Mathias Moser

Macroeconomic Applications of Bayesian Model Averaging

Thesis

WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

DOCTORATE OF SOCIAL AND ECONOMIC SCIENCES

1st Supervisor: **Univ.Prof. Dr. Jesús CRESPO CUARESMA**

2nd Supervisor: **Prof. Gernot DOPPELHOFER, PhD**

Submitted on:

DISSERTATION

Macroeconomic Applications of Bayesian Model Averaging

Mag. Mathias MOSER

matmoser@wu.ac.at

Submitted in Partial Fulfillment
of the Requirements for the Degree of
DOCTOR OF SOCIAL AND ECONOMIC SCIENCES
in the Subject of
ECONOMIC POLICY

Supervisors:

Univ.Prof. Dr. Jesús CRESPO CUARESMA
Institute for Macroeconomics
Vienna University of Economics and Business (WU)

Prof. Gernot DOPPELHOFER, PhD
Department of Economics
Norwegian School of Economics (NHH)

Vienna, February, 2015

I hereby declare that,

- I have written this dissertation independently and without the aid of unfair or unauthorized resources. Whenever content was taken directly or indirectly from other sources, this has been indicated and the source referenced.
- this dissertation has neither previously been presented for assessment, nor has it been published.
- this dissertation is identical with the assessed thesis and the thesis which has been submitted in electronic form.

Vienna, 12th February, 2015

Für Jasmin

Acknowledgments

I am highly indebted to a myriad of persons, all of whom have made this thesis possible. First and foremost my supervisor Jesús Crespo Cuaresma, who is and has been a constant inspiration to me and my work and has motivated me to finish this project more times than I can recall. My other supervisors, Gernot Doppelhofer, Bettina Grün and Silvia Frühwirth-Schnatter, who have contributed to this thesis through an uncountable number of discussions and comments.

Furthermore all persons who enabled me to work in this environment, especially Wilfried Altzinger, Manuela Lenk, Markus Marterbauer and Bernhard Salzger.

I am most thankful for the opportunity to work together with my coauthors Paul Hofmarcher, Stefan Humer and Matthias Schnetzer, who do not only share my passion for learning and knowledge but also for after-work beers.

My family has contributed to this thesis through social, financial and emotional support. I could not have finished this without their investment of time and commitment to my career choice, for which I am grateful.

Finally my partner Jasmin for her endless patience, support and faith in me.
She is the sole reason that this thesis does not end right here.

Kurzfassung

Bayesian Model Averaging (BMA) ist eine ökonometrische Methode um für Modellunsicherheit zu kontrollieren. Sie ermöglicht robuste Inferenz für Parameter durch das Schätzen einer großen Anzahl von möglichen Modellen, über die ein Durchschnitt gebildet werden kann. Dies ist speziell in Anwendungsgebieten relevant, in denen keine starken theoretischen Vorgaben bezüglich der Spezifizierung von Modellen vorliegen. Ein Beispiel hierfür ist die Analyse von Wirtschaftswachstum, auf die sich diese Arbeit konzentriert. Diese kumulative Dissertation widmet sich in drei Teilen den Abhängigkeitsstrukturen zwischen Variablen, die in solch großen Modellräumen auftreten können. In einer ersten Arbeit wird untersucht, wie sich a-priori Wahrscheinlichkeiten (*priors*) auswirken, die eine bestimmte Modellklasse bevorzugen. Dies ist speziell für Interaktionen zwischen Variablen und Daten mit hoher Multikollinearität relevant. Die Arbeit baut auf einer Diskussion im *Journal of Applied Econometrics* auf, in welcher der Frage nach einer unterschiedlichen Modellstruktur für Subsahara-Afrika nachgegangen wird. Der zweite Aufsatz widmet sich der Suche nach Modellklassen im Modellraum von BMA. Dabei wird das Werkzeug der Latenten Klassenanalyse verwendet, um unterschiedliche Gruppen von Modellen in zwei prominenten Wachstumsdatensätzen zu finden. Das letzte Papier beschäftigt sich mit der Analyse von *Jointness* (Gemeinsamkeiten) von Variablenpaaren in BMA. Dabei wird versucht die bisherige ökonometrische Literatur mit dem Feld des *Machine Learnings* zu verbinden, um *Substitute* und *Komplemente* zwischen Wachstumsfaktoren erkennen zu können.

