Learning the Macro-Dynamics of U.S. Treasury Yields with Arbitrage-free Term Structure Models

Discussion

Jessica A. Wachter

March 28, 2014

This paper

- This paper studies parameter uncertainty, learning, and forecasting with dynamic term structure models.
- The models in this paper are very rich. They provide an empirically plausible account of bond yields in a way that is consistent with no-arbitrage.
- This very richness makes studying parameter uncertainty, etc. a challenge.
- However, the benefits are that we learn more by looking at realistic models.

Model

- 3 factors Z_t:

$$Z_{t+1} = K_0^{\mathbb{P}} + K_Z^{\mathbb{P}} Z_t + \Sigma_Z^{1/2} e_{Z,t+1}^{\mathbb{P}},$$

where $e_{Z,t+1}^{\mathbb{P}} \overset{\text{iid}}{\sim} N(0, I)$.

- Short-rate process

$$r_t = \rho_0 + \rho_Z Z_t.$$

- Prices of risk

$$\Lambda_{Zt} = \Lambda_0 + \Lambda_1 Z_t.$$

- Stochastic Discount Factor

$$\log \mathcal{M}_{t+1} = -r_{t+1} - \Lambda_{Zt}^{\top} e_{t+1}^{\mathbb{P}} - \frac{1}{2} \Lambda_{Zt}^{\top} \Lambda_{Zt}.$$

Bond pricing

- Let $\Theta = \{K_0^{\mathbb{P}}, K_Z^{\mathbb{P}}, \Sigma_Z, \rho_0, \rho_Z, \Lambda_0, \Lambda_1\}$.
- Bond prices: $D_t^m = E_t\left[\mathcal{M}_{t+1} D_{t+1}^{m-1}\right]$ with boundary condition $D_t^0 = 1$.
- 3 factors implies that 3 bonds will be priced without error, but what about the others?
- Possibilities
  - 3 bonds priced without error, assume others are priced with error. Conditional on Θ, Z_t is observed.
  - All bonds priced with error, Z_t unobserved.

This paper: First 3 PCs are priced without error, other linear combinations priced with error. Conditional on Θ, Z_t is observed.
In fact, Z_t equals the 3 PCs.

P1: The naive econometrician forecasts bond yields

Let Z_1^t = history of Z_t, $\mathcal{O}_1^t$ = history of yields. At t, the forecaster

1. Maximizes the likelihood $f(Z_1^t, \mathcal{O}_1^t | \Theta, \Sigma_{\mathcal{O}})$, implying values $\hat{\Theta}_t, \hat{\Sigma}_{\mathcal{O},t}$.
2. Creates forecasts of Z_{t+h}

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

3. Which imply forecasts of yields

$$\hat{y}_{t+h}^{m} = A_m(\hat{\Theta}) + B_m(\hat{\Theta})\hat{Z}_{t+h}$$

"This is naive for both forward- and backward-looking reasons."

Why is this forecast naive?

Forecasts of future bond yields ... are based on the fitted vector-autoregression assuming that Θ is fixed at the current estimate $\hat{\Theta}_t$ *even though* $\hat{\Theta}_{t+1}$ *will in fact change with the arrival of new information*.

This learning rule is also naive looking backwards, because $\hat{\Theta}_t$ is updated by estimating a likelihood function over the sample up to date t presuming that Θ is fixed and *has never changed in the past even though* $\hat{\Theta}_t$ *did change every month*.

P2: A Bayesian econometrician forecasts bond yields

The Bayesian knows what he doesn't know.

1. Prior distribution over the parameters: $p(\Theta, \Sigma_{\mathcal{O}})$
2. Likelihood function as of time t: $f(Z_1^t, \mathcal{O}_1^t | \Theta, \Sigma_{\mathcal{O}})$
3. Posterior distribution

$$p_t(\Theta, \Sigma_{\mathcal{O}} | Z_1^t, \mathcal{O}_1^t) \propto f(Z_1^t, \mathcal{O}_1^t | \Theta, \Sigma_{\mathcal{O}}) p(\Theta, \Sigma_{\mathcal{O}}).$$

4. Predictive distribution:
   1. Draw $\tilde{\Theta}$ from the posterior
   2. Draw $\tilde{Z}_{t+h}$ from multivariate normal implied by VAR and $\tilde{\Theta}$
   3. Calculate yield as function of the $\tilde{\Theta}$ and $\tilde{Z}_{t+h}$

Comparing P1 (Naive) and P2 (Bayesian)

- P2 is harder, probably, and most likely implies forecasts similar to P1.
- Why? Uncertainty could enter through convexities in bond pricing. There's probably not enough convexity, and not enough parameter uncertainty, for this to make a big difference for first moments.
- Isn't the Bayesian econometrician also being a bit naive?

