BAYESIAN MODEL AVERAGING AND JOINTNESS MEASURES: THEORETICAL FRAMEWORK AND APPLICATION TO THE GRAVITY MODEL OF TRADE

Krzysztof Beck[1]

ABSTRACT

The following study presents the idea of Bayesian model averaging (BMA), as well as the benefits coming from combining the knowledge obtained on the basis of analysis of different models. The BMA structure is described together with its most important statistics, g prior parameter proposals, prior model size distributions, and also the jointness measures proposed by Ley and Steel (2007), as well as Doppelhofer and Weeks (2009). The application of BMA is illustrated with the gravity model of trade, where determinants of trade are chosen from the list of nine different variables. The employment of BMA enabled the identification of four robust determinants: geographical distance, real GDP product, population product and real GDP *per capita* distance. At the same time applications of jointness measures reveal some rather surprising relationships between the variables, as well as demonstrate the superiority of Ley and Steel's measure over the one introduced by Dopplehofer and Weeks.

Key words: Bayesian model averaging, jointness measures, multi-model inference, gravity model of trade.

1. Introduction

In economics, a situation often arises when a vast number of different theories attempt to explain the same phenomenon. Although these theories may complement each other, it is very common that they contradict one another or are even mutually exclusive. In such cases, basing empirical verification on one or a few specifications of an econometric model turns out to be insufficient. Moreover, researchers applying varying specifications will arrive at different, very often incoherent or even contradictory, conclusions. Testing hypotheses on the basis of various economic model specifications can result in a situation in which a variable that is statistically significant in one research specification, may prove to be not significant in another one.

[1] Lazarski University. E-mail: beckkrzysztof@gmail.com.

Brock and Durlauf (2001) draw attention to a problem they called theory open-endedness. It takes place in a situation where two or more competing models propose different explanations of the same phenomenon, and each of the variables proposed as an explanation can be expressed using a different measure. Moreover, some of the theories can complement each other, while other serve as substitutes or even contravene each other. In such a situation, inference based on a single model can lead to contradictory or false conclusions.

The above-mentioned problem is clearly present in the context of the research into the determinants of international trade. The vast body of trade theories offers a great variety of explanations for international trade flows, which can be seen in any international economics textbook. What is more, there is considerable dispute over potential effects of participation in free trade agreements as well as monetary unions on international trade. Even though the gravity model of trade has been the backbone of international trade empirics for over half the century, it is still rather unclear which variables should accompany the core of the model. The literature is full of competing specifications without much attention paid to robustness checks.

For these reasons, this paper pertains to the transition from statistical relevance to basing inference on the robustness of results against a change in the specifications of a model. However, in such a case it is necessary to apply inference and combination of knowledge coming from different model specifications. In such a situation, it is possible to apply BMA, i.e. Bayesian Model Averaging. Through the estimation of all the models within a given set of data, this procedure allows one to determine which variables are robust regressors regardless of the specification. It also allows one to unequivocally establish the direction and strength given regressors possess, and it makes it possible to choose the best models of all possible configurations. Furthermore, using the jointness measures that are available within the BMA framework enables the determination of the substitutional and complementary relationships between the studied variables.

Therefore, for the above-mentioned reasons, BMA and jointness measures are the subject of this study. Theory and structure of Bayesian model averaging is presented in the first section while in the second one jointness measures are discussed. The third section provides an example of BMA application in the analysis of the gravity model of trade and comprises four sub-sections. In the first one, the gravity model of trade is presented, whereas the second shows the variables employed in the verification of the model. The third sub-section presents the results of applying BMA, and the fourth one demonstrates the results of the analysis using jointness measures. The last section provides the summary and conclusions of the article.

2. BMA – Bayesian Model Averaging

For the space of all models, unconditional posterior distribution of coefficient β is given by:

$$P(\beta|y) = \sum_{j=1}^{2^K} P(\beta|M_j, y) * P(M_j|y) \tag{1}$$

where: y denotes data, j (j=1, 2,..,m) is the number of the model, K being the total number of potential regressors, $P(\beta|M_j, y)$ is the conditional distribution of coefficient β for a given model M_j, and $P(M_j|y)$ is the posterior probability of the model. Using the Bayes' theorem, the posterior probability of the model (PMP – Posterior Model Probability) $P(M_j|y)$ can be rendered as (Błażejowski *et al.*, 2016):

$$PMP = p(M_j|y) = \frac{l(y|M_j) * p(M_j)}{p(y)}, \tag{2}$$

where PMP is proportional to the product of $l(y|M_j)$ – model specific marginal likelihood – and $P(M_j)$ – model specific prior probability – which can be written down as $P(M_j|y) \propto l(y|M_j) * P(M_j)$. Moreover, because: $P(y) = \sum_{j=1}^{2^K} l(y|M_j) * P(M_j)$, weights of individual models can be transformed into probabilities through the normalization in relation to the space of all 2^K models:

$$P(M_j|y) = \frac{l(y|M_j) * P(M_j)}{\sum_{j=1}^{2^K} l(y|M_j) * P(M_j)}. \tag{3}$$

Applying BMA requires specifying the prior structure of the model. The value of the coefficients β is characterized by normal distribution with zero mean and variance $\sigma^2 V_{oj}$, hence:

$$P(\beta|\sigma^2, M_j) \sim N(0, \sigma^2 V_{oj}). \tag{4}$$

It is assumed that the prior variance matrix V_{oj} is proportional to the covariance in the sample: $(gX_j'X_j)^{-1}$, where g is the proportionality coefficient. The g prior parameter was put forward by Zellner (1986) and is widely used in BMA applications. In their seminal work on the subject of choosing the g prior Fernández *et al.* (2001) put forward the following rule, to choose the best g prior:

$$g = \frac{1}{\max(n, k^2)}, \tag{5}$$

where $\frac{1}{n}$ is known as UIP – unit information prior (Kass and Wasserman, 1995), whereas $\frac{1}{k^2}$ is convergent to RIC – risk inflation criterion (Foster and George, 1994). For further discussion on the subject of g priors see: Ley and Steel (2009, 2012); Feldkircher and Zeugner (2009); and Eicher *et al.* (2011).

