Model Comparison

Jesús Fernández-Villaverde
University of Pennsylvania

Model Comparison

- Assume models $1, 2, ..., I$ to explain Y^T. Let $M = \{1,\ 2, ...,\ I\}$.

- Let $\{\Theta_1, \Theta_2, ..., \Theta_I\}$ be associated parameter sets.

- Let $\left\{f(Y^T|\theta_1, 1), f(Y^T|\theta_2, 2), ..., f(Y^T|\theta_I, I)\right\}$ be associated likelihood functions.

- Let $\{\pi\left(\theta_1|1\right), \pi\left(\theta_2|2\right), ..., \pi\left(\theta_I|I\right)\}$ be associated prior distributions.

- Let $\{\pi\left(1\right), \pi\left(2\right), ..., \pi\left(I\right)\}$ be associated prior about the models.

Marginal Likelihood and Model Comparison

- Assume $\sum_{i=1}^{I} \pi(i) = 1$.

- Then Bayes rule implies posterior probabilities for the models:

$$\pi\left(i|Y^T\right) = \frac{\pi\left(i, Y^T\right)}{\sum_{i=1}^{k} \pi\left(i, Y^T\right)} = \frac{\pi(i)\, P(Y^T|i)}{\sum_{i=1}^{k} \pi(i)\, P(Y^T|i)}.$$

where $P\left(Y^T|i\right) = \int_{\Theta_{Mi}} f(Y^T|\theta_i, i)\pi(\theta_i|i)\, d\theta_i$

- This probability is the Marginal Likelihood.

Why is the Marginal Likelihood a Good Measure to Compare Models?

- Assume i^* is the true model, then:

$$\pi\left(i^*|Y^T\right) \longrightarrow 1 \text{ as } T \longrightarrow \infty.$$

- Why?

$$\pi\left(i^*|Y^T\right) = \frac{\pi\left(i^*\right) P(Y^T|i^*)}{\sum_{i=1}^{k} \pi\left(i\right) P(Y^T|i)} = \frac{\pi\left(i^*\right)}{\sum_{i=1}^{k} \pi\left(i\right) \frac{P(Y^T|i)}{P(Y^T|i^*)}}$$

- Under some regularity conditions, it can shown that:

$$\frac{P(Y^T|i)}{P(Y^T|i^*)} \longrightarrow 0 \text{ as } T \longrightarrow \infty \text{ for all } i \in M/\left\{i^*\right\}$$

An Important Point about Priors

- Priors need to be proper. Why?

- If priors are not proper then $P\left(Y^T|i\right)$ may not be proper, and it cannot be interpret as a probability.

- If priors are proper and likelihood is bounded, then the Marginal Likelihood exists.

- How do we compute it?

Approach I – Drawing from the Prior

- Let $\left\{\theta_{ij}\right\}_{j=1}^{M}$ be a draw from the prior of model i, $\pi\left(\theta_i|i\right)$.

- By Monte-Carlo integration: $P^*\left(Y^T|i\right) = \frac{1}{M}\sum_{j=1}^{M} f(Y^T|\theta_{ij}, i)$.

- Very inefficient if likelihood very informative.

$$Var\left[P^*\left(Y^T|i\right)\right] \simeq \frac{1}{M}\sum_{j=1}^{M}\left(f(Y^T|\theta_{ij}, i) - P^*\left(Y^T|i\right)\right)^2 \text{ very high.}$$

- Likelihood very informative if likelihood and prior far apart.

Example I – Drawing from the Prior

- Assume the true likelihood is $\mathcal{N}(0, 1)$.

- Let calculate the Marginal Likelihood for different priors.

- $\mathcal{N}(k, 1)$ for $k = 1, 2, 3, 4$, and 5.

