Learning the Dynamics of U.S. Treasury Yields With an Aribitrage-free Term Structure Model

Marco Giacoletti, Kristoffer T. Laursen
and Kenneth J. Singleton

Stanford University

June, 2014

Motivating Learning

1. Investors in Treasury bonds have experienced:
   - several major financial crises;
   - unforeseen changes in policies and transparency of the FRB;
   - lack of clarity on the future pathes of fiscal policies.

Motivating Learning

1. Investors in Treasury bonds have experienced:
   - several major financial crises;
   - unforeseen changes in policies and transparency of the FRB;
   - lack of clarity on the future pathes of fiscal policies.
2. We explore how learning about the risk profile of Treasury bonds affects:
   - the prices of bonds,
   - required compensations for bearing relevant factor risks,
   - (forecasts of) the future shapes of the term structure of yields.

Modeling Learning

- Endow agents with an yield-based *DTSM* that they use for updating their beliefs every month by ML.
- Based on this learning rule they price bond and forecast future yields (and compute market risk premiums).
- As naive as this rule is, it performs strikingly well against:
  1. the consensus forecasts of the BCFF survey professionals.
  2. the simple random walk model of bond yields.
  3. When macroeconomic information is incorporated, our *DTSM*-based learning rule outperforms other models, especially during the 2000's leading up to the current crisis.
- A computationally simple, naive and yet plausible, and remarkably effective learning rule. Why?

What Is Our Agent Learning About?

- *Our agent is not the professional forecaster.* No survey information is used in fitting our learning rules.
- Our agent updates her views about the (unknown?) risk structure of yields using an arbitrage-free *DTSM*.
- Agents are not learning about the state of the economy. Over 98% of the variation in Treasury yields is accounted for by the low-order PCs ($\mathcal{P}$) of yields, which are measured accurately.
- Agents are learning about how bond yields are related to $\mathcal{P}$ and about the dynamics of $\mathcal{P}$ over the business cycle.
- View updating the parameters of a *DTSM* as updating an approximation to the conditional distribution of bond yields.

Joslin, Priebsch, and Singleton (2013) Model of the Historical Distribution of Risk

- Macroeconomic information, over and above $\mathcal{P}$, is important for understanding risk compensation in bond markets.
- No macro factors in $\mathcal{P}$, because the resulting *DTSMs* do not accurately price bonds (Joslin, Le, and Singleton (2013)).
- Following JPS, $Z_t \equiv (\mathcal{P}_t, M_t)$ follows the Gaussian process

$$Z_t = K_0^{\mathbb{P}} + K_Z^{\mathbb{P}} Z_{t-1} + \Sigma_Z^{-1/2} \epsilon_{Zt}^{\mathbb{P}}.$$

- The market prices of risks $\mathcal{P}$: $\Lambda_{\mathcal{P}t} = \Lambda_0 + \Lambda_Z Z_t$.
- Agents are learning about $\Theta^{\mathbb{P}} = (K_0^{\mathbb{P}}, K_Z^{\mathbb{P}})$, along with the parameters $\Theta^{\mathbb{Q}}$ of the pricing distribution.

Sophisticated "Partially Bayesian" Learner

- Bayesian learning is a sophisticated calculation since agents are learning about a high dimensional $(\mathbb{P}, \mathbb{Q})$ parameter set.
- Consider the simpler *Partially Bayesian* $(\mathcal{PB})$ learner who updates on $\Theta^{\mathbb{P}}$ taking $\Theta^{\mathbb{Q}}$ as given:

$$f(Z_1^t, O_1^t) = \prod_{s=1}^{t} f(\mathcal{O}_s | Z_1^s, O_1^{s-1}; \Theta^{\mathbb{Q}}, \Sigma_e) \times \int f(Z_s | Z_1^{s-1}, \mathcal{O}_1^{s-1}, \Theta_{s-1}^{\mathbb{P}}; \Sigma_Z) f(\Theta_{s-1}^{\mathbb{P}} | Z_1^{s-1}, \mathcal{O}_1^{s-1}) d(\Theta_{s-1}^{\mathbb{P}}).$$

- Formaly learning about the historical distribution of Z.
- This $\mathcal{PB}$ case is interesting because:
  1. its structure can be reinterpreted as a constrained version of the fully Bayesian rule;
  2. the presumption that $\Theta^{\mathbb{Q}}$ is fixed and known turns out to be consistent with our empirical learning rules.

