BAYESIAN MODEL SEARCH AND MODEL AVERAGING

AAEC 6984
INSTRUCTOR: KLAUS MOELTNER

Textbooks:	Koop (2003), Ch.11; Koop et al. (2007), Ch.16
`Matlab` scripts:	`mod8_SSVS_data, mod8_SSVS, mod8_SSVS_modProbs,` `mod8_SSVS_fishing_data, mod8_SSVS_fishing, mod8_SSVS_fishing_modProb,` `mod8_SSVS_fishing_BMA` `mod8_MC3,mod8_MC3_convTest, mod8_MC3_modProb,` `mod8_MC3_growth,mod8_MC3_growth_convTest, mod8_MC3_growth_modProb,`
`Matlab` functions:	`gs_SSVS, gs_SSVS_fishing, graycode, gs_MC3`

Introduction

We have already seen that the Bayesian estimation framework allows for the pair-wise comparison of different models, even non-nested ones, via the computation of marginal likelihoods and Bayes Factors.

This module deals with situations where a whole array of candidate models must be considered, evaluated, and compared. In theory, one could estimate each model separately, compute the marginal likelihood for each case, and derive individual model probabilities as the ratio of the model-specific marginal likelihood (or the product of marginal likelihood and model prior) to the sum of marginal likelihoods over all models.

Specifically, denoting $p(\mathbf{y}|M_m)$ as the marginal likelihood for model m, $p(M_m)$ the model prior, and the complete *model space* as $\mathcal{M} = \{M_1, M_2, \ldots, M_M\}$, individual model probabilities can be derived as

$$p(M_m|\mathbf{y}) = \frac{p(\mathbf{y}|M_m)\, p(M_m)}{\sum_{j=1}^{M} p(\mathbf{y}|M_j)\, p(M_j)} \tag{1}$$

If model priors are the same for all models, this further simplifies to

$$p(M_m|\mathbf{y}) = \frac{p(\mathbf{y}|M_m)}{\sum_{j=1}^{M} p(\mathbf{y}|M_j)} \tag{2}$$

A model-averaged posterior distribution for parameters $\boldsymbol{\theta}$ can then be obtained via

$$p(\boldsymbol{\theta}|\mathbf{y}) = \sum_{m=1}^{M} p(\boldsymbol{\theta}|\mathbf{y}, M_m)\, p(M_m|\mathbf{y}) \tag{3}$$

In practice (i.e. in absence of an analytical solution to this expression), this is implemented by first obtaining R draws of $p(\boldsymbol{\theta}|\mathbf{y}, M_m)$ for each model, and then drawing from these model-specific posteriors with relative frequency dictated by the computed model weights. All subsequent inference, including posterior predictive densities, are then based on these model-weighted, or model-averaged draws.

While this approach is conceptually straightforward, its implementation becomes quickly infeasible with increasing model space. Therefore, Bayesians have developed model-search methods that, for specific applications, can "travel" quickly through model space and identify the most promising specifications. In most cases, the model search component simply becomes an extra step in a Gibbs Sampler.

In this module we will discuss the two most common approaches to model search when the task at hand is to identify "meaningful" explanatory variables in a regression model: Stochastic Search Variable Selection ($SSVS$) (George and McCulloch, 1993), and Markov-Chain Monte-Carlo Model Composition (MC^3) (Madigan and York, 1995).

Stochastic Search Variable Selection

The SSVS algorithm has the following salient features:

(1) Variable inclusion probabilities are directly modeled via a mixture prior for each coefficient
(2) Implementation proceeds via a standard Gibbs Sampler without MH steps
(3) At each iteration, the full model (with all possible variables) is evaluated
(4) Irrelevant variables will end up with a coefficient that has both posterior mean and variance close to zero
(5) At each iteration, the algorithm captures if a given coefficient was deemed "important" or "irrelevant"
(6) This information can then be used to compute posterior inclusion probabilities for each coefficient and posterior model probabilities
(7) The SSVS algorithm does *not* lend itself to a direct derivation of model-averaged posterior distributions.

