, the covariance $\text{cov}\left(i_t^{(1)}, i_{t-1}^{(1)}\right) = \rho \, \text{var}\left(i_t^{(1)}\right) = \rho \times 0.51 = 0.48$ which implies that the first-order autocorrelation is 0.94.

The statistical dynamics of the state variables are persistent. The largest eigenvalues of the matrix ϕ_x are complex with a modulus of 0.95, while the third eigenvalue is 0.70. In Figure 2, the
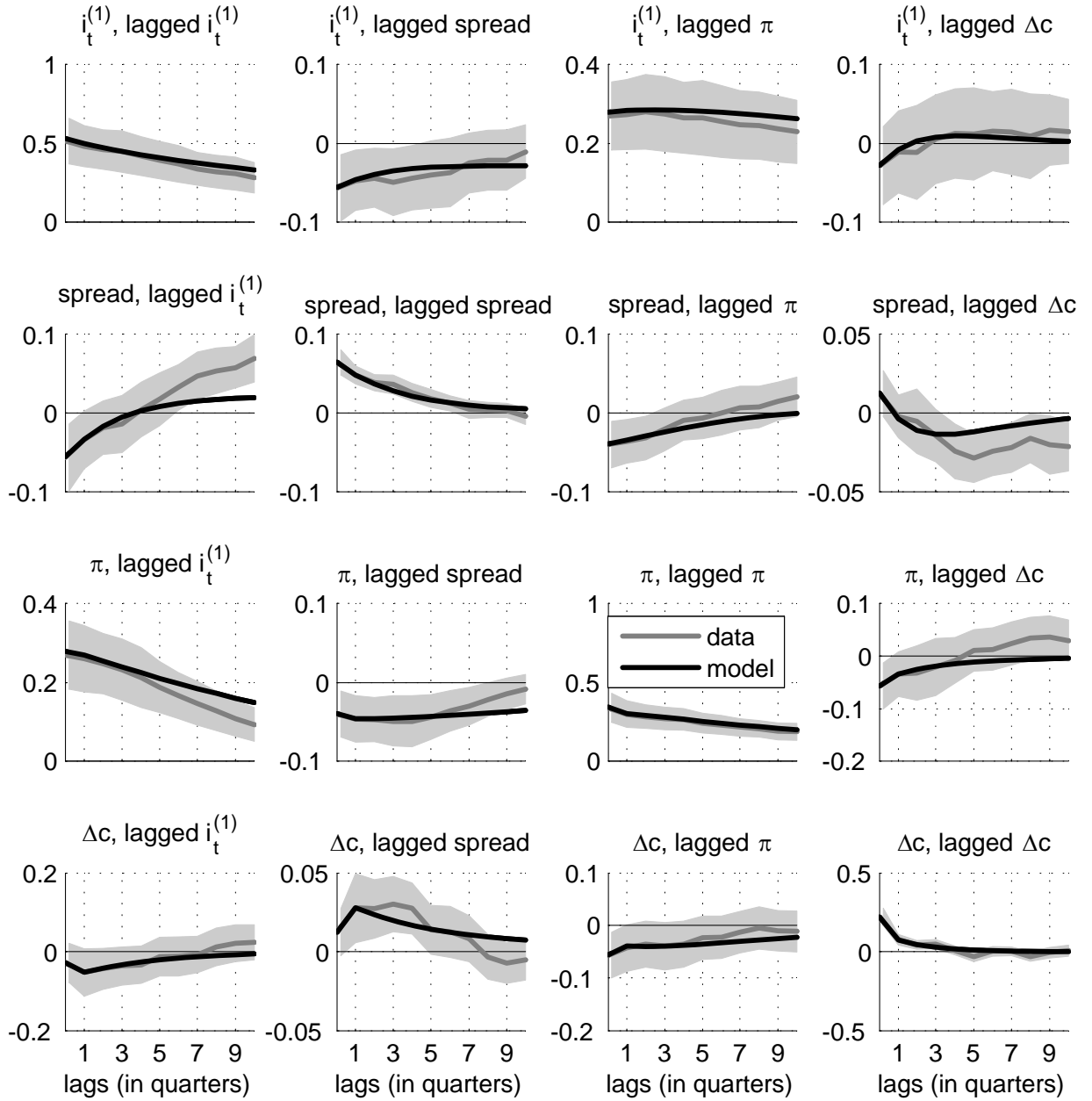
Figure 2: Covariance functions for the observables computed from the estimated state space system and from the raw data. Shaded areas indicate 2 × standard errors bounds around the covariance function from the data computed with GMM. For example, the graph titled "$i_t^{(1)}$, lagged $i_t^{(1)}$" shows the covariance of the current short rate with the short rate lagged τ quarters, where τ is measured on the horizontal axis.

autocovariance functions of the short rate and inflation are flatter than that of the spread, which indicates that they are more persistent. The short rate and the spread are contemporaneously

negatively correlated and the spread is negatively correlated with the short rate lagged less than year, and positively correlated with longer lags of the short rate. The short rate is negatively correlated with the lagged spread, even at long lags.

<div align="center">TABLE 3: ESTIMATION OF STATISTICAL MODEL</div>

<div align="center">Panel A: Loadings of expected excess returns on state variables</div>

		horizon $h = 1$ year			
	n	short rate	spread	exp π	exp Δc
maturity of the bond	2 year	1.31	2.46	-1.51	-1.73
	10 year	7.55	24.5	-9.33	-10.2

<div align="center">Panel B: Fitting errors for bond yields (annualized)</div>

	maturity						
	1 qrt	1 year	5 year	10 year	15 year	20 year	30 year
mean absolute errors (in %)	0	0.30	0	0.24	0.36	0.42	0.45

NOTE: Panel A reports the model-implied loadings of the function $\hat{E}_t\left[rx_{t+h}^{(n,h)}\right] + \frac{1}{2}\widehat{Var}_t\left[rx_{t+h}^{(n,h)}\right]$. Since the conditional variance of excess returns is constant, we can derive these loadings from

$$
\begin{aligned}
\hat{E}_t\left[rx_{t+h}^{(n,h)}\right] &= \hat{E}_t\left[p_{t+h}^{(n-h)}\right] - p_t^{(n)} - hi_t^{(h)} \\
&= -(n-h)\left(a_{n-h} + b_{n-h}^\top \hat{E}_t\left[x_{t+h}\right]\right) + n\left(a_n + b_n^\top x_t\right) - h\left(a_h + b_h^\top x_t\right)
\end{aligned}
$$

on the current factors x_t for a holding period of $h = 1$ year and bond maturities of $n = 2$ years, 10 years. Panel B reports mean absolute model fitting errors for yields.

To understand the implications of the estimated parameters l_0 and l_1, we investigate how expected excess returns depend on the state variables. This dependence can be derived from the state space system together with the yield coefficients. Table 2 reports the loadings of these conditional expected values on the state variables. For a 1-quarter holding period, these loadings are $-nb_{n-1}^\top \sigma_f \Omega l_1$. For longer holding periods, we can compute $\hat{E}_t\left[rx_{t+h}^{(n,h)}\right] + \frac{1}{2}\widehat{Var}_t\left[rx_{t+h}^{(n,h)}\right]$ using the recursions for the coefficients a_n and b_n. The coefficients in Table 3 indicate that the expected excess return on a 2-year bond is high in periods with high spreads. For example, a 1-percent increase in the spread (and everything else constant) leads to a 2.46 percent increase in the statistical premium.

This dependence on the spread captures is most important driving force of statistical premia and captures their countercyclical nature. The premium on the 10-year bond has larger loadings on all state variables. Roughly speaking, the coefficient for the 10 year bond are roughly 10 times higher than for the 2-year bond.Panel B of Table 3 reports by how much the model-implied yields differ from observed yields on average. By construction, the model hits the 1-quarter and 5-year interest rates exactly, because these rates are included as factors. For intermediate maturities, the error lies within the $24 - 45$ basis points range. We will see below that these errors are sufficiently small for our purposes.

Subjective vs. statistical dynamics

By construction, subjective forecasts are on average equal to the forecasts from our statistical model. At the same time, the factors are more persistent under the subjective system. The largest eigenvalues of the two matrices

$$
\widehat{\phi} = \begin{pmatrix} 0.89 & 0.04 & 0.16 & 0.31 \\ 0.04 & 0.83 & -0.07 & -0.21 \\ -0.02 & -0.08 & 0.98 & 0.11 \\ 0.02 & 0.08 & -0.07 & 0.58 \end{pmatrix}, \quad \phi^* = \begin{pmatrix} 1.00 & 0.14 & 0.001 & -0.06 \\ -0.04 & 0.88 & -0.01 & 0.05 \\ 0.10 & 0.11 & 0.86 & -0.06 \\ 0.01 & 0.09 & -0.04 & 0.71 \end{pmatrix}.
$$

are 0.948 and 0.954, respectively. The next highest eigenvalues are 0.70 for the statistical system and 0.86 for the subjective forecasts. Other things equal, a one-percent increase in the short rate (spread) increases the subjective forecast of the short rate (spread) next period by 1% (88%), as opposed to 89% (83%) under the statistical model.

