

Learning, Dispersion of Beliefs, and Risk Premiums in an Arbitrage-free Term Structure Model*

Marco Giacoletti Kristoffer T. Laursen Kenneth J. Singleton †

May 29, 2016

Abstract

We model a Bayesian learner who updates beliefs about risk premiums in the US Treasury market using a dynamic term structure model that conditions on belief dispersion. Learning has a significant effect on measured risk premiums. This is true even though she infers the pricing distribution from the current yield curve. Our *real-time* learning rule substantially outperforms the consensus forecasts of market professionals over the past twenty-five years, particularly during the years following recessions in the U.S. economy. The predictive power of dispersion in beliefs for future yields appears to be largely distinct from that of inflation and output growth.

*We are grateful for feedback from Darrell Duffee, Lars Lochstoer, Jules Van Binsbergen, and seminar participants at Columbia University, the University of Michigan, University of Southern California, and Washington University-St. Louis.

†Marco Giacoletti is at the Graduate School of Business, Stanford University, mgiacol@gsb.stanford.edu; Kristoffer T. Laursen is at AQR Capital Management, LLC; and Kenneth J. Singleton is at the Stanford University, Graduate School of Business, and NBER, kenneths@stanford.edu

1 Introduction

There is substantial evidence that risk premiums in the U.S. Treasury bond markets vary over time, both as the shape of the yield curve changes and with macroeconomic conditions.¹ The vast majority of this evidence is based on econometricians' *retrospective within-sample* estimates of (fixed) model parameters. More plausibly, the marginal investors in the Treasury market are forming *prospective real-time* views about their risks. Moreover, the survey evidence suggests that market professionals disagree, at times substantially, about how changing economic conditions impact the future paths of bond yields.² Most econometric implementations of arbitrage-free term-structure models (*DTSMs*) presume (by omission) that this belief heterogeneity has no effect on the conditional distribution of bond yields. Nor do the standard pricing kernels in *DTSMs* capture the rich state-dependent and nonlinear pricing that may arise from learning. As such, there is the empirically relevant possibility that variation in risk premiums in bond markets induced by belief heterogeneity and learning has been misattributed to variation in agents' risk preferences.

How might a market participant have prospectively formed real-time views about risks in the Treasury market? This paper provides an answer to this question from the perspective of a Bayesian agent, \mathcal{RA} , who is learning about the distribution of bond yields while recognizing that the U.S. economy may experience permanent structural changes (in macro policy rules, regulation of financial markets, etc.). Consistent with prior *DTSMs*, the priced risks are taken to be the low-order principal components (*PCs*) \mathcal{P} of bond yields. \mathcal{RA} prices \mathcal{P} -sensitive payoffs using a stochastic discount factor (SDF) \mathcal{M}^B that reflects learning within a conditionally Gaussian *DTSM*, under the presumption that both \mathcal{P} and the degree of belief heterogeneity H are informative about future yields. By analyzing *real-time* learning about the parameters governing the objective distribution of $Z \equiv (\mathcal{P}, H)$ we shed new light on the interplay between learning, belief heterogeneity, and risk premiums.

A central motivation for our analysis is that relatively little is known about how learning affects measured risk premiums, or about how knowledge today of the extent of disagreement among market participants would alter views (of either an econometrician or a market participant) about the future shape of the yield curve. Agents in our economy are (plausibly) presumed to know that the priced risks are \mathcal{P} , and to fully observe these *PCs*. The challenge they face, and the source of their disagreement, is learning about the dynamics of the state Z in an environment of bond-price relevant structural change.

Though \mathcal{M}^B is a reduced-form, bond-market specific SDF, our analysis reveals heretofore unexplored features of investors' learning problems that are likely to discipline the SDFs of

¹ See, for examples, Duffee (2002), Ludvigson and Ng (2010), and Joslin, Pribsch, and Singleton (2014).

² See, for examples, Dovern, Fritsche, and Slacalek (2012) and Andrade, Crump, Eusepi, and Moench (2014).

associated equilibrium models. Specifically, we find that agents effectively knew key risk-neutral (\mathbb{Q}) parameters and treated them as constants over the past twenty-five years.³ This knowledge impacts the structure of \mathcal{RA} 's forecasts of future yields and required risk premiums. In particular, it implies that recursive least-squares and constant-gain learning are special cases of \mathcal{RA} 's optimal Bayesian learning. Simple least-squares based learning rules (that practitioners might reasonably have adopted) arise as (nearly) optimal Bayesian rules.

When \mathcal{RA} conditions her SDF \mathcal{M}^B on the joint history of yield information (\mathcal{P}) and forecaster disagreement (H) she obtains substantially more accurate real-time forecasts of yields than when conditioning on past \mathcal{P} alone. Additionally, constant-gain learning outperforms recursive least-squares learning, because it is more responsive to perceived structural changes. Indeed, \mathcal{RA} 's constant-gain (Bayesian) learning rule systematically out-forecasts (in real time) both random-walk models for individual yields (especially for long-term bonds) and the median Blue-Chip Financial (BCFF) professional. The latter outperformance is particularly evident following NBER recessions, when the median BCFF forecaster repeatedly thought that long-bond yields would rise much faster than they actually rose.

Moreover, consistent with a central premise of this study, learning and conditioning on H materially affects model-implied risk premiums. There are two potential sources of this forecast power of H : (i) a direct effect on future \mathcal{P} through the law of motion of (\mathcal{P}, H) , and (ii) an indirect effect through Bayesian updates of the parameters governing the covariances of \mathcal{P}_t with future \mathcal{P}_{t+s} , $s > 0$. Both effects are operative for forecasting the dominant level factor for the yield curve, with the indirect effects being especially large following NBER recessions. When the U.S. economy is emerging from a recession, knowledge of the extent of disagreement among professionals is informative about how today's yield curve will impact its future shape.

Finally, we examine whether the forecasting power of the dispersion measures H is related to the predictive power of macro factors for risk premiums documented in prior studies (see [footnote 1](#)). We find that the forecasting power of belief dispersion is distinct from that of inflation and output growth. In fact, accommodating learning as we do here, the macro factors have only weak real-time predictive power, whereas conditioning on H substantially lowers out-of-sample mean-squared forecast errors.

Our analysis of learning in bond markets complements several related studies. There is a large literature incorporating survey information directly into *DTSMs*. Extending the frameworks of [Kim and Orphanides \(2012\)](#) and [Chun \(2011\)](#), [Piazzesi, Salomao, and Schneider \(2013\)](#) model survey forecasts as subjective views that are distinct from those of

³Most equilibrium models with belief heterogeneity focus on dynamically complete economies, in which cases agents agree on the risk-neutral distribution of priced risks (e.g., [Jouini and Napp \(2007\)](#)). The issue of how agents learn about the \mathbb{Q} distribution typically does not arise in such settings, because researchers focus directly on the SDF of the implied fictitious representative agent (e.g., [Xiong and Yan \(2009\)](#)).

the econometrician. In these models the median forecaster has full knowledge of risk-factor dynamics and her forecasts are spanned by the low-order PC s of bond yields. We extend these frameworks by introducing learning about parameters and allowing for belief heterogeneity to affect market prices of factor risks. Furthermore, survey data is not used in the estimation of our learning rules, because a large percentage of the variation in median Blue-Chip Financial (BCFF) forecasts is in fact not spanned by the yield PC s (the risk factors \mathcal{P}). Leaving the objective forecasts from our $DTSM$ unencumbered by a tight link to the median BCFF forecasts materially improves the accuracy of fitted risk premiums and real-time forecasts.

Collin-Dufresne, Johannes, and Lochstoer (2016) study equity risk premiums implied by a representative-agent, consumption-based model in which there is learning. We instead focus on a reduced-form SDF to ensure high accuracy in pricing of the entire yield curve (Dai and Singleton (2000), Duffee (2002)), while exploring learning about the objective distribution of the (in our case bond-relevant) state of the economy. Additionally, \mathcal{RA} 's SDF \mathcal{M}^B explicitly recognizes the prevalence of investor heterogeneity and the possibility that dispersion of beliefs is a source of priced risk and variation in risk premiums in Treasury markets.

Agents in the models of David (2008), Xiong and Yan (2009), and Buraschi and Whelan (2016) optimally filter for unknown state (e.g., aggregate output), while “agreeing to disagree” about the known values of the parameters governing the state process. Equilibrium bond prices depend on the pairwise relative beliefs across all agent types, thereby giving rise to a potentially high-dimensional factor space.⁴ Yet the low-order PC s account for the vast majority of the cross-sectional variation in bond yields. Therefore, we instead follow Joslin, Singleton, and Zhu (2011) (JSZ) and represent \mathcal{P} in terms of yield PC s, which market participants can reasonably be assumed to observe without error (Joslin, Le, and Singleton (2013)). This motivates a second key difference with our setting: \mathcal{RA} is learning about the unknown parameters governing a directly observed state process Z .

Section 2 sets forth our learning problem. The simplified case of learning from yield-curve information alone is discussed in Section 3 to provide a set of reference empirical results for our more general analysis. The impact of dispersion in beliefs on learning is explored in depth in Section 4. For comparability with the vast majority of macro-finance $DTSM$ s without learning, we focus on learning about a conditionally Gaussian state process experiencing structural change. As it turns out, an empirically tenuous aspect of the Bayesian learning rule underlying \mathcal{M}^B relates to learning about factor volatilities. Accordingly, in Section 7 we assess the robustness of our findings to introducing learning in the presence of stochastic volatility. A comparison of learning rules that condition on macroeconomic information versus

⁴Similarly, the setting of Barillas and Nimark (2014) gives rise to a “forecasting the forecast of others” problem (Townsend (1983), Singleton (1987)) which, in turn, leads to an infinite-dimensional set of higher-order beliefs affecting bond prices.

belief dispersion is presented in [Section 6](#). Concluding remarks are presented in [Section 8](#).

2 Learning with Dynamic Term Structure Models

Consider the space of future risky payoffs generated by portfolios of U.S. Treasury bonds with weights that are functions of the history $Z_1^t \equiv (Z_t, Z_{t-1}, \dots, Z_1)$ of a state vector Z_t .⁵ In the absence of arbitrage opportunities and under weak regularity properties of the portfolio payoff space, there exists a stochastic discount factor (SDF) $\mathcal{M}_{t+1}^{\mathcal{B}}$, and an associated equivalent martingale measure \mathbb{Q} that prices these Treasury portfolio payoffs ([Dalang, Morton, and Willinger \(1990\)](#)). $\mathcal{M}^{\mathcal{B}}$ can be generically represented in terms of the risk-neutral and historical (\mathbb{P}) conditional distributions of the priced risk factors \mathcal{P}_t in Treasury markets:

$$\mathcal{M}^{\mathcal{B}}(\mathcal{P}_{t+1}, Z_1^t) = e^{-r_t} \times \frac{f^{\mathbb{Q}}(\mathcal{P}_{t+1}|\mathcal{P}_t)}{f^{\mathbb{P}}(\mathcal{P}_{t+1}|Z_1^t)}. \quad (1)$$

Moreover we argue on both conceptual and empirical grounds that agents know and agree on the \mathbb{Q} distribution of \mathcal{P} , the numerator of $\mathcal{M}^{\mathcal{B}}$. Accordingly, in constructing $\mathcal{M}^{\mathcal{B}}$ in the presence of learning we first elaborate on our choices of the priced risks \mathcal{P} and the state Z . Then we provide empirical motivation for our specification of $f^{\mathbb{Q}}(\mathcal{P}_t|\mathcal{P}_{t-1})$. Finally, we specify the learning problem that gives rise to $f^{\mathbb{P}}(\mathcal{P}_t|Z_1^{t-1})$ in (1). The SDF $\mathcal{M}^{\mathcal{B}}$ is bond-market specific, owing to our focus on payoffs generated by Treasury bonds.

Bond prices are recovered by discounting Treasury coupons by the appropriate multiples of $\mathcal{M}^{\mathcal{B}}(\mathcal{P}_{t+s+1}, Z_1^{t+s})$ under the objective measure \mathbb{P} . As such, there are three complementary interpretations of our reduced-form $\mathcal{M}^{\mathcal{B}}$. First, since we endow the econometrician with $\mathcal{M}^{\mathcal{B}}$ and the associated likelihood function of bond yields and measures of belief dispersion, we are modeling how this econometrician learns and represents risk premiums in Treasury markets *in real time*. Second, if in fact there is a marginal agent \mathcal{RA} that prices bonds using $\mathcal{M}^{\mathcal{B}}$ (an active arbitrageur, for example) then we are also effectively modeling this agent’s SDF. Third, for the case of a dynamically complete economy, $\mathcal{M}^{\mathcal{B}}$ can be viewed as a bond-market specific SDF derived from the SDF of the fictitious “representative agent’s” SDF. After projecting the latter SDF onto \mathcal{P} , a change of measure from the representative agent’s beliefs to those of the econometrician gives $\mathcal{M}^{\mathcal{B}}$.⁶

⁵For any variable X , the notation X_i^k , $k > i$, is short-hand for $(X_k, X_{k-1}, \dots, X_i)$.

⁶This last step is necessary, because the beliefs of the representative agent (a confounding of individual agents’ views) typically do not correspond to the *objective* views of the econometrician.

2.1 Priced Risks and Conditioning Information

Before getting into the details of constructing \mathcal{M}^B we motivate our choices of the risk-factors \mathcal{P} and the composition of Z governing the market prices of these risks. There is substantial evidence that bond yields follow a low-dimensional factor structure and, in fact, such structures underlie the pricing and risk management systems of primary dealers. Moreover, over-fitting with high-dimensional \mathcal{P} can induce economically implausible Sharpe ratios on bond portfolios (Duffee (2010)). With these considerations in mind, we assume that \mathcal{P} is comprised of the known first three PC s of bond yields.⁷

Though investors observe \mathcal{P} today, they are unlikely to know, or indeed to agree on, the conditional distribution of future \mathcal{P} . The associated belief heterogeneity may be reflected in \mathcal{M}^B through several channels. First, dispersion of beliefs may be a priced risk, a source of variation in \mathcal{P} and hence bond prices. Second, the market prices of the risks \mathcal{P} may depend on the degree of disagreement across bond investors. Finally, when included in the state Z , measures of belief heterogeneity may be informative about the future paths of yields.⁸ As we now document, disagreement among professional forecasters varies substantially over time, is highly correlated with \mathcal{P} , and has considerable predictive power for future realized excess returns in Treasury markets.

Our measures of disagreement are constructed from interdecile ranges of forecasts by professionals as surveyed by the Blue Chip Financial Forecasts (BCFF).⁹ Specifically, we use the BCFF survey forecasts of yields over the period January, 1985 through March, 2012. This survey is typically released at the beginning of the month (usually the first business day), based on information collected over a two-day period (usually between the 20th and the 26th of the previous month). A total of 177 institutions provide forecasts during the period of our study. These institutions are divided by the BCFF survey provider into different broad sectors: Financial Services, Consulting, Business Associations, Manufacturing, Insurance Companies and Universities.¹⁰ Forecasts are averages over calendar quarters and cover horizons out to five quarters ahead. For example, in January, 1999, the two-quarter ahead forecast for a

⁷This places our agents in a very different learning environment than those in the the noisy- (Sims (2003)) and sticky-information (Mankiw and Reis (2002)) models of belief dispersion studied by Coibion and Gorodnichenko (2012) and Andrade, Crump, Eusepi, and Moench (2014), among others.

⁸Some of these features of our learning environment are shared by difference-of-opinion models (e.g., David (2008)) though, again, we emphasize that in our setting \mathcal{RA} is learning about the parameters governing the \mathcal{P} process and not about the values of these (observed) risk factors.

⁹Our subsequent analysis of learning is qualitatively robust to measuring dispersion in beliefs as the cross-sectional (point-in-time) volatility of professional forecasts (Patton and Timmerman (2010)) or the cross-sectional mean-absolute-deviation in forecasts (Buraschi and Whelan (2016)), and our measure is similar to that used by Andrade, Crump, Eusepi, and Moench (2014).

¹⁰In our sample there are 48 consulting companies, 111 financial services companies, and the remaining 18 are from insurance companies, business associations, manufacturers, and universities.

	\mathcal{P}			
	1Q	2Q	3Q	4Q
$ID(\mathcal{P}_1)$	41.74 %	55.71 %	56.83 %	55.50 %
$ID(\mathcal{P}_2)$	13.10 %	18.96 %	19.78 %	31.34 %
$ID(\mathcal{P}_3)$	31.98 %	38.14 %	43.75 %	49.25 %
$ID(y^{2y})$	44.36 %	52.95 %	52.63 %	49.87 %
$ID(y^{7y})$	37.54 %	50.12 %	57.63 %	59.29 %
$ID(y^{2y} - y^{6m})$	5.10 %	8.22 %	12.96 %	16.01 %

Table 1: Adjusted R^2 's from the projections of cross-forecasters inter-quantile differences for relevant yields and portfolios onto \mathcal{P} over forecast horizons of one through four quarters. The sample period is January, 1985 through March, 2011.

specific variable will be equal to its average value between April and June.

We use yields on zero-coupon bonds with maturities of 6 months and 1, 2, 3, 5, 7, and 10 years calculated from coupon-bond yields as reported in the CRSP database using the Fama-Bliss methodology for the sample period June, 1961 through March, 2012. The survey forecasts are for U.S. Treasury 6-month bill yield and par yields on coupon bonds with maturities of 1, 2, 3, 5, 7, and 10 years. To facilitate comparisons of forecasts from our *DTSMs* with those by the BCFP professionals, we use the survey-implied forecasts of averages of zero-coupon bond yields computed by [Le and Singleton \(2012\)](#).¹¹

To measure disagreement about the future path of the yield y^m on an m -year bond we compute, at each point in time and for each forecast horizon, the differences between the ninetieth and tenth percentiles of the cross-sectional distribution of forecasts by professionals, and denote this by $ID(y^m)$.¹² A similar methodology is used to compute $ID(PCi)$ for the dispersion of forecasts of future values of the i^{th} *PC*.

The ID series are highly correlated with \mathcal{P} , implying that belief heterogeneity is effectively a priced risk in bond markets. In [Table 1](#) we report the adjusted R^2 's from the projections of these dispersion measures onto the risk factors \mathcal{P}_t for forecast horizons of one through four quarters. Over 50% of the dispersion in beliefs about the “level” of yields is spanned by the first three *PCs* of Treasury yields. Correlations are notably lower for the measures $ID(PC2)$ and $ID(y^{2y} - y^{6m})$ of dispersions of views about the slopes of the entire and short-end of the yield curve, a pattern we return to subsequently.

These dispersion measures also have substantial predictive content for future excess returns.

¹¹Whereas forecasting zero-coupon yields in an affine *DTSM* is a linear forecasting problem (see below), par yields are nonlinear functions of zero-coupon yields. We avoid this complexity by interpolating the forecasts of par yields to obtain approximate forecasts of zero yields.

¹²In each month we check how many forecasters have published a forecast for the desired yield and predictive horizon. Out of the total 117 forecasters, we usually find approximately 45 forecasts.

\mathcal{P}_t	\mathcal{P}_2	\mathcal{P}_3	$ID(y^{6m})$	$ID(y^{2y})$	$ID(y^{5y})$	$ID(y^{7y})$	$adj R^2$
1.75 [2.56]	1.61 [2.51]	-0.82 [-1.17]					17.1%
4.21 [5.95]	1.86 [3.39]	-1.40 [-2.39]	-0.17 [-0.17]	3.59 [3.38]	0.88 [0.63]	-6.51 [-4.99]	36.4%

Table 2: Projections of the average one-year realized excess returns across bonds of maturities of two, three, five, seven, and ten years. The sample period is January, 1985 through March, 2011, and robust test statistics for the null hypothesis of zero are given in brackets.

Table 2 shows that the full-sample linear projections of the cross-sectional average realized excess returns over a one-year horizon (from t to $t + 1y$) onto \mathcal{P}_t and measures of forecast dispersion for yields on bonds of several maturities. The adjusted R^2 's more than double—from 17% to 36%—when the conditioning information is expanded from \mathcal{P}_t to also include the dispersion measures.¹³ Furthermore, the disagreements $H'_t = (ID(y^{2y}), ID(y^{7y}))$ about the two- and seven-year yields, displayed in Figure 1, drive this incremental predictability.

This descriptive evidence serves as strong motivation for our focus on belief heterogeneity as a major force in shaping \mathcal{RA} 's views about current and future Treasury yield curves, as well as in determining the market prices of the PC risks underlying variation in bond yields. At this juncture it remains only suggestive, however, as the impacts of H have yet to be assessed through the lens of a pricing model with *ex ante* learning. Towards this end, we focus on $DTSMs$ with priced risks \mathcal{P}_t and state vector $Z_t \equiv (\mathcal{P}_t, H_t)$.¹⁴

Fixing notation, when N yields y_t are used in modeling and estimation, the relevant conditioning information for learning is the history of Z_1^t plus the history of any $N - 3$ linearly independent combinations, say \mathcal{O}_t , of the yields such that $(\mathcal{P}_t, \mathcal{O}_t)$ fully span y_t .