P3: A Bayesian rep. agent prices bonds

- The agent observes factors Z_t and infers parameters through Bayesian updating from the VAR.
- Are r_t and Λ_{Zt} also unknown? Don't these depend at least partially on the agent's utility function?
- r_t and Λ_{Zt} are themselves equilibrium objects that will be affected by learning. The arrival of new information represents a risk to the agent that may be priced.
- Equilibrium bond prices:

$$D_t^m = E_t^{\mathrm{RA}}\left[\mathcal{M}_{t+1} D_{t+1}^{m-1} | Z_t^1\right]$$

where E^{RA} denotes expectations taken with respect to the posterior distribution of the representative agent.

An example of P3

- Assume a representative agent with power utility. Log endowment growth follows

$$\Delta c_{t+1} \overset{\text{iid}}{\sim} N(\mu, \sigma)$$

- Assume μ is unknown to the representative agent.
- Let $\hat{\mu}_t$ denote the mean of the agent's posterior distribution and $\hat{\sigma}_t$ the standard deviation of the predictive distribution for Δc_{t+1}.
- In equilibrium

$$r_t = -\log \beta + \gamma \hat{\mu}_t - \frac{1}{2}\hat{\sigma}_t^2$$

- Negative shocks to consumption lower $\hat{\mu}_t$, lower r_t, and raise bond prices. Thus bonds are a hedge, and learning lowers risk premia.

Comparing P2 and P3

- Both are Bayesian models in which agents learn about the parameters. They differ in what is being learned about and what information is being used.
- The learning model in this paper combines a bit of both.

What does this paper do?

1. The full Bayesian approach. The agent prices bonds using:

$$D_t^m = \int E^{\mathbb{Q}} \left[\prod_{s=1}^{m} e^{-r_{t+1}} \mid \Theta_t^{\mathbb{Q},t+m+1} \right] f^{\mathbb{Q}} \left(\Theta_t^{\mathbb{Q},t+m-1} \mid Z_1^t, \mathcal{O}_1^t \right),$$

and updates $\Theta^{\mathbb{P}} \subset \Theta$ using the VAR on Z_t.

- How does the agent form $f^{\mathbb{Q}} \left(\Theta_t^{\mathbb{Q},t+m-1} \mid Z_1^t, \mathcal{O}_1^t \right)$?
- Seems reasonable, but where does it come from?

What does this paper do? (cont.)

2. The naive approach.
3. In-between: the semi-consistent ($\mathcal{SC}$) learner.
   - Derive posterior distribution for $\Theta^{\mathbb{P}}$ using a VAR, as in P3 – except with yields.
   - Use the mean of this posterior distribution to calculate forecasts $\hat{Z}_{t+h}$.
   - Using these forecasts, and $\Theta^{\mathbb{Q}}$ from MLE (?), construct yield forecasts.

Comments:

- $\mathcal{SC}$ is a tractable way to bring in a degree of parameter uncertainty. However, I struggle with the economic interpretation of this learning framework.
- In the end, $\mathcal{SC}$ and Naive are similar for forecasting.

Root-mean-squared forecasting errors

	Panel (a): RMSE's (in basis points) for Quarterly Horizon						
Rule	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	51.6	52.4	54.3	49.5	47.9	44.8
	()	()	()	()	()	()	()
	[4.10]	[3.28]	[4.48]	[5.03]	[4.86]	[3.40]	[3.54]
$\ell(JSZ)$	39.7	41.8	45.2	44.6	43.0	41.2	37.7
	(−4.03)	(−3.07)	(−3.92)	(−5.28)	(−4.39)	(−3.92)	(−3.33)
	[1.96]	[0.76]	[2.85]	[1.31]	[0.65]	[0.08]	[0.27]
$\ell(JSZ_{CG})$	38.5	41.6	45.2	45.0	43.4	42.1	38.8
	(−4.36)	(−3.17)	(−3.80)	(−4.45)	(−4.10)	(−3.66)	(−2.96)
	[0.50]	[0.48]	[3.05]	[1.55]	[1.20]	[1.21]	[2.01]
$\ell(JPS)$	36.2	41.2	44.2	43.9	41.4	40.7	39.3
	(−3.96)	(−2.74)	(−2.99)	(−3.86)	(−4.71)	(−3.94)	(−2.64)
	[−0.78]	[0.04]	[0.57]	[0.13]	[−1.20]	[−0.41]	[1.26]

