Bayesian Model Averaging for Autoregressive Distributed Lag (BMA$_A$DL)ingretl

Blazejowski, Marcin and Kwiatkowski, Jacek

WSB University in Torun, Nicolas Copernicus University

17 January 2020

Bayesian Model Averaging for Autoregressive Distributed Lag (BMA_ADL) in **gretl**

Marcin Błażejowski

WSB University in Torun

marcin.blazejowski@wsb.torun.pl

Jacek Kwiatkowski

Nicolaus Copernicus University in Toruń

jkwiat@umk.pl

January 29, 2020

Abstract

This paper presents a software package that implements Bayesian Model Averaging for Autoregressive Distributed Lag models **BMA_ADL** ver. 0.9 in **gretl**.

Gretl (the GNU regression, econometrics and time-series library) is an increasingly popular free, open-source software for econometric analysis with an easy-to-use graphical user interface.

Bayesian Model Averaging (BMA) incorporates model uncertainty into conclusions about the estimated parameters. It is an efficient tool for discovering the most likely models and variables by obtaining estimates of their posterior characteristics.

The package code changelog:

0.9 Pre-release version.

1 Introduction

When we want to find the relationship between two or more variables by fitting a simple linear equation to observed data, we have to decide how to select a "good" subset of variables from a large set of regressors. When the number of possible exogenous variables is K, the number of possible linear models is 2^K. This means that it is very difficult, if not impossible, to find the estimates for all combinations. Moreover, standard variable selection procedures lead to different estimates and conflicting conclusions about the main questions of interest.

In seminal papers by Raftery et al. (1997); Fernández et al. (2001a,b) authors introduce a Bayesian Model Averaging, which enables to measure the importance of particular potential regressors. BMA is a standard Bayesian solution to model

uncertainty, where the inference on parameters is based on a weighted average over all possible models under consideration. These weights are the Bayesian posterior probabilities of the individual models.

2   Bayesian Model Averaging for ADL models

From Bayesian perspective, uncertainty is a natural way of decision making process and therefore it can be easily included in the model selection rules Zellner (1971); Koop (2003); Koop et al. (2007). Among the many seminal papers about Bayesian Model Averaging are Hoeting et al. (1999); Fernández et al. (2001a,b). The most recent detailed overview is presented in Steel (2019).

So, once again, we are dealing with a problem which model and variables are the most appropriate in the analysis of the dependencies, but in this case we use a natural and explicit way of combining prior information with data, without any approximation of marginal data density and Bayes factors. Moreover, we impose a stationary condition on autoregressive parameters.

Let's consider the following dynamic linear regression model M_j, $(j = 1, 2, \ldots, K)$:

$$y = X_j \beta_j + \varepsilon, \tag{1}$$

where y is a vector of T observations, X_j is $(T \times k_j)$ matrix, and β_j is a $(k_j \times 1)$ vector of parameters, ε is a vector of dimensions $(T \times 1)$ with a normal distribution $N(0, \sigma^2 I_T)$, where σ^2 is a variance of random error ε and I_T is an identity matrix of size T. Moreover $X_j = [Y_{j-} \ Z_j]$, where Y_{j-} is a $(T \times k_j^y)$ matrix containing k_j^y lagged values of dependent variable, while Z_j is $(T \times k_j^z)$ matrix of exogenous variables. Furthermore, $\beta_j = [\beta_j^y \ \beta_j^z]'$ is a vector of unknown parameters, where $\beta_j^y \in \Gamma \subseteq R^{k_j^y}$, $\beta_j^z \in R^{k_j^z}$ and Γ is stationary region for the parameters of autoregressive processes. We also assume that we observe initial values $y_{(0)}$.

Let us consider a prior density of the following form:

$$p\left(\beta_j, h \mid M_j\right) = p\left(\beta_j \mid h, M_j\right) p\left(h\right), \tag{2}$$

where:

$$p\left(\beta_j \mid h, M_j\right) \propto f_N\left(\beta_j \mid \underline{\beta}_j, h^{-1}\underline{V}_j\right) I\left(\beta_j^y \in \Gamma\right) \tag{3}$$

and $f_N(\beta_j \mid \mu, \Sigma)$ denotes the multivariate normal density with mean μ and covariance matrix Σ and $I(A)$ is the indicator function.