Example I – Drawing from the Prior

Marginal Likelihood

k	1	2	3	4
$P^* \left(Y^T \vert i\right)$	0.2175	0.1068	0.0308	0.0048
$\frac{Var[P^*(Y^T \vert i)]^{0.5}}{P^*(Y^T \vert i)}$	0.6023	1.1129	2.0431	4.0009

Example II – Drawing from the Prior

- Assume the likelihood is $\mathcal{N}(0, 1)$.
- Let us calculate the Marginal Likelihood for different priors.
- $\mathcal{N}(0, k)$ for $k = 1, 2, 3, 4$, and 5.

Example II – Drawing from the Prior

Marginal Likelihood

k	1	2	3	4
$P^* \left(Y^T \vert i \right)$	0.2797	0.1731	0.1303	0.0952
$\frac{Var[P^*(Y^T \vert i)]^{0.5}}{P^*(Y^T \vert i)}$	0.3971	0.8292	1.1038	1.4166

Approach II – Important Sampling

- Assume we want to compute $P\left(Y^T|i\right)$.

- Assume $j_i(\theta)$ is a probability density (not a kernel) which support is contained in Θ_i.

- Let $P\left(\theta|Y^T,i\right) \propto f\left(Y^T|\theta,i\right)\pi\left(\theta|i\right)$, both properly normalized densities (not kernels).

- Let $w\left(\theta\right) = f\left(Y^T|\theta,i\right)\pi\left(\theta|i\right)/j_i(\theta)$.

Approach I I– Important Sampling

- Let $\{\theta_{ij}\}_{j=1}^{M}$ be a draw from $j_i(\theta)$. It can be shown that:

$$w_M^* = \frac{\sum_{j=1}^{M} w\left(\theta_{ij}\right)}{M} \rightarrow \int \frac{f\left(Y^T|\theta_i, i\right)\pi(\theta_i|i)}{j_i(\theta_i)} j_i(\theta_i)\, d\theta_i = P\left(Y^T|i\right)$$

- If $w(\theta)$ is bounded above, then we also have:

$$\sigma^{*2} = \frac{\sum_{m=1}^{M}[w\left(\theta_{ij}\right) - w_M^*]^2}{M} \rightarrow \sigma^2$$

Approach I I– Important Sampling

- The problem is the common drawback of important sampling.
- To find $j_i(\theta)$ such that $w(\theta)$ is bounded and well-behaved.
- Alternative: use the posterior. How?

Approach III – Harmonic Mean

- Argument due to Gelfand and Dey (1994).

- Let $f_i(\theta)$ be a p.d.f. which support is contained in Θ_i.

- Then, it can be proved that:

$$\frac{1}{P(Y^T|i)} = \int_{\Theta_i} \frac{f_i(\theta_i)}{f\left(Y^T|\theta_i, i\right)\pi\left(\theta_i|i\right)} P(\theta_i|Y^T, i)d\theta_i$$

Proof

Since:

$$P(\theta_i|Y^T, i) = \frac{f\left(Y^T|\theta_i, i\right)\pi\left(\theta_i|i\right)}{\int_{\Theta_i} f\left(Y^T|\theta_i, i\right)\pi\left(\theta_i|i\right)d\theta_i}$$

$$\int_{\Theta_i} \frac{f_i(\theta_i)}{f\left(Y^T|\theta_i, i\right)\pi\left(\theta_i|i\right)} P(\theta_i|Y^T, i)d\theta_i =$$

$$= \int_{\Theta_i} \frac{f_i(\theta_i)}{f\left(Y^T|\theta_i, i\right)\pi\left(\theta_i|i\right)} \frac{f\left(Y^T|\theta_i, i\right)\pi\left(\theta_i|i\right)}{\int_{\Theta_i} f\left(Y^T|\theta_i, i\right)\pi\left(\theta_i|i\right)d\theta_i} d\theta_i =$$

$$= \frac{\int_{\Theta_i} f_i(\theta_i)d\theta_i}{\int_{\Theta_i} f\left(Y^T|\theta_i, i\right)\pi\left(\theta_i|i\right)d\theta_i} = \frac{1}{\int_{\Theta_i} f\left(Y^T|\theta_i, i\right)\pi\left(\theta_i|i\right)d\theta_i} = \frac{1}{P(Y^T|i)}$$

We Need to Find $f_i(\theta)$ I

As always, we need to find a $f_i(\theta)$ such that:

$$\frac{f_i(\theta)}{f\left(Y^T|\theta, i\right)\pi\left(\theta|i\right)}$$

bounded above.