Abstract

Bayesian Model Averaging (BMA) is a common econometric tool to assess the uncertainty regarding model specification and parameter inference and is widely applied in fields where no strong theoretical guidelines are present. Its major advantage over single-equation models is the combination of evidence from a large number of specifications. The three papers included in this thesis all investigate model structures in the BMA model space. The first contribution evaluates how priors can be chosen to enforce model structures in the presence of interactions terms and multicollinearity. This is linked to a discussion in the *Journal of Applied Econometrics* regarding the question whether being a Sub-Saharan African country makes a difference for growth modelling. The second essay is concerned with clusters of different models in the model space. We apply Latent Class Analysis to the set of sampled models from BMA and identify different subsets (kinds of) models for two well-known growth data sets. The last paper focuses on the application of “jointness”, which tries to find bivariate relationships between regressors in BMA. Accordingly this approach attempts to identify *substitutes* and *complements* by linking the econometric discussion on this subject to the field of Machine Learning.

Contents

Introduction

This thesis is concerned with the application of Bayesian Model Averaging (BMA) techniques to estimate determinants of economic growth. The papers which are presented in the following, all share this tool as a mutual methodological basis but address different practical issues. This first chapter introduces the shared topic of economic growth, gives a short overview of the method used, and motivates the link between the three essays, which form this cumulative thesis.

Economic growth is arguably one of the most pivotal topics in the field of economics. Substantial differences in growth performance, as measured by gross domestic products, have contributed to both convergence and global imbalances of economic and socioeconomic factors, such as trade, production structures, education and health outcomes. Understanding the various determinants of growth is a key precondition to predict growth rates, plan policy interventions and to support development in low-income countries.

Followingly, a major part of economic theory is devoted to explaining such differing outcomes. Classical (Malthusian) growth models have focused on the role of the primary production factors, labor and capital. Extensions to this approach have highlighted technological change (Neoclassical models), innovation (Schumpeterian growth), human capital and endogeneity (Lucas, 1988; Barro and Lee, 1993; Romer, 1994), and institutions (Acemoglu, Johnson, and Robinson, 2005) as important factors that can serve as (partial) explanations and predictors.

Based on all of these different theories it becomes clear that economics does not provide one *recipe* for explaining growth, but suggests a large set of possible sources of contribution. This issue has been termed *open-endedness* of growth theories (Brock and Durlauf, 2001) and poses challenges to empirical modeling of growth and the testing of theoretical models.

Especially standard frequentist econometric approaches are concerned with a single model, which has to be specified beforehand by the researcher. The choice of covariates to be included in such a model is mostly inferred from the varying theories or based on suggestions from related empirical literature. As a result, the researcher needs to focus on one specific model (or a small set), which is consistent with a specific theory and/or provides *reasonable* results. However alternative specifications, based on different theories or measurements thereof, might exist, which also explain the data well but deviate from the original model with regard to magnitude or signs of regressors. In such a case estimates are sensitive to the model specification and therefore to the subjective choice made by the researcher (Hoeting, Madigan, Raftery, and Volinsky, 1999).

To robustify such an analysis a researcher would in fact need to consult a number of different models instead of a single equation, to take the inherent model uncertainty into account. Based on economic theory these models can involve large sets of possible predictors (K), e.g. related to labour-force, education, health, institutions, trade, population and production structure for the case of economic growth. The evaluation of every possible model given this set of candidate covariates is however often not feasible, since the model space increases with 2^K.

A common approach to address this issue is Bayesian Model Averaging, which applies the idea of averaging over a large number of models by combining posterior distributions of all evaluated models for inference instead of just selecting one *true* model. Model averaging is a natural extension in a Bayesian framework and integrates the weighting scheme through Bayes' rule. Still, BMA is a comparably young field for econometrics and has only gained widespread interest due to the rise in computational power. Especially for large model spaces, which can not be evaluated as a whole, the availability of e.g. Markov Chain Monte Carlo Model Composition (MC^3) algorithms have made BMA a widely used tool to address model uncertainty (see e.g. Raftery, 1995; Fernández, Ley, and Steel, 2001b; Masanjala and Papageorgiou, 2008; Eicher, Papageorgiou, and Raftery, 2011; Hofmarcher, Crespo Cuaresma, Grün, and Hornik, 2014).

More technically, in BMA a model M_j describing a variable y with a set of regressors $Z_j \in Z$ may be written as

$$y = \alpha\iota_n + Z_j\beta_j + \epsilon\,, \qquad \epsilon \sim N(0, I\sigma^2)\,, \tag{1}$$

where α is the intercept, ι_n is an n-dimensional vector of ones and ϵ a disturbance term. Bayesian inference can be carried out by multiplying the likelihood with suitable priors on the parameters (α, β, σ) to derive the posterior distribution for e.g. the parameters as

$$\Pr(\beta|y, M_j) \propto \Pr(y|\beta, M_j)\Pr(\beta|M_j)\,. \tag{2}$$

Priors are often chosen in such a way, that the posterior density can be solved analytically, via so-called "natural conjugate priors". Furthermore, many empirical studies (see e.g. Ley and Steel, 2009) use non-informative priors for α and σ such as

$$\begin{aligned} \Pr(\alpha) &\propto 1 \\ \Pr(\sigma) &\propto \frac{1}{\sigma}\,. \end{aligned}$$