2.2 Risk-Neutral Pricing of Bonds

Absent arbitrage opportunities, and under regularity, market participants can reverse engineer the risk-neutral distribution \mathbb{Q} from the prices of traded bonds. This distribution will not in general be unique, unless agents live in a dynamically complete economy. Therefore, just as in prior studies of arbitrage-free $DTSMs$, our parametric specification of the \mathbb{Q} distribution $f^{\mathbb{Q}}(\mathcal{P}_t | \mathcal{P}_{t-1})$ is presumed to represent an econometrically identified member of the family of admissible distributions.

¹³Buraschi and Whelan (2016) provide complementary descriptive evidence that measures of disagreement have full- and within-sample predictive content for excess returns. They do not explore real-time learning.

¹⁴With this choice of Z_t one can view our analysis as extending the setting of Joslin, Pribsch, and Singleton (2014) to one with real-time learning and pricing conditional on belief heterogeneity. As in their analysis, Z could be augmented to include macroeconomic information relevant for assessing risk premiums. We defer this extension until Section 6 in order to focus on characterizing learning about the dynamics of the risk factors \mathcal{P}_t using information about belief heterogeneity.

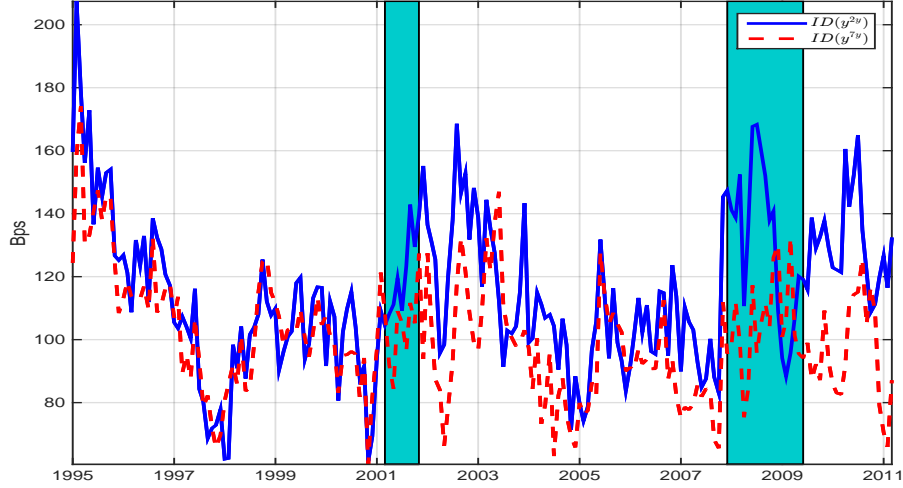


Figure 1: Historical measures of dispersions in professional forecasts one-year ahead for the two- and seven-year bond yields, $ID(2y)$ and $ID(7y)$

Concretely, as in a \mathbb{Q} -affine Gaussian *DTSM*, we assume that agents agree that the one-period riskless rate r_t follows the factor structure

$$r_t = \rho_0 + \rho_{\mathcal{P}} \mathcal{P}_t, \quad (2)$$

with \mathcal{P}_t following the autonomous Gaussian \mathbb{Q} process

$$\mathcal{P}_{t+1} = K_{0\mathcal{P}}^{\mathbb{Q}} + K_{\mathcal{P}\mathcal{P}}^{\mathbb{Q}} \mathcal{P}_t + \Sigma_{\mathcal{P}\mathcal{P}}^{1/2} e_{\mathcal{P},t+1}^{\mathbb{Q}}. \quad (3)$$

Then, the price D_t^m of a zero-coupon bond issued at date t and maturing at date $t + m$ is

$$D_t^m = E_t^{\mathbb{Q}} \left[\prod_{u=0}^{m-1} \exp(-r_{t+u}) | \Theta^{\mathbb{Q}}, \mathcal{P}_t \right]. \quad (4)$$

For econometric identification $\Theta^{\mathbb{Q}}$ can be normalized so that (JSZ and Appendix B) the parameters ρ_0 , $\rho_{\mathcal{P}}$, $K_{0\mathcal{P}}^{\mathbb{Q}}$, and $K_{\mathcal{P}\mathcal{P}}^{\mathbb{Q}}$ are known functions of $\Theta^{\mathbb{Q}} \equiv (k_{\infty}^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_{\mathcal{P}\mathcal{P}})$, with $k_{\infty}^{\mathbb{Q}}$ a scalar,¹⁵ $\Sigma_{\mathcal{P}\mathcal{P}}$ the upper $K \times K$ block of Σ_Z , and $\lambda^{\mathbb{Q}}$ the K -vector of eigenvalues of $K_{\mathcal{P}\mathcal{P}}^{\mathbb{Q}}$.

The JSZ normalization reveals clearly why it is reasonable to presume that agents effectively

¹⁵When \mathcal{P} follows a stationary process under \mathbb{Q} , $k_{\infty}^{\mathbb{Q}}$ is proportional to the risk-neutral long-run mean of r . We adopt this more robust normalization, since the shape of the yield curve may call for the largest eigenvalue $\lambda_1^{\mathbb{Q}}$ to be very close to or even larger than unity. See JSZ for details.

know and agree upon key elements of $\Theta^{\mathbb{Q}}$. Yields take the form

$$y_t^m = A_m(k_{\infty}^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_{\mathcal{PP}}) + B_m(\lambda^{\mathbb{Q}})\mathcal{P}_t. \quad (5)$$

Therefore, at each date t , the $B_m(\lambda^{\mathbb{Q}})$ (and hence $\lambda^{\mathbb{Q}}$) can be estimated very precisely from the cross-section of Treasury yields;¹⁶ agents will effectively know $\lambda^{\mathbb{Q}}$ from the date t cross-section y_t . The maturity-specific intercepts A_m depend also on $k_{\infty}^{\mathbb{Q}}$ and $\Sigma_{\mathcal{PP}}$. However, the impact of $\Sigma_{\mathcal{PP}}$ on A_m is through a convexity adjustment that is typically very small (see [Appendix A](#)). Therefore, knowledge of $\lambda^{\mathbb{Q}}$ and a tight prior on $k_{\infty}^{\mathbb{Q}}$ (also estimable from the cross-section of yields) mean that agents effectively know the A_m 's as well.

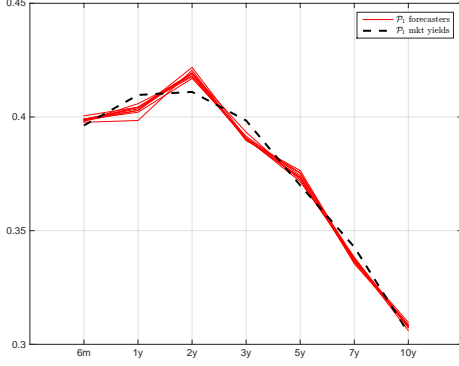
Using surveys of individual BCFF professionals, we can shed light on whether forecasters do in fact agree on $\lambda^{\mathbb{Q}}$. If they all believe that yields follow (5), then the yield forecasts for horizon h ordered by deciles, $y_{t,o_1}^h < \dots < y_{t,o_{10}}^h$, must satisfy

$$\hat{y}_{t,o}^{mh} = \bar{A}_m + \bar{B}_m \hat{\mathcal{P}}_{t,o}^h + e_{t,o}^{mh}, \quad (6)$$

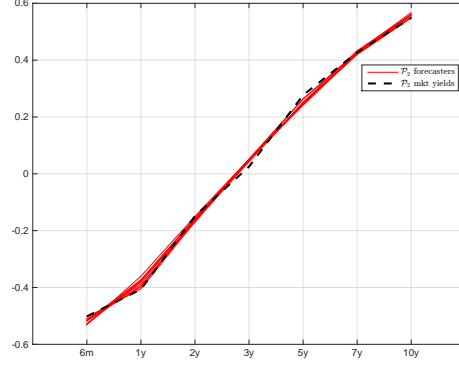
where y_{t,o_1}^h is the forecast of the professional falling at the tenth percentile, $y_{t,o_{10}}^h$ is the ninetieth percentile, etc. (We focus on order statistics, because the individual forecasters change over our sample.) Holding h fixed, the loadings should be common across the ordered professionals. [Figure 2](#) displays the full-sample estimates of these loadings by decile for *PC1* and *PC2* and h equal to one year (solid lines), along with the corresponding sample least-squares estimates of the loadings based on (5). The loadings are in fact very similar across forecaster deciles. Moreover, the professionals' values correspond very closely to the sample estimates. There are small differences for *PC1* over the one- to three-year maturity spectrum.

While this evidence supports the view that market professionals know $\lambda^{\mathbb{Q}}$, we are mindful of the phenomenon of trading desks recalibrating their yield curve models to new data on a regular basis. This recalibration, which gives updated hedge ratios $B_m(\lambda^{\mathbb{Q}})$ for bond portfolios, is perhaps motivated by the view that (5) is an approximation to the true pricing function. With this possibility in mind, for our subsequent econometric analysis, we allow \mathcal{RA} (equivalently, the econometrician) to update $\Theta^{\mathbb{Q}}$ monthly as new data becomes available, using the model-implied likelihood. However strikingly, even with this flexibility, she holds $(\lambda^{\mathbb{Q}}, k_{\infty}^{\mathbb{Q}})$ virtually fixed over our entire sample period (see [Section 2.5](#)).

¹⁶This is why the factor loadings B_m are reliably recovered from contemporaneous correlations among bond yields y_t^m and \mathcal{P}_t ([Duffee \(2011\)](#)). It also explains why, holding (K, N, Z, \mathcal{P}) fixed, estimates of $\lambda^{\mathbb{Q}}$ is *DTSMs* without learning are typically nearly invariant across specifications of the \mathbb{P} distribution of Z .



(a) Loadings on PC1



(b) Loadings on PC2

Figure 2: Time-series projections of the decile-ordered one-year ahead forecasts by BCFF forecasters of yields onto their forecasts of $PC1$ and $PC2$.

2.3 The Pricing Kernel \mathcal{M}^B In the Presence of Learning

Though much about Θ^Q can be inferred from the rich cross-sectional information in the Treasury yield curve, market participants must learn about the parameters governing the state-dependence of $f^{\mathbb{P}}(\mathcal{P}_{t+1}|Z_1^t)$ from historical time-series data. To motivate the specification of \mathcal{M}^B in the presence of such learning we adopt the perspective of a marginal agent \mathcal{RA} who prices portfolios of Treasury bonds using (5). Under the presumption that \mathcal{RA} treats Θ^Q as known, (1) can be expressed as

$$\mathcal{M}^B(\Theta^Q, \mathcal{P}_{t+1}, Z_1^t) = e^{-r_t} \times \frac{f^Q(\mathcal{P}_{t+1}|\mathcal{P}_t; \Theta^Q)}{f^{\mathbb{P}}(\mathcal{P}_{t+1}|Z_1^t)}, \quad (7)$$

with the numerator derived from (3) evaluated at Θ^Q .

The form of $f^{\mathbb{P}}(\mathcal{P}_{t+1}|Z_1^t)$ depends on \mathcal{RA} 's learning rule. We explore the pricing implications of the belief that $Z_t^t = (\mathcal{P}_t^t, H_t^t)$ follows the historical process

$$\begin{bmatrix} \mathcal{P}_{t+1} \\ H_{t+1} \end{bmatrix} = \begin{bmatrix} K_{\mathcal{P}0,t}^{\mathbb{P}} \\ K_{H0,t}^{\mathbb{P}} \end{bmatrix} + \begin{bmatrix} K_{\mathcal{P}\mathcal{P},t}^{\mathbb{P}} & K_{\mathcal{P}H,t}^{\mathbb{P}} \\ K_{H\mathcal{P},t}^{\mathbb{P}} & K_{HH,t}^{\mathbb{P}} \end{bmatrix} \begin{bmatrix} \mathcal{P}_t \\ H_t \end{bmatrix} + \Sigma_Z^{1/2} \begin{bmatrix} e_{\mathcal{P},t+1}^{\mathbb{P}} \\ e_{H,t+1}^{\mathbb{P}} \end{bmatrix}, \quad (8)$$

where $\Theta_t^{\mathbb{P}}$, the vectorized $(K_{0t}^{\mathbb{P}}, K_{Zt}^{\mathbb{P}})$ governing the conditional mean of Z_t , is time varying and unknown and the shocks are jointly Gaussian. In this environment \mathcal{RA} builds forecasts of future yields under the following three simplifying assumptions: (i) Θ^Q is treated as known and, period by period, is replaced by its maximum likelihood estimate $\tilde{\Theta}_t^Q$; (ii) learning about $\Theta_t^{\mathbb{P}}$ is summarized by \mathcal{RA} 's posterior density $f(\Theta_{t+1}^{\mathbb{P}}|Z_1^t, \mathcal{O}_1^t)$; and (iii) \mathcal{RA} invests assuming

that parameter variation per se is not a source of priced risk in bond markets.¹⁷

Given $\Theta^{\mathbb{Q}}$ and the specification in (8), the physical distribution of Z_t is determined by the evolution of the unknown and drifting parameters $\Theta_t^{\mathbb{P}}$. In Section 2.4 we show that if innovations to $\Theta_t^{\mathbb{P}}$ are normally distributed, then the posterior density $f(\Theta_{t+1}^{\mathbb{P}}|Z_1^t, \mathcal{O}_1^t)$ also follows a normal distribution. Furthermore, the conditional \mathbb{P} -distribution of Z_{t+1} given Z_1^t is:

$$f^{\mathbb{P}}(Z_{t+1}|Z_1^t) = \text{Normal}\left(\hat{K}_{0t}^{\mathbb{P}} + \hat{K}_{Zt}^{\mathbb{P}}Z_t, \Omega_t\right), \quad (9)$$

where $(\hat{K}_{0t}^{\mathbb{P}}, \hat{K}_{Zt}^{\mathbb{P}})$ denotes \mathcal{RA} 's posterior mean of $\Theta_t^{\mathbb{P}}$ and her one-period-ahead forecast covariance matrix Ω_t depends on Σ_Z and the uncertainty regarding $\Theta_t^{\mathbb{P}}$.

Within this learning environment \mathcal{RA} 's SDF \mathcal{M}^B takes the form:

$$\mathcal{M}(\Theta^{\mathbb{Q}}, \mathcal{P}_{t+1}, Z_1^t) = \exp\left\{-r_t - \frac{1}{2} \log |\Gamma_t| - \frac{1}{2} \hat{\Lambda}'_{\mathcal{P}t} \Gamma_t^{-1} \hat{\Lambda}_{\mathcal{P}t} - \hat{\Lambda}'_{\mathcal{P}t} \Gamma_t^{-1} \varepsilon_{t+1}^{\mathbb{P}} + \frac{1}{2} (\varepsilon_{t+1}^{\mathbb{P}})' (I - \Gamma_t^{-1}) \varepsilon_{t+1}^{\mathbb{P}}\right\}, \quad (10)$$

$$\begin{aligned} \Gamma_t &= \Omega_{\mathcal{P}\mathcal{P},t}^{-1/2} \Sigma_{\mathcal{P}\mathcal{P}} (\Omega_{\mathcal{P}\mathcal{P},t}^{-1/2})', \\ \Omega_{\mathcal{P}\mathcal{P},t}^{1/2} \hat{\Lambda}_{\mathcal{P}t} &= \hat{\Lambda}_{0t}(\Theta^{\mathbb{Q}}, \hat{\Theta}_t^{\mathbb{P}}) + \hat{\Lambda}_{1t}(\Theta^{\mathbb{Q}}, \hat{\Theta}_t^{\mathbb{P}}) Z_t, \end{aligned}$$

where the market price of risk $\hat{\Lambda}_{\mathcal{P}t}$ depends on the posterior mean $\hat{\Theta}_t^{\mathbb{P}}$ and, therefore, implicitly on the entire history Z_1^t (Appendix D). The quadratic form in (10) differs from that in the SDF of a standard Gaussian *DTSM* owing to learning. Though the form of $\Lambda_{\mathcal{P}t}$ is familiar from Duffee (2002), the weights are state-dependent, again owing to learning. Thus, under \mathcal{RA} 's beliefs, \mathcal{P} follows a nonlinear (in particular, non-affine) process and \mathcal{RA} 's optimal forecasts are computed accordingly.

The portfolios \mathcal{P}_t are priced perfectly by (4). The model is completed with the following pricing equation for the higher-order fourth through seventh *PCs* \mathcal{O}_t :

$$\mathcal{O}_t = A_{\mathcal{O}}(\Theta^{\mathbb{Q}}) + B_{\mathcal{O}}(\Theta^{\mathbb{Q}}) \mathcal{P}_t + \Sigma_{\mathcal{O}}^{-1/2} \varepsilon_{\mathcal{O},t},$$

where $(\mathcal{P}_t, \mathcal{O}_t)$ fully spans y_t . The errors $\varepsilon_{\mathcal{O},t}$ are assumed to be *iid Normal*(0, $\Sigma_{\mathcal{O}}$), with $\Sigma_{\mathcal{O}}$ diagonal (consistent with its sample counterpart from a regression of \mathcal{P}_t on \mathcal{O}_t).

At date t a Bayesian \mathcal{RA} , faced with new observations (Z_t, \mathcal{O}_t) and the past history $(Z_1^{t-1}, \mathcal{O}_1^{t-1})$, evaluates an (approximate) likelihood function by integrating out the uncertainty about $\Theta_t^{\mathbb{P}}$ using her posterior distribution. Thus, with \mathcal{RA} treating $(\Theta^{\mathbb{Q}}, \Sigma_{\mathcal{O}})$ as fixed and

¹⁷The absence of compensation for parameter risk is fairly standard in the literature on pricing with Bayesian learning, and it greatly simplifies what is already a challenging modeling problem. Collin-Dufresne, Johannes, and Lochstoer (2016) explore implications of priced parameter uncertainty in a single-agent, consumption based setting with a much lower dimensional state space than what is required to reliably price bonds. Accommodating priced parameter uncertainty within a higher dimensional *DTSM* is an interesting topic for further research.

known,

$$f(Z_1^t, \mathcal{O}_1^t) = \prod_{s=1}^t f(\mathcal{O}_s | Z_1^s, \mathcal{O}_1^{s-1}; \Theta^{\mathbb{Q}}, \Sigma_{\mathcal{O}}) \times \int f(Z_s | Z_1^{s-1}, \mathcal{O}_1^{s-1}, \Theta_{s-1}^{\mathbb{P}}; \Sigma_Z) f(\Theta_{s-1}^{\mathbb{P}} | Z_1^{s-1}, \mathcal{O}_1^{s-1}) d(\Theta_{s-1}^{\mathbb{P}}). \quad (11)$$

2.4 \mathcal{RA} 's Bayesian Learning Problem

For our empirical analysis we specialize \mathcal{RA} 's Bayesian learning problem as follows. To allow for constraints on the market prices of risk, we assume that $\Theta_t^{\mathbb{P}}$ can be partitioned as $(\psi^r, \psi_t^{\mathbb{P}})$, where $\psi_t^{\mathbb{P}}$ is the vectorized set of free parameters and ψ^r is the vectorized set of parameters that are fixed conditional on $\Theta^{\mathbb{Q}}$. Letting ι_r and ι_f denote the matrices that select the columns of $(I \otimes [1, Z'_{t-1}])$ corresponding to the restricted and free parameters, and collecting the known terms in (8) into $\mathcal{Y}_t = Z_t - (I \otimes [1, Z'_{t-1}]) \iota_r \psi^r$, we can rewrite the state equation as

$$\mathcal{Y}_{t+1} = \mathcal{X}_t \psi_t^{\mathbb{P}} + \Sigma_Z^{1/2} e_{Z,t+1}^{\mathbb{P}}, \quad (12)$$

where $\mathcal{X}_t = (I \otimes [1, Z'_t]) \iota_f$.