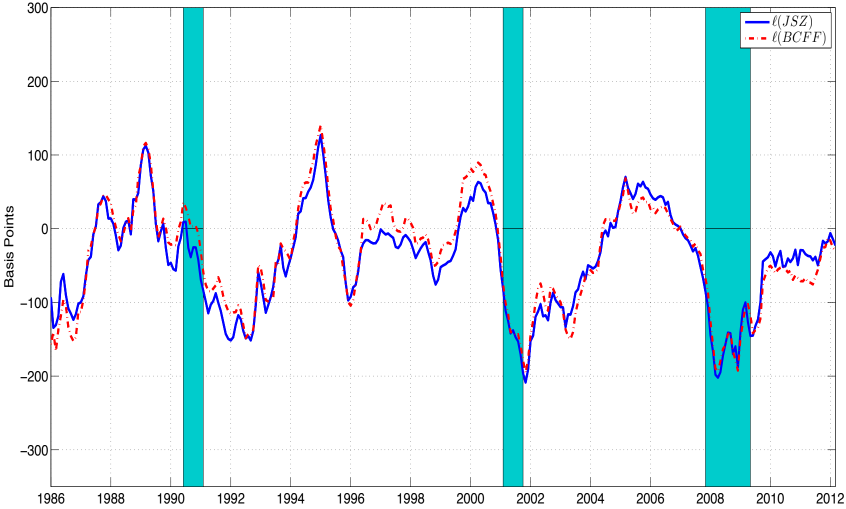
Root-mean-squared forecasting errors

	Panel (b): RMSE's (in basis points) for Annual Horizon						
Rule	6m	1Y	2Y	3Y	5Y	7Y	10Y
$\ell(RW)$	136.2	135.3	126.3	118.0	107.3	102.2	96.0
$\ell(BCFF)$	148.2	144.6	140.1	136.2	119.6	113.9	106.0
	()	()	()	()	()	()	()
	[1.18]	[0.90]	[1.59]	[2.28]	[2.30]	[2.40]	[2.56]
$\ell(JSZ)$	141.7	140.6	134.7	125.9	111.7	102.3	92.9
	(−1.07)	(−0.51)	(−0.84)	(−1.61)	(−1.22)	(−1.66)	(−1.63)
	[0.75]	[0.77]	[1.26]	[1.28]	[0.81]	[0.02]	[−0.58]
$\ell(JSZ_{CG})$	137.3	136.6	130.5	122.5	110.7	104.1	97.4
	(−1.33)	(−0.92)	(−1.38)	(−1.93)	(−1.65)	(−1.85)	(−1.49)
	[0.19]	[0.26]	[0.92]	[1.01]	[1.14]	[0.72]	[0.50]
$\ell(JPS)$	130.4	130.7	123.3	114.4	101.8	96.5	92.8
	(−1.51)	(−1.31)	(−1.80)	(−2.52)	(−2.37)	(−2.23)	(−1.48)
	[−0.47]	[−0.42]	[−0.43]	[−0.72]	[−1.44]	[−1.12]	[−0.51]

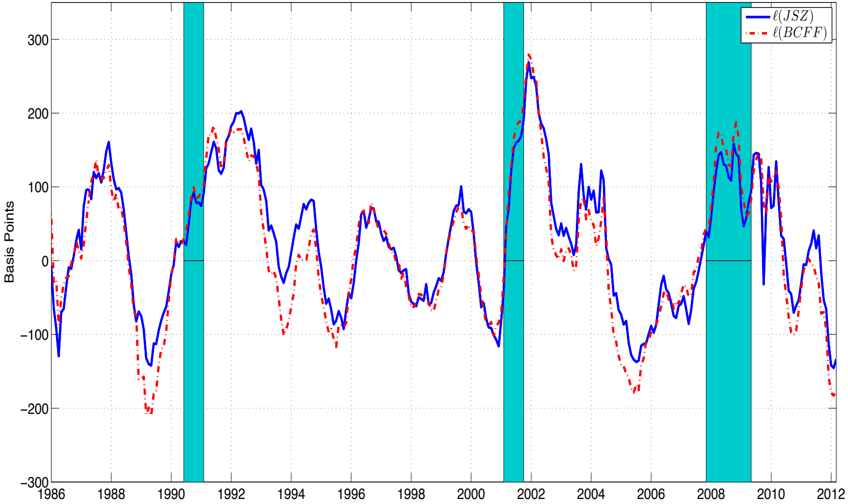
Results

- Learning rules from $\mathcal{SC}$ offer improvements, often significant ones, over professional forecasters.
- They do not offer significant improvements over the random walk model.
- Out-of-sample forecasting is interesting but may not be a powerful model diagnostic.

Forecasts of the level factor



Forecasts of the slope factor



Errors vs. shocks

Forecasting "errors" combine two quantities:

1. Errors in capturing the correct conditional distribution of yields
2. Not knowing the future.

If its only 2, then errors should be uncorrelated (might be difficult to assess in a finite sample).

Errors vs. shocks

- Note that even 2 is not measurement error in a traditional sense: shocks are correlated with future yields,
- Taking this into account affects inference from the VAR: Inference is non-standard and posterior distributions of parameters are no longer normally distributed.
- Standard normalizations, effectively taking the mean as known, may not be harmless.