Besides the specification of g prior, it is necessary to determine the prior model distribution while applying BMA. For binomial model prior (Sala-I-Martin *et al.*, 2004):

$$P\big(M_j\big) \propto \left(\frac{Em}{K}\right)^{k_j} * \left(1 - \frac{Em}{K}\right)^{K-k_j}, \tag{6}$$

where Em denotes the expected model size, while k_j the number of covariate in a given model. When $Em = \frac{K}{2}$ it turns into uniform model prior – priors on all the models are all equal $(P\big(M_j\big) \propto 1)$. Yet another instance of prior model probability is binomial-beta distribution (Ley, Steel, 2009):

$$P\big(M_j\big) \propto \Gamma\big(1 + k_j\big) * \Gamma\left(\frac{K - Em}{Em} + K - k_j\right). \tag{7}$$

In the case of binomial-beta distribution with expected model size *K/2*, the probability of a model of each size is the same $\left(\frac{1}{K+1}\right)$. Thus, the prior probability of including the variable in the model amounts to 0.5, for both binomial and binomial-beta prior with $Em = K/2$.

Using the posterior probabilities of the models in the role of weights allows one to calculate the unconditional posterior mean and standard deviation of the coefficient β_i. Posterior mean (PM) of the coefficient β_i, independent of the space of the models, is then given with the following formula (Próchniak, Witkowski, 2012):

$$PM = E(\beta_i|y) = \sum_{j=1}^{2^K} P\big(M_j|y\big) * \hat{\beta}_{ij}, \tag{8}$$

where $\hat{\beta}_{ij} = E\big(\beta_i|y, M_j\big)$ is the value of the coefficient β_i estimated with OLS for the model M_j. The posterior standard deviation (PSD) is equal to (Próchniak, Witkowski, 2014):

$$PSD = \sqrt{\sum_{j=1}^{2^K} P\big(M_j|y\big) * V\big(\beta_j|y, M_j\big) + \sum_{j=1}^{2^K} P\big(M_j|y\big) * \big[\hat{\beta}_{ij} - E(\beta_i|y, M_j)\big]^2}, \tag{9}$$

where $V(\beta_j|y, M_j)$ denotes the conditional variance of the parameter for the model M_j.

The most important statistic for BMA is posterior inclusion probability (PIP). PIP for the regressor x_i equals:

$$PIP = P(x_i|y) = \sum_{j=1}^{2^K} 1(\varphi_i = 1|y, M_j) * P(M_j|y) \qquad (10)$$

where $\varphi_i = 1$ indicates that the variable x_i is included in the model.

PM and PSD are calculated for all models, even those whose value $\varphi_i = 0$, which means that the variable is not present. Due to that fact the researcher can be interested in the value of the coefficient in the models in which a given variable is present. For that purpose, the value of the conditional posterior mean (PMC), that is the posterior mean, can be calculated on condition that a variable is included in the model:

$$PMC = E(\beta_i|\varphi_i = 1, y) = \frac{E(\beta_i|y)}{P(x_i|y)} = \frac{\sum_{j=1}^{2^K} P(M_j|y) * \hat{\beta}_{ij}}{P(x_i|y)}, \qquad (11)$$

whereas the conditional posterior standard deviation (PSDC) is given by:

$$PSDC = \sqrt{\frac{V(\beta_j|y) + [E(\beta_i|y)]^2}{P(x_i|y)} - [E(\beta_i|\varphi_i = 1|y)]^2}. \qquad (12)$$

Additionally, the researcher can be interested in the sign of the estimated parameter if it is included in the model. The posterior probability of a positive sign of the coefficient in the model [P(+)] is calculated in the following way:

$$P(+) = P[sign(x_i)|y] = \begin{cases} \sum_{j=1}^{2^K} P(M_j|y) * CDF(t_{ij}|M_j), & if\ sign[E(\beta_i|y)] = 1 \\ 1 - \sum_{j=1}^{2^K} P(M_j|y) * CDF(t_{ij}|M_j), & if\ sign[E(\beta_i|y)] = -1 \end{cases}$$

$$(13)$$

where CDF denotes cumulative distribution function, while $t_{ij} \equiv (\hat{\beta}_i/\widehat{SD}_i|M_j)$.

3. Jointness measures

All the statistics cited so far served to describe the influence of regressors on the dependent variable. However, the researcher should also be interested in relationships that emerge between the independent variables. To achieve that, one can utilize the measure of dependence between regressors, which is referred to as *jointness*.

Two teams of scientists came up with jointness measures at the same time. The article by Ley and Steel (2007) was published first; however, in this paper the concept of Doppelhofer and Weeks (2009) shall be presented first due to the fact that Ley and Steel's article constitutes by and large the critique of Dopplehofer and Weeks' concepts. Measures allow the determination of the substitution and complementary relationships between explanatory variables. Below, the focus will be put only on the jointness relationships between pairs of variables. It must also be mentioned, however, that testing the relationships between triplets or even more numerous sets of variables is possible.