An Illustrative Learning Environment

- $\mathcal{PB}$ agent learning about $\Theta^{\mathbb{P}}$ taking $(\Theta^{\mathbb{Q}}, \Sigma_e)$ as known.
- Suppose that $\Theta_t^{\mathbb{P}}$ can be partitioned as $(\psi^r, \psi_t^{\mathbb{P}})$, and that

$$\psi_t^{\mathbb{P}} = \psi_{t-1}^{\mathbb{P}} + \eta_t, \qquad \eta_t \overset{iid}{\sim} N(0, Q_t),$$

  Q_t denotes the (possibly) time-varying covariance matrix of η_t.
- Adopting a Gaussian prior on $\psi_0^{\mathbb{P}}$, the posterior distribution for $\psi_t^{\mathbb{P}}$ is Gaussian, $\psi_t^{\mathbb{P}} | Z_1^t \sim N(\mu_t, P_t)$, with the posterior mean

$$\mu_t = \mu_{t-1} + R_t^{-1} x'_{t-1} \Sigma_Z^{-1} (y_t - x_{t-1}\mu_{t-1}),$$

  where $R_t^{-1} \equiv P_t - Q_t$ and R_t satisfies the recursion

$$R_t = \left(I - P_{t-1}^{-1} Q_{t-1}\right) R_{t-1} + x'_{t-1} \Sigma_Z^{-1} x_{t-1}.$$

The $\mathcal{PB}$ Learner as a (Near Fully) Bayesian Learner

- Two special cases of Bayesian updating on $\psi_t^{\mathbb{P}}$:

  $\mathcal{B}\downarrow$CGLS: If $P_{t-1}^{-1}Q_{t-1} = (1-\gamma)\cdot I$, μ_t is a *constant gain least-squares (CG)* estimator of $\psi^{\mathbb{P}}$ with gain coefficient $\gamma \in (0,1]$.

  $\mathcal{B}\downarrow$RLS: If $\gamma = 1$, then $\psi_t^{\mathbb{P}} = \psi_{t-1}^{\mathbb{P}}$ and μ_t is the *recursive least-squares (RLS)* estimator of $\psi^{\mathbb{P}}$.

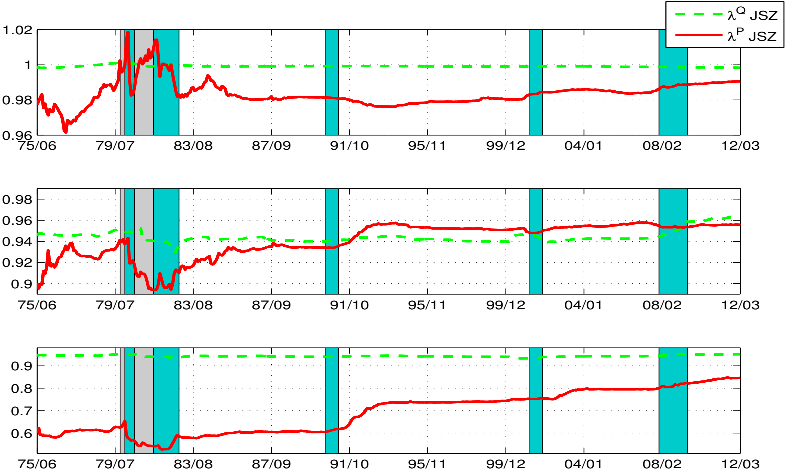
- RLS learning has a Bayesian interpretation when the agent believes that $\psi^{\mathbb{P}}$ is unknown, but is not changing over time.

- We search over γ in the CG case to minimize the RMSE of forecasts of $PC1$ one year ahead.

