Appendix to "A Bayesian Approach to Optimal Monetary Policy with Parameter and Model Uncertainty"

Timothy Cogley,[*] Bianca De Paoli,[†] Christian Matthes,[‡]
Kalin Nikolov,[§] and Tony Yates[¶]

July 2009

1 Model averaging when models involve different variables

Suppose that two models involve distinct lists of variables, Y_1 and Y_2, respectively. For model averaging, we want posterior model probabilities $p(M_1|Y)$ and $p(M_2|Y)$ for some conditioning set Y. In order for this comparison to be meaningful, the probabilities must be conditioned on a common list of variables. In principle, we can arrive at a common variable list either by taking the union of variables across models or the intersection. However, conditioning on the union of variables is problematic because ill-defined nuisance terms appear in various places. For that reason, we develop a strategy for working with the intersection of model variables.

Let X represent the intersection of variables across models, and suppose that X_i represents variables that appear in model i but not in the other model. Then $Y_i = [X, X_i]$. Also, let M_i index model i and let θ_i represent its parameters. Finally, suppose that ϕ represents Taylor-rule parameters and that $l(\phi|\theta_i, M_i)$ represents expected loss conditional on model i and a calibration of its parameters θ_i.

One coherent approach would be to condition the entire analysis on the intersection of variables X. A downside is that precision is lost when estimating parameters. I.e., the posterior $p(\theta_i|X, M_i)$ conditioned on the smaller set of common variables is

[*]New York University. Email: tim.cogley@nyu.edu
[†]Bank of England. Email: bianca.depaoli@bankofengland.co.uk.
[‡]New York University. Email: cm1518@nyu.edu.
[§]Bank of England. Email: kalin.nikolov@bankofengland.co.uk.
[¶]Bank of England. Email: tony.yates@bankofengland.co.uk.

likely to be more diffuse than the posterior $p(\theta_i|Y_i, M_i)$ conditioned on the complete list of observables for model i. For policy, it is desirable to estimate parameters as well as possible. Thus, we outline a strategy for using the full set of variables Y_i for estimation and the smaller set of common variables X for model averaging.

For estimation, the posterior $p(\theta_i|Y_i, M_i)$ for each model and the marginal data density $p(Y_i|M_i)$ can be calculated via standard methods. For a given ϕ, the model-specific expected loss is

$$l(\phi|Y_i, M_i) = \int l(\phi|\theta_i, M_i)p(\theta_i|Y_i, M_i)d\theta_i. \tag{1}$$

Assuming an evenly-weighted sample from $p(\theta_i|Y_i, M_i)$, this can be approximated as

$$l(\phi|Y_i, M_i) \approx N^{-1} \textstyle\sum_{j=1}^{N} l(\phi|\theta_{ij}, M_i), \tag{2}$$

where θ_{ij} is the jth draw in the Monte Carlo sample and N is the total number of draws. To condition down from Y_i to X, we use importance sampling. Thus, consider

$$\begin{aligned} l(\phi|X, \theta_i, M_i) &= \int l(\phi|\theta_i, M_i)\frac{p(\theta_i|X, M_i)}{p(\theta_i|Y_i, M_i)}p(\theta_i|Y_i, M_i)d\theta_i, \\ &\approx N^{-1} \textstyle\sum_{j=1}^{N} l(\phi|\theta_{ij}, M_i)w(\theta_{ij}), \end{aligned} \tag{3}$$

where $w(\theta_{ij}) \equiv p(\theta_i|X, M_i)/p(\theta_i|Y_i, M_i)$. Expected loss conditioned on X can be calculated by taking weighted averages using posterior draws for the big information set Y_i. Thus, there is no need to re-simulate the posterior for the smaller set X.