We shall define posterior probabilities for the model M_j as:

$$(M_j|y) = P(\varphi_1 = w_1, \varphi_2 = w_2, \dots, \varphi_K = w_K|y, M_j) \qquad (14)$$

where w_i can assume value 1 (if a variable is present in the model) and 0 if a variable is not present in the model. In the case of analysing two variables x_i and x_h the combined posterior probability of including two variables in the model can be expressed as follows:

$$P(i \cap h|y) = \sum_{j=1}^{2^K} 1(\varphi_i = 1 \cap \varphi_2 = 1|y, M_j) * P(M_j|y). \qquad (15)$$

Table 1. Points of probability mass defined on space $\{0,1\}^2$ for uniform distribution $P(\varphi_i, \varphi_l|y)$.

$P(\varphi_i, \varphi_l|y)$	$\varphi_h = 0$	$\varphi_h = 1$	Sum
$\varphi_i = 0$	$P(\bar{\imath} \cap \bar{h}|y)$	$P(\bar{\imath} \cap h|y)$	$P(\bar{\imath}|y)$
$\varphi_i = 1$	$P(i \cap \bar{h}|y)$	$P(i \cap h|y)$	$P(i|y)$
Sum	$P(\bar{h}|y)$	$P(h|y)$	1

Source: Doppelhofer, Weeks, 2009.

It can be thus stated that $P(i \cap h|y)$ is the sum of the posterior probability of the models, where variables marked by x_i and x_h appear. Doppelhofer and Weeks observe that the relationships between variables x_i and x_h can be analyzed by comparing posterior probabilities of including these variables separately $[P(i|y)$ and $P(h|y)]$ with probability of including and excluding both variables at the same time. The authors justify their reasoning by presenting an analysis of the case of a random vector (φ_i, φ_h) of the combined posterior distribution $P(\varphi_i, \varphi_l|y)$. The points of probability mass defined on space $\{0,1\}^2$ are shown in Table 1.

Table 1 shows distributions related to all the possible realizations of vector (φ_i, φ_h). It is easy to read from the table that the marginal probability of including variable x_i in the model can be calculated as:

$$P(i|y) = P(i \cap h|y) + P(i \cap \bar{h}|y), \qquad (16)$$

whereas the probability of excluding the variable x_i can be rendered as:

$$P(\bar{\imath}|y) \equiv 1 - P(i|y) = P(\bar{\imath} \cap \bar{h}|y) + P(\bar{\imath} \cap h|y). \qquad (17)$$

If there is a correlation between variables x_i and x_h, one should expect that expressions $P(i \cap h|y)$ and $P(\bar{\imath} \cap \bar{h}|y)$ will get higher values than expressions $P(i \cap \bar{h} \backslash y)$ and $P(\bar{\imath} \cap h|y)$. On that basis, to follow Whittaker (2009), the authors observe that the natural measure of correlation between two binary random variables φ_i and φ_h is the cross-product ratio (CPR), expressed as:

$$CPR(i, h|y) = \frac{P(i \cap h|y)}{P(i \cap \bar{h}|y)} * \frac{P(\bar{\imath} \cap \bar{h}|y)}{P(\bar{\imath} \cap h|y)}. \qquad (18)$$

As the realizations of the vector (φ_i, φ_h) for each of the variables can only amount to 1 or 0, $P(i \cap h|y)$ is the binomial distribution of the uniform posterior probability i, which can be rendered as follows:

$$P(\varphi_i, \varphi_h|y) = P(i \cap h|y)^{\varphi_i \varphi_h} * P(i \cap \bar{h}|y)^{\varphi_i(1-\varphi_h)} *$$
$$* P(\bar{\imath} \cap h|y)^{(1-\varphi_i)\varphi_h} * P(\bar{\imath} \cap \bar{h}|y)^{(1-\varphi_i)(1-\varphi_h)} \qquad (19)$$

Logarithmized and put in order, the expressions take the following form:

$$ln[P(\varphi_i, \varphi_h|y)] = ln[P(\bar{\imath} \cap \bar{h}|y)] + \varphi_h ln\left[\frac{P(\bar{\imath} \cap h|y)}{P(\bar{\imath} \cap \bar{h}|y)}\right] +$$

$$+\varphi_i ln\left[\frac{P(i \cap \bar{h}|y)}{P(\bar{\imath} \cap \bar{h}|y)}\right] + \varphi_i \varphi_h ln\left[\frac{P(i \cap h)}{P(i \cap \bar{h}\backslash|)} * \frac{P(\bar{\imath} \cap \bar{h}|y)}{P(\bar{\imath} \cap h\backslash|)}\right] \qquad (20)$$

The independence between variables x_i and x_h is possible if and only if $ln[P(\varphi_i, \varphi_h|y)]$ is additive for $P(\varphi_i|y)$ and $P(\varphi_h|y)$. Independence can therefore occur if and only if the natural logarithm of CPR is 0, which means CPR equals 1.