To accommodate the possibility of permanent structural change in the underlying economic environment, we assume that $\psi_t^{\mathbb{P}}$ evolves according to

$$\psi_t^{\mathbb{P}} = \psi_{t-1}^{\mathbb{P}} + Q_{t-1}^{1/2} \eta_t, \quad \eta_t \stackrel{iid}{\sim} Normal(0, I), \quad (13)$$

where Q_{t-1} denotes the (possibly) time-varying covariance matrix of η_t , with η_t independent of all past and future $(e_{\mathcal{P}t}^{\mathbb{P}}, e_{Ht}^{\mathbb{P}})$. \mathcal{RA} knows that $\psi_t^{\mathbb{P}}$ follows (13), but she does not observe the realized $\psi_t^{\mathbb{P}}$. Her Bayesian learning rule filters for $\psi_t^{\mathbb{P}}$ conditional on $(\Theta^{\mathbb{Q}}, \psi^r)$. Adopting a Gaussian prior on $\psi_0^{\mathbb{P}}$ leads to a posterior distribution for $\psi_t^{\mathbb{P}}$ that is also Gaussian, $\psi_t^{\mathbb{P}} | Z_1^t \sim Normal(\hat{\psi}_t^{\mathbb{P}}, P_t)$. In [Appendix C](#) we show that her posterior mean follows the recursion

$$\hat{\psi}_t^{\mathbb{P}} = \hat{\psi}_{t-1}^{\mathbb{P}} + R_t^{-1} \mathcal{X}'_{t-1} \Sigma_Z^{-1} (\mathcal{Y}_t - \mathcal{X}_{t-1} \hat{\psi}_{t-1}^{\mathbb{P}}), \quad (14)$$

which depends on the posterior variance P_t through $R_t^{-1} \equiv P_t - Q_t$, with R_t satisfying

$$R_t = (I - P_{t-1}^{-1} Q_{t-2}) R_{t-1} + \mathcal{X}'_{t-1} \Sigma_Z^{-1} \mathcal{X}_{t-1}. \quad (15)$$

This rule has a revealing interpretation within the class of *adaptive least-squares estimators* (*ALS*) of $\psi_t^{\mathbb{P}}$. We say that $\hat{\psi}_t^{\mathbb{P}}$ is an *ALS* estimator if there exists a sequence of scalars $\gamma_t > 0$

such that $\widehat{\psi}_t^{\mathbb{P}}$ can be expressed recursively as

$$\widehat{\psi}_t^{\mathbb{P}} = \widehat{\psi}_{t-1}^{\mathbb{P}} + R_t^{-1} \mathcal{X}'_{t-1} \Sigma_Z^{-1} (\mathcal{Y}_t - \mathcal{X}_{t-1} \widehat{\psi}_{t-1}^{\mathbb{P}}), \quad (16)$$

$$R_t = \gamma_{t-1} R_{t-1} + \mathcal{X}'_{t-1} \Sigma_Z^{-1} \mathcal{X}_{t-1}. \quad (17)$$

It follows immediately from (14) - (15) that the posterior mean in the Kalman filter used by \mathcal{RA} to update $\psi_t^{\mathbb{P}}$ can be represented as a generalized *ALS* estimator. Moreover, (15) reveals three special cases where the filtering underlying Bayesian learning reduces to an actual *ALS* estimator (that is, (15) reduces to (17)):¹⁸

B \downarrow ALS: Setting $P_{t-1}^{-1} Q_{t-2} = (1 - \delta_{t-1}) \cdot I$ ¹⁹ for some sequence of scalars $0 < \delta_t \leq 1$, $\widehat{\psi}_t$ becomes an *ALS* estimator of $\psi^{\mathbb{P}}$ with $\gamma_t = \delta_t$.

B \downarrow CGLS: Specializing further by setting $\delta_t = \delta$ to a constant leads to $\widehat{\psi}_t$ being a *constant gain least-squares (CGLS)* estimator of $\psi^{\mathbb{P}}$ with $\gamma = \delta$.

B \downarrow RLS: If the constant $\delta = 1$, then $\widehat{\psi}_t$ is the *recursive least-squares (RLS)* estimator of $\psi^{\mathbb{P}}$.

Among the insights that emerge from this construction is that a Bayesian agent whose learning rule specializes to the *RLS* estimator is not adaptive in the following potentially important sense. With $\gamma = 1$ it follows that $Q_t = 0$, so an agent following a *RLS* rule is learning about an unknown value of $\psi^{\mathbb{P}}$ that is presumed to be fixed over time. Consequently, sudden changes in market conditions that result in sharp movements in recent values of Z may have an imperceptible effect on $\widehat{\psi}_t^{\mathbb{P}}$ as updated by \mathcal{RA} . Indeed, in environments where the *ML* estimator converges to a constant for large T , an *RLS*-based \mathcal{RA} will be virtually non-adaptive on $\widehat{\psi}^{\mathbb{P}}$ to new information after a long training period.

A more adaptive rule, one that is more responsive to changes in the structure of the economy owing say to changes in government policies, is obtained by giving less weight to values of Z far in the past. Such down-weighting arises naturally when \mathcal{RA} 's learning specializes to Case **B \downarrow CGLS**. The constant-gain coefficient γ determines the “half-life” of the weight on past data. This follows from the observation that, conditional on $\Theta^{\mathbb{Q}}$, the first-order conditions to the likelihood function implied by Bayesian learning with *CGLS* updating (Appendix C) are identical to those of a likelihood with terms of the form $\gamma^t \epsilon_{Zt}^{\mathbb{P}} \Sigma_Z^{-1} \epsilon_{Zt}^{\mathbb{P}}$.²⁰

Expressions (12) and (13) are the measurement and transition equations in a Gaussian linear filter over the unknown parameters. Therefore, the distribution of \mathcal{Y}_{t+1} conditional

¹⁸See McCulloch (2007), and the references therein, for discussions of similar issues in a setting of univariate y_t and econometrically exogenous x_t .

¹⁹This condition can be obtained by recursively setting $Q_{t-1} = \frac{1}{\delta_t} (P_{t-1} - P_{t-1} x'_{t-1} \Omega_{t-1}^{-1} x_{t-1} P_{t-1})$.

²⁰The latter is the likelihood function of a naive learner who simply re-estimates the likelihood function of a fixed-parameter model every period using the latest data and with down weighting by γ^t .

on Z_1^t is distributed $f^{\mathbb{P}}(\mathcal{Y}_{t+1}|Z_1^t) = \text{Normal}(\mathcal{X}_t \hat{\psi}_t^{\mathbb{P}}, \Omega_t)$, with the one-step ahead forecast variance determined inductively by $\Omega_t = \mathcal{X}_t P_t \mathcal{X}_t' + \Sigma_Z$. The first term captures the uncertainty related to the unknown $\Theta^{\mathbb{P}}$ and the second term is the innovation variance of the state Z_t .

Throughout this construction the direct dependence Ω_t on Σ_Z is a consequence of \mathcal{RA} treating Σ_Z as known, not as an object to be learned. This is an admittedly strong assumption as, empirically, \mathcal{RA} 's learning rule shows sizable revisions in Σ_Z . Though revisions in $\Sigma_{\mathcal{PP}}$ through learning would be largely inconsequential for pricing (convexity effects are small), they could be material for how \mathcal{RA} updates beliefs about $\Theta^{\mathbb{P}}$. In [Section 7](#) we evaluate the robustness of our core findings to the introduction of learning about the structure of the conditional covariances of Z in a model with time-varying second moments.

2.5 Empirical Learning Rules

As a benchmark case, we estimate a three-factor *DTSM* in which \mathcal{RA} learns by conditioning on past information on \mathcal{P} alone, rule $\ell(\mathcal{P})$; measures of cross-forecaster dispersion are omitted from Z . The parameters of the pricing distribution are normalized as in *JSZ*, and the coefficients on \mathcal{P}_t in (11) that determine the market prices of risk $\hat{\Lambda}$ depend on \mathcal{P}_1^t . Absent learning (if \mathcal{RA} observed the true historical dynamics of \mathcal{P}) this benchmark case would specialize to the case of affine $\hat{\Lambda}_t$ with (known) coefficients that are changing over time.

Rule $\ell(\mathcal{P}, H)$ preserves the same factor structure for pricing and extends rule $\ell(\mathcal{P})$ by allowing the market prices of \mathcal{P}_t to depend on (\mathcal{P}_t, H_t) . Including H_t in Z_t may lead to substantial gains in forecast accuracy when there are economically important effects of belief heterogeneity on bond prices beyond what is captured by \mathcal{P}_t . We compare the yield forecasts and risk premiums implied by rules $\ell(\mathcal{P})$ and $\ell(\mathcal{P}, H)$ to those of the median professional forecaster (whose rule $\ell(\text{BCFF})$ is unknown). We also examine the simple yield-based rule that has each zero yield following a random walk, rule $\ell(\text{RW})$.

Initially, in [Section 3](#), we explore the properties of rule $\ell(\mathcal{P})$ initialized using *ML* estimates for the “training” period January, 1961 through December, 1984. Then every month, up through March, 2011, as new data becomes available, \mathcal{RA} updates her posterior and the associated forecasts of future \mathcal{P} . Moving through the sample, *DTSM*-based rules impose the *JSZ* normalizations based on current-month information about yields and the weights determining the first three *PCs* (updated monthly using the current-month sample covariance matrix of the yields). When the dispersion information H is introduced in [Section 4](#) we use the more recent training period January, 1985 through December, 1994 owing to the limited availability of historical *BCFF* forecasts. To highlight when we are using the earlier training period for $\ell(\mathcal{P})$ we use the superscript L , $\ell^L(\mathcal{P})$ (“*L*”ong sample).

The *DTSM*-based rules are quite highly parametrized, especially $\ell(\mathcal{P}, H)$. For parsimony,

Rule	<i>DTSM</i>	State Vector	Restrictions	γ
$\ell(RW)$	No	Own Yield	N/A	N/A
$\ell(\mathcal{P})$	Yes	\mathcal{P}	No-Arbitrage <i>PC3</i> unpriced	1
$\ell_{CG}(\mathcal{P})$	Yes	\mathcal{P}	No-Arbitrage + <i>PC3</i> unpriced	0.99
$\ell(\mathcal{P}, H)$	Yes	(\mathcal{P}, H)	No-Arbitrage + 15 constraints	1
$\ell_{CG}(\mathcal{P}, H)$	Yes	(\mathcal{P}, H)	No-Arbitrage + 15 constraints	0.99

Table 3: Description of models used as learning rules.

which is relevant for the subsequent out-of-sample assessments, the parameters governing the market prices of the risks \mathcal{P} are set to zero if their marginal significance level *during the training period* is larger than 0.3. As a consequence, the learning rules are not nested across conditioning information sets because both the information and constraints typically differ. Since $K_{\mathcal{P}\mathcal{P}}^{\mathbb{Q}}$ is presumed to be known by \mathcal{RA} , constraints on $\Lambda_{\mathcal{P}}$ effectively transfer *a priori* knowledge of $\lambda^{\mathbb{Q}}$ to (some) knowledge about $K_Z^{\mathbb{P}}$. All constraints on $\Lambda_{\mathcal{P}}$ selected during the training period are maintained throughout the remainder of the sample period.

From the fitted *DTSM* at date t , an h -period ahead forecast of Z is given by

$$\hat{Z}_{t+h} = \hat{K}_{0t}^{\mathbb{P}} + \left(\hat{K}_{Zt}^{\mathbb{P}}\right) \hat{K}_{0t}^{\mathbb{P}} + \dots + \left(\hat{K}_{Zt}^{\mathbb{P}}\right)^{h-1} \hat{K}_{0t}^{\mathbb{P}} + \left(\hat{K}_{Zt}^{\mathbb{P}}\right)^h Z_t. \quad (18)$$

This leads directly to the h -period ahead forecasts of yields:

$$\hat{y}_{t+h}^m = A_m \left(\tilde{K}_0^{\mathbb{Q}}, \tilde{K}_{\mathcal{P}\mathcal{P}}^{\mathbb{Q}}, \tilde{\Sigma}_{\mathcal{P}\mathcal{P}} \right) + B_m \left(\tilde{K}_{\mathcal{P}\mathcal{P}}^{\mathbb{Q}} \right) \hat{\mathcal{P}}_{t+h}, \quad (19)$$

where the tildes indicate maximum likelihood estimators as of the forecast date. If h is the last month in a quarter, then the average expected yield over the quarter is:

$$\hat{y}_{t+h:t+h-3}^m = \frac{1}{3} \sum_{i=1}^3 \hat{y}_{t+h-i}^m, \quad (20)$$

which is the construct that is comparable to the median of the forecasts reported by BCFF professionals. A summary of the learning rules based on Cases $\mathcal{B}\downarrow\mathbf{RLS}$ and $\mathcal{B}\downarrow\mathbf{CGLS}$ that are implemented empirically is displayed in [Table 3](#).

To determine the best replication of the implicit rule followed by BCFF professionals using our *DTSM*, we searched for the value of γ that gives the best match of the constant-gain rule

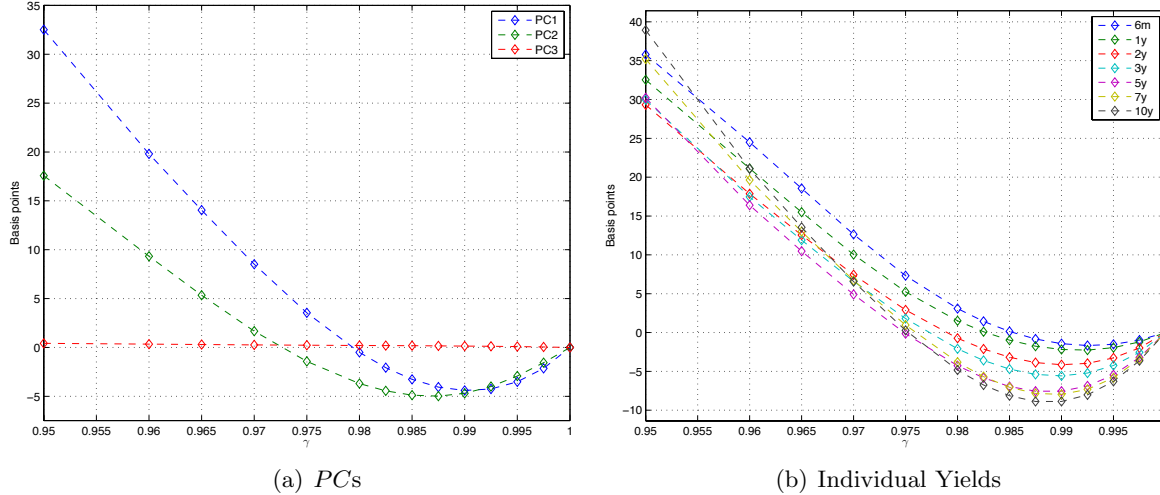


Figure 3: Relative RMSE's for one-year ahead forecasts based on rules $\ell(BCFF)$ and $\ell_{CG}(\mathcal{P}, H)$, from January, 1995 through March, 2011, as γ varies between 0.95 and 1.0.

$\ell_{CG}(\mathcal{P})$ to $\ell(BCFF)$. For each γ in the range $[0.93, 1.00]$ we computed the RMSE's of the differences in the errors in forecasting $PC1$ and individual bond yields one year ahead. In both cases $\gamma = 1.0$ (RLS updating) produces the smallest RMSE. This is true for the entire sample between 1985 and 2011, and also for the subsamples of 1985-1999 and 2000-2011. The unsubscribed rules in Table 3 set $\gamma = 1$.

Also of interest is whether there are rules with $\gamma < 1$ that systematically outperform rule $\ell(BCFF)$ in terms of RMSE's of forecasts. To assess this we search for the value of γ that gives the best relative performance of (largest negative difference between the RMSE's of) $\ell_{CG}(\mathcal{P}, H)$ over $\ell(BCFF)$). Figure 3 shows that best relative outperformance of $\ell_{CG}(\mathcal{P}, H)$ for forecasting both $PC1$ and the individual yields over a one-year horizon is approximately $\gamma = 0.99$, and henceforth $\ell_{CG}(\mathcal{P})$ and $\ell_{CG}(\mathcal{P}, H)$ refer to this choice. Interestingly, \mathcal{RA} 's outperformance using a CG rule is larger when conditioning on (\mathcal{P}, H) than on \mathcal{P} alone.

3 Learning from Information in the Yield Curve: Rule $\ell^L(\mathcal{P})$

As discussed in Section 2.2, agents can reasonably be modeled as knowing λ^Q . Nevertheless, in setting out to explore the properties of $\ell^L(\mathcal{P})$ we assume that \mathcal{RA} follows a Bayesian rule for updating Θ_t^P that treats Θ^Q as known and fixed, but then in fact Θ^Q is updated every month using the likelihood function (11). The time-series of real-time estimates of λ^Q from this learning scheme are displayed in Figure 4 for $\ell_{CG}^L(\mathcal{P})$ (the patterns for $\ell^L(\mathcal{P})$ are nearly identical). Notably, \mathcal{RA} holds λ^Q virtually fixed over the entire sample, consistent with the

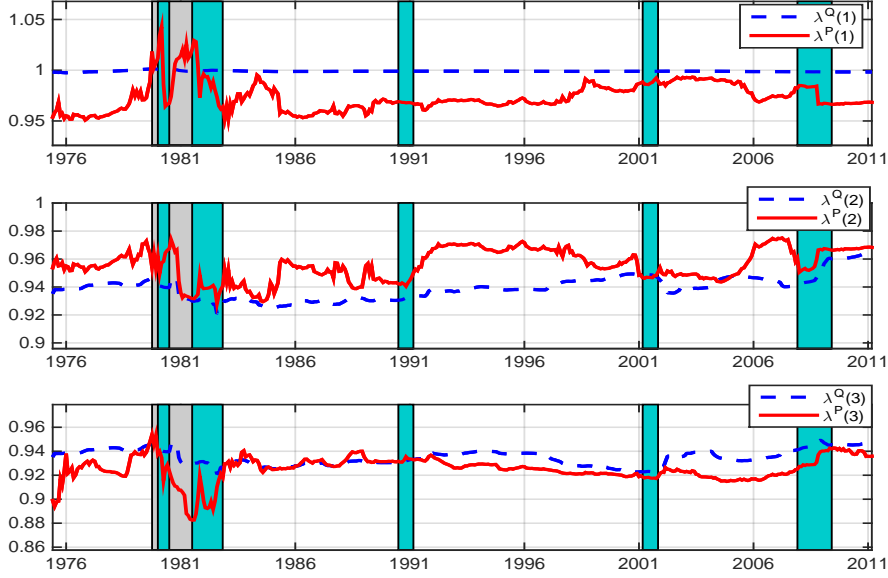


Figure 4: Estimates from model $\ell_{CG}^L(\mathcal{P})$ of the eigenvalues λ^Q (λ^P) of the feedback matrix K_1^Q (K_1^P) governing the persistence in \mathcal{P} . The estimates at date t are based on the historical data up to observation t , over the period July, 1975 to March, 2012.

presumption of her Bayesian rule for learning about Θ_t^P .²¹ Indeed, repeating our learning exercise with λ^Q fixed from the initial training period onward has a very small effect on the properties of the rule-implied prices or forecasts.

Pursuing this insight, if λ^Q is known and fixed over time, then so are the loadings B_m on \mathcal{P} in the affine pricing expression (5). Combining this with the fact that \mathcal{P}_t is measured with negligible error, the state-dependent components of bond yields that emerge from (5) with learning take the same form $B_m(\lambda^Q)\mathcal{P}_t$, just as in a *DTSM* without learning. Furthermore, agents will use fixed “hedge ratios” over time to manage the risks of their bond portfolios.²²

Views about the objective feedback matrix $K_{Z,t}^P$ do change over time and, thereby, learning can have a large effect on risk premiums (see Section 5). The largest changes in the eigenvalues λ_t^P of $K_{Z,t}^P$ (Figure 4) occur during the Fed experiment in the early 1980’s (when λ^Q remained remarkably stable). Even outside of this turbulent period there is substantial drift in λ_{1t}^P and λ_{2t}^P , the eigenvalues associated with the dominant first two *PCs* of yields. Whereas λ^Q is very similar across rules, λ_t^P for $\ell_{CG}^L(\mathcal{P})$ shows more temporal variation than for $\ell^L(\mathcal{P})$ owing to

²¹In all subsequent figures, light green shaded areas correspond to NBER recessions, and the grey shaded area corresponds to the period when the Federal Reserve followed a rule targeting the monetary base.

²²The estimates of the weights that define \mathcal{P} are also stable over time.

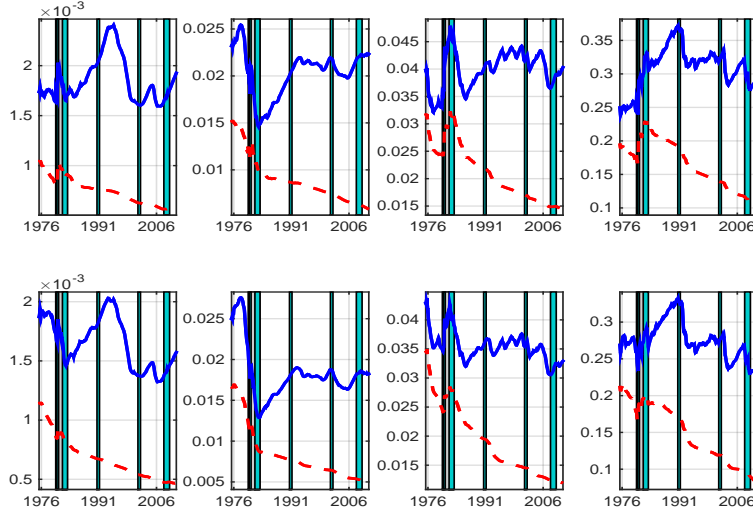


Figure 5: Posterior standard deviations of the first two rows of $[K_{0t}^{\mathbb{P}}, K_{1t}^{\mathbb{P}}]$ implied by the constant-gain ($\ell_{CG}^L(\mathcal{P})$, solid) and recursive least-squares ($\ell^L(\mathcal{P})$, dashed) learning rules.

the former's down-weighting of historical data.