The estimated subjective dynamics imply that subjective risk premia are less cyclical than statistical premia. The subjective loadings on the spread in Table 4 are smaller than those in Table 3. For long bonds, the loading on the short rate increases under subjective beliefs, while the loading on expected inflation is now also positive. This indicates that subjective premia on long bonds will reflect some of the low-frequency movements in expected inflation and nominal interest rates.

Panel B of Table 4 reports mean absolute distances between the survey forecasts and model-implied forecasts, for both the subjective belief and the statistical model. Comparison of these

errors provides a measure of how well the change of measure works to capture the deviation of survey forecasts from statistical forecasts.

Panel A: Loadings of expected excess returns on state variables

	n	short rate	spread	exp π	exp Δc
		horizon $h = 1$ year			
maturity	2 year	0.20	0.78	0.05	0.68
	10 year	6.38	7.86	1.53	-3.82

Panel B: Mean absolute fitting errors for yield forecasts (% p.a.)

	maturity									
	subjective model					statistical model				
	Bluechip sample, maturity of forecasted yield									
horizon	1 qrt	1 year	3 year	5 year	10 year	1 qrt	1 year	3 year	5 year	10 year
2 quarter	0.28	0.39	0.26	0.25	0.35	0.42	0.60	0.39	0.28	0.34
1 year	0.33	0.36	0.25	0.24	0.33	0.62	0.75	0.60	0.51	0.54
	Combined sample, maturity of forecasted yield									
	1 year			20 year		1 year				20 year
2 quarter	1.81			0.46		1.94				0.64

The results show that the improvement is small for short-horizon forecasts of short yields. However, there is a marked reduction of errors for 1-year forecasts, especially for the 10-year bond. Figure 3 shows where the improvements in matching the long-bond forecasts come from. The top panel shows one-year ahead forecasts of the 10-year zero coupon rate constructed from survey data in Section II.A, together with the corresponding forecasts from our subjective and statistical models, for the sample 1982:4-2007:3. All forecasts track the actual 10-year rate over this period, which is natural given the persistence of interest rates. The largest discrepancies between the survey forecasts and the subjective model on the one hand, and the statistical model on the other hand, occur during and after the recessions of 1990 and 2001. In both periods, the statistical model quickly forecasts a drop in the interest rate, whereas investors did not actually expect such a drop. The subjective model captures this property.

For our asset pricing application, we are particularly interested in how well the subjective model captures deviations of survey forecasts of long interest rates from their statistical forecasts over the

Survey forecasts and model implied forecasts; 10 year bond

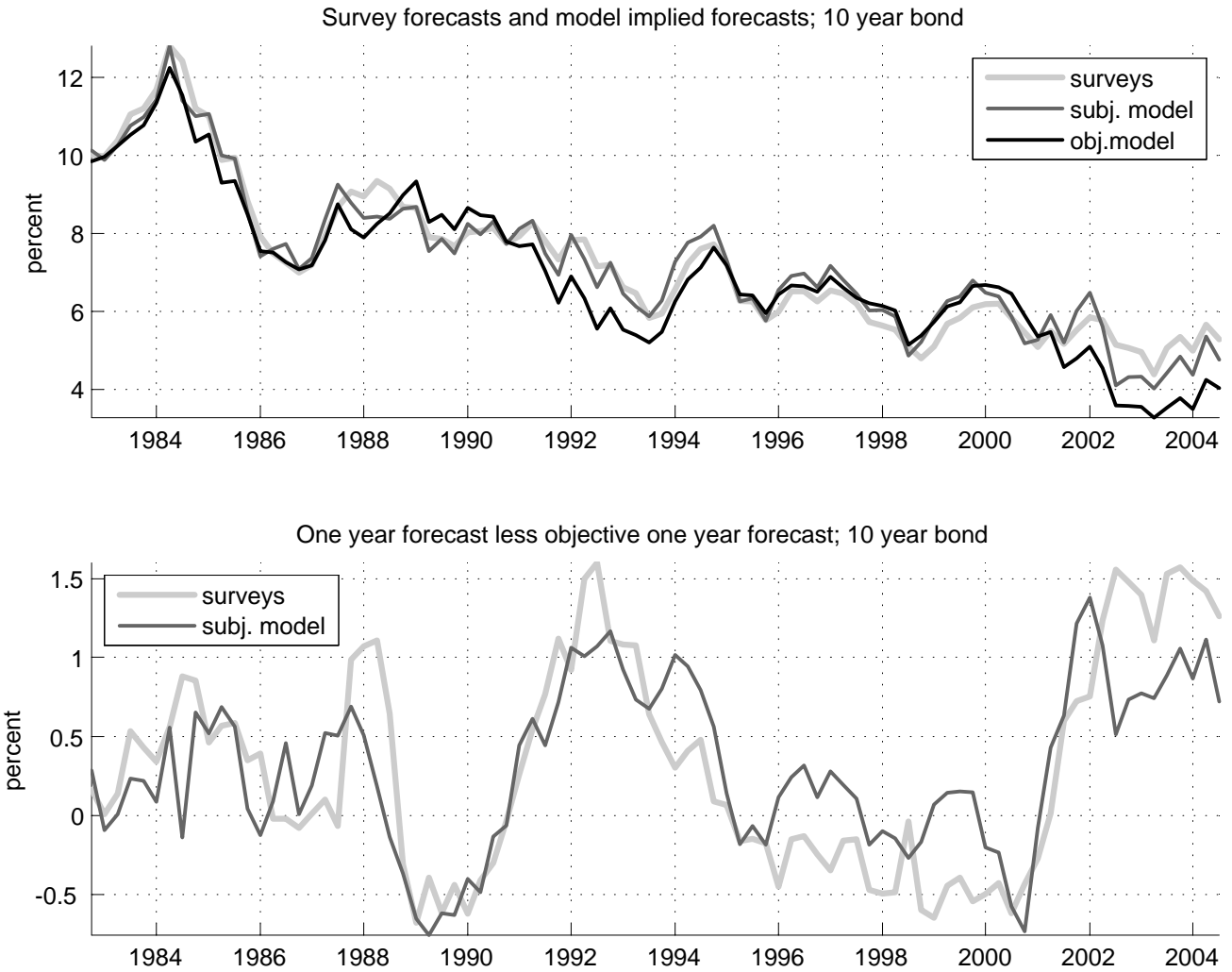One year forecast less objective one year forecast; 10 year bond

Figure 3: The top panel shows one-year ahead forecasts of the 10-year zero coupon rate constructed from survey data in Section II.A, together with the corresponding forecasts from our objective and subjective models. The bottom panel shows the difference between the survey forecast and the objective model forecast, as well as the difference between the subjective and objective model forecasts.

business cycle. As discussed in Section II.A, this forecast difference is closely related to measured expected excess returns. The bottom panel of the figure focuses again on forecasting a 10-year rate over one year, and plots the difference between the survey forecast and the statistical model forecast, as well as the difference between the subjective and statistical model forecasts. It is apparent that both forecast differences move closely together at business cycle frequencies, increasing during and after recessions. We thus conclude that the subjective model is useful to capture this key fact about subjective forecasts that matters for asset pricing.

E   Subjective risk premia

We now compare our estimated subjective risk premia to common statistical measures of risk premia. The motivation is that statistical measures of premia provide stylized facts that rational expectations asset pricing models try to match. In particular, empirical evidence of predictability of excess returns from standard predictability regressions has led to a search for sources of time varying risk or risk aversion. The preliminary results of section II.A suggest that less time variation in risk premia is required once investors' forecast errors are taken into account. Here we quantify how much time variation in expected excess returns is left once we move to subjective beliefs.

We focus on 1-year holding period returns on bonds with 2 and 10 years maturity. We compare our subjective premia to three statistical measures of statistical premia. The first is the fitted value of a regression of excess returns on a single yield spread, the 5-year-1-quarter spread, denoted the YS measure. In other words, we regress the excess return $rx_{t+4}^{(n,4)}$ on $i_t^{(20)} - i_t^{(1)}$ and a constant. This regression is closely related to that in the classic Fama-Bliss study of bond return predictability, which uses the forward-spot spread. The second measure (labelled CP measure) is the fitted value from a regression on five yields with maturities 1,2,3, 4 and 5 years. This follows Cochrane and Piazzesi (2005) who showed that this approach leads to higher R^2s. The third measure is, for each subjective model specification, the forecast from the corresponding statistical model which provides the conditional expectation of the excess return.