For the error precision h which is defined as $h = 1/\sigma^2$ we use noninformative prior:

$$p\left(h\right) \propto \frac{1}{\sigma^2}, \ h > 0, \tag{4}$$

where $\underline{\beta}_j$ is k_j-vector of prior means for regression coefficients and $\underline{V}_j$ is a $k_j \times k_j$ positive definite prior covariance matrix of the form:

$$\underline{V}_j = \left(g_j X_j' X_j\right)^{-1}. \tag{5}$$

The factor of proportionality g_j $(j = 1, 2, \ldots, K)$ is part of the so-called *g*-prior, as introduced in Zellner (1986). G-prior is a convenient way to specify the prior variance matrix, because it reduces the number of prior variance parameters and considerably simplifies posterior computations. The BMA_ADL package offers the five most popular alternative Zellner's g-priors (see Fernández et al., 2001a; Moral-Benito, 2010; Ley and Steel, 2009).

- Unit Information Prior (g-UIP), recommended by Kass and Wasserman (1995)

$$g = \frac{1}{N}. \tag{6}$$

- Risk Inflation Criterion (g-RIC), proposed by Foster and George (1994)

$$g = \frac{1}{K^2}. \tag{7}$$

- Benchmark Prior, recommended by Fernández et al. (2001a)

$$g = \left\{ \begin{array}{ll} \frac{1}{K^2} & \text{for } N \leq K^2; \\ \frac{1}{N} & \text{for } N > K^2 \end{array} \right. . \tag{8}$$

- g-HQ prior which mimics the Hannan and Quinn criterion, see Fernández et al. (2001a)

$$g = \frac{1}{(\ln N)^3}. \tag{9}$$

- Root of g-UIP, see Fernández et al. (2001a)

$$g = \sqrt{\frac{1}{N}}. \tag{10}$$

Assuming the prior structure (2) we obtain the following joint posterior density:

$$p\left(\beta_j, h \mid y, M_j\right) = c_j \cdot f_{NG}\left(\beta_j, h \mid \overline{\beta}_j, \overline{V}_j, \overline{s}_j^{-2}, \overline{v}_j\right) I\left(\beta_j^y \in \Gamma\right), \tag{11}$$

where c_j is normalizing constant and f_{NG} is normal-gamma density (see Koop et al., 2007). In our case constant c_j plays important role to obtain the Bayes factor between competitive models and can be easily obtained by Monte Carlo simulations.

Using the properties of normal-gamma density, equation (11) leads to:

$$p\left(\beta_j \mid h, y, M_j\right) \propto f_N\left(\overline{\beta}_j, \overline{V}_j\right), \tag{12}$$

$$p\left(h \mid y\right) = f_G\left(\overline{s}_j^{-2}, \overline{v}_j\right), \tag{13}$$

where

$$\overline{V}_j = \left(\underline{V}_j^{-1} + X_j' X_j\right)^{-1}, \tag{14}$$

$$\overline{\beta}_j = \overline{V}_j\left(\underline{V}_j^{-1} \underline{\beta}_j + X_j' X_j \widehat{\beta}_j\right), \tag{15}$$

and $\overline{v}_j = T$. We also have:

$$\widehat{\beta}_j = \left(X_j'X_j\right)^{-1} X_j'y, \tag{16}$$

$$s_j^2 = \frac{\left(y - X_j\widehat{\beta}_j\right)' \left(y - X_j\widehat{\beta}_j\right)}{v_j}, \tag{17}$$

$$\overline{v}_j \overline{s}_j^2 = v_j s_j^2 + \left(\widehat{\beta}_j - \underline{\beta}_j\right)' \left[\underline{V}_j + \left(X_j'X_j\right)^{-1}\right]^{-1} \left(\widehat{\beta}_j - \underline{\beta}_j\right), \tag{18}$$

where $v_j = T - k_j$.