An interesting perspective on \mathcal{RA} 's confidence in forecasts based on $\mathcal{M}^{\mathbb{B}}$ comes from examining the standard deviations of her posterior distributions of the non-zero entries in $[K_{0t}^{\mathbb{P}}, K_{\mathcal{PP},t}^{\mathbb{P}}]$ implied by rules $\ell_{CG}^L(\mathcal{P})$ and $\ell^L(\mathcal{P})$ (see Figure 5). For $\ell_{CG}^L(\mathcal{P})$, confidence in the fitted $\Theta^{\mathbb{P}}$ is low during the FRB experiment, but afterwards tends to fluctuate around a constant mean value. As shown above, $\ell^L(\mathcal{P})$ presumes—quite likely falsely—that $\Theta^{\mathbb{P}}$ is a constant that must be learned over time. An *RLS*-based \mathcal{RA} following $\ell^L(\mathcal{P})$ becomes increasingly confident about her knowledge of $\psi^{\mathbb{P}}$ as more data becomes available for estimation.

The relative accuracies of the rule-based forecasts, which depend primarily on the selected $\hat{\Theta}_t^{\mathbb{P}}$, can be assessed from the RMSE's displayed in Table 4. Below each RMSE are Diebold and Mariano (1995) (D-M) statistics for assessing whether two RMSE's are statistically the same, calculated as extended by Harvey, Leybourne, and Newbold (1997).²³ The first (in parentheses) tests against $\ell(RW)$, and the second (in brackets) tests against $\ell(BCFF)$. Both

²³Consider two series of forecast errors e_{1t} and e_{2t} , $t = \{1, 2, \dots, T\}$, define $d_t \equiv e_{1t}^2 - e_{2t}^2$, and let

$$\hat{\mu}_d = \frac{1}{T} \sum_{t=1}^T d_t \quad \text{and} \quad \hat{V}_d = \sum_{t=1}^T (d_t - \hat{\mu}_d)^2 + 2 \sum_{j=1}^h k(j/h) \sum_{t=1}^{T-j} (d_t - \hat{\mu}_d)(d_{t+j} - \hat{\mu}_d),$$

where $k(\cdot)$ is a Bartlett kernel that down-weights past lags to ensure that the variance of the difference in mean squared errors stays positive. The number of lags h is set to three for the one-quarter ahead forecasts and to twelve for the four-quarters ahead forecasts. Then the D-M statistic is equal to $\sqrt{T} \hat{\mu}_d / \hat{V}_d^{1/2}$.

Rule	RMSE's (in basis points) for Quarterly Horizon						
	6m	1Y	2Y	3Y	5Y	7Y	10Y
$\ell(RW)$	38.0	41.1	43.3	43.7	42.4	41.1	37.5
$\ell(BCFF)$	51.4 (\emptyset) [4.10]	51.6 (\emptyset) [3.28]	52.4 (\emptyset) [4.48]	54.3 (\emptyset) [5.03]	49.5 (\emptyset) [4.86]	47.9 (\emptyset) [3.40]	44.8 (\emptyset) [3.54]
$\ell^L(\mathcal{P})$	39.4 (-4.06) [1.74]	41.8 (-3.06) [0.77]	45.2 (-3.91) [2.83]	44.5 (-5.31) [1.15]	42.7 (-4.59) [0.36]	41.1 (-3.97) [-0.07]	37.8 (-3.31) [0.38]
$\ell_{CG}^L(\mathcal{P})$	38.4 (-4.29) [0.48]	41.7 (-3.11) [0.60]	45.3 (-3.70) [3.83]	44.9 (-4.54) [1.89]	43.3 (-4.21) [1.18]	42.0 (-3.61) [1.19]	39.0 (-2.91) [2.16]

Table 4: RMSE's for one-quarter ahead forecasts, January, 1985 to March, 2012. The D-M statistics for the differences between the *DTSM*- and *BCFF*-implied (*DTSM*- and *RW*-implied) forecasts are given in parentheses (brackets).

DTSM-based rules are statistically more accurate than the median professional $\ell(BCFF)$, while being very similar to $\ell(RW)$. The differences (in basis points) are largest for forecasting the short end of the yield curve. This outperformance of $\ell^L(\mathcal{P})$ over $\ell^L(BCFF)$ is notable given that the $\ell^L(\mathcal{P})$ learner is conditioning only on ex ante information about \mathcal{P}_t , whereas professionals can potentially use a wider variety of economic data. The RMSE's for rule $\ell_{CG}^L(\mathcal{P})$ are essentially the same as those for $\ell^L(\mathcal{P})$ over a quarterly forecast horizon. We examine longer horizons in the next section.

Not surprisingly, for this long sample period, the largest and most frequent revisions in $(K_{\mathcal{P}\mathcal{P},t}^{\mathbb{P}}, \Sigma_{\mathcal{P}\mathcal{P}})$ occurred during the FRB experiment of 1979 - 1982 when monetary policy induced a sharp rise in interest rates and flattening of the Treasury curve. [Figure 6](#) displays the views about the conditional standard deviations (diagonal) and correlations (off-diagonal) of the risk factors \mathcal{P} and [Figure 7](#) shows the revisions in $K_{\mathcal{P}\mathcal{P},t}^{\mathbb{P}}$. There is a notable and sharp increase in the perceived innovation variances of the yield *PC*s at the beginning of the FRB experiment. The perceived risks of *PC1* and *PC3* gradually decline after the announced return to an interest-rate targeting rule, while the perceived risk of the slope factor *PC2* remained high until 1985. Concurrently, $\mathcal{R}\mathcal{A}$'s views about the mean reversion parameters $K_{\mathcal{P}\mathcal{P},t}^{\mathbb{P}}$ were quite unstable. Increases in $\mathcal{R}\mathcal{A}$'s perceived volatilities of bond yields at the beginning of the FRB experiment come about through increases in both her perceived conditional variance of *PC1* and its degree of persistence (as measured by the (1, 1) element of $K_{\mathcal{P}\mathcal{P},t}^{\mathbb{P}}$). Similarly, $\lambda_{1t}^{\mathbb{P}}$ frequently exceeded unity during the FRB experiment ([Figure 4](#)).

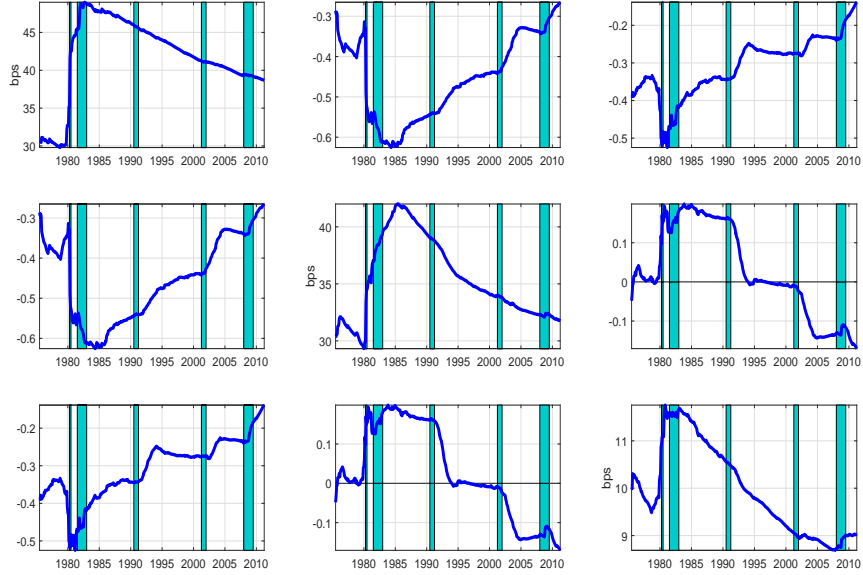


Figure 6: Conditional standard deviations on the diagonal and correlations off the diagonal implied by $\Sigma_{\mathcal{P}\mathcal{P}}$ from rule $\ell_{CG}^L(\mathcal{P})$ over the period June, 1975 to March, 2011.

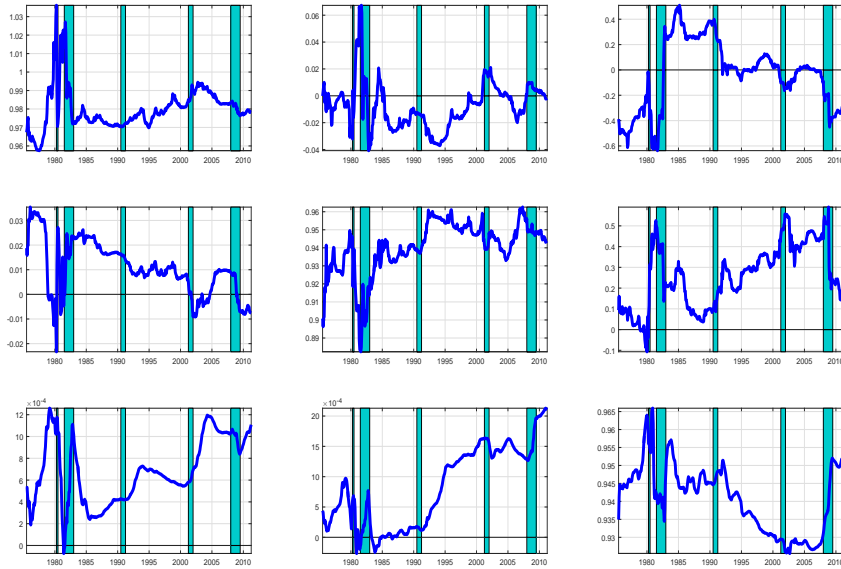


Figure 7: Posterior means of $K_{\mathcal{P}\mathcal{P},t}^{\mathbb{P}}$ implied by rule $\ell_{CG}^L(\mathcal{P})$.

Rule	RMSE's (in basis points) for Annual Horizon						
	6m	1Y	2Y	3Y	5Y	7Y	10Y
$\ell(RW)$	131.8 (-0.93) []	127.8 (-0.75) []	114.4 (-1.73) []	103.7 (-2.31) []	91.0 (-2.62) []	81.4 (-2.76) []	69.8 (-2.66) []
$\ell(BCFF)$	142.2 (0) [0.93]	136.6 (0) [0.75]	133.4 (0) [1.73]	130.8 (0) [2.31]	110.6 (0) [2.62]	100.4 (0) [2.76]	83.6 (0) [2.66]
$\ell_{CG}(\mathcal{P})$	124.2 (-1.32) [-1.03]	120.3 (-1.35) [-0.98]	112.5 (-1.89) [-0.31]	103.6 (-2.42) [-0.01]	93.9 (-2.39) [0.50]	85.6 (-2.31) [0.87]	76.6 (-1.39) [1.82]
$\ell_{CG}(\mathcal{P}, H)$	119.8 (-1.45) [-1.45]	115.1 (-1.54) [-1.56]	104.5 (-2.20) [-1.69]	93.0 (-2.80) [-1.70]	79.9 (-3.26) [-1.81]	69.7 (-3.72) [-2.29]	59.8 (-3.45) [-2.95]
$\ell(\mathcal{P}, H)$	119.8 (-1.67) [-2.06]	116.1 (-1.75) [-2.04]	107.1 (-2.53) [-1.81]	96.7 (-3.09) [-1.42]	85.3 (-3.39) [-1.00]	76.0 (-3.46) [-1.01]	67.8 (-2.49) [-0.41]

Table 5: RMSE's for one-year ahead forecasts, January, 1995 to March, 2011. The D-M statistics for the differences between the *DTSM*- and *BCFF*-implied (*DTSM*- and *RW*-implied) forecasts are given in parentheses (brackets).

4 Dispersion of Beliefs and Learning

We turn next to learning based on the expanded information set (\mathcal{P}, H) . There are three notable patterns in the RMSE's in Table 5 for the one-year forecasts of individual bond yields. First, conditioning on our measures of belief dispersion leads to a substantial improvement in forecast accuracy relative to the *DTSM*-based rules that condition only on \mathcal{P} . For the $\mathcal{B}\downarrow\text{CGLS}$ learner this pickup in accuracy occurs across the maturity spectrum with a slight tendency for larger gains at the long end of the Treasury curve. Second, $\ell_{CG}(\mathcal{P}, H)$ outperforms the random-walk rule to a statistically significant degree, and this outperformance is particularly strong for long-maturity bonds. Third, the median professional was the least accurate forecaster, relative to $\ell(RW)$ and especially relative to $\mathcal{R}\mathcal{A}$'s rule $\mathcal{B}\downarrow\text{CGLS}$. This underperformance of $\ell(BCFF)$ is statistically large beyond five-year maturities.

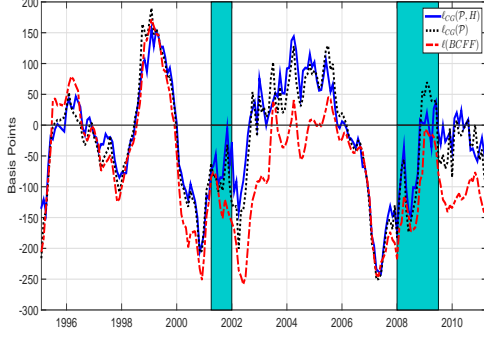
Table 6 provides a more nuanced view of the relative one-year forecast accuracies across sub-periods. Conditioning on H gives a sizable pickup in accuracy for the longer maturities in all three sub-periods. Interestingly, the most challenging period to forecast shorter term Treasury yields one year ahead was the early 2000's leading up to the global financial crisis. The outperformance of $\ell_{CG}(\mathcal{P}, H)$ relative to both $\ell(RW)$ and $\ell(BCFF)$ was especially large during this period. The poor relative performance of $\ell(BCFF)$ in forecasting the two- to three-year segment of the Treasury curve is interesting in light of the findings of Fleming and Remolona (1999) and Piazzesi (2005) that this segment of the yield curve shows the largest responses to surprise macroeconomic announcements.

Rule	RMSE's by Bond Maturity						
	6m	1Y	2Y	3Y	5Y	7Y	10Y
January, 1995 – December, 2000							
$\ell(RW)$	128	130	119	108	100	94	84
$\ell(BCFF)$	136	131	125	113	104	95	85
$\ell_{CG}(\mathcal{P})$	117	118	110	100	95	91	86
$\ell_{CG}(\mathcal{P}, H)$	111	112	101	89	81	76	72
$\ell(\mathcal{P}, H)$	114	114	102	90	82	79	78
January, 2001 – December, 2007							
$\ell(RW)$	154	144	127	114	90	71	56
$\ell(BCFF)$	157	149	142	136	110	95	77
$\ell_{CG}(\mathcal{P})$	145	137	126	114	96	81	69
$\ell_{CG}(\mathcal{P}, H)$	142	134	121	106	83	65	47
$\ell(\mathcal{P}, H)$	140	134	124	110	90	73	57
January, 2008 – March, 2012							
$\ell(RW)$	75	75	67	67	76	78	69
$\ell(BCFF)$	116	118	129	148	122	119	94
$\ell_{CG}(\mathcal{P})$	82	79	81	84	88	84	75
$\ell_{CG}(\mathcal{P}, H)$	73	68	65	67	70	67	59
$\ell(\mathcal{P}, H)$	73	69	70	74	80	77	69

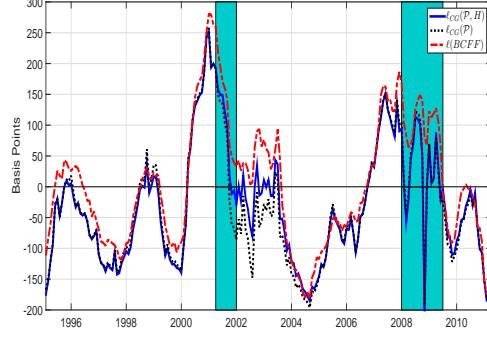
Table 6: RMSE's in basis points for one-year-ahead forecasts of individual bond yields over the indicated sample periods.

The portion of our sample covering the global financial crisis was a relatively easy period for forecasting Treasury yields. Nevertheless, again, $\ell(BCFF)$ underperformed, while $\ell_{CG}(\mathcal{P}, H)$ was the most accurate (especially for long-maturity bonds).

Figure 8 displays the errors in forecasting one-year ahead changes in the first two PC s of yields, which are largely governed by how agents learn about $\Theta^{\mathbb{P}}$ (conditioning on \mathcal{P} or (\mathcal{P}, H)). For the most part the higher forecasting power of $\ell_{CG}(\mathcal{P}, H)$ over $\ell_{CG}(\mathcal{P})$ arises from better forecasts of $PC1$, though $\ell_{CG}(\mathcal{P}, H)$ also outperforms for $PC2$ during the early 2000's. The $DTSM$ -based rules outperform $\ell(BCFF)$ in forecasting $PC1$, often by tens of basis points, especially following the dot-com bust and also during most of the recent financial crisis. The forecast accuracies of all three rules deteriorated during the periods of the dot-com bust and at the beginning of the crisis in 2007.



(a) Forecast Errors for $PC1$



(b) Forecast Errors for $PC2$

Figure 8: Errors from forecasting the twelve-month ahead change in $PC1$ and $PC2$ implied by $l(BCFF)$, $l_{CG}(\mathcal{P})$, and $l_{CG}(\mathcal{P}, H)$.

5 Bond Market Risk Premiums

Within reduced-form $DTSMs$, risk premiums have typically been computed under the presumption that market participants have full knowledge of the laws of motion of the risk factors (Duffee (2001), Dai and Singleton (2002)). Under real-time learning and drifting coefficients, there may be variation in $\mathcal{R}\mathcal{A}$'s expected excess returns induced by the learning process and drifting coefficients *per se* as contrasted with variation induced by movement in the underlying market prices of the risks \mathcal{P}_t . On top of this learning effect, a distinctive feature of our $DTSM$ is that measures of belief heterogeneity enter as unspanned risk factors.

Figure 9 displays the *expected* excess returns for one-year holding periods on ten-year bonds implied by $l(BCFF)$ and $l_{CG}^L(\mathcal{P})$ over our long sample period.²⁴ The premiums implied by these rules are strikingly different after every NBER recession in our sample. Key to understanding these differences is the strong positive correlation between the risk premium on ten-year bonds and the steepness of the yield curve. Precisely when the Treasury curve is relatively steep, the consensus professional forecaster believes that risk compensation is much lower than what is implied by our $DTSM$ -based learning rule. This finding is complementary to and distinct from Rudebusch and Williams (2009)'s finding that the slope of the yield curve gives more reliable forecasts of recessions than the one-year ahead recession probabilities from the Survey of Professional Forecasters. Our focus is on risk compensation in Treasury markets post-recessions as the U.S. economy entered recoveries.

Figure 10 shows the corresponding excess returns implied by $l_{CG}(\mathcal{P}, H)$ for the sample

²⁴We focus on yields on individual bonds so as to avoid the re-balancing and approximations involved in computing multi-period expected excess returns for the PC -mimicking portfolios.

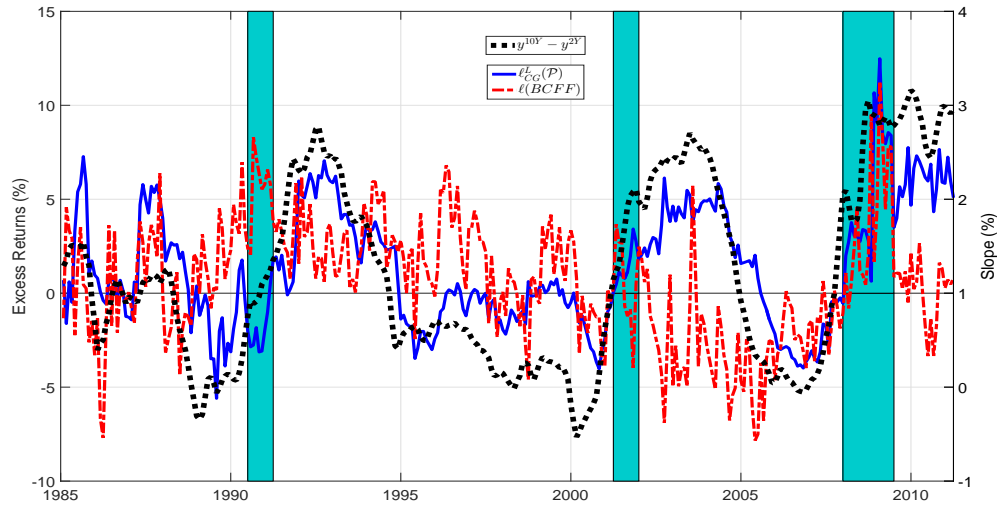


Figure 9: Average expected excess returns over holding periods of ten, eleven and twelve months for the ten-year bond based on $\ell_{CG}^L(\mathcal{P})$ and $\ell(BCFF)$ (left axis) and the slope of the Treasury curve measured as $y^{10y} - y^{2y}$ (right axis), January, 1985 to March, 2011.