A default choice for the prior on β is the g-prior (Zellner, 1986)

$$\Pr(\beta_j|\alpha, \sigma, M_j) \propto f^{k_j}(\beta_j|0, \sigma^2 g(Z_j'Z_j)^{-1})\,, \tag{3}$$

where the value of g shrinks the prior variance based on the empirical variance-covariance matrix. For the choice of g a number of suggestions can be found, such as the Unit Information-, Risk Criterion- or the Benchmark Prior (the so-called BRIC Prior), where g is $1/max(n,K^2)$ (Fernández, Ley, and Steel, 2001a). An alternative to specifying one g-value is the elicitation of a hyperparameter, as suggested by Liang et al. (2008). This approach can be used to set the prior expected shrinkage by imposing a Beta-distribution on the parameter g.

To address model uncertainty we can take the average over all (relevant) models, that is the distribution of the parameters times their posterior model probability. Applying Bayes rule, we can write for β

$$\Pr(\beta|y) \propto \sum_{j=1}^{2^K} \Pr(\beta|y, M_j)\Pr(M_j|y)\,. \tag{4}$$

The latter term is given by the marginal likelihood multiplied with the prior model probability

$$\Pr(M_j|y) \propto \Pr(y|M_j)\Pr(M_j)\,. \tag{5}$$

The prior model probability $\Pr(M_j)$ can be chosen in various ways. A popular choice is the uniform prior which assigns the same *ex ante* probability to each model. This however favours mid-sized models, due to their high relative frequency. Another solution by Ley and Steel (2009) introduces a hierarchical Binomial-Beta prior, which imposes an equal prior model probability over models of different size. More sophisticated prior choices include for example the class of Heredity and Dilution Priors, which are addressed in chapter I (see e.g. George, 1999; Chipman, 1996).

Inference for BMA models can not only be based on the posterior distribution parameters, but also on posterior inclusion probabilities (PIP). PIPs are a rank measure for the importance of a variable and are calculated as the share of posterior model mass in which a certain covariate is included. While the posterior density of a parameter gives insights in the magnitude of an effect, PIPs can be used by the practitioner to measure its relevance.

However, the reliability of this tool may be limited in certain cases where BMA averages over a number of very different models. For example, it may be the case that growth can be explained equally well by two disjoint sets of covariates. In such a case BMA could report an averaged value of 0.5 for the PIP of a variable, which is actually highly important for one type of model but unimportant for the other. Furthermore it may be the case, that a certain variable exhibits a rather high importance, but only in conjunction with another explanatory. In both cases the overall results from BMA have to be interpreted with care.

The three essays of this dissertation all address this problem of specific model structures in BMA from various perspectives. The first paper deals with specification issues in the presence of interaction terms. The second article takes a different approach and tries to find classes of models in the huge BMA model space, which vary in their properties and structure (i.e. parameter size and PIP). The last contribution focuses on "jointness" of predictors in the model space. This term describes, whether two variables do frequently occur together or independently. Such an analysis of jointness allows the researcher to highlight substitutes and complements among regressors. The following gives a brief overview for each paper.

Dilution Priors in BMA Growth Applications with Interactions This first paper is concerned with the implementation of suitable priors to deal with interaction terms in BMA for the case of Sub-Saharan African countries. In a study by Masanjala and Papageorgiou (2008), the authors focus on the issue, whether growth in Sub-Saharan African countries is related to other explanatories than in the rest of the world. To assess this effect, they include interaction terms for a Sub-Saharan country dummy in their analysis.

However, Crespo Cuaresma (2011) points out that in BMA this can lead to the inclusion of models, which contain an interaction term but not its main effects. Such models are considered not *well-specified* and are often avoided due to the lack of interpretability or the possible omitted-variable bias (Chipman, 1996).

In BMA this issue can be addressed by choosing appropriate priors which enforce a *correct* model structure. Such priors have been termed *Heredity Priors* and adapt prior model probabilities depending on the number of missing parent effects in a certain specification. Therefore, a model with an interaction but no parent effects will be a-priori less likely than a model which includes these effects (Chipman, 1996). Alternatively, it can be argued that such a misspecification is only relevant, depending on the multicollinearity between parents and interactions. The tessellation prior of George (2010) addresses this issue and penalizes high correlations between variables.

The contribution included in this thesis implements all of these priors in the R BMS package (Feldkircher and Zeugner, 2009) and provides a simulation study for the effects of these different prior choices. The results indicate that while such stricter priors influence the choice of models and PIPs in BMA, they do not affect predictive performance substantially.