Consider a standard linear regression model with an intercept α and a potentially large number of slope coefficients $\beta_j,\ j = 1 \ldots k-1$. The sample likelihood is given as:

$$\begin{aligned} &p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{X}) = (2\pi)^{-n/2} \left(\sigma^2\right)^{-n/2} exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{Z}\boldsymbol{\theta})'(\mathbf{y} - \mathbf{Z}\boldsymbol{\theta})\right), \quad \text{where} \\ &\mathbf{Z} = \begin{bmatrix}\mathbf{i} & \mathbf{X}\end{bmatrix} \\ &\boldsymbol{\theta} = \begin{bmatrix}\alpha & \beta_1 & \beta_2 & \ldots & \beta_{k-1}\end{bmatrix}' \end{aligned} \tag{4}$$

Assume that economic theory provides little or no guidance as to the inclusion or exclusion of the variables in $\mathbf{X}$. That is, the total number of possible models, each distinguished by a different combination of regressors, is 2^{k-1}. Even with just a "moderate" number of candidates, this model space quickly explodes and makes it infeasible to evaluate every single model.

The SSVS algorithm tackles this problem via a mixture prior for each β_j. It assigns a *point-mass* prior probability p_0 that β_j comes from a *healthy* normal density with mean zero and "large" variance $c^2 * t^2$. With probability $1 - p_0$, the prior density for β_j is "virtually degenerate", i.e. a normal density with mean zero and variance t^2, where t^2 is also close to zero. Thus:

$$p\left(\beta_j\right) = p_0\phi\left(0, c^2 * t^2\right) + (1 - p_0)\phi\left(0, t^2\right) \tag{5}$$

where ϕ denotes the normal pdf. In theory, the variance "tuners" c^2 and t^2 could be specific to each coefficient, perhaps based on "thresholds of practical significance" (George and McCulloch, 1997). For simplicity, we assign the same values of c^2 and t^2 to all coefficients. The larger the difference in magnitude between c and t, the more sharply the algorithm will be able to discriminate between relevant and irrelevant variables. However, George and McCulloch (1993) and George and McCulloch (1997) recommend against a variance ratio of higher than 10,000, and a value of t "too close to zero" to avoid convergence problems in the sampler.

The model is *augmented* with the introduction of indicator vector $\boldsymbol{\gamma} = \begin{bmatrix}\gamma_1 & \gamma_2 & \dots & \gamma_{k-1}\end{bmatrix}'$, where each element is a binary 0 / 1 indicator, with "1" signaling that the corresponding β_j comes from the "healthy" prior density. This augmentation facilitates model implementation and allows to capture inclusion counts for each coefficient as by-product of the Gibbs Sampler.

Thus, we can re-express the prior of β_j with the following hierarchical structure:

$$\begin{aligned} &p\left(\beta_j|\gamma_j\right) = \gamma_j\phi\left(0, c^2 * t^2\right) + (1 - \gamma_j)\phi\left(0, t^2\right), \\ &\gamma_j \sim Bin\left(p_0, 1\right) \end{aligned} \tag{6}$$

where $Bin\left(p, 1\right)$ indicates the Binomial distribution with parameter p_0 and a single trial (essentially a Bernoulli density). Choosing a normal prior for the intercept, and the standard inverse-gamma prior for the variance of the regression error with shape ν_0 and scale τ_0 yields

the following augmented joint posterior kernel:

$$
\begin{aligned}
&p\left(\boldsymbol{\theta}, \sigma^2, \boldsymbol{\gamma} | \mathbf{y}, \mathbf{Z}\right) \propto p\left(\alpha\right) p\left(\boldsymbol{\beta} | \boldsymbol{\gamma}\right) p\left(\boldsymbol{\gamma}\right) p\left(\sigma^2\right) p\left(\mathbf{y} | \boldsymbol{\theta}, \sigma^2, \mathbf{Z}\right) \propto \\
&exp\left(-\frac{1}{2V_\alpha}\left(\alpha - \mu_\alpha\right)^2\right) * \\
&\prod_{j=1}^{k-1}\left\{\gamma_j exp\left(-\frac{1}{2c^2t^2}\left(\beta_j - \mu_{\beta_j}\right)^2\right) + \left(1-\gamma_j\right) exp\left(-\frac{1}{2t^2}\left(\beta_j - \mu_{\beta_j}\right)^2\right)\right\} * \\
&\prod_{j=1}^{k-1} p_0^{\gamma_j}\left(1-p_0\right)^{\left(1-\gamma_j\right)} * \\
&\left(\sigma^2\right)^{\frac{-n-2\nu_0-2}{2}} exp\left(-\frac{1}{\sigma^2}\left(\tau_0\right)\right) * \\
&exp\left(-\tfrac{1}{2\sigma^2}\left(\mathbf{y} - \mathbf{Z}\boldsymbol{\theta}\right)'\left(\mathbf{y} - \mathbf{Z}\boldsymbol{\theta}\right)\right)
\end{aligned}
\tag{7}
$$