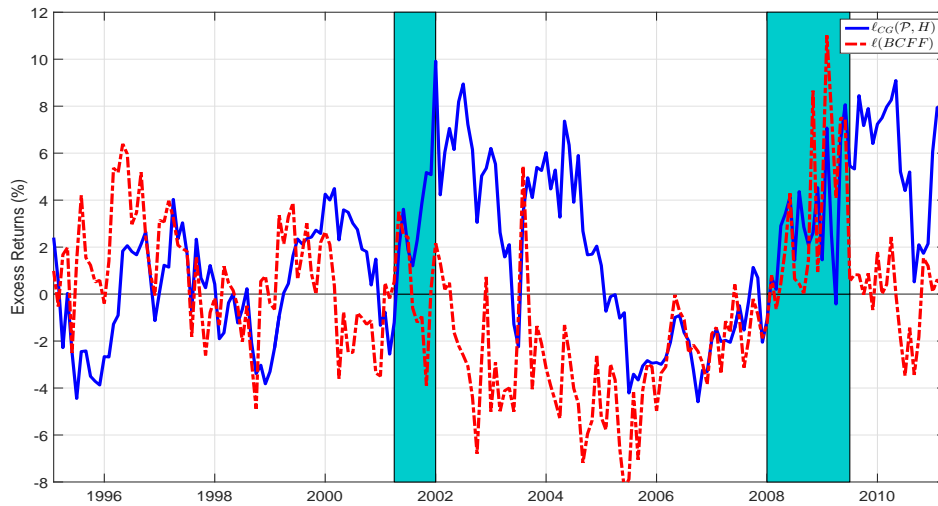


Figure 10: Average expected excess returns over holding periods of ten, eleven and twelve months for the ten-year bond based on $\ell_{CG}(\mathcal{P}, H)$ and $\ell(BCFF)$.

period for which dispersion information is available. Again there are large and systematic differences in risk premiums especially after recessions. To translate these findings into implications for views on the shape of the Treasury curve, consider the period from 2001 through 2005. Whereas the risk premiums implied by $(\ell_{CG}^L(\mathcal{P}), \ell_{CG}(\mathcal{P}, H))$ are highly correlated with the slope of the yield curve during this period (Figure 9), this is less so for $\ell(BCFF)$. The differences between the expected excess returns from $(\ell_{CG}^L(\mathcal{P}), \ell_{CG}(\mathcal{P}, H))$ and $\ell(BCFF)$ are attributable primarily to differences in forecasts of the ten-year yield coming out of recessions. Following the low (recession) levels of y^{10y} from late 2002 until 2004, the professional forecasters expected a much more rapid rise in y^{10y} than the *DTSM*-based learning rules. Therefore, $\ell(BCFF)$ implies relatively lower premiums for bearing ten-year price risk.

These differences underlie the greater accuracy of the *DTSM*-based rules. Over our sample the RMSE's in forecasting the realized excess returns on $(2y, 10y)$ bonds were (1.36%, 8.11%) for $\ell(BCFF)$, (1.20%, 7.24%) for $\ell_{CG}(\mathcal{P})$, and (1.15%, 5.73%) for $\ell_{CG}(\mathcal{P}, H)$ (see Table 7, Panel A). Rule $\ell_{CG}(\mathcal{P}, H)$ also outperforms $\ell_{CG}(\mathcal{P})$, with a notable 21% gain in accuracy on the ten-year bond. This improvement increases to 27% in the sub-period leading up to the crisis (Panel B). Moreover, $\ell_{CG}(\mathcal{P}, H)$ outperforms the random-walk based forecasts of excess returns in all sub-periods, with an especially large gain for the long-term bonds (over 14%).

We again stress that these improvements in forecast accuracy of $\ell_{CG}(\mathcal{P}, H)$ are obtained using a fully ex ante rule. Furthermore, $\ell_{CG}(\mathcal{P}, H)$ is well approximated by a recursively updated *VAR*-based rule (constant-gain, least-squares rule). As such it would have been fully available to market participants. Given that the median survey professional does not follow a comparably accurate learning rule, using median *BCFF* long-term bond forecasts to fit empirical learning rules would likely lead to distorted measures of required risk compensations.²⁵

Within our learning models there are two channels through which H can impact expected excess returns. The first is the direct effect that H_t has on forecasts of future *PC*s as components of Z_t in (18). The second is the indirect effect of H on how \mathcal{RA} updates the parameters $(\hat{K}_{0t}^{\mathcal{P}}, \hat{K}_{Zt}^{\mathcal{P}})$ by conditioning on H as part of the learning process. To frame this in terms of risk premiums, let $er_t^{10,1}(\mathcal{P}, H)$ ($er_t^{10,1}(\mathcal{P})$) denote \mathcal{RA} 's expected excess return from holding a ten-year bond for one year under rule $\ell_{CG}(\mathcal{P}, H)$ ($\ell_{CG}(\mathcal{P})$). Using (18) and the expression for realized excess returns in terms of y_{t+1y}^{10y} , we can write

$$er_t^{10,1}(\mathcal{P}, H) = \hat{a}_t^{\mathcal{P}, H} + \hat{b}_t^{\mathcal{P}, H} \mathcal{P}_t + \hat{c}_t H_t, \quad (21)$$

$$er_t^{10,1}(\mathcal{P}) = \hat{a}_t^{\mathcal{P}} + \hat{b}_t^{\mathcal{P}} \mathcal{P}_t. \quad (22)$$

²⁵Gains in forecast performance may come from using information embedded in survey forecasts of short-term rates and, indeed, Altavilla, Giacomini, and Ragusa (2014) present evidence consistent with this view.

Rule	Annual Excess Return RMSE's				
	2Y	3Y	5Y	7Y	10Y
Panel A: January, 1995 – March, 2011					
$\ell(RW)$	127.8 (-0.75) []	228.8 (-1.73) []	385.7 (-2.52) []	512.0 (-2.87) []	666.5 (-2.76) []
$\ell(BCFF)$	136.6 (0) [0.75]	266.9 (0) [1.73]	485.3 (0) [2.52]	637.8 (0) [2.87]	811.7 (0) [2.76]
$\ell_{CG}(\mathcal{P})$	120.3 (-1.35) [-0.98]	225.1 (-1.89) [-0.31]	390.4 (-2.53) [0.20]	541.8 (-2.39) [0.87]	723.7 (-1.82) [1.57]
$\ell_{CG}(\mathcal{P}, H)$	115.0 (-1.54) [-1.56]	208.9 (-2.20) [-1.69]	340.3 (-3.08) [-1.85]	451.45 (-3.49) [-1.69]	572.8 (-3.66) [-2.55]
Panel B: January, 2001 – December, 2007					
$\ell(RW)$	144	255	403	492	540
$\ell(BCFF)$	149	284	493	634	757
$\ell_{CG}(\mathcal{P})$	137	253	412	548	662
$\ell_{CG}(\mathcal{P}, H)$	134	242	370	461	483
Panel C: January, 2008 – March, 2011					
$\ell(RW)$	75	134	300	444	706
$\ell(BCFF)$	118	258	563	726	969
$\ell_{CG}(\mathcal{P})$	79	162	354	506	761
$\ell_{CG}(\mathcal{P}, H)$	68	131	285	399	615

Table 7: RMSE's in basis points for one-year excess returns, January, 1995 to March, 2011 and two sub periods. The D-M statistics for the differences between the *DTSM*- and BCFF-implied (*DTSM*- and *RW*-implied) forecasts are given in parentheses (brackets).

We separate the direct and indirect effects of H in (21) by estimating a rolling constant-gains least-squares projection of H_t onto \mathcal{P}_t :

$$H_t = \hat{\alpha}_t + \hat{\beta}_t \mathcal{P}_t + u_t, \quad (23)$$

and then construct the pseudo risk premium

$$er_t^{10,1}(\mathcal{P}, 0) = (\hat{a}_t^{\mathcal{P}, H} + \hat{c}_t \hat{\alpha}_t) + (\hat{b}_t^{\mathcal{P}, H} + \hat{c}_t \hat{\beta}_t) \mathcal{P}_t \equiv \hat{a}_t^{\mathcal{P}, 0} + \hat{b}_t^{\mathcal{P}, 0} \mathcal{P}_t. \quad (24)$$

The difference between $er_t^{10,1}(\mathcal{P}, 0)$ and $er_t^{10,1}(\mathcal{P})$ arises entirely from the effect that H has on the updating of the weights on \mathcal{P}_t when learning is conditioned on the full information generated by (\mathcal{P}, H) .²⁶

²⁶While it is true that (23) is fit outside of our *DTSM*, the $(\hat{\alpha}_t, \hat{\beta}_t)$ that we recover using monthly data

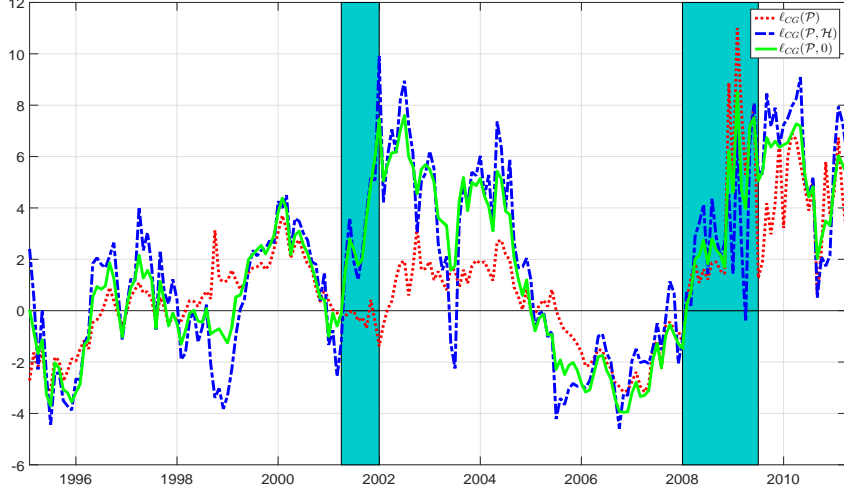


Figure 11: Comparing annual expected excess returns on the ten-year zero coupon bond in models $l_{CG}(\mathcal{P}, H)$, $l_{CG}(\mathcal{P}, 0)$, and $l_{CG}(\mathcal{P})$.

The close tracking of $er_t^{10,1}(\mathcal{P}, H)$ by $er_t^{10,1}(\mathcal{P}, 0)$ in Figure 11 implies that the vast majority of the impact of H on $er_t^{10,1}(\mathcal{P}, H)$ arises through its indirect effect on the weight on \mathcal{P}_t . That is, H primarily affects the broad cyclical patterns in risk premiums through $\mathcal{R}\mathcal{A}$'s updating of her views on how the current yield curve (\mathcal{P}_t) affects its future shape. Furthermore, the sizable differences between $er_t^{10,1}(\mathcal{P}, H)$ and the H -free construction $er_t^{10,1}(\mathcal{P})$ shows that conditioning on H had a material effect on measured risk premiums. This was especially the case following the two recessions covered by this sample period. Most of the indirect effect it through updating the forecasts of $PC1$ (see Figure 8(a)).

There is a small direct effect of H on risk premiums through the VAR for Z primarily around turning points (peaks and troughs) in premiums. Additionally, the gaps between $er_t^{10,1}(\mathcal{P}, H)$ and $er_t^{10,1}(\mathcal{P}, 0)$ were large in late 1998 following the start of the Asian crisis and the default by Russia, in early 2003 when the U.S. and its allies launched an invasion into Iraq, and again in early 2009 when the Treasury was actively purchasing preferred stocks of large U.S. banks and the Federal Reserve launched the Term Asset-Backed Securities Loan Facility. Bond yields fell dramatically in all three periods owing to flights to quality.

Digging deeper into the role of H , Figure 12(a) displays the time-series of the loadings $\hat{b}_{1t}^{\mathcal{P}}$ on $PC1$ associated with rule $l_{CG}(\mathcal{P})$ against the loadings $\hat{b}_t^{\mathcal{P},0}$ for the forecast $er_t^{10,1}(\mathcal{P}, 0)$ from

would be literally identical to those recovered within a $DTSM$ without constraints on the market prices of risk. This is an immediate implication of the propositions in JSZ. Therefore, we believe we are obtaining a reliable picture of the impact of H on the loadings on \mathcal{P} in the expression for $er_t^{10,1}(\mathcal{P}, 0)$.

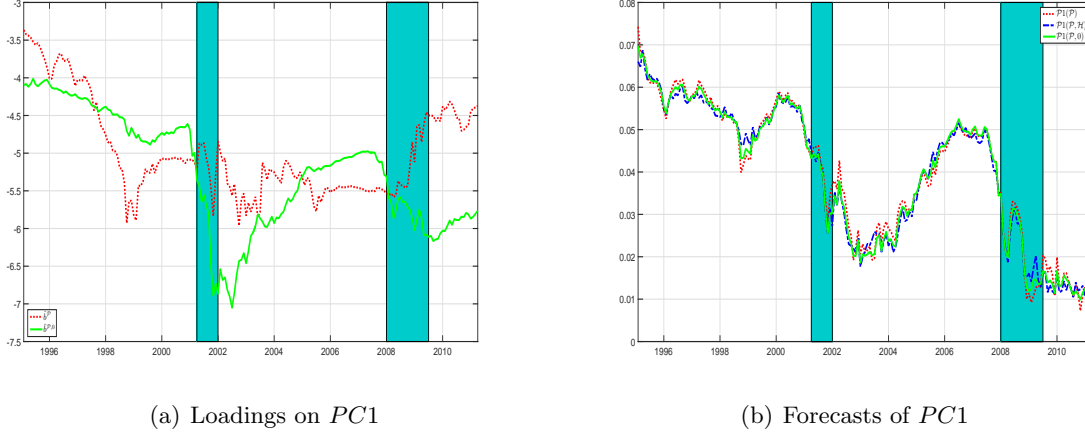


Figure 12: The left figure displays the loading on \mathcal{P}_1 from vector $\hat{b}^{\mathcal{P},0}$ in (24) against the loading on \mathcal{P}_1 from $\hat{b}^{\mathcal{P}}$ for rule $\ell_{CG}(\mathcal{P})$ from (22). The right figure compares the four-quarters ahead forecasts of \mathcal{P}_1 based on rules $\ell_{CG}(\mathcal{P})$ ($\mathcal{P}1(\mathcal{P})$) and $\ell_{CG}(\mathcal{P}, H)$ ($\mathcal{P}1(\mathcal{P}, H)$). $\mathcal{P}1(\mathcal{P}, 0)$ uses the PC loadings from $\ell_{CG}(\mathcal{P}, H)$ and the projection (23) analogously to (24).

(24). The loadings are very different during and exiting from recessions, when the loadings are more negative in $er_t^{10,1}(\mathcal{P}, 0)$. The indirect effect of H amplifies the effect of level on excess returns and makes them more cyclical. That is, \mathcal{RA} puts more weight on information about the level of the yield curve when updating during economic downturns.

We also compare the four-quarter ahead forecasts of the three PC s across the learning rules $\ell_{CG}(\mathcal{P})$, $\ell_{CG}(\mathcal{P}, H)$, and $\ell_{CG}(\mathcal{P}, 0)$. For each of $PC2$ and $PC3$ these forecasts are very similar, which suggests that conditioning on H is having little impact on how \mathcal{RA} forecasts these higher-order PC s. On the other hand, Figure 12(b) shows some sizable differences across rules for forecasting $PC1$. Thus, the (relatively small) direct effect of H on \mathcal{RA} 's forecasts is mostly reflected through the first principal component.

6 Belief Heterogeneity, Macroeconomic Risks, and Risk Premiums in Treasury Markets

Is the forecasting power of H distinct from the predictive power of macro factors for risk premiums in the Treasury market documented by Ludvigson and Ng (2010) and Joslin, Pribsch, and Singleton (2014) (JPS)? Might conditioning on output growth or inflation, in addition to H , materially improve the performance of \mathcal{RA} 's real-time learning rules?

To address this issue we constructed measures of inflation (INF) and real economic activity (REA) that market participants would have known in real time. Starting from the Archival

Federal Reserve Economic Data (ALFRED) database that reports the original releases of macroeconomic series, this data was updated and rescaled as more information became available. Letting $x_{s|t}$ denote an economic statistic indexed to time s and available at time $t \geq s$, and recognizing that most economic statistics are released with a one-month delay, an investor at time t can typically condition on

$$x_{t_0-1|t}, x_{t_0|t}, \dots, x_{t-2|t}, x_{t-1|t}$$

where t_0 indicates the start of the training sample. Importantly, this is the fully updated series through time t , and not the series as it was released in real time.²⁷ *INF* is the twelve-month log difference of the Consumer price index for all urban consumers that is available at the time of estimation. *REA* is the three-month moving average of the first principal component of six series related to real economic activity.²⁸

Recall that the learning rules are not nested across conditioning information sets because both the information and constraints typically differ. Under rule $\ell_{CG}(\mathcal{P}, REA, INF)$, $\mathcal{R}A$ updates her beliefs about the \mathbb{P} distribution of yields conditional on (\mathcal{P}, REA, INF) . Thus, it is the learning counterpart to the full-sample, fixed-parameter analysis in JPS.

Root mean-squared forecast errors for excess returns over one-year holding periods are displayed in Table 8. Focusing first on the long end of the yield curve (seven- and ten-year), our core learning rule $\ell_{CG}(\mathcal{P}, H)$ outperforms all of the other learning rules, including $\ell_{CG}(\mathcal{P}, REA, INF)$. Moreover, adding (INF, REA) to our core rule, which gives the most comprehensive rule $\ell_{CG}(\mathcal{P}, H, REA, INF)$, leads to a substantial *deterioration* in real-time out-of-sample accuracy. The higher accuracy of $\ell_{CG}(\mathcal{P}, H)$ is particularly striking for the post-2000 sample period, both pre-crisis and during the crisis.

Interestingly, the rule $\ell_{CG}(\mathcal{P}, REA)$ that conditions only on the yield *PCs* and real activity delivers the same out of sample forecasting performance of the rule $\ell_{CG}(\mathcal{P}, REA, INF)$. Moreover, both “macro rules” add little to the forecast accuracy of the basic yields-only rule $\ell_{CG}(\mathcal{P})$, other than during the subsample 1995-2000 (not shown). Information about real economic activity does improve learning for the short end of the yield curve (under two years) during the post-2000 run-up to the financial crisis. However, during the crisis, $\ell_{CG}(\mathcal{P}, H)$ provided more accurate forecasts than all of the other learning rules across the entire maturity

²⁷Prior studies using original release data have not always updated their series through time t as we do (e.g., Ghysels, Horan, and Moench (2014)). Such studies are using stale data relative to what market participants knew at the time they constructed their forecasts.

²⁸The series are the difference in the logarithm of Industrial production index (INDPRO), the difference in the logarithm of total nonfarm payroll (PAYEMS), the difference of the civilian unemployment rate (UNRATE), the difference of the logarithm of “All employees: Durable goods” (DMANEMP), the difference of the logarithm of “All employees: Manufacturing” (MANEMP), and the difference of the logarithm of “All employees: NonDurable goods” (NDMANEMP). The first PC is smoothed similarly to the Chicago Fed National Activity Index.