Table 4 summarizes the properties of the regression based measures of statistical premia. According to these measures, the volatility of the predictable part of 1-year holding period returns is below 1% per year for the 2-year bond, and around 5.3% for the 10-year bond. The regression based on five yields naturally delivers a higher R^2, 26% on both bonds. We are also interested in the frequency properties of premia. We use a band pass filter to decompose premia into three orthogonal components, a low frequency "trend" component (period > 8 years), a "cycle" component (period between 1.5 and 8 years), as well as high frequency noise. The columns labelled "trend" and "cycle" show the percentage of variance contributed by the respective components. Since the yield spread is a key business cycle indicator, the YS measure is particularly cyclical. The CP measure improves in part by including a larger trend component.

maturity 2 years				maturity 10 years			
Regression on yield spread (YS measure)							
volatility	% trend	% cycle	R^2	volatility	% trend	% cycle	R^2
0.61	14	56	0.12	5.27	14	56	0.25
Regression on five yields (CP measure)							
volatility	% trend	% cycle	R^2	volatility	% trend	% cycle	R^2
0.91	31	38	0.26	5.34	28	48	0.26

Table 5 presents a set of comparison statistics. The first line shows the properties of the subjective premium itself. For the baseline model, the standard deviation of the one-year premium on a 2-year bond is 29 basis points; it is 3% on the 10-year bond. These volatilities are substantially smaller than those of regression measures of premia. The frequency properties are also different: subjective premia tend to have larger trend components and smaller cyclical components than regression based premia. This is particularly pronounced for the longer 10 year bond: in regression models at least one half of the time variation in premia is cyclical, whereas for subjective premia the share of cyclical variation is only 15%.

Table 5: Subjective Risk Premia

	maturity 2 years			maturity 10 years		
	Baseline Model					
	volatility	% trend	% cycle	volatility	% trend	% cycle
	0.29	45	28	2.96	81	15
	volatilities relative to measures of statistical premia					
	total	trend	cycle	total	trend	cycle
YS measure	0.48	0.87	0.34	0.56	1.39	0.49
CP measure	0.33	0.39	0.28	0.55	1.64	0.53
state space system	0.35	0.27	0.38	0.49	0.86	0.48

The other rows in the table consider directly the change in volatility as one moves from statistical to subjective premia. All numbers are ratios of standard deviations, subjective divided by statistical. The columns labelled "total" report the volatility of a subjective premium as a fraction of the volatility of the statistical premium for the different measures. They range between 30 and 60 percent. Since the numerator is always the same, the volatility ratios are lower the better the comparison statistical model predicts excess returns. The largest ratios arise for the YS measure

which generates the least time variation in statistical premia. The columns labelled "trend" and "cycle" report volatility ratios for the trend and cycle components across models. For the regression based measures, the reduction in volatility is driven primarily by a reduction in the volatility of the cyclical component. In fact, for the YS measure, the trend component is less volatile than that of the subjective premium.

The row labelled "state space system" compares the subjective premium to the premium from the estimated state space system. For the 10-year bond, this system is a better predictor of expected excess returns than even the CP regression based measure. The reason is that the expected inflation state variable helps forecast returns. For the system, the move to subjective premia implies a substantial reduction in the volatility of the trend component. The system differs from the regression based measures in that it generates premia with somewhat larger trend components; the shares of the trend components is 72% for the 2 year bond and 59% for the 10 year bond. These components are due to the presence of expected inflation in the system.

Figure 4 plots subjective premia on the two bonds together with the respective CP measures as well as the statistical premia from the state space system. The properties from the table are also visible to the naked eye. Consider first the long, (10 year) bond in the bottom panel. Both measures of statistical premia show substantial cyclical movements: most recessions during the sample period can be identified as upward spikes in statistical premia, for example in 1970, 1991 and 2001. The CP measure also spikes in 1974 and 1979. Here the state-space system measure responds less to the business cycle; this is because it is driven more by expected inflation which lowers premia during this period. In contrast to both measures of statistical premia, the subjective premium on the long bond responds only weakly to recessions. The bulk of the movement in the subjective premium is at low frequencies: it was high in the late 1970s and early 1980s when the level of the yield curve was high, and low towards the beginning and end of our sample.

Consider next the medium (2 year) bond. It is clear again that the subjective premium is less volatile than the measures of statistical premia. At the same time, recession periods now register as upward spikes in all of the displayed lines. The main difference between statistical and subjective premia for the medium bond is in the trend: the subjective premium is much smaller in the late 1970s and early 1980s then the statistical premia. Comparing the statistical premia across
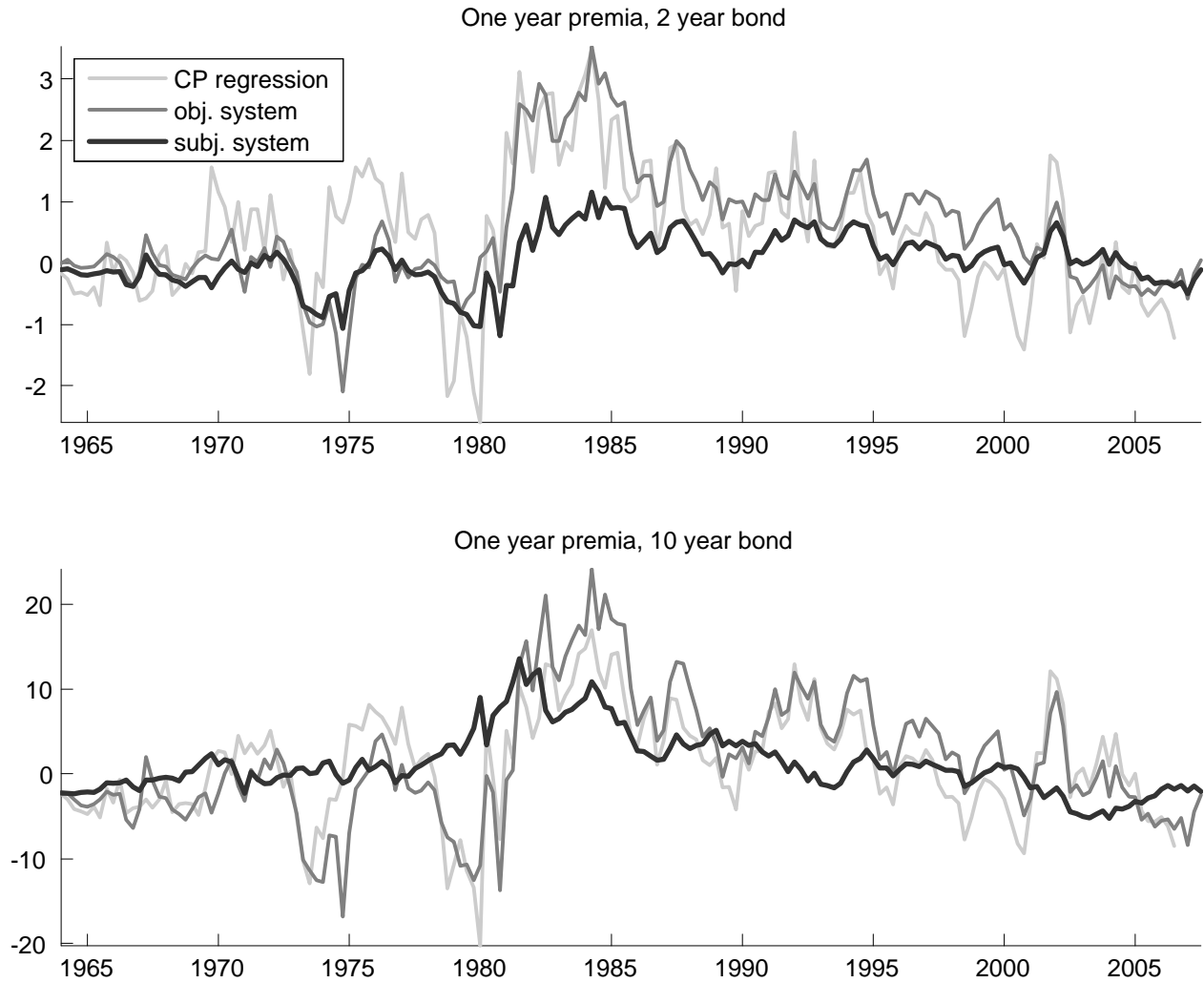
Figure 4: Subjective premia compared with measures of objective premia. Premia are expected excess holding period returns over one year, for a 2 year bond (top panel) and a 10 year bond (bottom panel). In both panels, black lines are estimated subjective premia, light gray lines are fitted values from CP regression, dark gray lines are objective premia from state space system.

maturities (across panels), it is also apparent that statistical premia on the long bond are more cyclical and exhibit less trend than statistical premia on the medium bond. For subjective premia, the situation is the reverse.

The intuition for these results comes from the differences between survey forecasts and forecasts derived from a statistical model. Under a statistical model, both the slope and the level are indicators of high expected excess returns. For example, Figure 4 shows that measures of statistical premia comove positively with both slope and level. We have seen in the previous section that survey

forecasters treat both the level and the slope of the yield curve as more persistent than what they are under a statistical model. This difference between survey forecasts and forecasts derived from a statistical model of future bond prices then weakens the effect of both indicators.

If the level of the yield curve is high, survey forecasters, who view the level as more persistent than does an statistical model, expect higher interest rates, and hence lower prices, than the statistical model. Lower expected prices means lower expected excess returns. Similarly, if the slope is high, survey forecasters, who view the slope as more persistent, expect higher spreads, and hence higher long interest rates, and lower long bond prices, than the statistical model. Again, lower expected prices means lower expected excess returns. Both situations (high level, high spread) which lead statistical models to indicate high expected excess returns thus induce survey forecasters to predict lower prices and returns than the statistical models. This generates the overall reduction in volatility.