Conditional on $\boldsymbol{\gamma}$, the intercept α and the coefficient vector $\boldsymbol{\beta}$ can be drawn jointly as vector $\boldsymbol{\theta}$ if we collect their variances into a single matrix $\mathbf{V}_0$, given as:

$$
\begin{aligned}
&\mathbf{V}_0 = diag\left[V_\alpha \quad V_1 \quad V_2 \quad \ldots \quad V_{k-1}\right], \quad \text{where} \\
&V_j = \gamma_j c^2 * t^2 + \left(1-\gamma_j\right) t^2
\end{aligned}
\tag{8}
$$

This then leads to the standard Gibbs Sampler for the linear regression model with two important modifications: (i) $\boldsymbol{\gamma}$ is drawn in a third step of the sampler, and (ii) the prior variance $\mathbf{V}_0$ is adjusted accordingly at every iteration. Specifically, we have:

$$
\begin{aligned}
&\boldsymbol{\theta} | \sigma^2, \boldsymbol{\gamma}, \mathbf{y}, \mathbf{Z} \sim n\left(\boldsymbol{\mu}_1, \mathbf{V}_1\right), \quad \text{with} \\
&\mathbf{V}_1 = \left(\mathbf{V}_0^{-1} + \tfrac{1}{\sigma^2}\mathbf{Z}'\mathbf{Z}\right)^{-1}, \quad \text{and} \\
&\boldsymbol{\mu}_1 = \mathbf{V}_1 * \left(\mathbf{V}_0^{-1}\boldsymbol{\mu}_0 + \tfrac{1}{\sigma^2}\mathbf{Z}'\mathbf{y}\right)
\end{aligned}
\tag{9}
$$

where $\mathbf{V}_0$ is a function of $\boldsymbol{\gamma}$, as shown in (8). Furthermore:

$$
\begin{aligned}
&\sigma^2 | \boldsymbol{\theta}, \mathbf{y}, \mathbf{Z} \sim ig\left(\nu_1, \tau_1\right), \quad \text{with} \\
&\nu_1 = \frac{2\nu_0 + n}{2}, \quad \text{and} \\
&\tau_1 = \tau_0 + \tfrac{1}{2}\left(\mathbf{y} - \mathbf{Z}\boldsymbol{\theta}\right)'\left(\mathbf{y} - \mathbf{Z}\boldsymbol{\theta}\right)
\end{aligned}
\tag{10}
$$

The third step of the Gibbs Sampler is the core of the SSVS engine: it updates the indicator vector $\boldsymbol{\gamma}$, which, in turn, feeds into an updated prior variance $\mathbf{V}_0$. Formally, the conditional

posterior kernel for γ_j is given as

$$\begin{aligned} &p\left(\gamma_j|\beta_j\right) \propto \\ &p_0^{\gamma_j}\left(1-p_0\right)^{(1-\gamma_j)} * \gamma_j \phi\left(0, c^2 * t^2\right) + \left(1-\gamma_j\right)\phi\left(0, t^2\right) \end{aligned} \tag{11}$$

Draws can be obtained via the following strategy: First, we update the probability that a given γ_j takes a value of 1, i.e that the associated β_j comes from the "healthy" normal density:

$$\begin{aligned} &p_{1,j} = pr\left(\gamma_j = 1|\beta_j\right) = \frac{pr\left(\gamma_j = 1, \beta_j\right)}{p\left(\beta_j\right)} = \\ &\frac{p\left(\beta_j|\gamma_j = 1\right) pr\left(\gamma_j = 1\right)}{p\left(\beta_j\right)} = \\ &\frac{p_0 * \phi\left(\beta_j; 0, c^2 t^2\right)}{p_0 * \phi\left(\beta_j; 0, c^2 t^2\right) + \left(1-p_0\right)\phi\left(\beta_j; 0, t^2\right)} \end{aligned} \tag{12}$$

We then draw a random uniform u_j and compare it to $p_{1,j}$. We set $\gamma_j = 1$ if $p_{1,j} > u_j$, and to zero otherwise. Formally, this step can be expressed as

$$\gamma_j|\boldsymbol{\beta} \sim Bin\left(p_{1,j}, 1\right) \tag{13}$$

with $p_{1,j}$ given in (12).