The marginal data density $p(y \mid M_j)$ as well as posterior means and standard deviations of regression coefficients can be calculated numerically using Monte Carlo integration by sampling from the posterior distribution (11). In our case, we first draw error precision h from (13) and then we draw β_j from (12). We only accept those candidate values which lie in stationary region for the parameters of autoregressive processes. The constant c_j can be calculated as an inverse of the acceptance rate i.e. inverse of the fraction of random numbers accepted in Monte Carlo simulation.

Let assume, that prior probability $\Pr(M_j)$ of model M_j is binomially distributed; that is,

$$\Pr(M_j) = \theta^{k_i}(1 - \theta)^{K-k_i}, \theta \in [0, 1]. \tag{19}$$

The binomial distribution implies that we only need to specify a prior expected model size $\mathsf{E}(\Xi) = K\theta$, where $\mathsf{E}(\Xi) \in (0, K]$. If we define the value of $\mathsf{E}(\Xi)$, then our BMA package will automatically produce a value of prior inclusion probability for all competitive models. If $\theta = 0.5$, then the prior expected model size is equal to the average of the number of potential regressors, and the model prior distribution is uniform $\left(\Pr(M_j) = 2^{-K}\right)$ and reflects a lack of previous knowledge about the models.

The posterior probability of any variant of regression model M_j can be calculated by the following formula, which is crucial for Bayesian model averaging:

$$\Pr\left(M_j \mid y\right) = \frac{\Pr\left(M_j\right)p\left(y \mid M_j\right)}{\sum_{j=1}^{2^K} \Pr\left(M_j\right)p\left(y \mid M_j\right)}, \tag{20}$$

where $\Pr(M_1), \Pr(M_2), \ldots, \Pr(M_K)$ denote the prior probabilities of competitive models (see Equation 19).

Using BMA, we can also easily evaluate the mean and variance of the posterior distribution of regression parameters β for the whole model space Leamer (1978); Koop et al. (2007):

$$E(\beta \mid y) = \sum_{i=1}^{2^K} \Pr(M_i \mid y)\hat{\beta}_i, \tag{21}$$

$$Var(\beta \mid y) = \sum_{i=1}^{2^K} \Pr(M_i \mid y)Var(\beta_i \mid y, M_i) + \sum_{i=1}^{2^K} \Pr(M_i \mid y)\left(\hat{\beta}_i - E(\beta \mid y)\right)^2, \tag{22}$$

where $\hat{\beta}_i = E(\beta_i \mid y, M_i)$ and $Var(\beta_i \mid y, M_i)$ are the OLS estimates of β_i from model M_i.

Another useful and popular characteristic of the BMA approach is posterior inclusion probability (PIP), which is defined as the posterior probability that the variable x_i is relevant in the explanation of the dependent variable Leamer (1978); Mitchell and Beauchamp (1988). In our case, the PIP is calculated as the sum of the posterior model probabilities for all of the models that include a specific variable:

$$\Pr(\beta_i \neq 0 \mid y) = \sum_{i=1}^{2^K} \Pr(M_i \mid y). \tag{23}$$

For model averaging, a Bayesian pooling strategy can also provide useful information about future observations of the dependent variable on the basis of the whole model space:

$$E(y_f \mid y) = \sum_{i=1}^{2^K} \Pr(M_i \mid y) E(y_f \mid y, M_i), \tag{24}$$

$$Var(y_f \mid y) = \sum_{i=1}^{2^K} \Pr(M_i \mid y) Var(y_f \mid y, M_i)$$
$$+ \sum_{i=1}^{2^K} \Pr(M_i \mid y) \left(E(y_f \mid y, M_i) - E(y_f \mid y) \right)^2, \tag{25}$$

where $E(y_f \mid y)$ and $Var(y_f \mid y)$ denote the mean and variance of future observations y_f.