The higher persistence of level and slope perceived by survey forecasters also helps explain the different frequency properties of premia on medium (for example 2 year) and long (for example 10 year) bonds. The level of the yield curve is always relatively more important for short bonds rather than for long bonds. This is true not only for yields themselves, but also for measures of statistical premia: those measures are driven relatively more by the slope for long bonds and relatively more by the level for medium bonds. A move to subjective premia weakens the effect of both indicators of high premia, the slope and the level. For a given maturity, it tends to weaken more the effect of the indicator that is more important. The move thus makes the premium on long bonds less responsive to the slope and it makes the premium on medium bonds less responsive to the level. This is the effect displayed in the figure.

III Structural Model

In this section, we consider a representative agent asset pricing model with an investor who learns adaptively about consumption, inflation and bond prices. The learning algorithm, described in detail below, gives rise to a sequence of probabilities P_t that describe investors' subjective beliefs about future consumption, inflation and bond prices conditional on the information at date t. At

every date t, agents plan ahead using the probability P_t for that date, and their Euler equations hold under that belief.

A   Preferences & Euler equations

Preferences over consumption streams $C = \{C_t\}_{t=1}^{\infty}$ are represented by a version of the recursive utility model (Epstein and Zin, 1989). The utility process V_t associated with the consumption stream C solves

$$(13) \qquad \log V_t = (1 - \beta) \log C_t + \beta \log E_t \left(V_{t+1}^{1-\gamma} \right)^{1/(1-\gamma)},$$

where $\beta \in (0, 1)$ and $\gamma > 0$. Riskless consumption streams are ranked according to a separable log utility criterion with discount factor β. For an uncertain stream, current utility depends on the certainty equivalent of future continuation utility. The certainty equivalent function takes a power form with coefficient of relative risk aversion γ. A separable (log) expected utility criterion arises as the special case $\gamma = 1$. More generally, Epstein-Zin preferences differ from separable expected utility because they do not impose indifference to the temporal distribution of risk.[4]

We assume that all shocks are homoskedastic and lognormal. This allows us to derive a linear recursion for utility as in Hansen, Heaton, and Li (2005). Using lognormality to evaluate the conditional expectation in (13), and writing lower case letters for logarithms, we have

$$(14) \qquad v_t - c_t = \beta \, E_t \left(v_{t+1} - c_{t+1} + \Delta c_{t+1} \right) + \beta \, \frac{1}{2} \left(1 - \gamma \right) \mathrm{var}_t \left(v_{t+1} \right).$$

With homoskedastic shocks, this recursion is a linear difference equation in $v_t - c_t$ that can be solved forward. We can then express the log ratio of continuation utility to consumption as an

---

[4]Consider two random consumption streams $c = \{c_t\}_{t=1}$ and $\tilde{c} = \{\tilde{c}_t\}_{t=1}$ such that the distribution of the date t flows c_t and $\tilde{c}_t$ is the same for every t (given information at date 0). Separable expected utility implies indifference between c and $\tilde{c}$, regardless of the correlation of consumption flows across time periods. For example, suppose the flows c_t are perfectly correlated, whereas the $\tilde{c}_t$ are serially independent. Epstein-Zin utility allows a strict preference for $\tilde{c}$, but time separable expected utility does not.

infinite sum of expected discounted future consumption growth,

$$(15) \qquad v_t - c_t = \sum_{\tau=1}^{\infty} \beta^{\tau} E_t \left(\Delta c_{t+\tau} \right) + \text{constant}.$$

Let R_{t+1} denote the nominal return on some asset between dates t and $t+1$, and let P_t denote the nominal price level. The agent's Euler equation is

$$(16) \qquad E_t \left[\beta \frac{C_t}{C_{t+1}} \frac{V_{t+1}^{1-\gamma}}{E_t \left(V_{t+1}^{1-\gamma} \right)} \frac{P_t}{P_{t+1}} \quad R_{t+1} \right] = 1$$

If $\gamma = 1$, this is the standard Euler equation for a log utility investor. More generally, the marginal rate of substitution between wealth at t and $t+1$ depends on the surprise in continuation utility. The term multiplying the return under the expectation is the nominal pricing kernel.

To evaluate Euler equations for nominal bonds, it is convenient to write them in logs. We first define the log nominal pricing kernel

$$\begin{aligned} m_{t+1} \quad : \quad & = \log \beta - \Delta c_{t+1} - \pi_{t+1} \\ (17) \qquad & - (\gamma - 1) \left(v_{t+1} - E_t \left(v_{t+1} \right) \right) - \frac{1}{2} (1-\gamma)^2 \operatorname{var}_t \left(v_{t+1} \right). \end{aligned}$$

The nominal pricing kernel is high in "bad" states of the world, where the investor has a high need for nominal payoff. Under log utility ($\gamma = 1$), these are states where nominal consumption is low. If $\gamma > 1$, a state can also be bad just because the outlook on the future, captured by continuation utility, has become surprisingly low.

The Euler equation (16) holds in particular for a zero coupon bond of maturity 1 that has log return $i_t^{(1)}$. The 1-period interest rate thus satisfies

$$\begin{aligned} i_t^{(1)} \quad = \quad & -E_t m_{t+1} - \frac{1}{2} var_t m_{t+1} \\ = \quad & -\log \beta + E_t [\Delta c_{t+1} + \pi_{t+1}] - \frac{1}{2} \operatorname{var}_t \left(\Delta c_{t+1} + \pi_{t+1} \right) \\ (18) \qquad & - (\gamma - 1) \operatorname{cov}_t \left(\Delta c_{t+1} + \pi_{t+1}, \sum_{\tau=1}^{\infty} \beta^{\tau-1} E_{t+1} \left(\Delta c_{t+\tau} \right) \right), \end{aligned}$$

As in the standard log utility model, the short rate moves with expected nominal consumption growth. More generally, Epstein-Zin utility modifies the precautionary savings motive. With log utility, the nominal interest rate is smaller the larger is the volatility of nominal consumption growth. If $\gamma > 1$, investors particularly dislike persistent consumption risk. As a result, the nominal rate is lower the more nominal growth covaries with expected future growth

The Euler equation (16) also holds for a zero coupon bonds of maturity n, with log return $\log R_{t+1} = p_{t+1}^{(n-1)} - p_t^{(n)}$. Combining the Euler equations for n-period and 1-period bonds, we obtain

$$(19) \qquad E_t p_{t+1}^{(n-1)} - p_t^{(n)} - i_t^1 + \frac{1}{2} var_t p_{t+1}^{(n-1)} = -cov_t \left(m_{t+1}, p_{t+1}^{(n-1)} \right)$$

The left hand side is the (log) risk premium on an n-period bond held over one period. It consists of the log expected excess return – the log expected capital gain less the log one period rate– plus a Jensen's inequality term. Here we follow convention and include the Jensen's inequality term in the risk premium.

The right hand side of (19) is the covariance between the nominal pricing kernel and the uncertain payoff on the long (n period) bond. If the long bond pays off more in bad states – when m_{t+1} is high – then it provides insurance against bad states, and should earn a return that is lower than the short (one period) rate. In other words, it should earn a negative risk premium. In contrast, if the long bond pays more in good states, when m_{t+1} is low, then it pays off exactly at the wrong time, and investors must be compensated in order to hold it. This compensation comes is provided by a positive risk premium.

To see how different shocks affect the risk premium, we use (15) to write

$$m_{t+1} = K_t \ - \ \Delta c_{t+1} - \pi_{t+1} - (\gamma - 1)(E_{t+1} - E_t) \sum_{\tau=1}^{\infty} \beta^\tau \ \Delta c_{t+\tau}$$

where K_t is predetermined as of date t. Viewed from the perspective of date t, movements in m_{t+1} can be due to innovations to nominal consumption growth (as in the standard log case), but also to "news" about future consumption growth. Positive bond risk premia could thus arise either

because bond prices are high when nominal growth is high, or, with $\gamma > 1$, because nominal bond prices are high when bad news about future consumption growth arrives. Time variation in bond risk premia (under the belief of the agent) requires changes in the covariance between bond prices and either nominal growth or news about growth.[5]

B Comparing model and data

In the previous subsection, we have used only the implication of the Euler equation for returns between dates t and $t+1$. More generally, there are Euler equations corresponding to interest rates and excess returns over many different horizons. In order to quantitatively evaluate our model, we consider two sets of conditions. First, we consider the Euler equation for horizon h and a h-period bond. This is a generalization of (18). It says that the h-period interest rate satisfies

$$(20) \qquad i_t^{(h)} = -\frac{1}{h}\left(E_t\left[\sum_{j=1}^{h} m_{t+j}\right] + \frac{1}{2}var_t\left(\sum_{j=1}^{h} m_{t+j}\right)\right).$$

This type of condition thus connects interest rates of any horizon to conditional moments of the pricing kernel.