Model-Based Learning Rules

Rule	*DTSM*	Information	Restrictions	γ
$\ell(RW)$	Random Walk	Own Yield	N/A	N/A
$\ell(JSZ)$	JSZ	$\mathcal{P}$	No-Arbitrage $PC3$ unpriced	1
$\ell(JSZ_{CG})$	JSZ	$\mathcal{P}$	No-Arbitrage + $PC3$ unpriced	0.99
$\ell(JPS)$	JPS	$(\mathcal{P}, M)$	No-Arbitrage + $PC3$ unpriced	1

(No) Learning About Eigenvalues $\lambda^{\mathbb{Q}}$ of $K^{\mathbb{Q}}_{\mathcal{PP}}$



RMSE's for one-quarter ahead forecasts, January, 1985 to March, 2012

	RMSE's by Bond Maturity						
Rule	6m	1Y	2Y	3Y	5Y	7Y	10Y
$\ell(BCFF)$	51.4	51.6	52.4	54.3	49.5	47.9	44.8
$\ell(JSZ_{LS})$	39.7 (−4.03)	41.8 (−3.07)	45.2 (−3.92)	44.6 (−5.28)	43.0 (−4.39)	41.2 (−3.92)	37.7 (−3.33)
$\ell(JSZ_{CG})$	38.5 (−4.36)	41.6 (−3.17)	45.2 (−3.80)	45.0 (−4.45)	43.4 (−4.10)	42.1 (−3.66)	38.8 (−2.96)
$\ell(JPS_{LS})$	36.2 (−3.96)	41.2 (−2.74)	44.2 (−2.99)	43.9 (−3.86)	41.4 (−4.71)	40.7 (−3.94)	39.3 (−2.64)

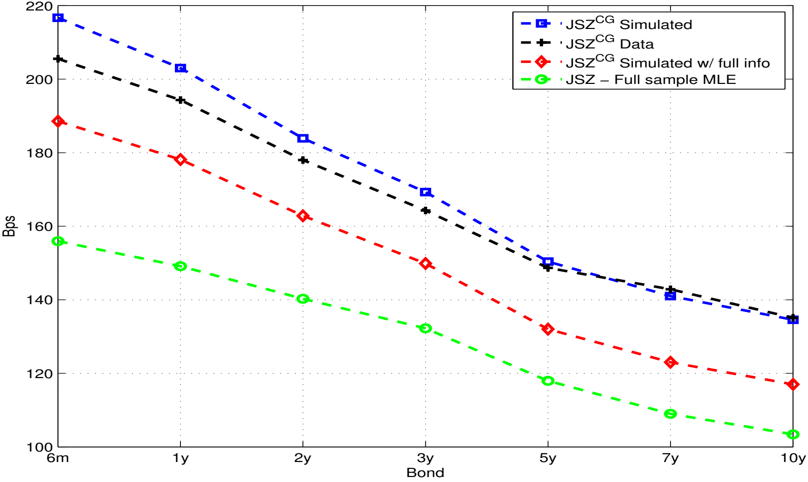
RMSE's One-Year Ahead Forecasts January, 2000 – December, 2007

	RMSE's by Bond Maturity						
Rule	6m	1Y	2Y	3Y	5Y	7Y	10Y
$\ell(RW)$	173	165	143	125	98	79	60
$\ell(BCFF)$	178	165	156	144	116	98	79
$\ell(JSZ)$	181	176	163	145	118	97	75
$\ell(JSZ_{CG})$	166	159	145	128	104	86	69
$\ell(JPS)$	141	138	125	109	86	71	64