3 Usage of the BMA_ADL package

3.1 BMA_ADL main function

The BMA_ADL package defines two public functions: `BMA_ADL_GUI()` and `BMA_ADL_Print()`. The `BMA_ADL_GUI()` takes following parameters:

- **Dependent variable** as series [required].

- **Y lags** as integer with zero (only for time-series).

- **List of independent variables (X)** as list [may be null].

- **Constant in model** as integer [required]:

    1. Never – all models without constant,
    2. Always – all models with constant,
    3. Can be dropped – constant may be removed from or added to any model.

- **Model prior** as integer [required]:

    1. Binomial – binomial distribution,

    2. Binomial-Beta – binomial-beta distribution[1].

- **Prior average model size** as scalar with zero [required].

- **Significance level for the initial model** as scalar from $(0, 1)$ [required].

- **Number of the top ranked models** as integer [required].

- **Jointness analysis** as integer with zero [required][2]:

    1. None – we do not perform any jointness analysis,

    2. Ley-Steel Measure – we perform jointness analysis with Ley-Steel measure (Ley and Steel, 2007),

    3. Doppelhofer-Weeks Measure – we perform jointness analysis with Doppelhofer-Weeks measure (Doppelhofer and Weeks, 2009).

- **Number of out-of-sample forecasts** as integer with zero [required].

- **Total number of replications** as integer [required].

- **Percentage of burn-in draws** as integer from $(0, 99)$ [required].

- **Verbosity** as integer [required].

- **Show overall progress** as Boolean.

The `BMA_ADL_GUI()` function returns a bundle with all information needed to print results anytime. One can print the results by calling `BMA_ADL_Print(&RES)`, where `RES` is the bundle returned by the `BMA_ADL_GUI()` function.

3.2   The GUI way

Once you start **gretl**, you must open a data file and then you can load the relevant **BMA_ADL** package from the **gretl** server. In the main window, go to File → Function packages → On server heading. By selecting **BMA_ADL**, you will download the package on your local machine. Next, go to File → Function packages → On local machine and by selecting BMA_ADL from the list open a window similar to the one shown in Figure 1:

    According to Figure 1, we can specify the following entries in the GUI **BMA_ADL** window

- **Dependent variable** – The dependent variable.

- **Y lags** – number of lags for dependent variable (for time-series only).

- **List of independent variables** – Loading variables from the database, which must have been opened previously. Note that by default we assume that you want to estimate an intercept; therefore, a constant is implicitly included in the list of the variables.

- **Constant in model** – Do we include intercept? It can be: always, never or can be dropped.

---

[1]Note: this option may generate errors as this distribution follows Occam's razor principle in a very aggressive way.