The second type of condition come from the Euler equation for horizon h and a bond of maturity $n > h$. The risk premium earned for holding an n-period bond over h periods is denoted $rp_{t,t+h}^{(n)}$. The Euler equation, a generalization of (19), is

$$rp_{t,t+h}^{(n)} := E_t p_{t+h}^{(n-h)} - p_t^{(n)} - i_t^{(1)} + \frac{1}{2}var_t p_{t+h}^{(n-h)} = -cov_t\left(\sum_{j=1}^{h} m_{t+j}, p_{t+h}^{(n-h)}\right).$$

It can be rewritten as a condition on the forward rate:

$$(21) \qquad f_t^{(n-h,h)} = -\frac{1}{n-h}\left(E_t p_{t+h}^{(n-h)} + \frac{1}{2}var_t p_{t+h}^{(n-h)} + cov_t\left(\sum_{j=1}^{h} m_{t+j}, p_{t+h}^{(n-h)}\right)\right)$$

In this form, the condition relates the forward rate to conditional moments of the pricing kernel as

---

[5] The importance of news shocks for risk premia when agents have Epstein-Zin utility is is directly related to agents' concern with the temporal distribution of risk. If $\gamma > 1$, agents dislike persistence in consumption streams and fear downward revisions in consumption expectations.

well as of bond payoffs at date $t + h$.

We want to use the family of conditional Euler equations (20)-(21) to compare the quantitative implications of the model to data. One difficulty we face is that the moments (20)-(21) must be computed under the agent's subjective belief. The subjective belief will generally differ from the empirical distribution of the data. As a result, standard instrumental variables techniques for the evaluation of conditional Euler equations do not apply. Instead, we directly evaluate the Euler equations date by date by computing the conditional moments in (20)-(21) under the subjective belief implied by the agent's learning algorithm. We can then check the resulting sequences of Euler equation errors.

Explaining statistical risk premia

We also want to study standard measures of risk premia derived from statistical models. Let $\widehat{rp}_{t,t+h}^{(n)}$ denote the risk premium implied by a statistical model, for example the model estimated in section II.B. It is defined in the same way as the subjective risk premium above, except that the conditional moments are evaluated under the statistical model, rather than the investor's subjective belief. We can decompose this statistical premium in a way that is analogous to the decomposition of the expected excess return in (4) – the only difference is the presence of Jensen's inequality terms in the premia:

$$(22) \qquad \widehat{rp}_{t,t+h}^{(n)} = \underbrace{\widehat{rp}_{t,t+h}^{(n)} - rp_{t,t+h}^{(n)}}_{\substack{\text{adjusted} \\ \text{forecast difference}}} + \underbrace{cov_t\left(\sum_{j=1}^{h} m_{t+j}, p_{t+h}^{(n-h)}\right)}_{\substack{\text{subjective} \\ \text{premium}}}$$

The results of section II.B suggest that both terms of the decomposition move around and contribute to predictability, with the forecast difference being more important at business cycle frequencies, and the subjective premium being more important at lower frequencies. We want to find out whether learning can explain this finding. To this end, we compute both terms in equation (22) – the difference in risk premia and the subjective premium – under the belief of a learning agent. We compare both terms to their counterparts observed in data. The observed difference in

risk premia is the difference between statistical premium and the premium computed using surveys. The observed subjective premium is the subjective premium computed from surveys in section II.B.

C   Learning algorithm

Most asset pricing studies assume that investor beliefs are conditionals of a stationary stochastic process that is estimated using all data through the date when the study is undertaken. This approach has two unattractive properties. First, it ignores the fact that investors in, say, 1980 only had access to data up to 1980. Second, it assumes that agents believed in the same stationary model throughout the postwar period. This is problematic given that the 1970s are often viewed as a period of structural change. Indeed, the decade witnessed the first ever peacetime inflation in the US, the breakdown of leading macroeconomic models, as well as significant innovation in bond markets.

We construct a sequence of beliefs for investors who learn from real time data and are concerned with structural change. We assume that at every date t, investors form beliefs based on a state space system of the form (5)-(8). The system includes prices, because we want to evaluate the price moments in conditional Euler equations of the type (21). We reestimate the system for every date t using only data up to date t. To accommodate concern with structural change, we maximize a modified likelihood function that puts more weight on more recent observations. To allow a sufficiently large initial sample for the estimation, and because we want to compare the results to our survey forecast data, the first belief is constructed for 1970:1. The analysis in this section will thus be restricted to the period since 1970.

Let $z_t = \left(i_t^{(1)}, i_t^{(20)} - i_t^{(1)}, \pi_t, \Delta c_t \right)^\top$. The system to be estimated can be written in compact notation as

$$
\begin{aligned}
z_t &= \mu_z + \phi_z x_{t-1} + e_t \\
x_t &= \phi_x x_{t-1} + \sigma_x e_t. \\
i_t^{(n)} &= a_n + b_n^\top x_t ; \quad n > 1, n \neq 20.
\end{aligned}
$$

(23)

We impose the same parameter restrictions on $\phi_x, \phi_z, \sigma_x, \Omega$ and (a_n, b_n) as in section II.

The estimation for every date t follows the same two steps as the estimation of the reduced form model in section II. In step 1, we estimate the dynamics of the factors together with the vector z_t, that is, the first two equation of (23). Let $\upsilon \in (0, 1)$ denote a "forget factor" that defines a sequence of geometrically declining sample weights. The weighted sample mean for date t is

$$(24) \qquad \hat{\mu}_z(t) = \left(\sum_{i=0}^{t-1} \upsilon^i \right)^{-1} \sum_{i=0}^{t-1} \upsilon^i z_{t-i}.$$

The sequence of estimated means picks up a low frequency component in z_t. We estimate (23) using data up to date t by minimizing the criterion

$$(25) \qquad -\frac{1}{2} \sum_{i=0}^{t-1} \upsilon^i \left[\log \det(\Omega) + (z_{t-i} - \hat{\mu}_z(t) - \phi_z x_{t-1-i})^{\top} \Omega^{-1} (z_{t-i} - \hat{\mu}_z(t) - \phi_z z_{t-1-i}) \right]$$

starting at $x_0 = 0$. The special case $\upsilon = 1$ corresponds to standard maximum likelihood estimation: it minimizes the equally weighted sum of squared in-sample forecast errors. In contrast, the criterion (25) penalizes recent forecast errors more heavily than those in the distant past. Ljung and Soderstrom (1987) and Sargent (1993) advocate this approach to adaptive learning in situations where the dynamics of a process may change over time. As in Section II, the estimation step not only delivers estimates for the matrices ϕ_x, ϕ_z, σ_x and Ω, but also estimates for the sequence of states $(x_\tau)_{\tau=1}^t$, starting from $x_0 = 0$. In particular, we obtain an estimate of the current state x_t that can be taken as the basis for forecasting future fundamentals under the system estimated with data up to date t.

The second step in the estimation is to find the parameters (μ_Q, ϕ_Q) that determine interest rate coefficients (a_n, b_n) in the third equation of (23). As in section II, this step minimizes the sum of squared differences between interest rates in the data and interest rates predicted by the model. In this step we also weight the squared errors in the same way as the forecast errors in (25).

A byproduct of the date t estimation is an estimate of the date t factors as perceived by the agent at date t. Computing conditional distributions given x_t date by date produces a sequence of investor beliefs. The subjective belief at date t determines investors' evaluation of future utility and

asset payoffs at date t. We thus use this belief below to calculate expectations of the pricing kernel, that is, yields, for date t. In contrast to the benchmark approach, the exercise of this section does not impose any direct restriction on beliefs across different dates; for example, it does not requires that all beliefs are conditionals of the same probability over sequences of data. The updating of beliefs is thus implicit in the sequential estimation.

The model also does not impose a direct link between investor beliefs and some "true data generating process", as the benchmark approach does by imposing rational expectations. The belief at date t captures investors' subjective distribution over fundamentals at date t. It is constrained only by past observations (via the estimation step), and not by our (the modelers') knowledge of what happened later.

D   Subjective forecasts

The learning algorithm performs two jobs in our model. First, it delivers samples of subjective forecasts that can be compared to survey forecasts. Second, it generates the complete conditional distribution of interest rates and the pricing kernel. In particular, the recursive estimation will generally lead to changes in conditional second moments, and hence to changes in subjective risk premia. The two jobs performed by the learning algorithm correspond closely to the two terms in the decomposition of statistical risk premia (22). In this subsection, we consider subjective forecasts.

Summary statistics

Table 6 shows summary statistics on subjective forecasts by adaptive learners. Its basic structure is similar to Table 3: for every variable (inflation and yields of different maturities) and every forecast horizon for which we have data, we compare mean absolute differences between subjective forecasts and actual survey forecasts. We report two models where more recent data are weighted more heavily, with forget factors $v = .9$ and $v = .95$. For comparison, we also report recursive maximum likelihood estimation ($v = 1$), which weighs all data up to t equally, as well as the forecast from the baseline statistical model of section II.B.

