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 $Z_t$:

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

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

- 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

$$\log M_{t+1} = -r_{t+1} - \Lambda_{Zt}^T e_{t+1}^P - \frac{1}{2} \Lambda_{Zt}^T \Lambda_{Zt}.$$
Bond pricing

- Let $\Theta = \{K_0^P, K_Z^P, \Sigma_Z, \rho_0, \rho_Z, \Lambda_0, \Lambda_1\}$.
- Bond prices: $D_t^m = E_t \left[ M_{t+1} D_{t+1}^{m-1} \right]$ with boundary condition $D_0^0 = 1$.
- 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 $\Theta$, $Z_t$ 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 $\Theta$, $Z_t$ is observed.
- In fact, $Z_t$ equals the 3 PCs.
The naive econometrician forecasts bond yields

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

1. Maximizes the likelihood $f(Z_t^t, \mathcal{O}_t^t | \Theta, \Sigma_\mathcal{O})$, implying values $\hat{\Theta}_t, \hat{\Sigma}_\mathcal{O}, t$.

2. Creates forecasts of $Z_{t+h}$

$$\hat{Z}_{t+h} = \hat{K}_0^0 + \left( \hat{K}_{Zt}^P \right)^0 \hat{K}_0^0 + \cdots + \left( \hat{K}_{Zt}^P \right)^{h-1} \hat{K}_0^0 + \left( \hat{K}_{Zt}^P \right)^h Z_t$$

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 $\Theta$ 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 $\Theta$ 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: \( p(\Theta, \Sigma) \)
2. Likelihood function as of time \( t \): \( f(Z_1^t, O_1^t|\Theta, \Sigma) \)
3. Posterior distribution

\[
p_t(\Theta, \Sigma | Z_1^t, O_1^t) \propto f(Z_1^t, O_1^t|\Theta, \Sigma)p(\Theta, \Sigma).
\]

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 $\Lambda_{Z_t}$ also unknown? Don’t these depend at least partially on the agent’s utility function?
- $r_t$ and $\Lambda_{Z_t}$ 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:

$$D^m_t = E^{RA}_t \left[ M_{t+1} D^{m-1}_{t+1} | Z^1_t \right]$$

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

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

- Assume \( \mu \) 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 \( \Delta c_{t+1} \).

- 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?

The full Bayesian approach. The agent prices bonds using:

\[
D_t^m = \int E^Q \left[ \prod_{s=1}^{m} e^{-r_{t+1} | \Theta_{t}^{Q,t+m+1}} \right] f^Q \left( \Theta_{t}^{Q,t+m-1} | Z_t^1, O_t^1 \right),
\]

and updates \( \Theta^P \subset \Theta \) using the VAR on \( Z_t \).

- How does the agent form \( f^Q \left( \Theta_{t}^{Q,t+m-1} | Z_t^1, O_t^1 \right) \)?
- Seems reasonable, but where does it come from?
What does this paper do? (cont.)

2 The naive approach.

3 In-between: the semi-consistent (SC) learner.
   - Derive posterior distribution for $\Theta^P$ 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 $\Theta^Q$ from MLE (?), construct yield forecasts.

Comments:

- 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, SC and Naive are similar for forecasting.
### Root-mean-squared forecasting errors

<table>
<thead>
<tr>
<th>Rule</th>
<th>6m</th>
<th>1Y</th>
<th>2Y</th>
<th>3Y</th>
<th>5Y</th>
<th>7Y</th>
<th>10Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ell(RW)$</td>
<td>38.0</td>
<td>41.1</td>
<td>43.3</td>
<td>43.7</td>
<td>42.4</td>
<td>41.1</td>
<td>37.5</td>
</tr>
<tr>
<td>$\ell(BCFF)$</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td></td>
<td>4.10</td>
<td>3.28</td>
<td>4.48</td>
<td>5.03</td>
<td>4.86</td>
<td>3.40</td>
<td>3.54</td>
</tr>
<tr>
<td>$\ell(JSZ)$</td>
<td>39.7</td>
<td>41.8</td>
<td>45.2</td>
<td>44.6</td>
<td>43.0</td>
<td>41.2</td>
<td>37.7</td>
</tr>
<tr>
<td></td>
<td>-4.03</td>
<td>-3.07</td>
<td>-3.92</td>
<td>-5.28</td>
<td>-4.39</td>
<td>-3.92</td>
<td>-3.33</td>
</tr>
<tr>
<td></td>
<td>1.96</td>
<td>0.76</td>
<td>2.85</td>
<td>1.31</td>
<td>0.65</td>
<td>0.08</td>
<td>0.27</td>
</tr>
<tr>
<td>$\ell(JSZ_{CG})$</td>
<td>38.5</td>
<td>41.6</td>
<td>45.2</td>
<td>45.0</td>
<td>43.4</td>
<td>42.1</td>
<td>38.8</td>
</tr>
<tr>
<td></td>
<td>-4.36</td>
<td>-3.17</td>
<td>-3.80</td>
<td>-4.45</td>
<td>-4.10</td>
<td>-3.66</td>
<td>-2.96</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.48</td>
<td>3.05</td>
<td>1.55</td>
<td>1.20</td>
<td>1.21</td>
<td>2.01</td>
</tr>
<tr>
<td>$\ell(JPS)$</td>
<td>36.2</td>
<td>41.2</td>
<td>44.2</td>
<td>43.9</td>
<td>41.4</td>
<td>40.7</td>
<td>39.3</td>
</tr>
<tr>
<td></td>
<td>-3.96</td>
<td>-2.74</td>
<td>-2.99</td>
<td>-3.86</td>
<td>-4.71</td>
<td>-3.94</td>
<td>-2.64</td>
</tr>
<tr>
<td></td>
<td>-0.78</td>
<td>0.04</td>
<td>0.57</td>
<td>0.13</td>
<td>-1.20</td>
<td>-0.41</td>
<td>1.26</td>
</tr>
</tbody>
</table>
Root-mean-squared forecasting errors