On that basis, Doppelhofer and Weeks derive their jointness measure, which they define as:

$$J_{Dw(ih)} = ln[CPR(i,h|y)] = ln\left[\frac{P(i \cap h|y)}{P(i \cap \bar{h}|y)} * \frac{P(\bar{\imath} \cap \bar{h}|y)}{P(\bar{\imath} \cap h|y)}\right] =$$

$$= ln\left[\frac{P(i|h,y)}{P(\bar{\imath}|h,y)} * \frac{P(\bar{\imath}|\bar{h},y)}{P(i|\bar{h},y)}\right] = ln[PO_{i|h} * PO_{\bar{\imath}|\bar{h}}]. \qquad (21)$$

The expression $ln[PO_{i|h} * PO_{\bar{\imath}|\bar{h}}]$ is the natural logarithm of the product of two quotients of posterior odds, where $PO_{i|h}$ indicates posterior odds of including the variable x_i to the model on condition that x_h is included, while $PO_{\bar{\imath}|\bar{h}}$ indicates posterior odds of excluding the variable x_i from the model on condition that the variable x_h is excluded.

At this moment, it is worth pointing out that if the probability product of including and excluding both variables $[(P(i \cap h|y) * P(\bar{\imath} \cap \bar{h}|y)]$ is greater than the probability product of including each of the variables one at a time $[P(i \cap \bar{h}|y) * P(\bar{\imath} \cap h|y)]$, then the logarithm assumes positive values. Thus, for the positive values of the measure, complementary relationship has to occur: models that include both variables at the same time or reject both variables at the same time are characterized by the highest posterior probability. If the product of probabilities of including the variables separately is greater than the product of including both or neither at the same time, the logarithm takes negative values. In such an event, a substitutional relationship occurs. To sum up, Doppelhofer and Weeks' jointness measure assumes positive values if there is a complementary relationship between variables, whereas it assumes negative values when this relationship is of substitutional character.

Ley and Steel (2007) set out to develop a jointness measure that would possess the following characteristics:
1) Interpretability – a measure should have a formal statistical or intuitive interpretation.
2) Calibration – values of a measure should be determined on a clearly defined scale based on formal statistical or intuitive interpretation.
3) Extreme jointness – in a situation when two variables appear in all the analyzed models together (e.g. in the case of using MC[3] methods), the maximum value of jointness measure should occur;
4) Definability – jointness should be defined always if at least one of the considered variables is characterized by positive inclusion probability.

Ley and Steel claimed that Doppelhofer and Weeks' jointness measure is faulty as it is not defined in a situation when both regressors are included in all models and when one of the regressors is not taken into consideration in any of the models. Moreover, when the probability of including a variable in the model approaches 1, then the value of the measure is by and large dependent on the limit of the expression $\left[P(\bar{\iota} \cap \bar{h}|y)\right]/\left[P(\bar{\iota} \cap h|y)\right]$. This means that a few models, excluding the variable x_i, that are characterized by a very low probability can strongly influence the value of the measure: both in the direction of 0 (if they include the variable x_h) or ∞ (if they do not include the variable x_h). Thus, the measure $J_{Dw(ih)}$ does not contain features 1) and 4).

What is more, the authors pointed out that the interpretation of Doppelhofer and Weeks' measure is not clear enough and, due to this fact, they proposed an alternative measure. This measure is the ratio of probability of including two variables simultaneously to the sum of probabilities of including each of the variables separately, with the exclusion of the probability of including two variables at the same time. This measure meets all the criteria laid out by the authors. Ley and Steel's jointness measure is given by:

$$J_{LS(ih)} = ln\left[\frac{P(i \cap h|y)}{P\left(i \cap \bar{h}|y\right) + P(\bar{\iota} \cap h|y)}\right]$$

$$= ln\left[\frac{P(i \cap h|y)}{P(i|y) + P(h|y) - 2P(i \cap h|y)}\right]. \tag{22}$$

The advantage of this measure is its interpretative clarity. The expression inside the natural logarithm represents the quotient of posterior odds of models including both variables to the models including each of them separately. Again, the logarithm of this expression takes positive values if the probability of the models including both variables is dominant, which testifies to the complementary relationship. The measure takes negative values if posterior odds of the models including variables separately are higher than in the case where variables appear in the model simultaneously, which testifies to a substitutional relationship.

Doppelhofer and Weeks calculated the limit values of jointness measures, which allow qualifying variables to one of five categories. These values also hold in the case of Lay and Steel's jointness measure. The limit values of jointness measures with their corresponding classifications of relationships between variables are presented in Table 2.

Table 2. Limit values of jointness measures and classification of relationships
between variables

Type of the relationship between the variables	Value of the jointness measure (*J*)
Strong substitutes	*J < (-2)*
Significant substitutes	*(-2) < J < (-1)*
Unrelated variables	*(-1) < J < 1*
Significant complements	*1 < J < 2*
Strong complements	*2< J*

Source: Błażejowski, Kwiatkowski, 2015.

4. Application on the example of the gravity model of trade

All the empirical analyses employing BMA were carried out using BMS
package for R environment (Zeugner and Feldkircher, 2015). Jointness measures
were computed using a package for gretl (Błażejowski and Kwiatkowski, 2015).

4.1. Gravity model of trade

In the simplest form, the equation describing the gravity model of trade
(Anderson, 1979, 2011; Egger, 2002; Anderson, Wincoop, 2003) can be shown
as:

$$TRADE = \alpha \frac{(RGDPprod)^{\beta_1}}{DIST^{\beta_2}}, \qquad (23)$$

which can be easily transformed into a log-linear form:

$$\ln(TRADE) = \ln(\alpha) + \beta_1 \ln(RGDPprod) - \beta_2 \ln(DIST), \qquad (24)$$

where *TRADE* stands for the amount of international trade, *RGDPprod* – product
of real GDP of the two countries, *DIST* – distance between the countries, whereas
α, β_1, β_2 are parameters in the model. However, the model can be expanded by
including additional explanatory variables, which was performed in this paper.