We can make further progress by noting that

$$\frac{p(\theta_i|X, M_i)}{p(\theta_i|Y_i, M_i)} = \frac{p(X|\theta_i, M_i)p(\theta_i|M_i)/p(X|M_i)}{p(Y_i|\theta_i, M_i)p(\theta_i|M_i)/p(Y_i|M_i)} = \frac{p(X|\theta_i, M_i)}{p(Y_i|\theta_i, M_i)}\frac{p(Y_i|M_i)}{p(X|M_i)}. \tag{4}$$

After the second equality, the first term is the ratio of likelihoods and the second is a ratio of marginalized likelihoods. The latter are unnecessary at this stage because they are independent of θ_i and wash out when normalizing importance weights so that they sum to 1. Thus, consider the unnormalized importance weights

$$\tilde{w}(\theta_i) = \frac{p(X|\theta_i, M_i)}{p(Y_i|\theta_i, M_i)}. \tag{5}$$

Since the normalizing constants $p(X|M_i)$ and $p(Y_i|M_i)$ are independent of θ_i, it follows that $\tilde{w}(\theta_i) = kw(\theta_i)$ for some constant k. Hence, after re-normalizing,

$$\frac{\tilde{w}(\theta_{ij})}{\sum_j \tilde{w}(\theta_{ij})} = w(\theta_{ij}). \tag{6}$$

The advantage of working with $\tilde{w}(\theta_i)$ follows from the fact that the numerator and denominator of (5) are easy to calculate. For a log-linearized system in state-space form, $\tilde{w}(\theta_i)$ can be evaluated using the prediction-error decomposition of the two likelihood functions and two passes of the Kalman filter.

The other ingredient needed for model averaging are the posterior model probabilities, $p(M_i|X)$. Suppose that $p(M_i)$ is the prior probability on model i. According to Bayes' theorem, the posterior model probability is

$$p(M_i|X) \propto p(X|M_i)p(M_i), \tag{7}$$

where

$$p(X|M_i) = \int p(X|\theta_i)p(\theta_i|M_i)d\theta_i. \tag{8}$$

Thus we need to calculate the marginal data density $p(X|M_i)$ for the common variable set. As stated above, $p(Y_i|M_i)$ can be approximated using standard methods. To condition down from Y_i to X, we apply a change of variables and integrate numerically. Start by re-writing (8) as

$$\begin{aligned} p(X|M_i) &= \int \frac{p(X|\theta_i)}{p(Y_i|\theta_i)} p(Y_i|\theta_i)p(\theta_i|M_i)d\theta_i, \\ &= \int \tilde{w}(\theta_i)\frac{p(Y_i|\theta_i)p(\theta_i|M_i)}{p(Y_i|M_i)} p(Y_i|M_i)d\theta_i, \\ &= p(Y_i|M_i)\int \tilde{w}(\theta_i)p(\theta_i|Y_i, M_i)d\theta_i. \end{aligned} \tag{9}$$

Hence the ratio of the small- and big-information set marginal data densities can be expressed as the posterior mean of unnormalized importance weights,

$$\frac{p(X|M_i)}{p(Y_i|M_i)} = \int \tilde{w}(\theta_i)p(\theta_i|Y_i, M_i)d\theta_i, \tag{10}$$

where the expectation is taken with respect to the big-information posterior. This can be approximated as

$$\frac{p(X|M_i)}{p(Y_i|M_i)} \approx N^{-1}\textstyle\sum_{j=1}^{N} \tilde{w}(\theta_{ij}). \tag{11}$$

In fact, this is proportional to the denominator in (6), so no new calculations are involved. To transform from $p(Y_i|M_i)$ to $p(X|M_i)$, we just multiply the former by the mean of the unnormalized importance weights,

$$p(X|M_i) \approx p(Y_i|M_i)\textstyle\sum_{j=1}^{N} \tilde{w}(\theta_{ij})/N. \tag{12}$$

Unnormalized model probabilities are found by substituting $p(X|M_i)$ into (7). To normalize, we sum the right-hand side of (7) across models and divide by the result,

$$p(M_i|X) = \frac{p(X|M_i)p(M_i)}{\sum_i p(X|M_i)p(M_i)}. \tag{13}$$

All the ingredients for model averaging are now at hand. To account for model uncertainty, we average the model-specific losses $l(\phi|X, M_i)$ using posterior model probabilities as weights,

$$l(\phi) = \sum_i l(\phi|X, M_i) p(M_i|X). \tag{14}$$

A policy rule robust to both model and parameter uncertainty can be found by choosing ϕ to minimize $l(\phi)$.