The main result from the table is that forecasts made by adaptive learners are closer to survey

forecasts than forecasts from a statistical model. For essentially all interest rates as well as for inflation, the mean absolute difference between the learning forecast and the survey forecast is smaller than the difference between the statistical forecast and the survey forecast. The only exception come for short (2 quarter) horizon forecasts of medium term (e.g. 20 year) bonds. Overall, survey forecasters thus seem to behave more like adaptive learners than like agents who have been given the statistical model that we have today.

Differences in risk premia

We now turn to the role of adaptive learning for statistical risk premia. We know from Figure 1 seen that there are systematic differences between statistical and survey forecasts that account for a significant part of statistical risk premia. The question here is whether learning forecasts share this property of survey forecasts. In other words, can adaptive learning be the reason why survey forecasters differ systematically from statistical forecasts, and in a way that helps understand statistical risk premia?

We consider the risk premium for holding a 10-year bond over 1 year, and fix the forget factor to $v = .95$. Unfortunately, we do not have actual survey data for a forecast horizon of one year over the whole sample. We thus use forecasts computed from the subjective state space system in section II.B as a proxy. As we have seen, those forecasts provide a good approximation to the actual survey forecasts.

Figure 5 plots the difference in risk premia – statistical premium minus subjective premium – for the 10 year bond held over one year. For the gray line, the subjective premium is from surveys, whereas for the black it comes from the learning model. The two lines share two properties. First, both are negative most of the time before 1980, and positive thereafter. In times of increasing inflation and interest rates before 1980, statistical forecasts of interest rates were typically higher than survey forecasts, which were "lagging behind" the increase in rates. As a result, survey forecasters predicted higher bond prices and hence premia. In contrast, in times of decreasing inflation and interest rates after 1980, statistical forecasts of interest rates were typically lower than survey forecasts – the survey forecasters, again "lagging behind" were still predicting higher rates. As a result, survey forecasters were also predicting lower bond prices and premia. The adaptive

nature of the learning algorithm captures this basic pattern of forecasters "lagging behind".

TABLE 6: FORECASTS FROM LEARNING ALGORITHM

Panel A: Mean absolute fitting errors for Bluechip yield forecasts

(annualized basis points)

horizon	2 quarters					4 quarters				
maturity (qtrs)	1	4	20	40	120	1	4	20	40	120
learning $v = .9$	36	47	35	42	46	54	59	47	52	50
learning $v = .95$	30	47	33	44	58	49	56	51	62	69
rec. MLE ($v = 1$)	51	54	33	44	72	72	72	64	77	103
stat. model	46	49	32	36	57	67	65	57	65	83

Panel B: Mean absolute fitting errors for other forecasts

(annualized basic points)

survey	SPF	Goldsmith-Nagan	
variable	inflation	4 qtr yield	80 qtr yield
horizon	4 quarters	2 quarters	
learning $v = .9$	74	196	38
$v = .95$	85	202	57
$v = 1$	95	204	71
stat. model	99	211	61

The second property of both lines is that they tend to rise after recessions. This property is more pronounced for the surveys; in particular, the learning model "misses" increases in the premia difference after the 1990 recession, and also does not replicate all the increases after the 1970s recessions. Nevertheless, there is a clear cyclical pattern in the premia differences also for the learning model.
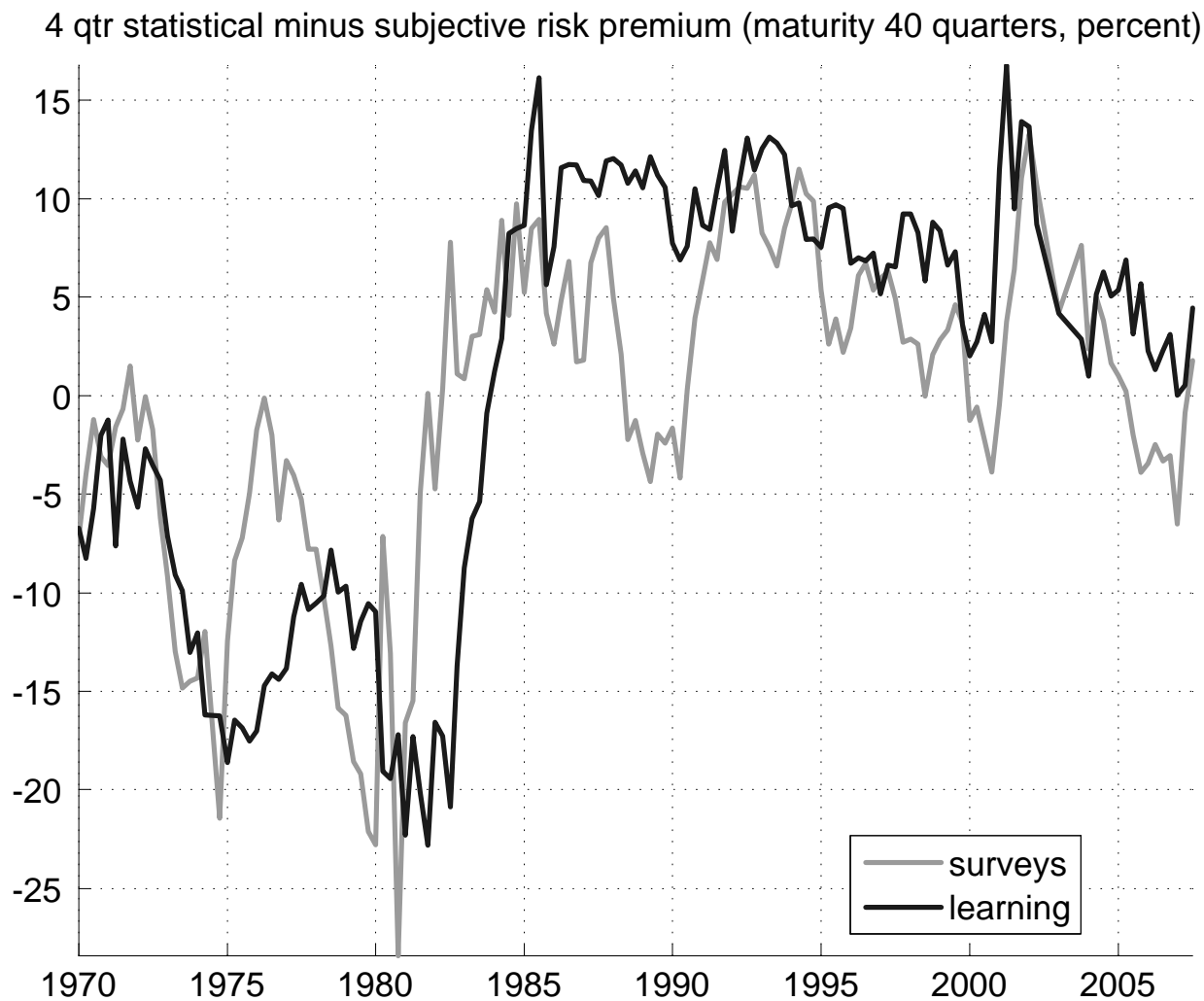
4 qtr statistical minus subjective risk premium (maturity 40 quarters, percent)

Figure 5: Differences in risk premia

E   Risk premia

In this subsection, we evaluate conditional Euler equations. We focus on three equations. First, we consider the interest rate equation (20) for horizons (and maturities) $h = 4$ quarters and $h = 40$ quarters. The Euler equation error in these interest rate equations tells us how well the model describes the dynamics of the 1 year and 10 year yields to maturity. Second, we consider a forward rate equation of the type (21) for $n = 40$ quarters and $h = 4$ quarters. The Euler equation error for this equation tells us how well the model captures the risk premium on a 10 year bond held over 1 year.

We fix the forget parameter $\upsilon = .95$ and derive the associated sequence of agent beliefs. We then select the preference parameters β and γ in order to minimize the equally weighted squared sum of the Euler equation errors on the three conditional Euler equations. To provide context, we perform this exercise not only for our learning agent with Epstein-Zin utility, but for two agents with beliefs given by the benchmark statistical model: an agent with log utility and one with Epstein-Zin utility. Table 7 reports, for every model, the preference parameters that minimize the Euler equation errors, as well as the mean absolute Euler equation errors for each of the three equations. Here we have labelled the three equation by the interest rate or forward rate appearing on the left hand side of (20) or (21), respectively.

The properties of the benchmark log utility case are well known. With log utility and homoskedastic shocks – as under the statistical belief – interest rates are, up to constant, equal to expected nominal consumption growth. As a result, the log utility model has problems generating sufficient volatility in yields. Moreover, the model cannot generate an upward sloping average yield curve.[6] Intuitively, in a log model, long bonds are less risky since they provide insurance against times of low consumption growth.

Retaining the statistical belief, but moving to Epstein-Zin utility does not change the Euler equation errors much, even though risk aversion increases now to 26. The new feature with Epstein-Zin utility is that covariance with expected future growth now matters for yield spreads and risk premia. On the one hand, the yield curve slopes upwards if inflation delivers bad news about future consumption growth, as in Piazzesi and Schneider (2006). On the other hand, Euler equations for bond risk premia can implies a positive bond risk premium if a high bond price is good news for future consumption growth. Under the statistical belief, the former effect is present for high enough γ. However, the second effect is not present under the statistical belief. This leads our fitting procedure to select low risk aversion, and a small improvement in fit relative to the log case.