4.2. Variables and source of data

Data for 19 European Union countries was used, namely: Austria, Belgium,
Cyprus, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy,
Luxembourg, Malta, the Netherlands, Poland, Portugal, Spain, Sweden and the
UK. All the variables are expressed bilaterally and as a result the size of the

sample for each variable amounts to 171 pairs of countries. The period of analysis spans the years between 1999 and 2007 for all the variables.

Bilateral trade, which is expressed as logarithmized trade between partners, constitutes the response variable in the model:

$$TRADE_{ij} = ln\left(\frac{1}{T}\sum_{t=1}^{T} Import_{ijt} + Export_{ijt}\right), \qquad (25)$$

where i and j are indexes of partner countries, and the measure itself is a mean for the entire analyzed period (1, 2, …, T). The data on bilateral trade are taken from IMF Directions of Trade.

In the BMA analysis, 9 variables were employed. The first one constitutes the logarithm of the product of real GDPs:

$$RGDPprod_{ij} = ln\left(\frac{1}{T}\sum_{t=1}^{T} GDP_{it} * GDP_{jt}\right), \qquad (26)$$

also treated as a mean for the whole period. Data on the subject of real GDP are taken from the Penn World Table. The second of the main gravity variables is the natural logarithm of the distance between the capitals of the countries under consideration, which is marked as *DIST*.

The basic explanatory variables in the gravity model of trade were complemented by additional 7. The first one is the similarity of the production structures measured by Krugman specialization index (1991):

$$KSI_{ij} = \frac{1}{T}\sum_{t=1}^{T}\sum_{l}^{17} |v_{it}^{l} - v_{jt}^{l}|, \qquad (27)$$

where v_{it}^{l} is the value added in the sector l expressed as the percentage of the value added in the entire economy of the country i in the period t, v_{it}^{l} and is the value added in the sector l expressed as the percentage of the value added in the entire economy of a country j in the period t. The mean for the entire period and the division of the economy into 17 sectors were used, whereas the data on them were taken from EU KLEMS. The measure takes values from the interval [0,2], while the growth of the value of the measure is accompanied by the decrease in similarity of production structures.

The next variable added to the gravity model is the average absolute value of the difference of natural log of GDP *per capita* for each pair of countries in the period between 1999 and 2007:

$$RGDPdist_{ij} = \frac{1}{T}\sum_{t=1}^{T} |\ln(GDPpercapita_{it}) - \ln(GDPpercapita_{jt})|. \qquad (28)$$

The data on GDP *per capita* comes from the Penn World Table. The similarity of production structures and the distance of GDP *per capita* can be justified by the theory of monopolistic competition adopted by Linder (1961). The theory assumes that there is a tendency that, together with the increasing industrialization, the structures of consumption/production become more similar, which leads to a situation where countries at similar level of affluence will display a high level of intra-industry trade. These conclusions are supported by the works of: Grubel (1971), Grubel and Loyd (1975), Dixit and Stiglitz (1977), Krugman (1979, 1980), Lancaster (1980), Helpman (1981) and Gray (1980).

What is more, averaged binary variables were used in the models in order to reflect the influence of participation in the European Union and Economic and Monetary Union. For the participation in the monetary union (*MU*), the variable takes the value equal to 1 if in a given year both countries were members of the Eurozone, and 0 for other years. Then, a mean for the whole period is calculated. Analogical construction was applied for the participation in the European Union (*EU*):

Another potential determinant of bilateral trade is the natural logarithm of the population product of two analyzed EU countries in the period between 1990 and 2007 – *POPprod*. The data on the size of population come from the Penn World Table. One can expect substitutional relationship between *POPprod* and *RGDPprod*.

Moreover, two additional binary variables were used. They are: *BORDER* - a dummy variable assuming 1 if two countries share a common border, and *LANG* – a binary variable assuming 1 if a pair of countries share at least one official language.

4.3. The results of applying BMA

Below one can find the results of applying BMA after employing Fernández *et al.* (2001) *Benchmark Prior*, which dictated the choice of unit information prior (UIP). Additionally, uniform model size prior was applied. This combination of priors was recommended by Eicher *et al.* (2011). The prior probability of including a given regressor is 0.5. As 9 regressors were used, the space of the model consists of $2^K=2^9=512$ elements, and the inference itself was carried out on the basis of all models. The results of applying BMA are presented in Table 3.

The results indicate that 5 variables were qualified as robust determinants of bilateral trade: geographical distance, product of real GDPs, population product, GDP *per capita* distance, and common language. The remaining four display lower posterior than the prior probability of inclusion, which is 0.5. A stable sign of the coefficient among all the analyzed models also characterizes all the variables that were qualified as robust, and it is in accordance with expectations of the theory, with an exception of population product, which is characterized by negative posterior mean. *DIST* and *RGDPprod* turned out to be the most robust

determinants of trade – models including these variables take the lion's share of posterior probability mass. This ascertains the gravity model of trade capacity to explain international trade flows. *RGDPpc* has a negative impact on trade. This gives support to the theories that suggest a positive relationship between GDP *per capita* and the volume of intra-industry trade. On the other hand, similarity of the production structure is marked as fragile. It will be instructive to look at the value of the jointness measures for *RGDPdist* and *KSI*.

Table 3. BMA statistics with the use of uniform prior model size distribution (dependent variable - bilateral trade).