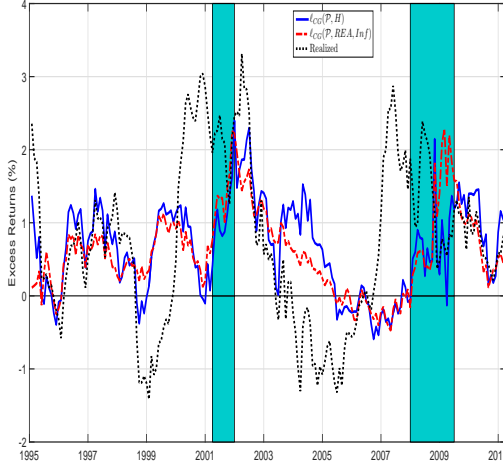
	6m	1Y	2Y	3Y	5Y	7Y	10Y	PC1	PC2	PC3
Panel A: January, 1995 – March, 2011										
$\ell_{CG}(\mathcal{P})$	124	120	113	104	94	86	77	96	102	15
$\ell_{CG}(\mathcal{P}, H)$	120	115	104	93	80	70	60	86	99	15
$\ell_{CG}(\mathcal{P}, REA)$	118	114	108	101	94	86	79	94	97	15
$\ell_{CG}(\mathcal{P}, REA, INF)$	118	114	108	101	94	86	79	93	97	15
$\ell_{CG}(\mathcal{P}, H, REA)$	117	112	104	95	86	77	69	88	97	15
$\ell_{CG}(\mathcal{P}, H, REA, INF)$	134	132	125	115	101	91	84	106	99	15
Panel B: January, 2001 – December, 2007										
$\ell_{CG}(\mathcal{P})$	145	137	126	114	96	81	69	105	108	16
$\ell_{CG}(\mathcal{P}, H)$	142	134	121	106	83	65	47	98	101	16
$\ell_{CG}(\mathcal{P}, REA)$	127	121	117	110	100	90	82	100	97	16
$\ell_{CG}(\mathcal{P}, H, REA)$	127	119	113	104	90	79	68	95	94	16
Panel C: January, 2008 – March, 2011										
$\ell_{CG}(\mathcal{P})$	82	79	81	84	88	84	75	73	91	16
$\ell_{CG}(\mathcal{P}, H)$	73	68	65	67	70	67	59	56	88	16
$\ell_{CG}(\mathcal{P}, REA)$	92	84	82	84	89	86	78	73	105	16
$\ell_{CG}(\mathcal{P}, H, REA)$	88	78	72	73	78	76	71	62	109	16

Table 8: RMSE’s for one-year ahead forecasts with conditioning on both disagreement and macroeconomic information.

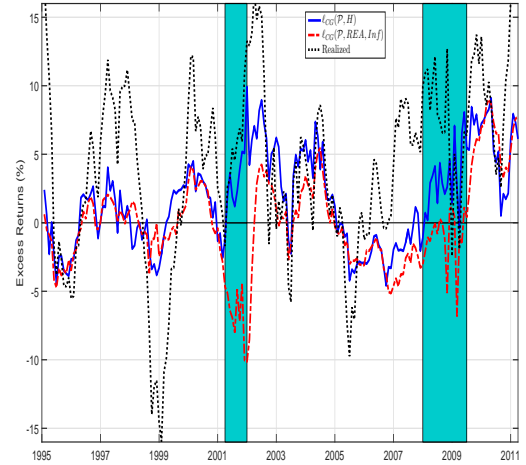
spectrum.

Elaborating, [Figure 13](#) displays the expected excess returns for one-year holding periods on two- (left) and ten-year (right) bonds for the rules $\ell_{CG}(\mathcal{P}, H)$ and $\ell_{CG}(\mathcal{P}, REA, INF)$, overlaid with the subsequent realized excess returns. For the ten-year bond there are notable differences in expected excess returns around turning points of the rates cycles. Especially leading up to and during recessions, $\ell_{CG}(\mathcal{P}, H)$ gives much more accurate forecasts than $\ell_{CG}(\mathcal{P}, REA, INF)$. These rules performed comparably well from late 2002 through early 2005, leading up to Greenspan’s “conundrum” remarks. However, overall and from a truly ex-ante perspective, the unspanned (by Treasury yields) information in H is more informative than unspanned information in (REA, INF) about future long-term bond yields.

At the two-year maturity point the implied risk premiums track each other quite closely with two notable exceptions. First, during the period from early 2003 to early 2005 conditioning on macro information led to more accurate forecasts of excess returns. On the other hand, during late 2008 and early 2009, in the midst of the financial crisis, $\ell_{CG}(\mathcal{P}, H)$ gave substantially more accurate forecasts. For the former sub-period, the outperformance of $\ell_{CG}(\mathcal{P}, REA, INF)$ is consistent with the findings of JPS, as this is the period when they found the largest gap



(a) Two-Year Bond



(b) Ten-Year Bond

Figure 13: Average expected excess returns over holding periods of ten, eleven and twelve months for the two- and ten-year bonds based on $\ell_{CG}(\mathcal{P}, H)$ and $\ell_{CG}(\mathcal{P}, REA, INF)$, overlaid with the following realized return.

between term premiums implied by (non-learning) models with and without conditioning on macro information. For the latter sub-period the dominance of $\ell_{CG}(\mathcal{P}, H)$ may in part reflect a flight-to-liquidity phenomenon as noted previously in regards to Figure 11.

Is H is informative also about the business cycle (the macroeconomy) or might its affects on forecasts arise from phenomena internal to the Treasury market? That is, does the component of H that predicts the level of the yield curve in our models have predictive power for real economic growth and inflation? To address this question we estimate the linear projection

$$\mathcal{M}_{t+12} = \alpha_t + \mathcal{B}_{\tilde{\mathcal{M}},t} \tilde{\mathcal{M}}_t + \mathcal{B}_{\mathcal{P},t} \mathcal{P}_{\mathcal{P}_t}^{f12} + \mathcal{B}_{(\mathcal{P},H),t} \left(\mathcal{P}_{(\mathcal{P},H)_t}^{f12} - \mathcal{P}_{\mathcal{P}_t}^{f12} \right) + v_{t+12}, \quad (25)$$

where \mathcal{M}_t denotes either INF or REA ; $\tilde{\mathcal{M}}_t$ denotes the current one-month inflation rate or our proxy for real activity;²⁹ and $\mathcal{P}_{\mathcal{P}_t}^{f12}$ and $\mathcal{P}_{(\mathcal{P},H)_t}^{f12}$ are the date t one-year ahead forecasts of $(PC1, PC2)$ according to the rules $\ell_{CG}(\mathcal{P})$ and $\ell_{CG}(\mathcal{P}, H)$, respectively. A statistically significant $\mathcal{B}_{(\mathcal{P},H),t}$ would suggest that information in H_t about the future shape of the yield curve is also informative about the future macroeconomy. The coefficients are updated monthly following the training period January 1985 through January 1995.

For the case of \mathcal{M}_t set to INF_t , the coefficients $\mathcal{B}_{(\mathcal{P},H),t}$ are statistically insignificant

²⁹That is, there is no temporal averaging in constructing $\tilde{\mathcal{M}}_t$. Replacing $\tilde{\mathcal{M}}_t$ with \mathcal{M}_t leads to a small deterioration in the $RMSE$'s of the forecasts.

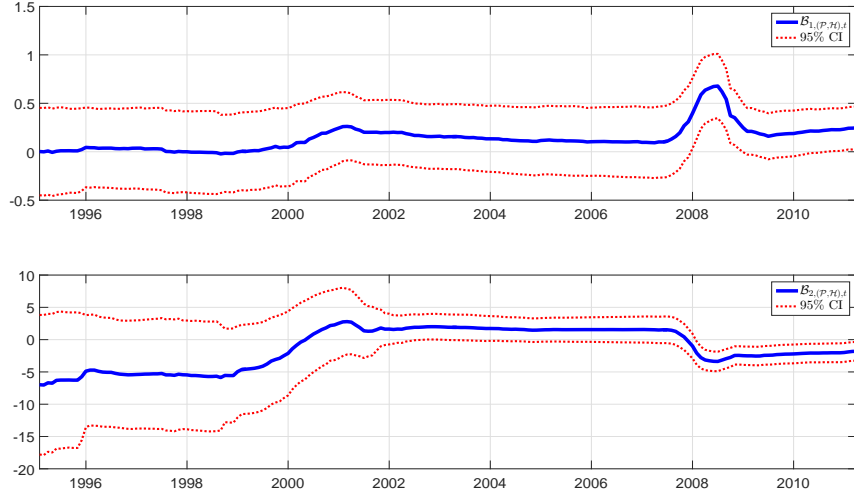


Figure 14: Time-series of fitted $\mathcal{B}_{(\mathcal{P},H),t}$ on $(\mathcal{P}_{(\mathcal{P},H)t}^{f12} - \mathcal{P}_{\mathcal{P}t}^{f12})$ in the forecast equation for REA_{t+12} . The upper solid line is for $PC1$ and the lower is for $PC2$. The dashes delineate two-standard error confidence intervals.

throughout our entire sample period. The analogous coefficients for \mathcal{M} set to REA are displayed in Figure 14. The coefficient on the difference in forecasts for $PC1$ remains near zero for the entire sample except during the depth of the financial crisis in 2008. The corresponding coefficients for $PC2$ are consistently negative during the second half of the 1990's, but they are not statistically significant. They are close to zero in the 2000's leading up to the crisis, and then turn significantly negative during 2008. Thus, only during the recent crisis, and arguably only during the depth of this crisis in 2008, is there evidence that the information in H about the future yield curve is also informative about future real economic activity.

Summarizing, for most sample periods and along the entire maturity spectrum, H has much more real-time predictive power for bond yields than the macro information in (INF, REA) . Moreover, the component of H that is informative about future bond yields is evidently largely uninformative about the future path of the macroeconomy, at least incrementally after conditioning on current inflation and real economic activity.

7 Robustness to Time-Varying Volatility

A less constrained Bayesian (relative to \mathcal{RA}) would formally build updating of $\Sigma_{\mathcal{PP}}$ (Figure 6) into her learning rule. A priori, we would not expect this generalization of our learning rules to materially affect \mathcal{RA} 's conditional forecasts of bond yields, our primary focus for modeling

risk premiums. Updating of $\Sigma_{\mathcal{P}\mathcal{P}}$ would only change the posterior conditional means indirectly through interactions with $\Theta^{\mathbb{Q}}$, passed onto the \mathbb{P} -feedback parameters by the restrictions on the market price of risk. In our current setting \mathcal{RA} keeps the \mathbb{Q} parameters $(k_{\infty}^{\mathbb{Q}}, \lambda^{\mathbb{Q}})$ nearly constant. Therefore, it seems unlikely that formally introducing learning about $\Sigma_{\mathcal{P}\mathcal{P}}$ would lead to large changes in the inferred posterior conditional \mathbb{P} -means of bond yields.

To provide further reassurance on this front, we proceed to investigate learning within a setting of stochastic volatility. Suppose there are three risk factors consisting of a univariate volatility factor V_t and a bivariate X_t that is Gaussian conditional on V_t . We adopt the following normalized just-identified representation of the state under \mathbb{Q} :

$$V_{t+1}|V_t \sim CAR(\rho^{\mathbb{Q}}, c^{\mathbb{Q}}, v^{\mathbb{Q}}), \quad (26)$$

$$X_{t+1} = K_V^{\mathbb{Q}}V_t + \text{diag}(\lambda^{\mathbb{Q}})X_t + \sqrt{\Sigma_{0X} + \Sigma_{1X}V_t}\varepsilon_t^{\mathbb{Q}}, \quad (27)$$

$$r_t = r_{\infty}^{\mathbb{Q}} + \rho_V V_t + 1'X_t, \quad (28)$$

where CAR denotes a compound autoregressive gamma process ([Gourieroux and Jasiak \(2006\)](#)) and $\Theta^{\mathbb{Q}} \equiv (r_{\infty}^{\mathbb{Q}}, \rho_V, \rho^{\mathbb{Q}}, c^{\mathbb{Q}}, v^{\mathbb{Q}}, K_V^{\mathbb{Q}}, \lambda^{\mathbb{Q}})$. As before, we assume that \mathcal{RA} treats $\Theta^{\mathbb{Q}}$ as constant and known, which implies that yields are given by

$$y_t = A(\Theta^{\mathbb{Q}}, \Sigma_{0X}, \Sigma_{1X}) + B_V(\Theta^{\mathbb{Q}}, \Sigma_{1X})V_t + B_X(\Theta^{\mathbb{Q}})X_t,$$

and the principal components $\mathcal{P}_t = W y_t$ are affine in (V_t, X_t) (see [Appendix F](#) for details). The market prices of risk are assumed to be such that, under \mathbb{P} , the state follows the process

$$V_{t+1}|V_t \sim CAR(\rho^{\mathbb{P}}, c^{\mathbb{P}}, v^{\mathbb{P}}), \quad (29)$$

$$X_{t+1} = K_{0t}^{\mathbb{P}} + K_{Vt}^{\mathbb{P}}V_t + K_{Xt}^{\mathbb{P}}X_t + \sqrt{\Sigma_{0X} + \Sigma_{1X}V_t}\varepsilon_{t+1}^{\mathbb{P}}, \quad (30)$$

where $\varepsilon_{t+1}^{\mathbb{P}}$ is independent of V_{t+1} and we let $\Theta_t^{\mathbb{P}} = (\rho^{\mathbb{P}}, c^{\mathbb{P}}, v^{\mathbb{P}}, K_{0t}^{\mathbb{P}}, K_{Vt}^{\mathbb{P}}, K_{Xt}^{\mathbb{P}})$. \mathcal{RA} presumes that the volatility parameters $(\rho^{\mathbb{Q}}, c^{\mathbb{Q}}, v^{\mathbb{Q}}, \Sigma_{0X}, \Sigma_{1X})$ are constant, while those governing the conditional means of X_t are unknown and drifting. In [Appendix F](#) we show that the conditional first moments of the principal components are given by

$$\mathbb{E}_t^{\mathbb{Q}}(\mathcal{P}_{t+1}) = K_{0\mathcal{P}}^{\mathbb{Q}} + K_{1\mathcal{P}}^{\mathbb{Q}}\mathcal{P}_t \quad \text{and} \quad \mathbb{E}_t^{\mathbb{P}}(\mathcal{P}_{t+1}) = K_{0\mathcal{P},t}^{\mathbb{P}} + K_{1\mathcal{P},t}^{\mathbb{P}}\mathcal{P}_t,$$

where $(K_{0\mathcal{P}}^{\mathbb{Q}}, K_{1\mathcal{P}}^{\mathbb{Q}}, K_{0\mathcal{P},t}^{\mathbb{P}}, K_{1\mathcal{P},t}^{\mathbb{P}})$ are known functions of $(\Theta^{\mathbb{Q}}, \Sigma_{X0}, \Sigma_{1X}, \Theta_t^{\mathbb{P}})$ from the rotation of $(V_t, X_t)'$ to \mathcal{P}_t . As before, a subset of the parameters in $[K_{0\mathcal{P},t}^{\mathbb{P}}, K_{1\mathcal{P},t}^{\mathbb{P}}]$ is constrained based on the training sample.

[Figure 15](#) plots the eigenvalues of the feedback matrices $K_{1\mathcal{P}}^{\mathbb{Q}}$ and $K_{1\mathcal{P},t}^{\mathbb{P}}$ from the perspective

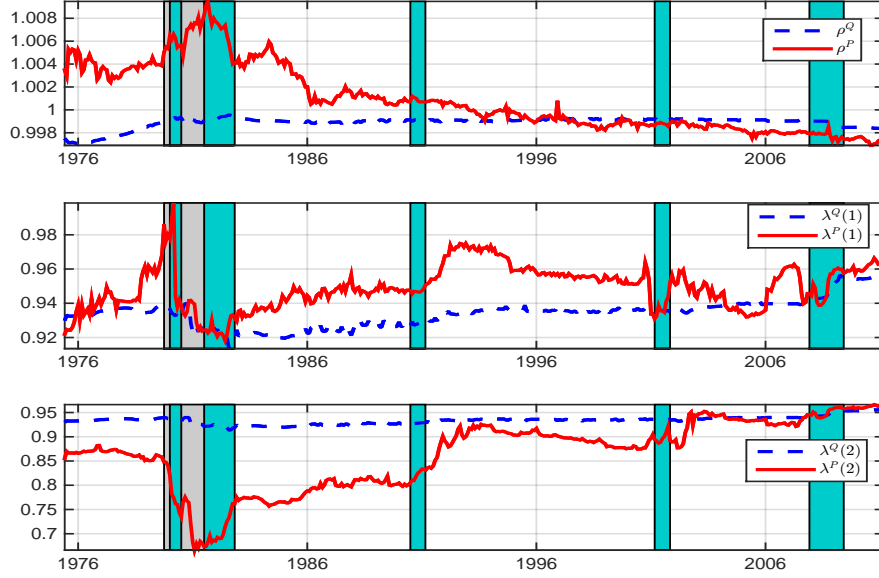


Figure 15: Estimates from model $\ell_{CG}^{1,3}(\mathcal{P})$ of the eigenvalues of the feedback matrix $K_{1\mathcal{P}}^{\mathbb{Q}}$ ($K_{1\mathcal{P},t}^{\mathbb{P}}$). The eigenvalues of $K_{1\mathcal{P}}^{\mathbb{Q}}$ are $(\rho^{\mathbb{Q}}, \lambda^{\mathbb{Q}})$ and the eigenvalues of $K_{1\mathcal{P},t}^{\mathbb{P}}$ are $(\rho^{\mathbb{P}}, \text{eig}(K_{X_t}^{\mathbb{P}}))$.

of \mathcal{RA} 's real-time learning rule in the presence of V_t and conditioning only on the history of the PC s. The eigenvalues of $K_{1\mathcal{P}}^{\mathbb{Q}}$ are $(\rho^{\mathbb{Q}}, \lambda^{\mathbb{Q}})$ and the eigenvalues of $K_{1\mathcal{P},t}^{\mathbb{P}}$ are $(\rho^{\mathbb{P}}, \text{eig}(K_{X_t}^{\mathbb{P}}))$.³⁰ Relaxing the assumption of constant conditional volatility as above does not alter our prior finding that the \mathbb{Q} eigenvalues are nearly constant over the entire sample period. The variation in the eigenvalues of $K_{1\mathcal{P},t}^{\mathbb{P}}$ reflects the substantial variation in the market prices of risk.

Figure 16 offers an interesting perspective on the degree to which the learning rule $\ell_{CG}(\mathcal{P})$ (that presumes constant $\Sigma_{\mathcal{P}\mathcal{P}}$) captures the swings in the conditional covariance matrix that would be perceived by an agent learning in the presence of stochastic volatility. On the diagonal are the estimated conditional standard deviations from models both with and without stochastic volatility. Rule $\ell_{CG}(\mathcal{P})$ captures the overall evolution of the conditional standard deviations, but fails to pick up the huge increment in volatilities during the Fed experiment. Perceptions about volatility under $\ell_{CG}(\mathcal{P})$ also decay relatively slowly during the great moderation. The constant conditional correlations are updated by $\ell_{CG}(\mathcal{P})$ in a manner very similar to the learning rule for the stochastic volatility model.

³⁰The feedback matrices in the conditional first moments of \mathcal{P}_t and $(V_t, X_t)'$ will have equal eigenvalues, as \mathcal{P}_t is an affine function of $(V_t, X_t)'$.

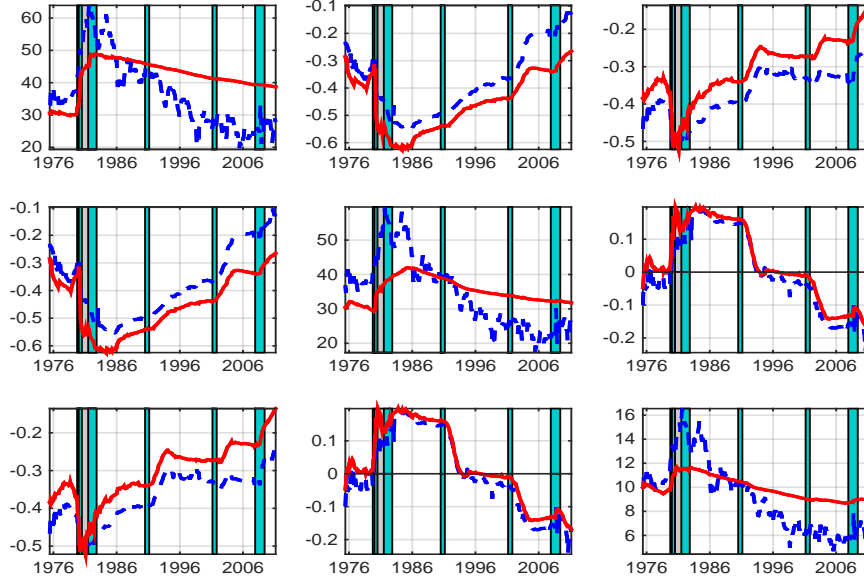


Figure 16: Conditional standard deviation and correlation estimates from learning models with (blue line) and without (red line) stochastic volatility. The estimates at date t are based on the historical data up to observation t , over the period July, 1975 to March, 2011.

8 Concluding Remarks

Three notable patterns emerge from our analysis: (i) \mathcal{RA} effectively treats the parameters governing the risk-neutral distribution of the pricing factors \mathcal{P} as known and constant over time (they were held virtually constant over the past thirty years); (ii) given this finding, a constrained version of the optimal Bayesian learning rule specializes to constant-gain, least-squares learning; and (iii) implementing the latter rule $\ell_{CG}(\mathcal{P}, H)$ in real time gives rise to forecasts of future bond yields and risk premiums that substantially outperform both the analogous rule $\ell_{CG}(\mathcal{P})$ based on yield curve information alone and the learning rule implicitly followed by the median BCFE professional forecaster. This outperformance is especially large following recessions when disagreement about future two-year Treasury yields is high.