We need to Find $f_i(\theta)$ II

- The following proposal is due to Geweke (1998).

- Let $\{\theta_{ij}\}_{j=1}^{M}$ be a draw from the posterior.

- Then we can write:

$$\theta_{iM} = \frac{\sum_{j=1}^{M} \theta_{ij}}{M}$$

and

$$\mathbf{\Sigma}_{iM} = \frac{\sum_{j=1}^{M} (\theta_{ij} - \theta_{iM})(\theta_{ij} - \theta_{iM})'}{M}$$

We need to find $f_i(\theta)$ III

- Define now the following set:

$$\Theta_{iM} = \{\theta : (\theta - \theta_{iM})' \Sigma_{iM}^{-1} (\theta - \theta_{iM}) \leq \chi^2_{1-p}(k)\}$$

- Define $f_i(\theta)$ to be:

$$f_i(\theta) = \frac{(2\pi)^{-k/2} |\Sigma_{iM}|^{-1/2} \exp[-\frac{(\theta - \theta_{iM})' \Sigma_{iM}^{-1} (\theta - \theta_{iM})}{2}]}{p} \psi_{\Theta_{iM}}(\theta)$$

We need to check the two conditions:

- Is $f_i(\theta)$ a p.d.f?

- Does the support of $f_i(\theta)$ belong to Θ_i?

Is $f_i(\theta)$ a p.d.f?

- Remember that $f(\theta_i)$ equals:

$$f_i(\theta) = \frac{(2\pi)^{-k/2}|\Sigma_{iM}|^{-1/2}\exp[-\frac{(\theta-\theta_{iM})'\Sigma_{iM}^{-1}(\theta-\theta_{iM})}{2}]}{p}\psi_{\Theta_{iM}}(\theta) \geq 0$$

- And, since:

$$\int_{\Theta_{iM}}(2\pi)^{-k/2}|\Sigma_{iM}|^{-1/2}\exp[-\frac{(\theta-\theta_{iM})'\Sigma_{iM}^{-1}(\theta-\theta_{iM})}{2}] = p$$

it does integrates to one.

- Therefore, $f_i(\theta)$ is a p.d.f

Does the Support of $f_i(\theta)$ Belong to Θ_i?

- The support of $f_i(\theta)$ is Θ_{iM}.

- In general we cannot be sure of it.

- If $\Theta_i = R^{k_i}$ there is no problem. This is the case of unrestricted parameters. Example: a VAR.

- If $\Theta_i \subset R^{k_i}$, maybe there is a problem. If $\Theta_{iM} \subsetneq \Theta_i$, we need to redefine the domain of integration to be $\Theta_{iM} \cap \Theta_i$.

- As a consequence, we also need to find the new normalization constant for $f_i(\theta)$. This is the typical case for DSGE models.

Recalculating the Constant for $f(\theta_i)$

- If $\Theta_{iM} \subsetneq \Theta_i$.

- We redefine $f(\theta_i)$ as $f^*(\theta_i)$ in the following way:

$$f_i^*(\theta) = \frac{1}{p^*}\frac{(2\pi)^{-k/2}|\Sigma_{iM}|^{-1/2}\exp[-\frac{(\theta-\theta_{iM})'\Sigma_{iM}^{-1}(\theta-\theta_{iM})}{2}]}{p}\psi_{\Theta_{iM}\cap\Theta_i}(\theta)$$

- Where $p^* = 1$ for the case that $\Theta_{iM} \subseteq \Theta_i$.

Recalculating the Constant for $f(\theta_i)$ II

How do we calculate p^*?