Matlab scripts `mod8_SSVS_data, mod8_SSVS` and function `gs_SSVS` show the implementation of this approach. The derivation of posterior inclusion probabilities and posterior model probabilities is illustrated in script `mod8_SSVS_modProb`.

Markov-Chain Monte Carlo Model Composition (MC^3)

This is an alternative approach to model search and selection that is based on comparing models with different parameter space. It has the following salient features:

(1) Models are defined as different combinations of included variables.
(2) Each model receives its own *model prior*.
(3) Implementation proceeds via a standard Gibbs Sampler *with* an MH step for model selection
(4) At each iteration, a candidate model is drawn from the *neighborhood* of the current model. It is accepted or rejected with a specific probability (as is the standard case in a MH algorithm).
(5) The algorithm is designed to "visit" models with many relevant variables more often.
(6) Posterior model probabilities can be computed empirically and analytically. The resulting comparison is used as one of the diagnostic tools to assess convergence.
(7) Model-averaged results can be obtained directly from the output of the Gibbs Sampler.

As before, the constant term will be included in every model. This poses a bit of a problem as it is difficult to assign a prior to it without using the data. For example, the algorithm may choose a model that has no explanatory variables. In that case the intercept, call it α, becomes the prior expectation of the dependent variable. What should we use for that value?

As a result, two preparatory steps are taken. First, the intercept is given an improper (= not integrating to one) prior, i.e.

$$p(\alpha) \propto 1 \tag{14}$$

However, this would break the conjugacy of the prior for the remaining coefficients $\boldsymbol{\beta}$ and the error variance σ^2, which is needed to make the MH step work. This can be overcome via the second preparatory intervention, de-meaning the regressors (= subtracting the mean from all of the candidate explanatory variables). This then implies that the "variable" associated with the intercept, i.e. the column of ones, is orthogonal to the remaining regressor matrix, i.e.

$$\mathbf{i}'\mathbf{X} = \mathbf{0} \tag{15}$$

Note that de-meaning does not change the interpretation of the slope coefficients $\boldsymbol{\beta}$.

Letting $\boldsymbol{\theta} = \begin{bmatrix}\alpha & \boldsymbol{\beta}'\end{bmatrix}'$ and $\mathbf{Z} = \begin{bmatrix}\mathbf{i} & \mathbf{X}\end{bmatrix}$, we can now write

$$\begin{aligned}
&(\mathbf{y} - \mathbf{Z}\boldsymbol{\theta})'(\mathbf{y} - \mathbf{Z}\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{i}\alpha - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{i}\alpha - \mathbf{X}\boldsymbol{\beta}) = \\
&(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - 2\alpha\mathbf{i}'\mathbf{y} + 2\alpha\mathbf{i}'\mathbf{X}\boldsymbol{\beta} + \mathbf{i}'\mathbf{i}\alpha^2 = \\
&(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - 2\alpha\mathbf{i}'\mathbf{y} + \mathbf{i}'\mathbf{i}\alpha^2
\end{aligned} \tag{16}$$

The second-to-last term in the second line drops out due to the imposed orthogonality via de-meaning. This is crucial in de-linking the posterior of the intercept from the rest of the model.

We follow standard convention (see e.g. Fernández et al. (2001a)) and choose an improper prior to the error variance and a conjugate g-prior for the $k-1$ by 1 coefficient vector $\boldsymbol{\beta}$. Such a g-prior greatly reduces the number of parameters that need to be determined by the researcher a priori,

facilitates the interpretation of posterior results, and assures speedy model evaluation as part of the search process.[1] Thus, we have

$$\begin{aligned} &p\left(\sigma^2\right) \propto \frac{1}{\sigma^2} \\ &p\left(\boldsymbol{\beta}|\sigma^2\right) = n\left(\boldsymbol{\mu}_0, g\sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right) \end{aligned} \tag{17}$$

where n denotes the $k-1$-variate normal density. We set $\boldsymbol{\mu}_0 = \mathbf{0}$.