Variable	PIP	PM	PSD	CPM	CPSD	P(+)
DIST	1.000	-0.879	0.097	-0.879	0.097	0.000
RGDPprod	1.000	1.169	0.180	1.169	0.180	1.000
POPprod	0.827	-0.311	0.176	-0.376	0.114	0.000
RGDPdist	0.739	-0.336	0.242	-0.455	0.159	0.000
LANG	0.627	0.299	0.275	0.476	0.190	1.000
BORDER	0.380	0.139	0.209	0.365	0.180	1.000
KSI	0.369	-0.465	0.723	-1.260	0.645	0.000
EU	0.244	0.162	0.364	0.662	0.461	0.916
MU	0.152	0.022	0.069	0.146	0.116	1.000

A cultural similarity captured by the common language dummy proved to have a robust and positive impact on trade. Unexpected result was obtained for the population product. The variable is robust but is characterized by a negative posterior mean. This result is especially surprising when we look at correlation coefficient between *RGDPprod* and *POPprod* – 0.96. This suggests a substitutional dependence between those two variables.

The common border dummy was classified as fragile. This might be explained by potential substitutional relationship with geographical distance or language dummy – these variables most certainly carry the same information. Similarly, the membership in the European Union and the Eurozone are considered fragile. In instances of both of these variables one might expect a substitutional relationship with other regressors, e.g. *RGDPdist* (European Union/Eurozone members are characterized by lower GDP *per capita* distances compared with pairs with countries outside these entities) or *BORDER*.

The next step requires an inquiry on whether the conclusions rely upon the undertaken assumptions. Impact of changing g prior, as well as, model size prior is depicted in the Figure 1. No matter what prior model specification is chosen *DIST*, *RGDPprod*, *POPPprod* and *RGDPdist* are robust determinants of international trade. *LANG* depends on the chosen prior combination, which deem questioning robustness of this variable.

This point shows the superiority of BMA over the classical methods. Applying BMA allows one not only to use knowledge coming from many models but also to check the robustness of the results over the changes in prior specification: both in terms of g prior and model size prior. The classical approach based on statistical significance relies upon the knowledge coming from just one model. Model averaging procedures used in classical econometrics rely on a given specific set of prior assumptions, yet one more time making entire analysis more limited and vulnerable to criticism.
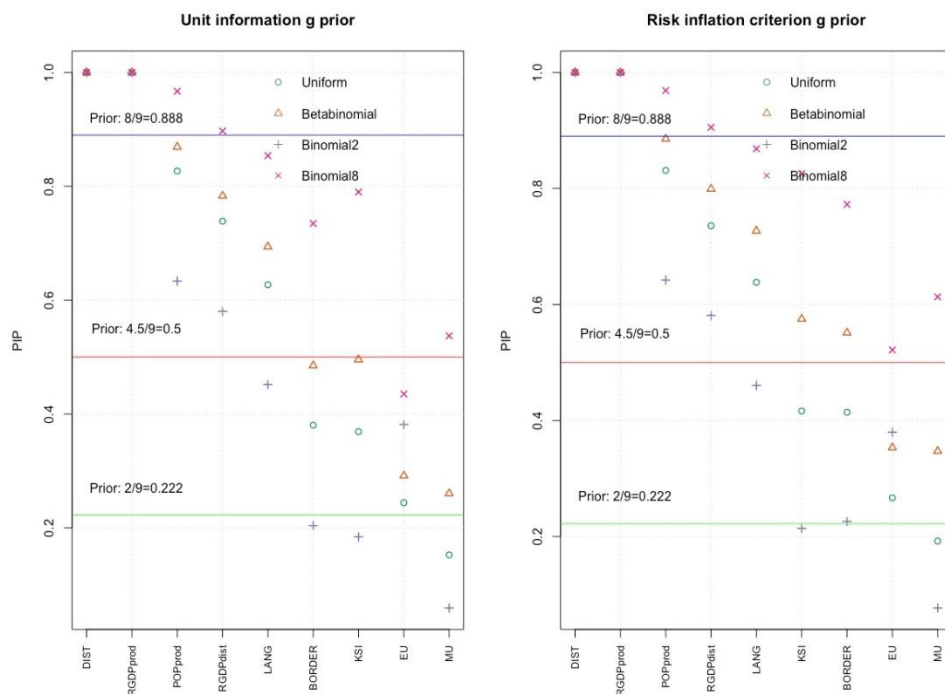
* Uniform, Betabinomial, Binomial2, Binomial8 – denotes uniform, binomial-beta with $Em = 4.5$, binomial with $Em = 2$ and binomial with $Em = 8$ model prior respectively.

Figure 1. Posteriori inclusion probabilities in different specifications of g prior and model size prior

4.4. Jointness measures

To uncover the character of the correspondence between regressors, jointness measures were employed. They were calculated for BMA with unit information prior and uniform model size prior. Results for both measures are shown in Table 4. The values of Doppelhofer and Weeks measures (J_{DW}) are located above the primary diagonal and for Ley and Steel's measure (J_{LS}) above.