Under learning, the errors on all Euler equations are significantly smaller than under the statistical belief. Risk aversion is now 80. The reason for this result is that, over much of the sample, the distributions estimated by the learning algorithm have both high inflation predict low consumption

---

[6]Mechanically, an upward sloping yield curve requires that the term in brackets on the right hand side of 20 grow faster than linearly with maturity h. Since nominal growth is persistent, this cannot be true with log utility.

growth, and high interest rates predict low growth. The first effect contributes to an upward sloping yield curve, whereas the second effect helps generate positive bond premia.

Table 7 does not directly show how the statistical belief and learning model differ in terms of time variation in risk premia. Since our statistical system is homoskedastic, the model under the statistical belief does not give rise to any time variation in statistical premia. In contrast, the learning model leads to time variation, both because of forecast differences, as discussed in the previous subsection, and because of changes in risk brought about by the learning process. We now describe this time variation.

<div align="center">

TABLE 7: EULER EQUATIONS ERRORS

</div>

model	Preferences		Euler equation errors		
			Mean absolute error (% p.a.)		
	β	γ	$i_t^{(4)}$	$i_t^{(40)}$	$f_t^{(36,4)}$
statistical belief					
log utility	.9970	1	1.97	1.70	0.83
Epstein-Zin	.9961	26	1.96	1.69	0.86
learning $\upsilon = .95$					
Epstein-Zin	.9966	80	1.76	1.21	0.62

Time variation in risk premia

Figure 6 plots subjective and statistical risk premia for the 10 year bond held over 1 year under the learning model, together with the data counterparts (derived from surveys). The top panel contains subjective premia, that is, the covariance between the pricing kernel and the bond payoff. As discussed in section II.B, the subjective premium implied by survey forecasts exhibits a large low frequency component that is particularly high around 1980. The figure shows that the subjective premium under learning also exhibits a low frequency component that is high around 1980. This movement is due to slow changes in the covariance between bond payoffs and expected future consumption growth.

The bottom panel of Figure 6 adds together the two components of the statistical risk premium
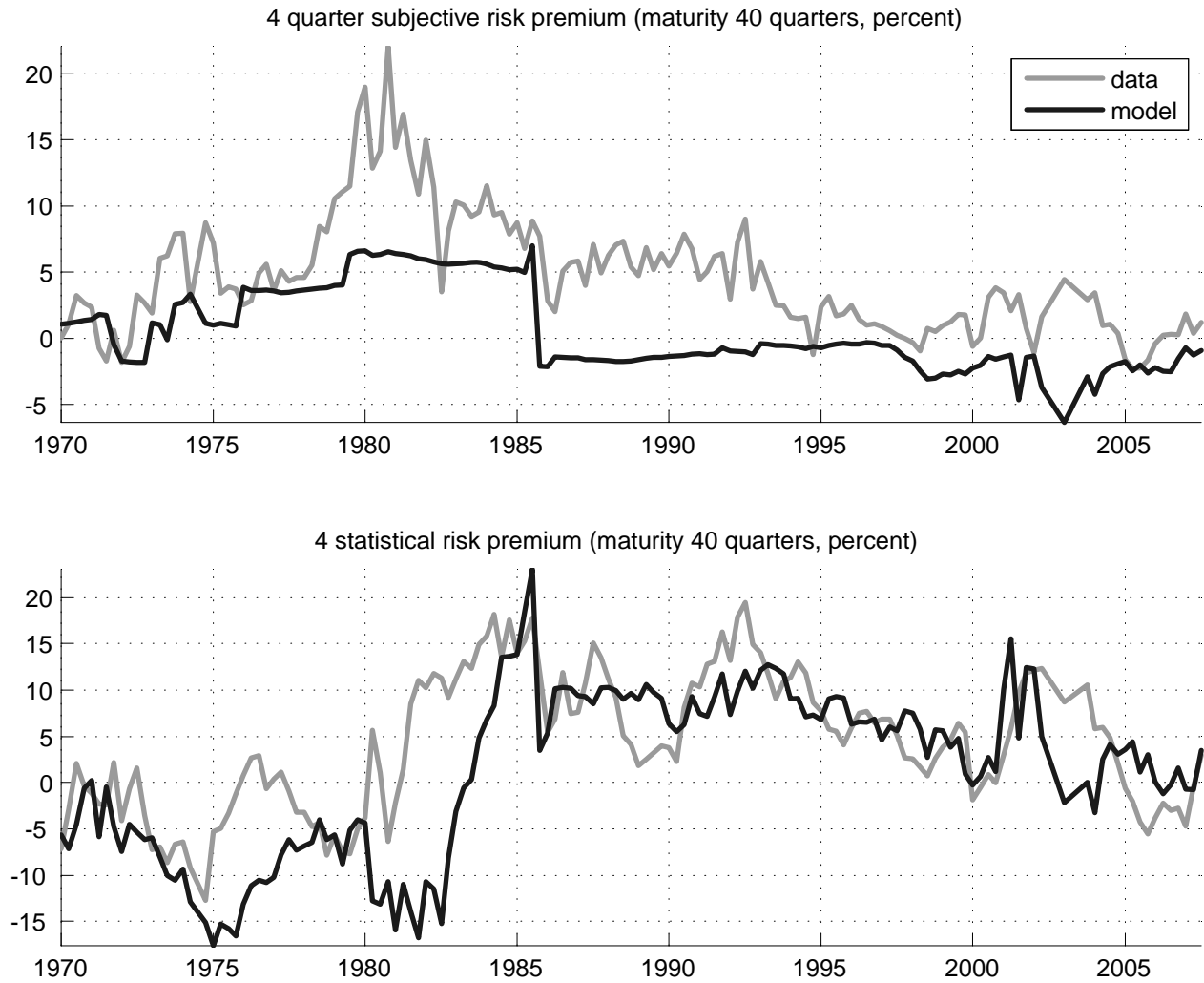
Figure 6: Time variation in risk premia

(22) under the learning model, and compares the sum to the statistical risk premium from the data. Both the forecast difference from Figure 5 and the subjective premium from the top panel of Figure 6 contribute to generate time variation in statistical premia under the learning model.

References

Ang, A., Bekaert, G., Wei, M. 2007. Do Macro Variables, Asset Markets or Surveys Forecast Inflation better? *Journal of Monetary Economics* 54, 1163-1212.

Bacchetta, P., Mertens, E., van Wincoop, E. 2008. Predictability in Financial Markets: What do Survey Expectations tell us? Manuscript, Lausanne.

Backus, D., Gregory, A., Zin, S. 1989. Risk Premiums in the Term Structure: Evidence from Artificial Economies. *Journal of Monetary Economics* 24, 371-399.

Bansal, R., A. Yaron 2004. Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance* 59, 1481-1509.

Campbell, J.Y., Chan Y.L., Viceira, L.M. 2003. A Multivariate Model of Strategic Asset Allocation. *Journal of Financial Economics* 67(1), 41-80.

Campbell, J.Y., Cochrane, J. 1995. By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. *NBER Working paper* 4995.

Campbell, J.Y., Viceira, L.M. 2001. Who should buy long-term bonds? *American Economic Review* 91(1), 99-127.

Chernov, M., Mueller, P. 2008. The Term Structure of Inflation Expectations. Manuscript, LBS.

Chun, A.L. 2006. Expectations, Bond Yields and Monetary Policy. Manuscript, University of Montreal.

Cochrane, J.H., M. Piazzesi 2005. Bond Risk Premia. *American Economic Review* 95:1, 138-160.

Cogley, T. 2005. Changing Beliefs and the Term Structure of Interest Rates: Cross-Equation Restrictions with Drifting Parameters. *Review of Economic Dynamics* 8, pp.

Dai, Q., K.J. Singleton 2002. Expectation Puzzles, Time-varying Risk Premia, and Affine Models of the Term Structure *Journal of Financial Economics* 63, 415-441.

Epstein, L., Zin, S. 1989. Substitution, risk aversion and the temporal behavior of consumption and asset returns: a theoretical framework. *Econometrica* 57, 937-69

Fama, E.F., Bliss, R.R.. The Information in Long-Maturity Forward Rates. *American Economic Review*, 1987, (77), 680-92.

Frankel, J.A., Froot, K. 1987. Using Survey Data to Test Standard Propositions Regarding Exchange Rate Expectations. *American Economic Review* 77, 133-153.

Friedman, B. 1980. Survey Evidence on the 'Rationality' of Interest-Rate Expectations. *Journal of Monetary Economics* 6, 453-465.

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

Gallmayer, M., Hollifield, B., Palomino, F., Zin, S. 2007. Arbitrage-Free Bond Pricing with Dynamic Macroeconomic Models. Manuscript, Carnegie-Mellon.