Let $\boldsymbol{\gamma}$ be a $k-1$ by 1 vector of binary indicators that signal if a given covariate $\mathbf{x}_j$ is included ($\gamma_j = 1$) or excluded ($\gamma_j = 0$) for a given model. Thus, we can express conditionality on a specific model, i.e. a specific mix of regressors, as conditionality on $\boldsymbol{\gamma}$.

Building on the results of the conjugate normal regression model (and letting $\mathbf{V}_0 = g\left(\mathbf{X}'\mathbf{X}\right)^{-1}$), the conditional posterior for $\boldsymbol{\beta}$ is again a "well-behaved" normal density, i.e

$$\begin{aligned} &\boldsymbol{\beta}|\sigma^2, \gamma, \mathbf{y}, \mathbf{X} \sim n\left(\boldsymbol{\mu}_1, \mathbf{V}_1\right) \quad \text{with} \\ &\mathbf{V}_1 = \sigma^2 \frac{g}{1+g}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ &\boldsymbol{\mu}_1 = \frac{g}{1+g}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \end{aligned} \tag{18}$$

Similarly, the conditional posterior distribution of α can be immediately derived as

$$\begin{aligned} &p\left(\alpha|\sigma^2, \mathbf{y}\right) \propto \\ &exp\left(-\frac{1}{2\sigma^2}\left(n\alpha^2 - 2\alpha\sum_{i=1}^{n} y_i\right)\right) = \\ &exp\left(-\tfrac{1}{2}\left(\tfrac{\sigma^2}{n}\right)^{-1}\alpha^2 + \alpha\left(\tfrac{\sigma^2}{n}\right)^{-1}\bar{y}\right) \end{aligned} \tag{19}$$

We recognize this as the kernel of the normal density with mean $\bar{y}$ and variance $\frac{\sigma^2}{n}$.

[1] Specifically, the conjugate g-prior, along with our other prior choices, assures an analytical expression for the the marginal likelihood. This, in turn, is an integral component of the model selection step of the posterior simulator, as discussed below in more detail.

The *marginal* posterior for the error variance follows again an inverse-gamma distribution. Specifically,

$$\begin{aligned}
&\sigma^2|\gamma,\mathbf{y},\mathbf{X} \sim ig\left(\nu_1,\tau_1\right) \quad \text{with} \\
&\nu_1 = \frac{n-1}{2} \\
&\tau_1 = \tfrac{1}{2}\left(\frac{g}{1+g}\mathbf{y}'\mathbf{M}_X\mathbf{y} + \frac{1}{1+g}\left(\mathbf{y}-\mathbf{i}\bar{y}\right)'\left(\mathbf{y}-\mathbf{i}\bar{y}\right) - \frac{ng}{1+g}\bar{y}^2\right) \quad \text{where} \\
&\mathbf{M}_X = \mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'
\end{aligned} \tag{20}$$

The marginal, model-conditioned likelihood is proportional to the following expression:

$$\begin{aligned}
&p\left(\mathbf{y}|\boldsymbol{\gamma},\mathbf{X}\right) \propto \left(\frac{1}{1+g}\right)^{k/2} \tau_1^{-\nu_1} = \\
&\left(\frac{1}{1+g}\right)^{k/2}\left(\tfrac{1}{2}\left(\frac{g}{1+g}\mathbf{y}\mathbf{M}_X\mathbf{y} + \frac{1}{1+g}\left(\mathbf{y}-\mathbf{i}\bar{y}\right)'\left(\mathbf{y}-\mathbf{i}\bar{y}\right) - \frac{ng}{1+g}\bar{y}^2\right)\right)^{-\left(\frac{n-1}{2}\right)}
\end{aligned} \tag{21}$$

where k is the dimension of β under model γ. The key step in the MC^3 algorithm is the model selection, implemented through draws of $\boldsymbol{\gamma}$. This requires a Metropolis-Hastings (MH) step with acceptance probability

$$a = min\left(1, \frac{q\left(\boldsymbol{\gamma}_0|\boldsymbol{\gamma}_c\right)p\left(\mathbf{y}|\boldsymbol{\gamma}_c,\mathbf{X}\right)p\left(\boldsymbol{\gamma}_c\right)}{q\left(\boldsymbol{\gamma}_c|\boldsymbol{\gamma}_0\right)p\left(\mathbf{y}|\boldsymbol{\gamma}_0,\mathbf{X}\right)p\left(\boldsymbol{\gamma}_0\right)}\right) \tag{22}$$

where, as before, "0" denotes the old or current model, and "c" denotes the new or candidate model. The function $q\left(\boldsymbol{\gamma}_i|\boldsymbol{\gamma}_j\right)$ is the candidate generating density for model $\boldsymbol{\gamma}_i$, given model $\boldsymbol{\gamma}_j$.