1. Fix N and let $j = 0$ and $i = 1$.

2. Draw θ_i from $f_i(\theta)$ and let $i = i + 1$.

3. If $\theta_i \in \Theta_i$, then $j = j + 1$ if $i < N$ got to 2, else $p^* = \frac{j}{N}$ and exit.

Compute the Marginal Likelihood

- Let $\left\{\theta_{ij}\right\}_{j=1}^{N}$ be a draw from the posterior of model i, $P(\theta_i|Y^T, i)$.

- Then, we can approximate $P(Y^T|i)$ using simple Monte Carlo integration:

$$\frac{1}{P^*(Y^T|i)} = N^{-1} \sum_{j=1}^{N} \frac{f_i(\theta_{ij})}{f\left(Y^T|\theta_{ij}, i\right) \pi \left(\theta_{ij}|i\right)}$$

- Notice that we have to evaluate $f_i(\theta_{ij})$ for every draw θ_{ij} from the posterior.

Algorithm

1. Let $j = 1$.

2. Evaluate $f_i(\theta_{ij})$.

3. Evaluate $\frac{f_i(\theta_{ij})}{f\left(Y^T|\theta_{ij},i\right)\pi\left(\theta_{ij}|i\right)}$

4. If $j \leq M$, set $j \rightsquigarrow j + 1$ and go to 2

5. Calculate $\frac{1}{P^*(Y^T|i)} = M^{-1} \sum_{j=1}^{M} \frac{f_i(\theta_{ij})}{f\left(Y^T|\theta_{ij},i\right)\pi\left(\theta_{ij}|i\right)}$.

Example

- Imagine you want to compare how a VAR(1) and a VAR(2) explain $\log y_t$ and $\log i_t$.

- Let us define a VAR(p) model.

$$x_t = C + \sum_{\ell=1}^{p} A(\ell) x_{t-\ell} + \varepsilon_t$$

- Where $x_t = (\log y_t \log i_t)'$, C is a 2×1 matrix, $A(\ell)$ is a 2×2 matrix for all ℓ, and ε_t is iid normally distributed with mean zero and variance-covariance matrix Σ.

Example II

- The likelihood function of a VAR(p) is:

$$L(x^T|\Xi(p)) = (2\pi)^{-T}|\Sigma|^{-T/2}\exp^{-\frac{\varepsilon_t'\Sigma\varepsilon_t}{2}}$$

where $\Xi(p) = \{C, A(1), \ldots, A(p)\}$.

- (Bounded) Flat and independent priors over all the parameters.

Example III - Drawing from the posterior

1. Set $p = 1$, $j = 1$ and set $\Xi(1)_1$ equal to the MLE estimate.

2. Generate $\Xi(1)^*_{j+1} = \Xi(1)_j + \xi_{j+1}$, where ξ_{j+1} is an iid draw from a normal distribution with mean zero and variance-covariance matrix Σ_ξ and generate ν from uniform $[0, 1]$.

3. Evaluate $\alpha(\Xi(p)^*_{j+1}, \Xi(p)_j) = \frac{L(x^T|\Xi(p)^*_{j+1})}{L(x^T|\Xi(p)_j)}$if $\alpha(\Xi(p)^*_{j+1}, \Xi(p)_j) < \nu$. Then $\Xi(1)_{j+1} = \Xi(1)^*_{j+1}$, otherwise $\Xi(1)_{j+1} = \Xi(1)_j$.

4. If $j \leq M$, set $j \rightsquigarrow j + 1$ and go to 2, otherwise exit.

Example IV - Evaluating the Marginal Likelihood

- Since priors are flat, the posterior is proportional to the likelihood $L(x^T|\Xi(p))$ for all p.

- Repeat the algorithm for $p = 2$.

- Let $\{\Xi(1)_j\}_{j=1}^M$ and $\{\Xi(2)_j\}_{j=1}^M$ be draws from the posterior of the VAR(1) and VAR(2) respectively.