[2]Values ± 999.000000 should be interpreted as $\pm\infty$.
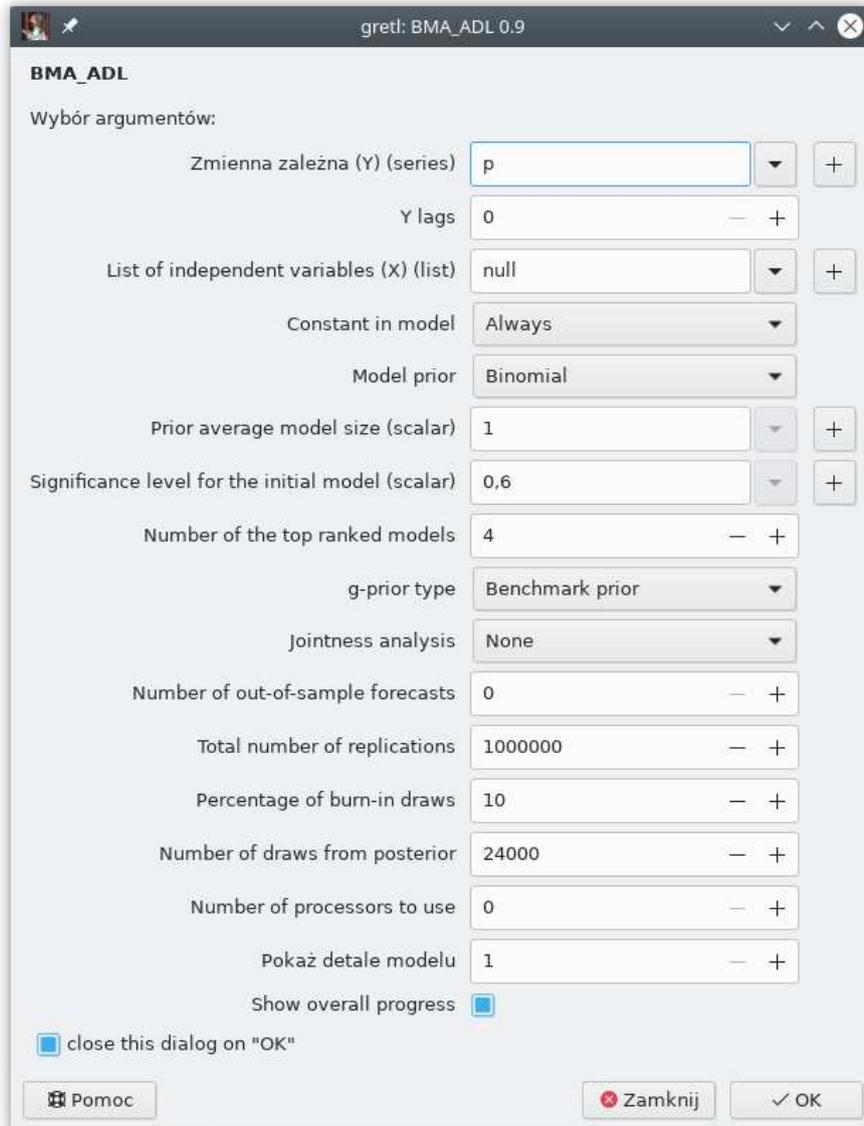
Figure 1: Main window for **BMA_ADL**.

- **Model prior** – Indicates the choice of model prior. One can employ the binomial model prior or the binomial-beta model prior. Note that the uniform model prior is a special case of the binomial model prior. Therefore, in fact, our package allows for three types of priors.

- **Prior average model size** – Specifies the prior expected model size $\mathsf{E}(\Xi)$. The expected model size may range from 0 to K. The default value equals 1. It is the smallest reasonable prior expected model size. It means that we penalize large models and we assign high probability to small ones

following Occam's razor principle.

Note that for the binomial model prior and with $E(\Xi) = 0.5K$ one can define the uniform prior on the model space. For example, for $K = 10$ regressors we can define the uniform prior with $E(\Xi) = 5$. For $E(\Xi) < 5$ we assign high probability to small models. The smaller the prior expected model size $E(\Xi)$, the less probable are larger models.

- **Significance level for the initial model** – Defines the significance level which is used to build the initial model. An explanatory variable enters the initial model if its p value is less than the significance level. If the significance level equals 1, the initial model will be randomly chosen (with equal probability) from all available models. Note that if all available explanatory variables enter the initial model, you will get the following **gretl** error message "No independent variables were omitted".

- **Number of top ranked models** – Specifies the number of best models for which detailed information is stored.

- **g-prior type** - Here one can choose between four Zellner's g-prior forthe regression coefficients. Choices include: Benchmark prior, Unit Information Prior, Risk Inflation Criterion, Hannan and Quinn prior, Root of g-UIP.

- **Jointness analysis** – If `None` (the default), the jointness analysis is omitted. Alternatively, one can choose the jointness measures of Ley and Steel (2007) or Doppelhofer and Weeks (2009).

- **Number of out-of-sample forecasts** – Defines the total number of out-of-sample forecasts of the dependent variable.

- **Total number of replications** – Defines the total number of iteration draws to be sampled.

- **Percentage of burn-in draws** – Specifies the number of burn-in replications, calculated as the percentage of the total number of iteration draws.

- **Number of draws from posterior** – Defines the number of draws samples of slope parameters from posterior distribution.

- **Number of processors to use** – Defines the number of CPUs (phisical or logical) to use when drawing from posterior. Default value 0 means that package use maximum available CPUs.

