Discussion of Doshi, Jacobs and Liu
"Loss Functions for Forecasting Treasury Yields"

Francis X. Diebold
University of Pennsylvania

November 4, 2015

There's a Lot to Like in This Paper

– Prediction under the relevant loss function
deserves lots of attention

– The yield curve model used for prediction
deserves lots of attention

– Maybe even yield curve curvature
deserves lots of attention

Prediction Under the Relevant Loss Function

Prediction is Key in an Evidence-Based Macro-Finance

History: $\{y_t\}_{t=1}^{T}$

Realization and prediction: $y_{T+h},\ \hat{y}_{T+h,T}$

Error: $e_{T+h,T} = y_{T+h} - \hat{y}_{T+h,T}$

Loss: $L(e_{T+h,T})$

Accuracy comparison via expected loss: $E(L(e_{T+h,T}))$

What is the Relevant Loss Function, $L(e_{T+h,T})$?

What is the horizon, h? Short term? Long term?

– Doshi et al. (2015)
(this paper)

What is the loss function L? $L(e) = e^2$? $L(e) = |e|$?

– Diebold and Shin (2015),
"Assessing Point Forecast Accuracy by Stochastic Error Distance"

Estimation Under the Relevant Loss Function: Shines in Principle

Correct specification:

– We learn the truth asymptotically

– It is best for all purposes

Incorrect specification:

– We never learn the truth, even asymptotically
Instead we learn a "best approximation," induced by L

– MLE effectively ties our hands and picks L

– Instead, think hard about the relevant loss function

– Best approximation for one purpose generally very different from (and not implied by) best approximation for another purpose
e.g., mis-specified $AR(1)$: $\hat{\rho}^{10} \neq \widehat{\rho^{10}}$, even as $T \rightarrow \infty$

Estimation Under the Relevant Multi-Step Loss Function: Flops in Practice

Of course everyone knows Weiss (1996, *J. Applied Econometrics*)

But there's a "file drawer problem"

Marcellino-Stock-Watson (2006, *Journal of Econometrics*)
is very clearly negative and not cited
("A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series")

The Yield Curve Model Used for Prediction

Successful Time-Series Prediction Requires *Parsimony*

– Selection

– Bayesian shrinkage

– Lasso

"The Parsimony Principle"

For prediction,
"maximally-flexible" models are not appealing

An Appealing Predictive Model

Arbitrage-Free Nelson-Siegel (AFNS)
(Christensen et al., 2011, *Journal of Econometrics*)

$$y_t = \Lambda f_t + \varepsilon_t$$

$$f_t = \Phi f_{t-1} + \eta_t$$

– Three (latent) factors; provably level, slope, curvature

– Factors are latent but estimation is trivial and reliable

– Easily accommodates the zero lower bound,non-spanning, etc.

– *Structure placed on factor loadings (Λ matrix)*
[Equivalently, structure on Duffie-Kan state-transition dynamics]
[Equivalently, structure on maximally-flexible $A_0(3)$]

– Joslin et al. (2011, *Review of Financial Studies*)
test the restrictions and find $p \approx 1/2$

Yield Curve Curvature

???

In Conclusion: What I'd Like to See

– 1-step *vs.* h-step estimation

– Squared-error *vs.* absolute-error loss

– AFNS $A_0(3)$ *vs.* JSZ maximally-flexible $A_0(3)$
(and drop the latent-state maximally-flexible affine models)

– Robustness to sample start date, sample end date,
in-sample / out-of-sample split, estimation method, etc.

– Progress in understanding curvature