Example V - Evaluating the Marginal Likelihood

Calculate:

$$\Xi(p)_M = \frac{\sum_{j=1}^{M} \Xi(p)_j}{M}$$

and

$$\Sigma(p)_M = \frac{\sum_{j=1}^{M} (\Xi(p)_j - \Xi(p)_M)(\Xi(p)_j - \Xi(p)_M)'}{M}$$

for $p = 1$ and $p = 2$.

Example VI - Evaluating the Marginal Likelihood

- Calculate $\{f_i(\Xi(p)_j)\}_{j=1}^{M}$ for $p = 1$ and $p = 2$.

- Calculate:

$$\frac{1}{P^*(x^T|p)} = M^{-1} \sum_{j=1}^{M} \frac{f_i(\Xi(p)_j)}{L\left(x^T|\Xi(p)_j\right)}$$

A Problem Evaluating the Marginal Likelihood

- Sometimes, $L\left(x^T|\Xi(p)_j\right)$ is a to BIG number.

- For example: The log likelihood of the VAR(1) evaluated at the MLE equals $1,625.23$. This means that the likelihood equals $\exp^{1,625.23}$. In `Matlab`, $\exp^{1,625.23} = Inf$.

- This implies that:

$$\frac{1}{P^*(x^T|p)} = M^{-1}\sum_{j=1}^{M}\frac{f_i(\Xi(p)_j)}{L\left(x^T|\Xi(p)_j\right)} = 0$$

Solving the Problem

- In general, we want to compute

$$\frac{1}{P^*(Y^T|i)} = M^{-1} \sum_{j=1}^{M} \frac{f_i(\theta_{ij})}{f\left(Y^T|\theta_{ij}, i\right) \pi\left(\theta_{ij}|i\right)}$$

- Instead of evaluating $f\left(Y^T|\theta_{ij}, i\right)$ and $\pi\left(\theta_{ij}|i\right)$, we evaluate $\log f\left(Y^T|\theta_{ij}, i\right)$ and $\log \pi\left(\theta_{ij}|i\right)$ for all $\{\theta_{ij}\}_{j=1}^{M}$ and for each of the models i.

- For each i, we compute $\wp_i = \max_j\{\log f\left(Y^T|\theta_{ij}, i\right) + \log \pi\left(\theta_{ij}|i\right)\}$.

- Then, we compute $\wp = \max_i\{\wp_i\}$.

- Compute:

$$\log \widetilde{f}\left(Y^T|\theta_{ij}, i\right) = \log f\left(Y^T|\theta_{ij}, i\right) + \log \pi\left(\theta_{ij}|i\right) - \wp.$$

- Compute

$$\widetilde{f}\left(Y^T|\theta_{ij}, i\right) = \exp \log \widetilde{f}\left(Y^T|\theta_{ij}, i\right).$$

- Finally, compute

$$\frac{1}{\widetilde{P}(Y^T|i)} = M^{-1} \sum_{j=1}^{M} \frac{f_i(\theta_{ij})}{\widetilde{f}\left(Y^T|\theta_{ij}, i\right)}$$

- And note that

$$\log \widetilde{P}(Y^T|i) - \log \widetilde{P}(Y^T|s) = \log P^*(Y^T|i) - \log P^*(Y^T|s)$$

- Why?

- Note that

$$\frac{1}{\widetilde{P}(Y^T|i)} = M^{-1} \sum_{j=1}^{M} \frac{f_i(\theta_{ij})}{\widetilde{f}\left(Y^T|\theta_{ij}, i\right)} = M^{-1} \sum_{j=1}^{M} \frac{f_i(\theta_{ij})}{\frac{f\left(Y^T|\theta_{ij},i\right)\pi\left(\theta_{ij}|i\right)}{\wp}}$$

- Therefore

$$\frac{1}{\widetilde{P}(Y^T|i)} = \frac{\wp}{P*(Y^T|i)}$$