Since our learning rules are inherently reduced-form, we cannot reach definitive conclusions regarding the economic mechanisms through which the dispersion measures H gain their predictive power. Nevertheless, several intriguing patterns emerge that are suggestive of fruitful directions for theoretical modeling. Under the premise that there is heterogeneity of views in the U.S. Treasury markets, the most likely source of disagreement is regarding the future paths of bond yields and not about the connection between the current state of the

economy and current yields. Financial institutions have long recognized that the cross-section of bond yields is well described by the low-order PCs which are readily observable. Indeed, most use pricing and risk management systems that presume that current economic conditions are fully reflected in the PCs .

A notable feature of $\ell_{CG}(\mathcal{P}, H)$ highlighted in [Figure 11](#) is that, to a substantial degree, the measures of dispersion in beliefs about future bond yields (H) affect risk premiums through the weights that \mathcal{RA} assigns to the current PCs when forecasting future yields. \mathcal{RA} finds it optimal to adjust the predictive content of \mathcal{P}_t about future \mathcal{P}_{t+s} depending on the degree of disagreement in the market, and this is especially the case after NBER recessions. Additionally, the effects of H on forecasts persist over several months. This leads us to doubt that the primary impact of H is through high-frequency episodes of flight-to-quality.

H could be proxying for omitted macroeconomic information. However, when \mathcal{RA} also conditions on the past histories of real economic activity and inflation, her forecasts of future excess returns actually *deteriorate* quite substantially. Moreover, her forecasts based on a learning rule that conditions on \mathcal{P}_t and this macro information are typically less accurate than those conditioned on (\mathcal{P}_t, H_t) . Of particular note, the component of H that is informative about future yields is largely uninformative about future inflation and output growth.

Pursuing this further, we computed the projection of $er_t^{10,1}$ implied by rule $\ell_{CG}(\mathcal{P}, H)$ onto the dispersion variables $ID_t^{lvl} \equiv 1/2(ID_t^{2y} + ID_t^{7y})$ and $ID_t^{slp} \equiv 1/2(ID_t^{7y} - ID_t^{2y})$, after controlling for a second-order polynomial in \mathcal{P}_t . This gives a (within-sample) R^2 of over 90% and reveals that it is ID_t^{slp} that is driving the predictable variation in risk premiums on the ten-year bond. Minus the “slope of dispersion” ($-ID_t^{slp}$) tends to be large coming out of recessions ([Figure 1](#)). This is when $y^{10y} - y^{2y}$ is increasing and $\ell_{CG}(\mathcal{P}, H)$ substantially outperforms $\ell(BCFF)$.

We computed the correlations between ID_t^{slp} and the measures of “economic policy uncertainty” constructed by Baker, Bloom, and Davis, both their overall index and twelve of their subcategories of policy uncertainty. In all cases these correlations were very small.³¹ This is further evidence that the predictive power of H is at most weakly tied to uncertainty about the macroeconomy as conventionally measured.

³¹This data was downloaded from www.policyuncertainty.com.

Appendices

A Bond Pricing in *GTSMs*

Suppose that bonds are priced under the presumption that $\Theta^{\mathbb{Q}}$ is fixed.

$$\begin{aligned} r_s &= \rho_0 + \rho_1 \mathcal{P}_s, \\ \mathcal{P}_{s+1} &= K_{0\mathcal{P}}^{\mathbb{Q}} + K_{\mathcal{P}\mathcal{P}}^{\mathbb{Q}} \mathcal{P}_s + \Sigma_{\mathcal{P}\mathcal{P}}^{1/2} e_{\mathcal{P},s+1}^{\mathbb{Q}}, \end{aligned}$$

The price of a zero coupon bond is then given by:

$$D_s^m = e^{\mathcal{A}_m + \mathcal{B}_m \mathcal{P}_s},$$

Where m denotes the maturity of the bond. We can calculate \mathcal{A}_m and \mathcal{B}_m by solving the first order difference equation:

$$\begin{aligned} \mathcal{A}_{m+1} - \mathcal{A}_m &= \left(K_0^{\mathbb{Q}}\right)' \mathcal{B}_m + \frac{1}{2} \mathcal{B}_m' \Sigma_{\mathcal{P}} \mathcal{B}_m - \rho_0, \\ \mathcal{B}_{m+1} - \mathcal{B}_m &= \left(K_1^{\mathbb{Q}} - I\right)' \mathcal{B}_m - \rho_1. \end{aligned}$$

With initial conditions $\mathcal{A}_0 = 0$ and $\mathcal{B}_0 = 0$. The loadings for the corresponding bond yields are $A_m = -\mathcal{A}_m/m$ and $B_m = -\mathcal{B}_m/m$.

B The Canonical Model

Under the assumption that $\Theta^{\mathbb{Q}}$ is fixed, we can proceed by adopting the computationally convenient [Joslin, Singleton, and Zhu \(2011\)](#) normalization scheme. Specifically, let X_s denote a set of latent risk factors with

$$\begin{aligned} r_s &= 1' X_s, \\ X_{s+1} &= \begin{bmatrix} k_{\infty}^{\mathbb{Q}} \\ 0 \\ 0 \end{bmatrix} + J(\lambda^{\mathbb{Q}}) X_s + \Sigma_{XX}^{1/2} e_{X,s+1}^{\mathbb{Q}}, \end{aligned}$$

with bond loadings:

$$y_s^m = A_{X,m}(k_{\infty}^{\mathbb{Q}}, \lambda^{\mathbb{Q}}, \Sigma_{XX}) + B_{X,m}(\lambda^{\mathbb{Q}}) X_s.$$

Under these normalizations [Joslin, Singleton, and Zhu \(2011\)](#) show that there exists a unique rotation of X_s so that the factors are the first three principal components of bond yields:

$$\mathcal{P}_s = v(k_\infty^\mathbb{Q}, \lambda^\mathbb{Q}, \Sigma_{XX}, W) + L(\lambda^\mathbb{Q}, W) X_s,$$

where $v = W[A_{X,m_1}, \dots, A_{X,m_J}]$ and $L = W[B_{X,m_1}, \dots, B_{X,m_J}]$, and W denote the principal component weights (that are computed at each estimation but we suppress any time indexes for ease of notation). The bond loadings A_m and B_m in the expression

$$y_s^m = A_m(\Theta^\mathbb{Q}) + B_m(\lambda^\mathbb{Q})\mathcal{P}_s,$$

are fully determined by the parameters $\Theta^\mathbb{Q} = (k_\infty^\mathbb{Q}, \lambda^\mathbb{Q}, \Sigma_{\mathcal{P}\mathcal{P}})$ and the principal component weights W . Furthermore, the parameters $K_{0\mathcal{P}}^\mathbb{Q}$, $K_{\mathcal{P}\mathcal{P}}^\mathbb{Q}$, ρ_0 , ρ_1 , and Σ_{XX} are all functions of the elements in $\Theta^\mathbb{Q}$, with transformations given by

$$\begin{aligned} K_{0\mathcal{P}}^\mathbb{Q} &= L \begin{bmatrix} k_\infty^\mathbb{Q} \\ 0 \\ 0 \end{bmatrix} - LJ(\lambda^\mathbb{Q})L^{-1}v, \\ K_{\mathcal{P}\mathcal{P}}^\mathbb{Q} &= LJ(\lambda^\mathbb{Q})L^{-1}, \\ \rho_0 &= -1'L^{-1}v, \\ \rho_1 &= (L^{-1})'1, \\ \Sigma_{\mathcal{P}\mathcal{P}} &= L\Sigma_{XX}L^{-1}. \end{aligned}$$

See [Joslin, Singleton, and Zhu \(2011\)](#) for details and proofs.

C Log likelihood function

We begin by noting that when $(\Theta^\mathbb{Q}, \Sigma_{\mathcal{O}})$ and Σ_Z are presumed to be constant, equation (11) implies that we can decompose the log likelihood function into a \mathbb{P} and \mathbb{Q} part

$$-2 \log L = -2 \log L^\mathbb{Q}(\Theta^\mathbb{Q}, \Sigma_{\mathcal{P}\mathcal{P}}, \Sigma_{\mathcal{O}}) - 2 \log L^\mathbb{P}(\Theta_t^\mathbb{P}, \Sigma_Z, Q_t).$$

$\log L^\mathbb{Q}$ denotes the part of the likelihood function associated with pricing errors and $\log L^\mathbb{P}$ the likelihood function of the dynamic evolution of Z_t ,

$$Z_{t+1} = K_{Z,0t}^\mathbb{P} + K_{Z,1t}^\mathbb{P} Z_t + \Sigma_Z^{1/2} e_{Z,t+1}^\mathbb{P}, \quad (31)$$

where $Z'_t = (P'_t, H'_t)'$ and $\Theta_t^{\mathbb{P}} = [K_{Z,0t}^{\mathbb{P}}, K_{Z,1t}^{\mathbb{P}}]$ denotes the drifting parameters. We assume $\Theta_t^{\mathbb{P}}$ can be partitioned as $(\psi^r, \psi_t^{\mathbb{P}})$, where $\psi_t^{\mathbb{P}}$ is the vectorized set of free parameters and ψ^r is the vectorized set of parameters that are fixed conditional on $\Theta^{\mathbb{Q}}$. The unrestricted parameters, $\psi_t^{\mathbb{P}}$, evolve according to a random walk

$$\psi_t^{\mathbb{P}} = \psi_{t-1}^{\mathbb{P}} + Q_{t-1}^{1/2} \eta_t \quad \eta_t \stackrel{iid}{\sim} N(0, I), \quad (32)$$

with stochastic covariance matrix Q_{t-1} . By moving terms that involve known parameters and observable states to the left hand side we can rewrite equation (31) into

$$\mathcal{Y}_t = \mathcal{X}_{t-1} \psi_{t-1}^{\mathbb{P}} + \Sigma_Z^{1/2} e_{Z,t}^{\mathbb{P}}, \quad (33)$$

where

$$\begin{aligned} \mathcal{Y}_t &= Z_t - (I \otimes [1, Z'_{t-1}]) \iota_r \psi^r, \\ \mathcal{X}_t &= (I \otimes [1, Z'_t]) \iota_f, \end{aligned}$$

with ι_r and ι_f denoting the matrices that select the columns of $(I \otimes [1, Z'_{t-1}])$ corresponding to the restricted and free parameters respectively. With normally distributed innovations to the latent parameter states (32) (the transition equation) and to the factor dynamics (33) (the measurement equation) we have a well-defined linear Kalman filter.³² Conditional on $(\Theta^{\mathbb{Q}}, \Sigma)$ the solution to the Kalman filter is given by recursively updating the posterior mean $\hat{\psi}_t^{\mathbb{P}} = \mathbb{E}^{\mathbb{P}}(\psi_t^{\mathbb{P}} | Z_1^t)$, posterior variance $P_t = \mathbb{V}^{\mathbb{P}}(\psi_t^{\mathbb{P}} | Z_1^t)$, and forecast variance $\Omega_t = \mathbb{V}^{\mathbb{P}}(Z_{t+1} | Z_1^t)$ according to:

$$\hat{\psi}_t^{\mathbb{P}} = \hat{\psi}_t^{\mathbb{P}} + P_{t-1} \mathcal{X}'_{t-1} \Omega_{t-1}^{-1} (\mathcal{Y}_t - \mathcal{X}_{t-1} \hat{\psi}_{t-1}^{\mathbb{P}}), \quad (34)$$

$$P_t = P_{t-1} + Q_{t-1} - P_{t-1} \mathcal{X}'_{t-1} \Omega_{t-1}^{-1} \mathcal{X}_{t-1} P_{t-1}, \quad (35)$$

$$\Omega_{t-1} = \mathcal{X}_{t-1} P_{t-1} \mathcal{X}'_{t-1} + \Sigma_Z, \quad (36)$$

with \mathbb{P} log likelihood function given by

$$\begin{aligned} -2 \log L^{\mathbb{P}} &= (t-1)N \log(2\pi) + \sum_{s=2}^t \log |\Omega_{s-1}| \\ &+ \frac{1}{2} \sum_{s=2}^t (\mathcal{Y}_s - \mathcal{X}_{s-1} \hat{\psi}_{s-1}^{\mathbb{P}})' \Omega_{s-1}^{-1} (\mathcal{Y}_s - \mathcal{X}_{s-1} \hat{\psi}_{s-1}^{\mathbb{P}}). \end{aligned} \quad (37)$$

³²Note that the latent states in the filtering problem are the parameters and not the factors.

Reworking equation (34) gives³³

$$\hat{\psi}_t^{\mathbb{P}} = \hat{\psi}_t^{\mathbb{P}} + (P_t - Q_{t-1})\mathcal{X}'_{t-1}\Sigma_Z^{-1} \left(\mathcal{Y}_t - \mathcal{X}_{t-1}\hat{\psi}_t^{\mathbb{P}} \right). \quad (38)$$

Letting $R_t = (P_t - Q_{t-1})^{-1}$, (38) reduces to the first equation in the definition of an adaptive least squares estimator (see (16)). Equation (35) can then be rewritten as³⁴

$$\begin{aligned} (P_t - Q_{t-1})^{-1} &= P_{t-1}^{-1} + \mathcal{X}'_{t-1}\Sigma_Z^{-1}\mathcal{X}_{t-1} \\ &= (I - P_{t-1}^{-1}Q_{t-2})(P_{t-1} - Q_{t-2})^{-1} + \mathcal{X}'_{t-1}\Sigma_Z^{-1}\mathcal{X}_{t-1}, \end{aligned} \quad (39)$$

which reduces to (17) if $Q_{t-2} = (1 - \gamma_{t-1})P_{t-1}$, for a sequence of scalars $0 < \gamma_t \leq 1$. Using (35) it follows that this condition is satisfied by choosing

$$Q_{t-1} = \frac{1 - \gamma_t}{\gamma_t} (P_{t-1} - P_{t-1}\mathcal{X}'_{t-1}\Omega_{t-1}^{-1}\mathcal{X}_{t-1}P_{t-1}).$$

From this expression it also follows that Q_{t-1} is measurable with respect to $\sigma(Z_1^{t-1})$ as long as γ_t is measurable. We can summarize the preceding calculations as:

$$R_t\hat{\psi}_t^{\mathbb{P}} = \gamma_{t-1}R_{t-1}\hat{\psi}_t^{\mathbb{P}} + \mathcal{X}'_{t-1}\Sigma_Z^{-1}\mathcal{Y}_t, \quad (40)$$

$$R_t = \gamma_{t-1}R_{t-1} + \mathcal{X}'_{t-1}\Sigma_Z^{-1}\mathcal{X}_{t-1}, \quad (41)$$

$$\hat{\psi}_t^{\mathbb{P}} = R_t^{-1}R_t\hat{\psi}_t^{\mathbb{P}}, \quad (42)$$

$$P_t = \frac{1}{\gamma_t}R_t^{-1}, \quad (43)$$

$$\Omega_{t-1} = \mathcal{X}_{t-1}P_{t-1}\mathcal{X}'_{t-1} + \Sigma_Z, \quad (44)$$

with log likelihood function given by (37). The constant gain estimator corresponds to the special case where $\gamma_t = \gamma$ for all t .

D Pricing Kernel

The pricing kernel can be expressed as

$$\mathcal{M}_{t,t+1} = e^{-r_t} \times \frac{f_{t,t+1}^{\mathbb{Q}}(\mathcal{P}_{t+1})}{f_{t,t+1}^{\mathbb{P}}(\mathcal{P}_{t+1})}.$$

³³Substitute (36) into (35) and the resulting equation into (34).

³⁴This expression is obtained by substituting (36) into (35), plugging the resulting equation back into (35), and multiplying by $(P_t - Q_{t-1})^{-1}$ from the left and P_{t-1}^{-1} from the right.

Since the distributions are conditionally normal under both measures, they have equal support. Then, $\mathcal{M}_{t,t+1}$ defines a strictly positive pricing kernel. We can rewrite the conditional distributions as

$$\begin{aligned} f_{t,t+1}^{\mathbb{P}} &= N(\hat{K}_{\mathcal{P}0,t}^{\mathbb{P}} + [\hat{K}_{\mathcal{P}\mathcal{P},t}^{\mathbb{P}}, \hat{K}_{\mathcal{P}H,t}^{\mathbb{P}}]Z_t, \Omega_{\mathcal{P}\mathcal{P},t}) = N(\hat{\mu}_t^{\mathbb{P}}, \Omega_{\mathcal{P}\mathcal{P},t}), \\ f_{t,t+1}^{\mathbb{Q}} &= N(K_0^{\mathbb{Q}} + K_1^{\mathbb{Q}}\mathcal{P}_t, \Sigma_{\mathcal{P}\mathcal{P}}) = N(\mu_t^{\mathbb{Q}}, \Sigma_{\mathcal{P}\mathcal{P}}), \end{aligned}$$

where $(\hat{K}_{\mathcal{P}0,t}^{\mathbb{P}}, [\hat{K}_{\mathcal{P}\mathcal{P},t}^{\mathbb{P}}, \hat{K}_{\mathcal{P}H,t}^{\mathbb{P}}])$ denote the posterior means of the latent parameters states, and $\Omega_{\mathcal{P}\mathcal{P},t}$ the upper left 3×3 entries of the conditional covariance matrix Ω_t given in equation (44). We can reduce this expression as follows. We will use the notation c_t to terms that are \mathcal{F}_t measurable but not of direct interest

$$\begin{aligned} \log \mathcal{M}_{t,t+1} + r_t &= c_t + \frac{1}{2}(\mathcal{P}_{t+1} - \hat{\mu}_t^{\mathbb{P}})' \Omega_{\mathcal{P}\mathcal{P},t}^{-1} (\mathcal{P}_{t+1} - \hat{\mu}_t^{\mathbb{P}}) - \frac{1}{2}(\mathcal{P}_{t+1} - \mu_t^{\mathbb{Q}})' \Sigma_{\mathcal{P}\mathcal{P}}^{-1} (\mathcal{P}_{t+1} - \mu_t^{\mathbb{Q}}) \\ &= c_t' - \left(\Omega_{\mathcal{P}\mathcal{P},t}^{-1} \hat{\mu}_t^{\mathbb{P}} - \Sigma_{\mathcal{P}\mathcal{P}}^{-1} \mu_t^{\mathbb{Q}} \right)' \mathcal{P}_{t+1} + \frac{1}{2} \mathcal{P}_{t+1}' (\Omega_{\mathcal{P}\mathcal{P},t}^{-1} - \Sigma_{\mathcal{P}\mathcal{P}}^{-1}) \mathcal{P}_{t+1} \\ &= c_t'' - \Lambda_{\mathcal{P}t}' \Gamma_t^{-1} \varepsilon_{t+1}^{\mathbb{P}} + \frac{1}{2} (\varepsilon_{t+1}^{\mathbb{P}})' (I - \Gamma_t^{-1}) \varepsilon_{t+1}^{\mathbb{P}}, \end{aligned}$$

where

$$\begin{aligned} \Lambda_{\mathcal{P}t} &= \Omega_{\mathcal{P}\mathcal{P},t}^{-1/2} (\hat{\mu}_t^{\mathbb{P}} - \mu_t^{\mathbb{Q}}) \\ \Gamma_t &= \Omega_{\mathcal{P}\mathcal{P},t}^{-1/2} \Sigma_{\mathcal{P}\mathcal{P}} (\Omega_{\mathcal{P}\mathcal{P},t}^{-1/2})' \\ c_t'' &= -\frac{1}{2} \log |\Gamma_t| - \frac{1}{2} \Lambda_{\mathcal{P}t}' \Gamma_t^{-1} \Lambda_{\mathcal{P}t} \end{aligned}$$

Thus the stochastic discount factor resembles a stochastic discount factor under full information, though with the parameters determining the market price of risks replaced by their posterior means, and with an additional stochastic convexity term and matrix Γ_t representing the change of conditional covariance matrix from \mathbb{P} to \mathbb{Q} .