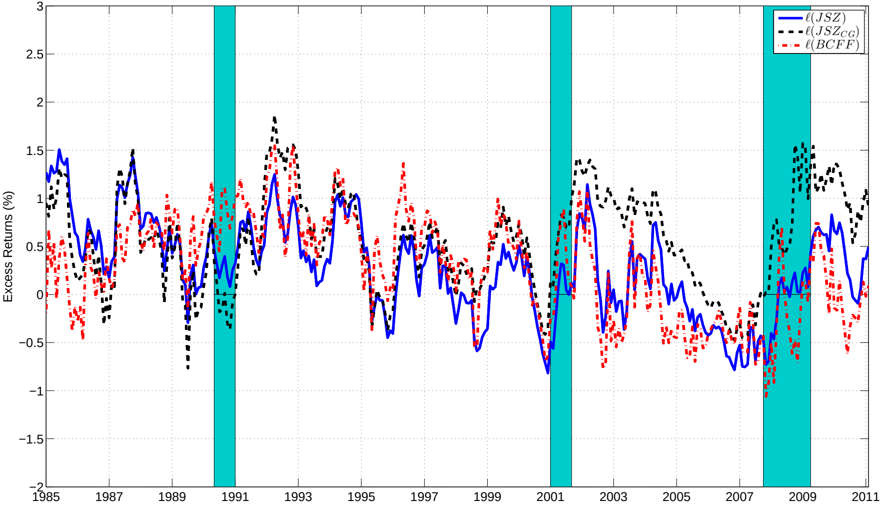
RMSE's One-Year Ahead Forecasts January, 2008 – March, 2012

	RMSE's by Bond Maturity						
Rule	6m	1Y	2Y	3Y	5Y	7Y	10Y
$\ell(RW)$	75	75	67	67	76	78	69
$\ell(BCFF)$	116	118	129	148	122	119	94
$\ell(JSZ)$	100	97	102	103	98	85	67
$\ell(JSZ_{CG})$	78	76	76	79	82	79	71
$\ell(JPS)$	92	87	79	75	77	76	78

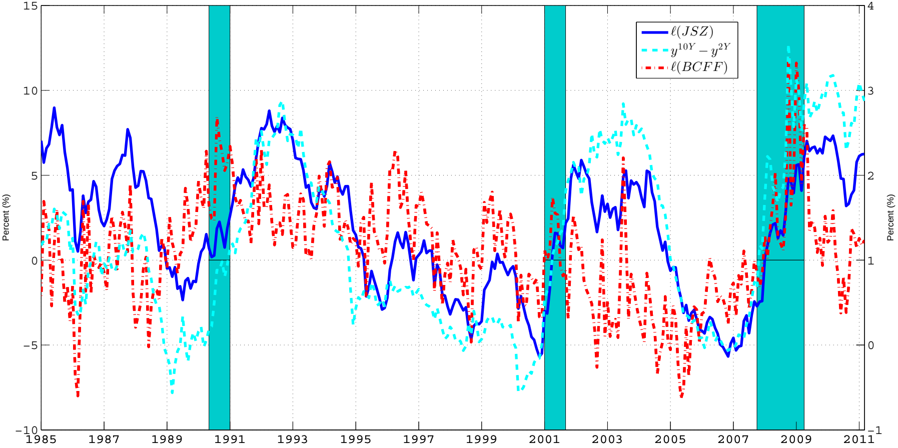
Imprecision with Learning
January, 1975 – March, 2011



Expected Excess Returns on Two-Year Treasury Bonds



Expected Excess Returns on Ten-Year Treasury Bonds



Why is $\ell(JSZ)$ Different From $\ell(BCFF)$?

- Post recessions *BCFF* forecasters incorrectly predict rising 10-year yields. Partly a consequence of *BCFF* forecasters predicting that slope will be more persistent than it is.
- Notably, less than 25% of the variation of *BCFF*-implied expected excess returns are explained by variations in $\mathcal{P}$.
- At the same time, 25% of the variation of expected excess returns in JSZ_{CG} are orthogonal to $\mathcal{P}$.

Which Forecasters Were More Accurate?

- Full sample: RMSE's in forecasting the realized excess returns for bearing $(2y, 10y)$ bond risks were:
  - $(1.55\%, 9.68\%)$ for $\ell(BCFF)$ and
  - $(1.50\%, 8.43\%)$ for $\ell(JSZ)$.
- For the specific episode over January, 2001 through January, 2006, the corresponding RMSE's were:
  - $(1.34\%, 7.62\%)$ for $\ell(BCFF)$ and
  - $(1.40\%, 4.60\%)$ for $\ell(JSZ)$.