Table 4. Jointness measures: J_{DW} (below primary diagonal) and J_{LS} (above primary diagonal)

x	MU	EU	RGDPdist	RGDPprod	POPprod	BORDER	DIST	LANG	KSI
MU	x	-2.48	-1.76	-1.73	-1.72	-2.30	-1.73	-1.97	-1.88
EU	-0.38	x	-1.96	-1.13	-2.48	-1.99	-1.13	-1.09	-1.84
RGDPdist	0.11	-1.85	x	1.03	1.14	-0.72	1.03	-0.02	-1.31
RGDPprod	nan	nan	nan	x	1.56	-0.48	0.00	0.53	-0.53
POPprod	0.24	-5.68	2.03	nan	x	-0.41	1.56	0.07	-0.49
BORDER	-0.45	-0.58	-0.13	nan	0.88	x	-0.48	-1.49	-0.98
DIST	nan	nan	nan	nan	nan	nan	x	0.53	-0.53
LANG	-0.28	0.50	-0.22	nan	-0.74	-1.57	nan	x	-0.86
KSI	0.14	-0.38	-1.72	nan	0.73	0.37	nan	-0.09	x

In Table 4, strong substitutes are highlighted in dark grey, whereas light grey indicates relevant substitutes. Employing the measure J_{DW} allowed the establishing of four pairs of substitutes, one pair of strong substitutes and one pair of complements. *EU* is a strong substitute of *POPprod* and a significant one of *RGDPdist* Border and language dummies are also substitutes, which might be reasonably explained in the following way: countries that are located closer to each other tend to share the same language more often. *KSI* exhibits substitutional relationship with *RGDPpc*. This result might be explained by U-shaped relationship between GDP *per capita* and degree of specialization described by Imbs and Wacziarg (2003): differences in GDP *per capita* are determining specialization patterns, and those in turn determine the patterns of trade. Moreover, using J_{DW} allowed for the identification of one pair of complements marked with the grey font: *POPprod* and *RGDPdist*.

Results in Table 3 reveal a few weaknesses related to the application of J_{DW}, which were mentioned in section 3. First, the measure did not identify many relationships between the variables. Second, an abbreviation "nan" (not a number), which denotes an undefined numeric value, is given in the table. In this case it is the result of the operations in the form of x/0. For that reason, it is worth

employing Ley and Steel's measure (J_{LS}), for which such problems are not present. The values of J_{LS} are located above the primary diagonal in Table 4.

The values of measure J_{LS} better justify the results obtained in section 4.3. The measure identifies 3 pairs of strong substitutes, 14 pairs of significant substitutes and 5 of significant complements. The J_{LS} measure indicates that the participation in the European Union and the Eurozone are either strong or significant substitutes for all the remaining variables. It explains why those variables themselves, despite their strong position in the literature and empirical analyses in the past, turned out to be fragile in the analysis described in section 4.3. Similarly to the J_{DW} measure, J_{LS} classified border and language dummy, as well as real GDP *per capita* and similarity of production structures as significant substitutes. Geographical distance was labelled complement of *POPprod* and *RGDPdist*.

Finally, J_{LS} captured the complementary relationship between *RGDPprod*, *POPprod* and *RGDPpc*. This might help provide two explanations for the negative coefficient on *POPprod*. Firstly, the higher the real GDP product, the bigger the economies and the greater their capacity to trade. At the same time, the higher the population product, the lower GDP *per capita*, and capacity for purchasing of individuals, which could explain negative coefficient on *POPprod*. This effect is present only if *RGDPprod* and *POPprod* are both present in the model. In this instance, *RGDPdist* allows one to control for structural similarity (in terms of both production and consumption) and participation in the EU or the Eurozone.

The second explanation relies upon economies of scale: the bigger the countries, the higher their capacities to explore economies of scale internally and lower the need to trade with outside world. In that instance, *RGDPprod* captures countries capacity to trade and *POPprod* captures their capacity to explore economies of scale internally. In this case, *RGDPdist* additionally allows for controlling differences in welfare between nations.

Therefore, the application of the measure allows one to explain all the results that defy the predictions made according to the theory. It also confirms the criticism levelled against Dopplehofer and Weeks' measure by Ley and Steel. J_{LS} is not only free form computational difficulties of J_{DW}, but also provides better explanations to the obtained results.

5. Conclusions

The following study presents the idea of Bayesian approach to statistics and econometrics, as well as the benefits coming from combining knowledge obtained on the basis of analysis of different models. In the first part, the BMA structure was described together with its most important statistics and g prior, as well as prior model proposals. The second part outlined jointness measures that were put forward by Ley and Steel, as well as Dopplehofer and Weeks.

The empirical part presents the results obtained from the analysis of the determinants of bilateral international trade. The application of Bayesian Model Averaging enabled the identification of four robust determinants: geographical distance, real GDP product, population product and real GDP *per capita* distance. Those four variables are robust to changes in both g prior and model size prior. Language and border dummy, similarity of production structures and participation in the EU were classified as robust for some prior specifications of BMA.

The applied procedure also showed that the model that is the closest to the true one is the model containing the following five independent variables: geographical distance, real GDP and population product, real GDP *per capita* distance and the language dummy. All variables, except for population product, have coefficient signs predicted by the theory. Owing to the application of Ley and Steel's jointness measure, it was possible to explain why some variables firmly rooted in theory were classified as fragile. Participation in the EU and the Eurozone are characterized by substitutional relationship with all other variables. Fragile border dummy and similarity of production structures are substitutes with language dummy and real GDP *per capita* distance respectively, *ergo* contained the same information as the variables classified as robust.

Finally, the complementary relationship between real GDP product and population product enabled two possible explanations of the negative sign of the population product coefficient to be proposed. The first uses the welfare effect reflected in real GDP *per capita*, and the second points to the exploitation of internal economies of scale. It is worth mentioning that the performed exercise demonstrated the superiority of Ley and Steel's jointness measure over the one introduced by Dopplehofer and Weeks.

REFERENCES

ANDERSON, J., (1979). A Theoretical Foundation for the Gravity Model, The American Economic Review. 69 (1), pp. 106–116.

ANDERSON, J., (2011). The Gravity Model, Annual Review of Economics, 3, pp. 133–160.