To show that $\Lambda_{\mathcal{P}t}$ is naturally interpreted as the market prices of risk in our learning setting, consider an asset with log total-return spanned by the factors \mathcal{P}_t : $r_t^a = \alpha + \beta' \mathcal{P}_t$ and satisfying $\mathbb{E}_t [e^{r_t^a + 1} \mathcal{M}_{t,t+1}] = 1$. Using the fact that $\mathbb{E}[e^{\theta' \varepsilon + \frac{1}{2} \varepsilon' (I - \Gamma^{-1}) \varepsilon}] = e^{\frac{1}{2} \theta' \Gamma \theta + \frac{1}{2} \log |\Gamma|}$, for $\varepsilon \sim N(0, I)$, the left-hand side of the last expression can be rewritten as

$$\begin{aligned} &\exp\{\alpha + \beta' \hat{\mu}_t^{\mathbb{P}} - r_t + c_t'' + \frac{1}{2} (\beta' \Omega_t^{1/2} - \Lambda_{\mathcal{P}t}' \Gamma_t^{-1}) \Gamma_t (\beta' \Omega_t^{1/2} - \Lambda_{\mathcal{P}t}' \Gamma_t^{-1})' + \frac{1}{2} \log |\Gamma_t|\} \\ &= \exp\{\mathbb{E}_t[r_{t+1}^a] - r_t + \frac{1}{2} \beta' \Omega_t \beta - \beta' \Omega_t^{1/2} \Lambda_{\mathcal{P}t}\}. \end{aligned}$$

This leads to

$$\mathbb{E}_t[r_{t+1}^a] - r_t + \frac{1}{2}\mathbb{V}[r_{t+1}^a] = \beta'\Omega_t^{1/2}\Lambda_{\mathcal{P}t};$$

the expected log excess return equals the quantity of risk times the market price of risk (after adjusting for a convexity term).

E Constraints on $\Lambda_{\mathcal{P}t}$ for Rule $\ell(\mathcal{P}, H)$ and $\ell_{CG}(\mathcal{P}, H)$

We use the training sample to reduce the dimension of $\Theta^{\mathbb{P}}$. For models evaluated out of sample between January 1995 and March 2011, the training sample consists of the prior 10 years from January 1985 to December 1994. Our dimension reduction strategy is based on restricting the physical measure towards the risk neutral. First we estimate a model without restrictions imposed, and then we inspect the statistic significance of each of the parameters in $PmQ_t = \begin{pmatrix} \hat{K}_{\mathcal{P}0,t}^{\mathbb{P}} - \hat{K}_{\mathcal{P}0}^{\mathbb{Q}} & \hat{K}_{\mathcal{P}\mathcal{P},t}^{\mathbb{P}} - \hat{K}_{\mathcal{P}\mathcal{P}}^{\mathbb{Q}} & \hat{K}_{\mathcal{P}H,t}^{\mathbb{P}} \\ \hat{K}_{H0,t}^{\mathbb{P}} & \hat{K}_{H\mathcal{P},t}^{\mathbb{P}} & \hat{K}_{HH,t}^{\mathbb{P}} \end{pmatrix}$. If the p-value, induced by the posterior variance, at the end of the training sample is above 0.3 the corresponding coefficient in $K_{Z,t}^{\mathbb{P}}$ is concentrated out such that the corresponding entry in PmQ_t is zero. The only exception to this rule is if the coefficient is of particular economic importance, in which case we leave the coefficient unrestricted also when the p-value is just above 0.3. Model $\ell_{CG}(\mathcal{P}, H)$ illustrates the procedure. We begin by estimating an unrestricted version of the model from January 1985 to December 1994 where we are learning about all coefficients in $K_{Z,t}^{\mathbb{P}}$. [Table 9](#) reports posterior means and p-values for PmQ_t on December 1994. We then identify all coefficients that have an implied p-value above 0.3 as candidates for restriction. Among those coefficients is PC2's coefficient on its own lagged value (the (2,3) entry in [Table 9](#)). This coefficient governs much of the persistence of the second factor, and could conceivably be important for forecasts of the slope of the term structure. As such, we allow it to be unrestricted on a discretionary basis. Two other examples are the insignificant slope loadings on the two dispersion variables. While there could be interesting direct feedbacks from belief heterogeneity measures to PC2, p-values of 0.80 and 0.71 seem to be too high to justify inclusion. Therefore, we restrict the (2,4) and (2,5) entries in $K_{Z,t}^{\mathbb{P}}$. [Table 10](#) displays the restrictions on the autoregressive feedback matrix that we end up imposing in the learning rule $\ell(\mathcal{P}, H)$. Similarly, [Figure 11](#) contains the restrictions for rule $\ell(\mathcal{P})$. Note that both rules include the constraint that the market price of *PC3* risk is zero.

	Λ_{0t}	Λ_{1t}				
		PC_1	PC_2	PC_3	ID_{2y}	ID_{7y}
PC_1	1.73	-0.051	-0.25	0.284	-0.008	0.01
(t)	(1.95)	(-1.54)	(-1.84)	(0.532)	(-1.47)	(1.72)
(p)	0.052	0.124	0.066	0.595	0.142	0.086
PC_2	0.424	0.001	-0.036	0.278	0	-0.001
(t)	(1.44)	(0.112)	(-0.806)	(1.57)	(0.26)	(-0.368)
(p)	0.149	0.911	0.42	0.115	0.795	0.713
PC_3	0.007	0.003	-0.001	0.039	0	0
(t)	(0.062)	(0.763)	(-0.062)	(0.565)	(0.473)	(-0.533)
(p)	0.95	0.445	0.951	0.572	0.636	0.594
ID_{2y}	-5	1.25	4.36	15.1	0.586	0.255
(t)	(-0.218)	(1.46)	(1.25)	(1.1)	(4.25)	(1.74)
(p)	0.827	0.145	0.213	0.272	0	0.083
ID_{7y}	3.04	0.815	2.01	5.03	0.043	0.807
(t)	(0.142)	(1.02)	(0.618)	(0.393)	(0.337)	(5.9)
(p)	0.887	0.306	0.537	0.694	0.736	0

Table 9: Posterior mean estimates of the parameters PmQ_t in rule $\ell(\mathcal{P}, H)$ and their posterior variance implied t-stats and p-values.

	Λ_{0t}	Λ_{1t}				
		PC_1	PC_2	PC_3	ID_{2y}	ID_{7y}
PC_1	*	*	*	0	*	*
PC_2	*	0	*	*	0	0
PC_3	0	0	0	0	0	0
ID_{2y}	0	*	*	*	*	*
ID_{7y}	0	*	0	0	0	*

Table 10: Restrictions applied in rule $\ell(\mathcal{P}, H)$ to the parameters in PmQ_t .

	const	PC_1	PC_2	PC_3
PC_1	*	*	*	*
PC_2	*	*	*	*
PC_3	0	0	0	0

Table 11: Restrictions applied in rule $\ell(\mathcal{P})$ to the parameters of the market price of \mathcal{P} risk.

F Stochastic Volatility Model

Suppose that there exist a 3-dimensional state-variable, consisting of a univariate volatility factor V_t , and 2 conditionally gaussian factors X_t . Following our specification of the gaussian models with learning, we assume that the parameters governing the risk neutral measure are known and constant. [Joslin and Le \(2014\)](#) show that an econometrically exactly identified specification is given by

$$\begin{aligned}
V_{t+1}|V_t &\sim CAR(\rho^{\mathbb{Q}}, c^{\mathbb{Q}}, v^{\mathbb{Q}}), \\
X_{t+1} &= K_{XV}^{\mathbb{Q}} V_t + J(\lambda^{\mathbb{Q}}) X_t + \sqrt{\Sigma_0 + \Sigma_1 V_t} \cdot \varepsilon_t^{\mathbb{Q}}, \\
r_t &= r_{\infty}^{\mathbb{Q}} + \rho_V V_t + 1' X_t,
\end{aligned}$$

where CAR is short for the compound autoregressive gamma process. The CAR process has a conditional Laplace transform that is exponentially affine and first and second moments given by

$$\begin{aligned}
\log \mathbb{E}^{\mathbb{Q}}(e^{uV_{t+1}}|V_t) &= -v^{\mathbb{Q}} \log(1 - uc^{\mathbb{Q}}) + \frac{\rho^{\mathbb{Q}} u}{1 - uc^{\mathbb{Q}}} V_t, \\
\mathbb{E}_t^{\mathbb{Q}}(V_{t+1}|V_t) &= v^{\mathbb{Q}} c^{\mathbb{Q}} + \rho^{\mathbb{Q}} V_t, \\
\mathbb{V}_t^{\mathbb{Q}}(V_{t+1}|V_t) &= v^{\mathbb{Q}^2} c^{\mathbb{Q}} + 2\rho^{\mathbb{Q}} V_t.
\end{aligned}$$

The innovation to the non-volatility factors, $\varepsilon_{t+1}^{\mathbb{Q}}$, is assumed to be normally distributed and independent of V_{t+1} . It follows that zero coupon bond prices are exponentially affine, $D_t^n = e^{A_n + B_{n,V} V_t + B_{n,X} X_t}$, with loadings that satisfy the recursions

$$\begin{aligned}
A_{n+1} &= A_n + \frac{1}{2} B'_{n,X} \Sigma_0 B_{n,X} - v^{\mathbb{Q}} \log(1 - c^{\mathbb{Q}} B_{n,V}) - r_{\infty}^{\mathbb{Q}}, \\
B_{n+1,X} &= J(\lambda^{\mathbb{Q}})' B_{n,X} - 1, \\
B_{n+1,V} &= B'_{n,X} K_{XV} + \frac{1}{2} B'_{n,X} \Sigma_1 B_{n,X} + \frac{\rho^{\mathbb{Q}} B_{n,V}}{1 - c^{\mathbb{Q}} B_{n,V}} - \rho_V.
\end{aligned}$$

Under the physical measure we assume that parameters that govern the dynamics of the volatility factor is known and constant, while the parameters that govern the conditional gaussian factors are drifting and unknown

$$V_{t+1}|V_t \sim CAR(\rho^{\mathbb{P}}, c^{\mathbb{P}}, v^{\mathbb{P}}), \quad (45)$$

$$X_{t+1} = K_{X0,t}^{\mathbb{P}} + K_{XV,t}^{\mathbb{P}}V_t + K_{XX,t}^{\mathbb{P}}X_t + \sqrt{\Sigma_{0X} + \Sigma_{1X}V_t} \cdot \varepsilon_t^{\mathbb{P}}. \quad (46)$$

As yields are affine in the state-variables,

$$y_t = A(\Theta^{\mathbb{Q}}, \Sigma_{0X}, \Sigma_{1X}) + B_V(\Theta^{\mathbb{Q}}, \Sigma_{0X}, \Sigma_{1X})V_t + B_X(\lambda^{\mathbb{Q}})X_t,$$

the PCs $\mathcal{P}_t = Wy_t$ are also affine in the state. This in turn implies that V_t can be written as an affine function of f_t :

$$V_t = \alpha(\Theta^{\mathbb{Q}}, \Sigma_{0X}, \Sigma_{1X}) + \beta(\Theta^{\mathbb{Q}}, \Sigma_{0X}, \Sigma_{1X})' \mathcal{P}_t.$$

Joslin and Le (2014) show that we can rewrite and reparameterize equation (46) with

$$\mathcal{P}_{t+1}^{2:3} - W^{2:3}B_VV_{t+1} = \tilde{K}_{\mathcal{P}0,t}^{\mathbb{P}} + \tilde{K}_{\mathcal{P}V,t}^{\mathbb{P}}V_t + \tilde{K}_{\mathcal{P}\mathcal{P},t}^{\mathbb{P}}\mathcal{P}_t^{2:3} + \sqrt{\tilde{\Sigma}_{0\mathcal{P}} + \tilde{\Sigma}_{1\mathcal{P}}V_t} \cdot \varepsilon_t^{\mathbb{P}}, \quad (47)$$

where the superscripts 2 : 3 refer to the second and third PCs, and the tilde is used to indicate that these are parameters governing the dynamics of $(V_t, (\mathcal{P}_t^{2:3})')$ (and not \mathcal{P}_t). Therefore, the model's parameters can be decomposed into constant and known \mathbb{Q} -parameters $(r_{\infty}^{\mathbb{Q}}, \rho_V^{\mathbb{Q}}, \rho^{\mathbb{Q}}, c^{\mathbb{Q}}, v^{\mathbb{Q}}, K_{XV}^{\mathbb{Q}}, \lambda^{\mathbb{Q}})$, constant and known covariance matrices $(\tilde{\Sigma}_{\mathcal{P}0}, \tilde{\Sigma}_{1\mathcal{P}})$, constant and known \mathbb{P} -parameters $(\rho^{\mathbb{P}}, c^{\mathbb{P}}, v^{\mathbb{P}})$, and unknown drifting \mathbb{P} parameters $(\tilde{K}_{\mathcal{P}0,t}^{\mathbb{P}}, \tilde{K}_{\mathcal{P}V,t}^{\mathbb{P}}, \tilde{K}_{\mathcal{P}\mathcal{P},t}^{\mathbb{P}})$. We impose that $c^{\mathbb{P}} = c^{\mathbb{Q}}$ and $v^{\mathbb{P}} = v^{\mathbb{Q}}$. These two conditions guarantee diffusion invariance of V_t , and that the market prices of risks are non-exploding in the continuous time limit (see Joslin and Le (2014)). From equations (45) - (47) it is seen that the conditional first and second moments of the principal components are given by

$$\mathbb{E}_t^{\mathbb{P}}(\mathcal{P}_{t+1}) = K_{0,t}^{\mathbb{P}} + K_{1,t}^{\mathbb{P}}\mathcal{P}_t \quad (48)$$

$$\mathbb{V}_t^{\mathbb{P}}(\mathcal{P}_{t+1}) = \Sigma_0^{\mathbb{P}} + \Sigma_1^{\mathbb{P}}\mathcal{P}_t \quad (49)$$

where $(K_{0,t}^{\mathbb{P}}, K_{1,t}^{\mathbb{P}}, \Sigma_0^{\mathbb{P}}, \Sigma_1^{\mathbb{P}})$ are known functions of $(\Theta^{\mathbb{Q}}, \tilde{\Sigma}_{\mathcal{P}0}, \tilde{\Sigma}_{1\mathcal{P}}, \Theta_t^{\mathbb{P}})$ induced by rotating $(V_t, (\mathcal{P}_t^{2:3})')$ to \mathcal{P}_t . Similar to the gaussian learning model we impose restrictions on $[K_{0,t}^{\mathbb{P}}, K_{1,t}^{\mathbb{P}}]$ based on a training sample. These restrictions as $vec\left([\tilde{K}_{\mathcal{P}0,t}^{\mathbb{P}}, \tilde{K}_{\mathcal{P}V,t}^{\mathbb{P}}, \tilde{K}_{\mathcal{P}\mathcal{P},t}^{\mathbb{P}}]\right) = R\psi_t + q$, where

ψ evolves according to a random walk

$$\psi_t = \psi_{t-1} + Q_{t-1}^{1/2} \eta_t.$$

A set of sufficient conditions that guarantees that the innovation co-variance matrix of ψ_t is proportional to the posterior co-variance matrix will ensure that the posterior means of ψ is given by a constant gain estimator. The proof is similar to the derivations for the Gaussian learning model.

References

- Altavilla, C., R. Giacomini, and G. Ragusa, 2014, Anchoring the Yield Curve Using Survey Expectations, Working paper, Working Paper No 1632, European Central Bank.
- Andrade, P., R. Crump, S. Eusepi, and E. Moench, 2014, Fundamental Disagreement, Working paper, Federal Reserve Bank of New York Staff Report No. 655.
- Barillas, F., and K. Nimark, 2014, Speculation, Risk Premia and Expectations in the Yield Curve, Working paper, Cornell University.
- Buraschi, A., and P. Whelan, 2016, Speculation, Hedging, and Interest Rates, Working paper, Imperial College.
- Chun, A., 2011, Expectations, Bond Yields and Monetary Policy, *Review of Financial Studies* 24, 208–247.
- Coibion, O., and Y. Gorodnichenko, 2012, What Can Survey Forecasts Tell Us About Informational Rigidities?, *Journal of Political Economy* 120, 116–159.
- Collin-Dufresne, P., M. Johannes, and L. Lochstoer, 2016, Parameter Learning in General Equilibrium: The Asset Pricing Implications, *American Economic Review* 106, 664–698.
- Dai, Q., and K. Singleton, 2000, Specification Analysis of Affine Term Structure Models, *Journal of Finance* 55, 1943–1978.
- Dai, Q., and K. Singleton, 2002, Expectations Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure, *Journal of Financial Economics* 63, 415–441.
- Dalang, R., A. Morton, and W. Willinger, 1990, Equivalent Martingale Measures and No-arbitrage in Stochastic Securities Market Models, *Stochastics Stochastics Rep.* 29 2, 185–201.
- David, A., 2008, Heterogeneous Beliefs, Speculation, and the Equity Premium, *Journal of Finance* 63, 41–83.
- Diebold, F., and R. Mariano, 1995, Comparing Predictive Accuracy, *Journal of Business and Economic Statistics* 13, 253–263.
- Dovern, J., U. Fritsche, and J. Slacalek, 2012, Disagreement Among Forecasters in G7 Countries, *Review of Economics and Statistics* 94, 1081–1096.
- Duffee, G., 2001, Term Premia and Interest Rates Forecasts in Affine Models, forthcoming, *Journal of Finance*.

- Duffee, G., 2002, Term Premia and Interest Rate Forecasts in Affine Models, *Journal of Finance* 57, 405–443.
- Duffee, G., 2010, Sharpe Ratios in Term Structure Models, Working paper, Johns Hopkins University.
- Duffee, G., 2011, Forecasting with the Term Structure: the Role of No-Arbitrage, Working paper, Johns Hopkins University.
- Fleming, M., and E. Remolona, 1999, Price Formation and Liquidity in the US Treasury Market: The Response to Public Information, *Journal of Finance* 54, 1901–1915.
- Ghysels, E., C. Horan, and E. Moench, 2014, Forecasting through the Rear-View Mirror: Data Revisions and Bond Return Predictability, Working paper, FRB of New York Staff Report No. 581.
- Gourieroux, C., and J. Jasiak, 2006, Autoregressive Gamma Processes, *Journal of Forecasting* 25, 129–152.
- Harvey, D., S. Leybourne, and P. Newbold, 1997, Testing the Equality of Prediction Mean-Squared Errors, *International Journal of Forecasting* 13, 281–291.
- Joslin, S., and A. Le, 2014, Interest Rate Volatility and No Arbitrage Affine Term Structure Models, Working paper, Pennsylvania State University.
- Joslin, S., A. Le, and K. Singleton, 2013, Why Gaussian Macro-Finance Term Structure Models are (Nearly) Unconstrained Factor-VARs, *Journal of Financial Economics* 109, 604–622.
- Joslin, S., M. Pribsch, and K. Singleton, 2014, Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks, *Journal of Finance* LXIX, 1197–1233.
- Joslin, S., K. Singleton, and H. Zhu, 2011, A New Perspective on Gaussian Dynamic Term Structure Models, *Review of Financial Studies* 24, 926–970.
- Jouini, E., and C. Napp, 2007, Consensus Consumer and Intertemporal Asset Pricing with Heterogeneous Beliefs, *Review of Economic Studies* 74, 1149–1174.
- Kim, D., and A. Orphanides, 2012, Term Structure Estimation with Survey Data on Interest Rate Forecasts, *Journal of Financial and Quantitative Analysis* 47, 241–272.
- Le, A., and K. Singleton, 2012, A Robust Analysis of the Risk-Structure of Equilibrium Term Structures of Bond Yields, Working paper, University of North Carolina.

- Ludvigson, S., and S. Ng, 2010, Macro Factors in Bond Risk Premia, *Review of Financial Studies* 22, 5027–5067.
- Mankiw, N. G., and R. Reis, 2002, Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve, *Quarterly Journal of Economics* 117, 1295–1328.
- McCulloch, J., 2007, The Kalman Foundations of Adaptive Least Squares, With Applications to U.S. Inflation, Working paper, Ohio State University.
- Patton, A., and A. Timmerman, 2010, Why do Forecasters Disagree? Lessons from the Term Structure of Cross-Sectional Dispersion, *Journal of Monetary Economics* 57, 803–820.
- Piazzesi, M., 2005, Bond Yields and the Federal Reserve, *Journal of Political Economy* 113, 311–344.
- Piazzesi, M., J. Salomao, and M. Schneider, 2013, Trend and Cycle in Bond Premia, Working paper, Stanford University.
- Rudebusch, G., and J. Williams, 2009, Forecasting Recessions: The Puzzle of the Enduring Power of the Yield Curve, *Journal of Business and Economic Statistics* 27, 492–503.
- Sims, C. A., 2003, Implications of Rational Inattention, *Journal of Monetary Economics* 50, 665–690.
- Singleton, K., 1987, Asset Prices in a Time Series Model with Disparately Informed, Competitive Traders, in W. Burnett, and K. Singleton, eds.: *New Approaches to Monetary Economics* (Cambridge University Press,).
- Townsend, R., 1983, Forecasting the Forecasts of Others, *Journal of Political Economy* 91, 546–588.
- Xiong, W., and H. Yan, 2009, Heterogeneous Expectations and Bond Markets, *Review of Financial Studies*.