Learning About Volatility

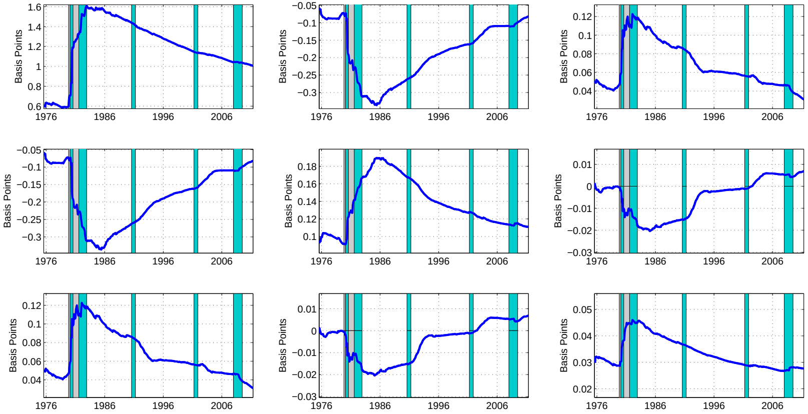
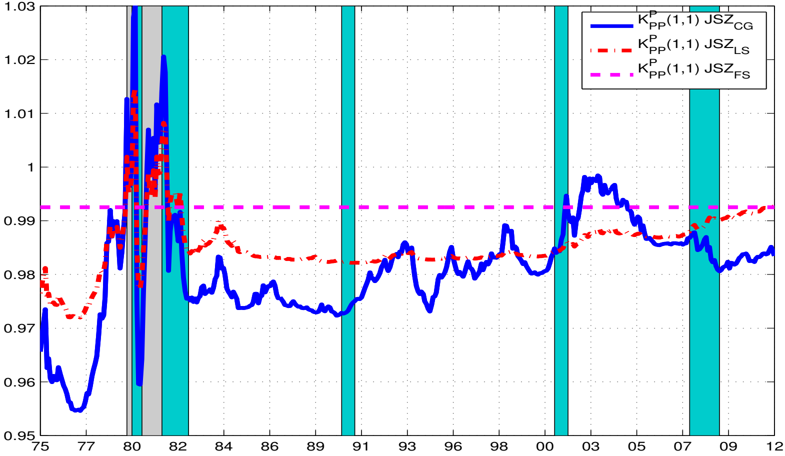


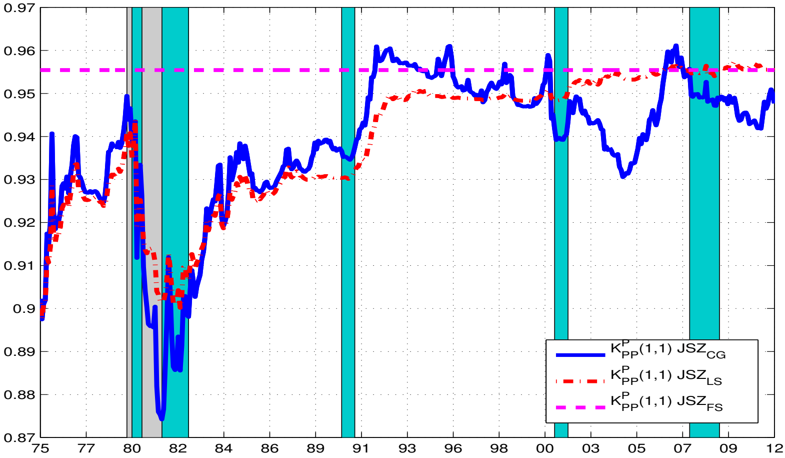
Figure: Estimates from $\ell(JPS)$ of $\Sigma_{\mathcal{P}}$, the innovation covariance matrix for $\mathcal{P}_t$, over the period June, 1975 to March, 2011.

Learning About the Drift: $K^{\mathbb{P}}_{\mathcal{PP}}(1,1)$.

