“Dynamic Prediction Pools: An Investigation of Financial Frictions and Forecasting Performance” by Marco Del Negro, Raiden Hasegawa, and Frank Schorfheide

Discussion by Bartosz Maćkowiak, *European Central Bank and CEPR*

Frankfurt, 13 June 2014

Combining forecasts from different models

The predictive density of y' given data y and models $\mathcal{M}_k$, $k = 1, ..., K$:

$$p\left(y' \mid y\right) = \sum_k \pi\left(\mathcal{M}_k \mid y\right) p\left(y' \mid \mathcal{M}_k, y\right) \tag{1}$$

where

$$\pi\left(\mathcal{M}_k \mid y\right) = \frac{p\left(y \mid \mathcal{M}_k\right) \pi\left(\mathcal{M}_k\right)}{\sum_k p\left(y \mid \mathcal{M}_k\right) \pi\left(\mathcal{M}_k\right)}$$

The weights in (1) are the models' posterior probabilities, determined by the marginal likelihoods and the models' prior probabilities.

Marginal likelihood example

If $\mathcal{M}_k$ is an AR model of y,

$$p\left(y \mid \mathcal{M}_k\right) = p\left(y_T, ..., y_1 \mid y^0, \mathcal{M}_k\right)$$

$$= \prod_{t=0}^{T-1} p\left(y_{t+1} \mid y^t, \mathcal{M}_k\right)$$

$$= \prod_{t=0}^{T-2} p\left(y_{t+2}, y_{t+1} \mid y^t, \mathcal{M}_k\right) = \prod_{t=0}^{T-3} p\left(y_{t+3}, y_{t+2}, y_{t+1} \mid y^t, \mathcal{M}_k\right) = \; ...$$

Marginal likelihood measures the overall out-of-sample predictive performance of a model (*not* predictive performance at a particular horizon).

Challenges

1. Each model being compared must be a model of *the same data* y.

2. It must be that "each of the discrete [models] makes scientific sense, and there are no (...) models in between." Gelman et al. (1995), p.176.

   (a) If the space of models is too sparse, posterior probabilities of models tend to come out implausibly decisive and to display a bang-bang pattern over time.

It is clear *in principle* how to confront these challenges.

Confronting the challenges in practice

- Geweke and Amisano (2011) form a weighted sum of predictive densities (here, $K = 2$):

$$\lambda * p\left(y_{1,t+h} \mid y_1^t, y_2^t, \mathcal{M}_1\right) + (1-\lambda) * p\left(y_{1,t+h} \mid y_1^t, y_3^t, \mathcal{M}_2\right)$$

where $\lambda \in [0,1]$, $h \geq 1$, and maximize the product of these sums w.r.t. λ.

- This paper proposes:

$$\lambda_t * p\left(y_{1,t+h} \mid y_1^t, y_2^t, \mathcal{M}_1\right) + (1-\lambda_t) * p\left(y_{1,t+h} \mid y_1^t, y_3^t, \mathcal{M}_2\right)$$

plus a law of motion for λ_t.

Implementation

- Nonlinear stace-space system: the 2nd expression on the previous slide is the measurement equation, the law of motion for λ_t is the transition equation.

- Use a particle filter to infer $\lambda_{1:T}$, also infer parameters of the law of motion for λ_t.

- When inferring parameters of $\mathcal{M}_1$ and $\mathcal{M}_2$, neglect information in $\lambda_{1:T}$. This is reasonable.

Application

- Forecast growth rates of Y and P using DSGE models: $SW\pi$ and $SWFF$.
  - $SWFF$ has an extra observable: corporate bond spread.
  - Sample starts in 1964Q1, forecast evaluation in 1992Q1-2011Q2, real-time data.

- The dynamic pool yields good forecasts.

- λ_t varies considerably over time, while staying away from 0 and 1.

Takeaways

- This is a very useful methodology.
- Paying *some* attention to the corporate bond spread was a good idea throughout the evaluation period.
- Paying *a lot* of attention to the corporate bond spread was a good idea already before the Lehman crisis.
- Let's not stop here, let's learn from the evidence and improve our models.

Back to the challenges: sparsity

- The space of models seems too sparse.

- I agree that a nonlinear encompassing model (e.g., a DSGE with regime switches) seems worth exploring in the future.

- One could also use that model as a prior for a less restricted model (e.g., a VAR with regime switches), with the weight of the prior distributed continuously and inferred rather than fixed.

  – In analogy to the DSGE-VAR of Del Negro and Schorfheide (2004).

Back to the challenges: modelling all the data y

- In principle, it is possible to form an encompassing model and to think of models omitting particular elements of y as restricted versions of the encompassing model.

- Jarociński and Maćkowiak (2014) show that the posterior probability of the relevant restriction can be expressed analytically in a Gaussian VAR with a conjugate prior.
  - Like this paper, we find that the corporate bond spread was useful for forecasting Y and P already before the Lehman crisis.