Gourinchas, P.0., Tornell A. 2005. Exchange Rate Dynamics, Learning, and Misperception. *Journal of International Economics* 64, pp. 303-333.

Greenwood, J. and D. Vayanos 2008, Bond Supply and Expected Excess Returns. Manuscript, HBS and LSE.

Gürkaynak, R.S., Sack, B., Wright, J.H., 2006. The U.S. Treasury Yield Curve: 1961 to the Present. Manuscript, Federal Reserve Board of Governors.

Hansen, L.P., Heaton, J.C., and Li, N. 2008. Consumption strikes back? Measuring Long Run Risk. *Journal of Political Economy.* 116(2), pp. 260-302.

Hansen, L.P., Heaton, J.C., Lee, J.H., and Roussanov, N. 2007. "Chapter 61. Intertemporal Substitution and Risk Aversion" in *Handbook of Econometrics Volume* 6 Part I., ed. by J.Heckman and E. Leamer, pp. 3967-4056.

Kim, D., A. Orphanides 2007. Term Structure Estimation with Survey Data on Interest Rate Forecasts. Manuscript, Federal Reserve Board.

Krishnamurthy, A., Vissing-Jorgensen, A. 2007. The Demand for Treasury Debt. *NBER Working Paper* #12881.

Lucas, R.E. Jr. 1978. Asset Prices in an Exchange Economy. *Econometrica* 46, 1429-1445.

McCulloch, J.H., Kwon, H-C. 1993. U.S. Term Structure Data, 1947-1991, Ohio State University Working Paper #93-6.

Ophanides, A., Wei, M. 2008. Evolving Macroeconomic Perceptions and the Term Structure of Interest Rates. Manuscript, Federal Reserve Board.

Piazzesi, M., Schneider, M. 2007. Equilibrium Yield Curves. *NBER Macroannual 2006.*

Veronesi, P., Yared, F. 2000. Bond Prices under Regime Shifts. Manuscript, Chicago GSB.

Wachter, J. 2006. A Consumption-Based Model of the Term Structure of Interest Rates. *Journal of Financial Economics* 79, 365-399.

A   Appendix

A.1 Estimation Procedure for state space system

We use a two-step procedure to estimate the dynamics of consumption growth, inflation, and interest rates. The parameters θ can be partitioned into the parameters of the state space $(\mu_z, \eta_z, \Omega, \phi_x, \sigma_x)$ and market prices of risk (l_0, l_1). In the first step, we estimate the state-space parameters with maximum likelihood. In the second step, we estimate the parameters l_0 and l_1 given the parameters $\mu_z, \eta_z, \Omega, \phi_x$, and σ_x estimated in the first step. This is done by minimizing the sum of squared fitting errors of the model. More precisely, we compute yields according to the fomrula $\widehat{i}_t = a_n + b_n^\top x_t$ for given values of the state at time t, and then solve

$$(\text{A-1}) \qquad \min_{\{l_0, l_1\}} \sum_{t=1}^{T} \sum_{n=1}^{N} \left(\widehat{i}_t^{(n)} - i_t^{(n)} \right)^2$$

for the N yields in our database.

We compute standard errors for our parameter estimates using GMM, with moments from both stages of our estimation procedure. The MLE moments are the scores of the likelihood function for $\theta_1 = (\mu_z, \eta_z, \Omega, \phi_x, \sigma_x)$, and the NLS moments are the first-order conditions of the minimization (A-1) for $\theta_2 = (l_0, l_1)$. The standard errors we compute adjust for the two-state nature of the estimation. More formally, the scores in the first step are

$$(\text{A-2}) \qquad g_1(\theta_1) = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \ln l\left(h_t | h_{t-1}, h_{t-2}, ..., x_0 = 0 \right)}{\partial \theta_1}$$

and the first-order conditions in the second step are

$$(\text{A-3}) \qquad g_2\left(\theta_2, \widehat{\theta}_1 \right) = 2\frac{1}{T} \sum_{t=1}^{T} \sum_{n=1}^{N} \left(\widehat{i}_t^{(n)} - i_t^{(n)} \right) \frac{\partial \widehat{i}_t^{(n)}}{\partial \theta_2}$$

where $\widehat{\theta}_1$ is fixed at the solution of the minimization problem (A-1). We define $G_{i,j} = \partial g_i / \partial \theta_j^\top$.

To obtain GMM standard errors, we expand the function g_1 and g_2 around the true parameter

values using a first-order Taylor approximation

$$
\begin{aligned}
\sqrt{T}g_1\left(\widehat{\theta}_1\right) &= \sqrt{T}g_1\left(\theta_1\right) + G_{1,1}\sqrt{T}\left(\widehat{\theta}_1 - \theta_1\right) \\
\sqrt{T}g_1\left(\widehat{\theta}_1,\widehat{\theta}_2\right) &= \sqrt{T}g_1\left(\theta_1,\theta_2\right) + G_{2,1}\sqrt{T}\left(\widehat{\theta}_1 - \theta_1\right) + G_{2,2}\sqrt{T}\left(\widehat{\theta}_2 - \theta_2\right).
\end{aligned}
$$

The left hand side of these equations are zero, so that we can stack these conditions in matrix form

$$
\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \sqrt{T}g_1\left(\theta_1\right) \\ G_{2,2}^{\top}\sqrt{T}g_1\left(\theta_1,\theta_2\right) \end{bmatrix} + \begin{bmatrix} G_{1,1} & 0 \\ G_{2,2}^{\top}G_{2,1} & G_{2,2}^{\top}G_{2,2} \end{bmatrix} \begin{bmatrix} \sqrt{T}\left(\widehat{\theta}_1 - \theta_1\right) \\ \sqrt{T}\left(\widehat{\theta}_2 - \theta_2\right) \end{bmatrix}
$$

Using the rule of partitioned matrix inversion, we get

$$
\begin{bmatrix} G_{1,1} & 0 \\ G_{2,2}^{\top}G_{2,1} & G_{2,2}^{\top}G_{2,2} \end{bmatrix}^{-1} = \begin{bmatrix} G_{1,1}^{-1} & 0 \\ -\left(G_{2,2}^{\top}G_{2,1}\right)^{-1}G_{2,2}^{\top}G_{2,1}G_{1,1}^{-1} & \left(G_{2,2}^{\top}G_{2,2}\right)^{-1} \end{bmatrix}.
$$

This allows us to solve for

$$
\begin{bmatrix} \sqrt{T}\left(\widehat{\theta}_1 - \theta_1\right) \\ \sqrt{T}\left(\widehat{\theta}_2 - \theta_2\right) \end{bmatrix} = \begin{bmatrix} G_{1,1}^{-1} & 0 \\ -\left(G_{2,2}^{\top}G_{2,1}\right)^{-1}G_{2,2}^{\top}G_{2,1}G_{1,1}^{-1} & \left(G_{2,2}^{\top}G_{2,2}\right)^{-1} \end{bmatrix} \begin{bmatrix} -\sqrt{T}g_1\left(\theta_1\right) \\ -G_{2,2}^{\top}\sqrt{T}g_1\left(\theta_1,\theta_2\right) \end{bmatrix}
$$

The asymptotic variances of the parameter estimates are thus

$$
\begin{aligned}
\operatorname{var}\left(\sqrt{T}\left(\widehat{\theta}_1 - \theta_1\right)\right) &= G_{1,1}^{-1}V_{11}G_{11}^{-1\top} \\
\operatorname{var}\left(\sqrt{T}\left(\widehat{\theta}_2 - \theta_2\right)\right) &= \begin{bmatrix} -\left(G_{2,2}^{\top}G_{2,1}\right)^{-1}G_{2,2}^{\top}G_{2,1}G_{1,1}^{-1} & \left(G_{2,2}^{\top}G_{2,2}\right)^{-1} \end{bmatrix} V \\
&\quad \begin{bmatrix} -\left(G_{2,2}^{\top}G_{2,1}\right)^{-1}G_{2,2}^{\top}G_{2,1}G_{1,1}^{-1} & \left(G_{2,2}^{\top}G_{2,2}\right)^{-1} \end{bmatrix}^{\top}
\end{aligned}
$$

where $V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$ is the variance of the sample mean of the moment conditions. The upper-left matrix V_{11} is the simply the information matrix.

A.2 Estimation procedure for the subjective state space system

The subjective state space system can be derived from the objective state space system using the parameters (k_0, k_1). Therefore, we estimate the subjective state space system in three steps. The first two steps are as before; in the first step, we obtain the state space parameters $(\mu_z, \eta_z, \Omega, \phi_x, \sigma_x)$ with MLE, and in the second step, we obtain market prices of risk (l_0, l_1) using NLS. In the third step, we estimate parameters (k_0, k_1) by minimizing the squared differences between survey forecasts and subjective forecasts. More precisely, we compute forecasts $E_t^* h_{t+n} = \mu_h^* + \phi^* x_t$ and use survey data to form:

$$\min_{\{k_0, k_1\}} \sum_{t=1}^{T} \sum_{n=1}^{N} \left(\widehat{h}_t^{(n)} - i_t^{(n)} \right)^2.$$