ANDERSON, J., VAN WINCOOP, E., (2003). Gravity with Gravitas: A Solution to the Border Puzzle, The American Economic Review, 93 (1), pp. 170–192.

BŁAŻEJOWSKI, M., KWIATKOWSKI, J., (2015). Bayesian Model Averaging and Jointness Measures for gretl, Journal of Statistical Software, 68 (5), pp. 1–24.

BŁAŻEJOWSKI, M., KWIATKOWSKI, J., (2016). Bayesian Model Averaging in the Studies on Economic Growth in the EU Regions – Application of the gretl BMA package, Economics and Sociology, 9(4), pp. 168–175.

BROCK, W. A., DURLAUF, S. N., (2001). Growth Empirics and Reality, World Bank Economic Review, 15 (2), pp. 229–272.

DOPPELHOFER, G., WEEKS, M., (2009). Jointness of Growth Determinants, Journal of Applied Econometrics, 24 (2), pp. 209–244.

DIXIT, A., STIGLITZ, J., (1997). Monopolistic Competition and Optimum Product Diversity, The American Economic Review, 67 (3), pp. 297–308.

EGGER, P., (2002). An Econometric View on the Estimation of Gravity Models and the Calculation of Trade Potentials, The World Economy, 25 (2), pp. 297–312.

EICHER, T., PAPAGEORGIOU, C., RAFTERY, A. E., (2011). Determining Growth Determinants: Default Priors and Predictive Performance in Bayesian Model Averaging, Journal of Applied Econometrics, 26 (1), pp. 30–55.

FERNÁNDEZ, C., LEY, E., STEEL, M., (2001). Benchmark priors for Bayesian model averaging, Journal of Econometrics, 100 (2), pp. 381–427.

FOSTER, D., GEORGE, E., (1994). The Risk Inflation Criterion for Multiple Regression, The Annals of Statistics, 22 (4), pp. 1947–1975.

FELDKIRCHER, M., ZEUGNER, S., (2009). Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging, IMF Working Paper, 202, pp. 1–39.

GRAY, H., (1980). The Theory of International Trade among Industrial Countries, Weltwirtschaftliches Archiv, 16 (3), pp. 447–470.

GRUBEL, H., LLOYD, P., (1971). The Empirical Measurement of Intra-Industry Trade, The Economic Record, 47 (120), pp. 494–517.

GRUBEL, H., LLOYD, P., (1975). Intra-industry trade: the theory and measurement of international trade in differentiated products, Wiley, New York.

HELPMAN, E., (1981). International trade in the presence of product differentiation. Economies of scale and monopolistic competition: Chamberlin-Heckscher-Ohlin approach, Journal of International Economics, 11, pp. 305–340.

HESTON. A., SUMMERS, R., ATEN, B., (2012). Penn World Table Version 7.1, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

IMBS, J., WACZIARG, R., (2003). Stages of Diversification, The American Economic Review, 93 (1), pp. 63–86.

KASS, R., WASSERMAN, L., (1995). A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion, Journal of the American Statistical Association, 90 (431), pp. 928–934.

KRUGMAN, P., (1979). Increasing Returns, Monopolistic competition, and International Trade, Journal of International Economics, 9, pp. 469–479.

KRUGMAN, P., (1980). Scale Economies, Product Differentiation and the Pattern of Trade, The American Economic Review, 70 (5), pp. 950–959.

KRUGMAN, P., (1991). Geography and Trade, The MIT Press, Cambridge, MA.

LANCASTER, K., (1980). Intra-Industry Trade under Perfect Monopolistic Competition, Journal of International Economics, 10 (2), pp. 151–175.

LEY, E., STEEL, M., (2007). Jointness in Bayesian variable selection with applications to growth regression, Journal of Macroeconomics, 29 (3), pp. 476–493.

LEY, E., STEEL, M., (2009). On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regressions, Journal of Applied Econometrics, 24 (4), pp. 651–674.

LEY, E., STEEL, M., (2012). Mixtures of g-priors for Bayesian model averaging with economic applications, Journal of Econometrics, 171 (2), pp. 251–266.

LINDER, S., (1961). An Essay on Trade Transformation, Wiley, New York.

MIN, C., ZELLNER, A., (1993). Bayesian and non-Bayesian methods for combining models and forecasts with application to forecasting international growth rates, Journal of Econometrics, 56 (1-2), pp. 89–118.

PRÓCHNIAK, M., WITKOWSKI, B., (2012). Konwergencja gospodarcza typu β w świetle bayesowskiego uśredniania oszacowań, Bank i Kredyt, 43 (2), pp. 25–58.

PRÓCHNIAK, M., WITKOWSKI, B., (2014). The application of Bayesian model averaging in assessing the impact of the regulatory framework on economic growth, Baltic Journal of Economics, 14 (1-2), pp. 159–180.

SALA-I-MARTIN, X., DOPPELHOFER, G., MILLER, R., (2004). Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach, The American Economic Review, 94, (4), pp. 813–835.

WHITTAKER, J., (2009). Graphical Models in Applied Multivariate Statistics, Wiley, Chichester.

ZELLNER, A., (1986). On Assessing Prior Distributions and Bayesian Regression Analysis with g Prior Distributions. In: Goel PK, Zellner A (eds.). Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti. Studies in Bayesian Econometrics 6. Elsevier, New York, pp. 233–243.

ZEUGNER, S., FELDKIRCHER, M., (2015). Bayesian Model Averaging Employing Fixed and Flexible Priors: The BMS Package for R, Journal of Statistical Software, 68 (4), pp. 1–37.