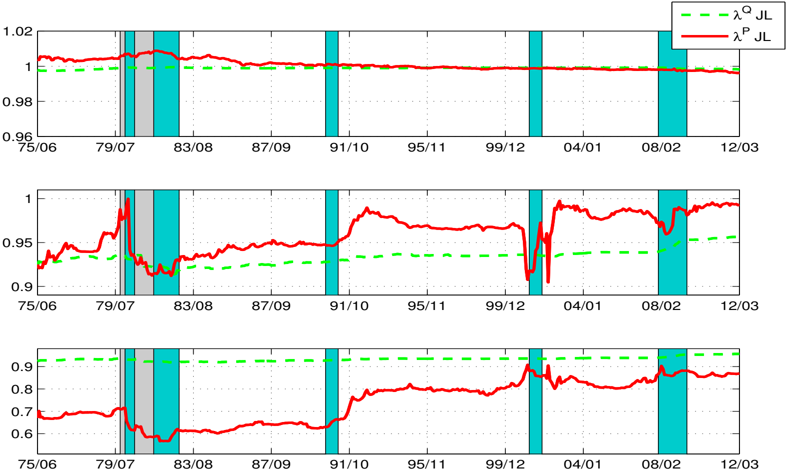


STANFORD UNIVERSITY

Learning About the Drift: $K^{\mathbb{P}}_{\mathcal{PP}}(2,2)$.

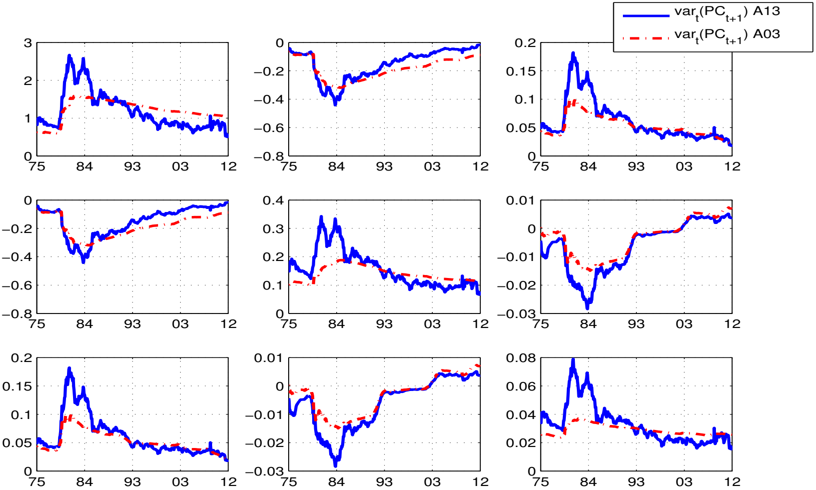


Learning In the Presence of Stochastic Volatility

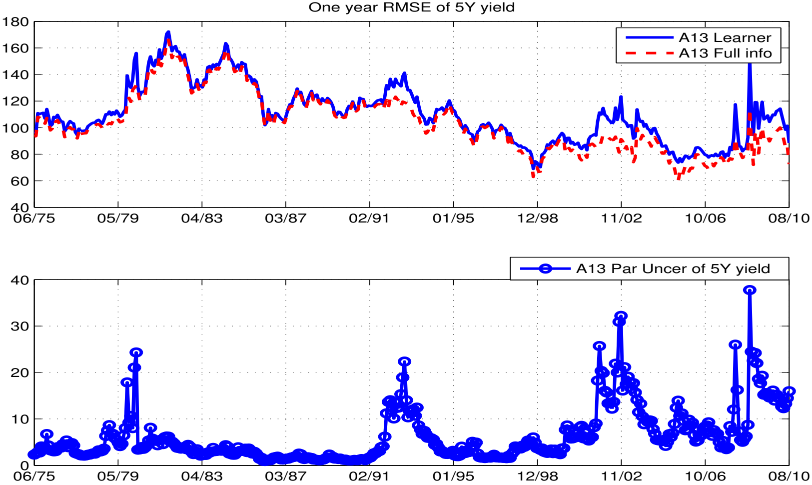


STANFORD UNIVERSITY

Learning About Volatility



Parameter Updating with Stochastic Volatility



STANFORD UNIVERSITY

Joslin, S., A. Le, and K. Singleton, 2013, Gaussian Macro-Finance Term Structure Models with Lags, *Journal of Financial Econometrics* 11, 581–609.

Joslin, S., M. Priebsch, and K. Singleton, 2013, Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks, Working paper, forthcoming, *Journal of Finance*.