<table>
<thead>
<tr>
<th>Rule</th>
<th>Panel (b): RMSE’s (in basis points) for Annual Horizon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6m</td>
</tr>
<tr>
<td>$\ell(RW)$</td>
<td>136.2</td>
</tr>
<tr>
<td>$\ell(BCFF)$</td>
<td>148.2</td>
</tr>
<tr>
<td>$\ell(JSZ)$</td>
<td>141.7</td>
</tr>
<tr>
<td>$\ell(JSZ_{CG})$</td>
<td>137.3</td>
</tr>
<tr>
<td>$\ell(JPS)$</td>
<td>130.4</td>
</tr>
</tbody>
</table>

Also, after the Federal Reserve's shift away from their experimental monetary rule in the early January, 1985 to March, 2012. The D-M statistics for the di

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 41.6 40.7

$\ell(BCFF)$ 38.3 39.1 40.6 41.5 40.2 39.4 38.5

$\ell(JSZ)$ 37.6 38.4 39.5 40.7 40.1 39.3 38.5

$\ell(JPS)$ 37.1 38.0 39.2 40.3 39.7 39.0 38.2

Wachter (Wharton)  Macro-Dynamics discussion  March 28, 2014  15 / 20
Results

- Learning rules from 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

Figure 3: Comparison of forecast errors (realization minus forecast) for PC1 and PC2 of rule (BCFF) against rule (JSZ) for the horizon of one year. The sample is January, 1985 through March, 2012.

There is a larger gain for the one-year horizon (which was used to select), particularly for bonds with maturities under three years. These RMSE's are, however, misleading summaries of the degree to which the forecasts from (BCFF) and (JSZ) track each other. In fact, the tracking for both PC1 and PC2 is much inferior for (JSZ) than for (JSZ). This deterioration in tracking of BCFF forecasts comes with the benefit of much more accurate forecasts by (JSZ) du ring crisis, especially from mid-2009 onwards.

5 Bayesian Learning with Consistent Pricing

Our Bayesian learning rules bring greater sophistication to the consensus agent, a conceptual improvement over our naive learning rules, with some potential costs associated with added structure on agents' beliefs. The Bayesian agent specifies the joint distribution of her beliefs for a high-dimensional $\theta$ and consistently updates these beliefs as new information becomes available. This requires specifying which parameters are known, which are unknown but constant, and which are drifting or state dependent, and for the latter parameters she must specify their laws of motion. With these beliefs in place, she then solves for prices that are consistent with the assumptions on parameter uncertainty.

Two learning schemes are explored in depth. In the first, $\theta$ is presumed fixed over time and the consensus agent is learning its value. The second environment has the agent learning about the parameters governing the conditional $P$-mean ($\theta_P$) as $\theta_P$ drifts according to the known law of motion (18) – a random walk reflecting permanent structural changes – presuming that the remaining parameters are fixed (but still unknown). Our interest in the second case is motivated in part by the debate in the macroeconomics literature about whether changes in monetary policies in the U.S. had material e↵ects on the dynamics of VAR models of the macroeconomy (see, e.g., Cogley and Sargent (2005) and Sims and Zha (2006)).

Even for the simplest of these learning problems where all of the parameters are fixed there...
Forecasts of the slope factor

Figure 3: Comparison of forecast errors (realization minus forecast) for PC1 and PC2 of rule \( BCFF \) against rule \( JSZ \) for the horizon of one year. The sample is January, 1985 through March, 2012.

Even for the simplest of these learning problems where all of the parameters are fixed there...
Errors vs. shocks

Forecasting “errors” combine two quantities:

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

If it's 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.