- **Verbosity** – An integer value of 1 or 2; the default is 1, which allows to see the basic Bayesian Averaging of Classical Estimates results. If Verbosity equals 2, a more detailed description of the analysis is provided (initial model, speed of convergence, estimation results for top ranked models).

- **Show overall progress** – The output is flushed every 10% of MC^3 and the progress is printed. Note: this consumes a lot of CPU time.

3.3 Output

If you select the appropriate entries in the GUI **BMA_ADL** window, our package returns: Posterior Inclusion Probabilities (PIP), the averages of the posterior means and standard deviations of parameters ("Mean" and "Std.Dev.", respectively) and the posterior mean and standard deviation of each coefficient conditional on the variable being included in the model ("Cond.Mean" and "Cond.Std.Dev."). Furthermore, the package also returns the predictive results for the dependent variable ("Mean" and "Std.Dev.").

Let us consider the data used in previous example (`australia.gdt`). This dataset contains the Australian macro data used in Johansen (1995) and originally supplied by Tony Hall. The variables used in the text are:

- lpau – log of Australian CPI (dependent variable),

- lpus – log of US CPI,

- le – log of exchange rate (price of US dollars in Aus. dollars),

- iau2 – 5-year Treasury bond rate, Australia,

- ius2 – 5-year Treasury bond rate, USA.

Suppose we want to perform BMA_ADL analysis up to $ADL(2,2)$ specification, i.e. all possible combinations from $ADL(0,0)$ to $ADL(2,2)$. The example script would be as follows:

```
set verbose off
include BMA_ADL.gfn
open australia.gdt
genr time

# We create lagged values
adl_lags = 2
lags(adl_lags, ius2)
lags(adl_lags, iau2)
lags(adl_lags, e2)
lags(adl_lags, lpus)
lags(adl_lags, le)

# We define list with all X-es (including lags)
list X_list = time ius2 ius2_* iau2 iau2_* e2 e2_* lpus lpus_* le le_*

set seed 1000000
scalar y_lags = 2
scalar with_const = 2
scalar model_prior = 1
scalar gprior = 1
scalar prior_avg_model_size = (nelem(X_list) + y_lags + (with_const > 0)) / 2
scalar alpha = 0.6
scalar top_ranked_model = 4
scalar joint = 0
scalar forecasts = 0
```

```
scalar Nrep = 1000
scalar Nburn = 10
scalar Ndraws = 2000
scalar Ncpus = 0
scalar verbosity = 2
scalar progress = 0

RES = BMA_ADL_GUI(lpau, y_lags, X_list, with_const, model_prior,
prior_avg_model_size, alpha, top_ranked_model, gprior, joint,
forecasts, Nrep, Nburn, Ndraws, Ncpus, verbosity, progress)
```

In the above analysis all variables (including dependent variable $lpau_t$) are lagged by 2 periods. Additionally we compute 4 forecasts:

```
Posterior average model size: 6.284444


Posterior moments (unconditional and conditional on inclusion):
              PIP         Mean     Std.Dev.    Cond.Mean   Cond.Std.Dev
ius2      1.000000    -0.377824    0.066905    -0.377824      0.066905
lpau_1    1.000000     0.904691    0.078254     0.904691      0.078254
lpus_1    0.996667     0.533282    0.158362     0.535065      0.155589
iau2      0.982222     0.359475    0.095730     0.365981      0.083360
lpus_2    0.966667    -0.463944    0.146446    -0.479942      0.120448
e2_2      0.222222    -0.002856    0.008493    -0.012853      0.014004
le_1      0.194444    -0.003440    0.009450    -0.017692      0.014391
lpau_2    0.184444     0.029732    0.076984     0.161198      0.104594
le_2      0.138889    -0.002219    0.007811    -0.015980      0.014811
iau2_1    0.100000     0.012392    0.058551     0.123915      0.143048
e2        0.083333     0.000425    0.005157     0.005094      0.017187
ius2_2    0.082222    -0.002623    0.024260    -0.031898      0.078893
le        0.067778     0.000709    0.006339     0.010462      0.022153
iau2_2    0.055556     0.002842    0.022305     0.051162      0.080519
ius2_1    0.055556     0.005347    0.031404     0.096237      0.094892
time      0.055556     0.000001    0.000052     0.000009      0.000222
lpus      0.046667     0.002892    0.036589     0.061968      0.158197
e2_1      0.032222    -0.000094    0.002737    -0.002931      0.014975
const     0.020000    -0.000551    0.009674    -0.027558      0.062733


---------------------------------


Posterior probability of models:
Model 1: 0.274444
Model 2: 0.082222
Model 3: 0.061111
Model 4: 0.035556
Total probability of the models in ranking (numerical): 0.453333


Correlation coefficient between the analytical
and numerical probabilities of the above models: 0.978674


---------------------------------
```