In our case, this boils down to the discrete probability of drawing model $\boldsymbol{\gamma}_i$ from the neighborhood of current model $\boldsymbol{\gamma}_j$. Thus, the intuition for this proposal is similar to that behind the random-walk MH algorithm encountered in the previous chapter. In our simple case of regressor selection, there are $k+1$ candidate models (including the current one) at every iteration. This implies that $q\left(\boldsymbol{\gamma}_i|\boldsymbol{\gamma}_j\right) = 1/\left(k+1\right)$ for any i,j, and the $q\left(.\right)$ terms drop out of the ratio in (22).

If, in addition, the model priors $p\left(\boldsymbol{\gamma}_i\right)$ are equal for all models, the MH acceptance formula reduces to a simple Bayes Factor of model c versus model 0:

$$a = min\left(1, \frac{p\left(\mathbf{y}|\boldsymbol{\gamma}_c,\mathbf{X}\right)}{p\left(\mathbf{y}|\boldsymbol{\gamma}_0,\mathbf{X}\right)}\right) \tag{23}$$

Matlab script `mod8_MC3` with function `gs_MC3` illustrate this procedure. Since g becomes the only tuning parameter in the algorithm, care must be taken in its choice. Fernández et al. (2001a)

discuss several options. They find that setting $g = n$ performs well in most general settings. We follow this suggestion in our code.

The programming code includes two other important features: (i) after each draw of $\boldsymbol{\gamma}$, the regressor matrix must be adjusted accordingly (add or delete columns of $\mathbf{X}$), and (ii) all collected draws of $\boldsymbol{\beta}$ should be $(k-1)\,x1$, with the actually drawn β's placed in the correct positions and zeros in all other positions. The resulting sequence of draws will then reflect how often a given coefficient was set to zero (i.e. how often a given regressor was omitted) by the algorithm.

Conveniently, and in contrast to the SSVS method, the posterior draws of model parameters are already "model-averaged", and model-averaged inference can be drawn by evaluating the entire set of posterior draws as usual.

The draws of $\boldsymbol{\gamma}$ can be used to (i) examine the posterior inclusion probabilities for each coefficient, (ii) identify the most frequently visited models, and (iii) perform a convergence check as suggested by Fernández et al. (2001a) by comparing empirical model probabilities to analytical probabilities (based on the known form of the marginal likelihood). Specifically, the correlation coefficient between these to sets of probabilities for, say, the top 1000 visited models should be close to 1 if the algorithm has converged. Naturally, our parameter-specific convergence tools (IEF, CD) still apply.

Matlab script `mod8_MC3_modProb` shows the derivation of empirical model probabilities. Script `mod8_MC3_convTest` shows the derivation of *analytical* model probabilities and the implementation of this convergence check.

Matlab scripts `mod8_MC3_growth`, `mod8_MC3_growth_modProb`, and `mod8_MC3_growth_convTest` apply this framework to the macroeconomic growth regression discussed in Fernández et al. (2001b).

References

Fernández, C., Ley, E. and Steel, M. (2001a). Benchmark priors for Bayesian model averaging, *Journal of Econometrics* **100**: 381–427.

Fernández, C., Ley, E. and Steel, M. (2001b). Model uncertainty in cross-country growth regressions, *Journal of Applied Econometrics* **16**: 563–576.

George, E. and McCulloch, R. (1993). Variable selection via Gibbs Sampling, *Journal of the American Statistical Association* **88**: 881–889.

George, E. and McCulloch, R. (1997). Approaches for Bayesian variable selection, *Statistica Sinica* **7**: 339–373.

Koop, G. (2003). *Bayesian Econometrics*, Wiley.

Koop, G., Poirier, D. and Tobias, J. (2007). *Bayesian Econometric Methods*, Cambridge University Press.

Madigan, D. and York, J. (1995). Bayesian graphical models for discrete data, *International Statistics Review* **63**: 215–232.