4 Conclusions

This paper has outlined the new software package that implements Bayesian Model Averaging for ADL models in gretl, which is a free, open-source software for econometric analysis with an easy-to-use GUI. Our goal was to familiarize potential users with the features and the different options that our package has to offer. We described how our package implements the Bayesian Model Averaging, as well as the outputs that are returned. However, the BMA_ADL package is still under development and needs intensive optimizations and improvements. The future development of the BMA_ADL package will be focused on speed improvements and features extension.

References

Doppelhofer G, Weeks M. 2009. Jointness of Growth Determinants. *Journal of Applied Econometrics* **24**: 209–244.

Fernández C, Ley E, Steel MFJ. 2001a. Benchmark Priors for Bayesian Model Averaging. *Journal of Econometrics* **100**: 381–427. ISSN 0304-4076.

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

Foster DP, George EI. 1994. The risk inflation criterion for multiple regression. *The Annals of Statistics* **22**: 1947–1975.
URL http://dx.doi.org/10.1214/aos/1176325766

Hendry DF. 2015. *Introductory Macro-econometrics: A New Approach.* London: Timberlake Consultants. ISBN 978-0-9932738-0-3.

Hoeting JA, Madigan D, Raftery AE, Volinsky CT. 1999. Bayesian model averaging: A tutorial. *Statistical Science* **14**: 382–401. ISSN 08834237.

Johansen S. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* Oxford University Press.

Kass RE, Wasserman L. 1995. A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the American Statistical Association* **90**: 928–934.
URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476592

Koop G. 2003. *Bayesian Econometrics.* Chichester: John Wiley & Sons Ltd. ISBN 0-470-84567-8.

Koop G, Poirier DJ, Tobias JL. 2007. *Bayesian econometric methods.* New York: Cambridge University Press. ISBN 978-0-521-67173-6.

Krolzig HM, Hendry DF. 2001. Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control* **25**: 831–866. ISSN 0165-1889.

Leamer E. 1978. *Specification Searches.* New Jersey, USA: John Wiley & Sons. ISBN 0471015202.

Ley E, Steel M. 2007. Jointness in Bayesian Variable Selection with Applications to Growth Regression. *Journal of Macroeconomics* **29**: 476–493.

Ley E, Steel MFJ. 2009. On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics* **24**: 651–674.

Mitchell TJ, Beauchamp JJ. 1988. Bayesian variable selection in linear regression. *Journal of the American Statistical Association* **83**: 1023–1032.

Moral-Benito E. 2010. Model averaging in economics. Working Papers wp2010_1008, CEMFI.
URL http://ideas.repec.org/p/cmf/wpaper/wp2010_1008.html

Raftery AE, Madigan D, Hoeting JA. 1997. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* **92**: 179–191.

Steel MFJ. 2019. Model Averaging and Its Use in Economics. *Journal of Economic Literature* (**forthcoming**).

Zellner A. 1971. *An Introduction to Bayesian Inference in Econometrics*. New York: John Wiley & Sons.

Zellner A. 1986. On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions. In Goel P, Zellner A (eds.) *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*. Amsterdam, Holland: Elsevier. ISBN 978-04-4487-712-3.