Unveiling Covariate Inclusion Structures The second paper deals with *typical* classes of models in BMA model spaces. Such model spaces consist of a potentially large number of models, which are averaged to make inference on their joint posterior distribution. Combining all these different models may involve very different types of specifications. While many of these specifications can have high posterior model probabilities, aggregated inference can be misleading, depending on the covariate structure of models.

To mitigate these effects we try to cluster the matrix of sampled models and base inference on groups of specifications where given covariates are independent conditional on class membership, instead of the heterogeneous model space. This allows a practitioner to disentangle effects for specific model classes with similar posterior model probability. The paper uses Dirichlet Process Clustering to categorize the binary matrix of variable inclusion profiles for the top models chosen by the BMA MC^3 sampler. This method has the merit that it does not depend on a pre-defined number of clusters, and therefore adds flexibility given the prior setting for the LCA algorithm.

By applying this method to the two growth data sets of Fernández, Ley, and Steel (2001b) and Sala-i-Martin, Doppelhofer, and Miller (2004) we find that the overall BMA results average-out parts of the underlying model structures. Our approach results in a number of clusters, with differing variable PIP. This result indicates competing model structures, which can not be identified from the overall BMA results.

A Comprehensive Approach to Posterior Jointness Analysis in Bayesian Model Averaging Applications The third paper investigates so-called "jointness" among growth determinants. This term was coined by Doppelhofer and Weeks (2005) and refers to the extent, to which two variables appear (dis-)jointly in model specifications. In BMA this analysis is based on the set of sampled models from an MC^3 exercise. Analyzing a relevant set of visited models, a researcher might be interested in the question, whether variables appear together frequently or seem to be sampled mutually exclusively. Doppelhofer and Weeks (2009) refer to variables in these two scenarios as "complements" or "substitutes" respectively. This classification can be especially important for practitioners, who need to take such dependencies between variables into account for appropriate policy advice.

To analyze jointness, a number of different measures have been proposed in the context of BMA (Ley and Steel, 2007; Doppelhofer and Weeks, 2009; Strachan, 2009). However, this discussion disregarded the wide field of machine learning, and especially that of association rules analysis. This field of research is concerned with similar issues, for example market shopping basket analysis. The focus of such an application is to find structures in the types of products that customers buy in combination. Similar to the BMA discussion, a large number of different measures have been proposed in this context, which can be used to analyze such joint occurrences.

The contribution of this third paper lies in the integration of insights from the association rules analysis literature into the BMA context. We review different measures and their characteristics and propose a set of properties that jointness indicators should fulfill for the application to BMA model spaces. We find that especially the null-invariance property — which was also heavily discussed in the BMA literature — plays an important role in selecting suitable measures. Additionally to simulation results for different measures, this paper also provides an empirical exercise for the FLS data set.

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2005): "Institutions as a fundamental cause of long-run growth". In: *Handbook of Economic Growth* 1, pp. 385–472.

Barro, R. J. and J.-W. Lee (1993): "International comparisons of educational attainment". In: *Journal of Monetary Economics* 32 (3), pp. 363–394.

Brock, W. A. and S. N. Durlauf (2001): "What have we learned from a decade of empirical research on growth? Growth empirics and reality". In: *World Bank Economic Review* 15 (2), pp. 229–272.

Chipman, H. (1996): "Bayesian variable selection with related predictors". In: *Canadian Journal of Statistics* 24, pp. 17–36.

Crespo Cuaresma, J. (2011): "How different is Africa? A comment on Masanjala and Papageorgiou". In: *Journal of Applied Econometrics* 26 (6), pp. 1041–1047.

Doppelhofer, G. and M. Weeks (2005): *Jointness of growth determinants*. Cambridge Working Papers in Economics.

Doppelhofer, G. and M. Weeks (2009): "Jointness of growth determinants". In: *Journal of Applied Econometrics* 24 (2), pp. 209–244.

Eicher, T., C. Papageorgiou, and A. Raftery (2011): "Default priors and predictive performance in Bayesian model averaging, with application to growth determinants". In: *Journal of Applied Econometrics* 26, pp. 30–55.

Feldkircher, M. and S. Zeugner (2009): "Benchmark priors revisited: On adaptive shrinkage and the supermodel effect in Bayesian model averaging". In: *IMF Working Papers*.

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

— (2001b): "Model uncertainty in cross–country growth regressions". In: *Journal of Applied Econometrics* 16 (5), pp. 563–576.

George, E. (1999): "Discussion of Bayesian model averaging and model search strategies by M.A. Clyde". In: *Bayesian Statistics 6*. Ed. by J. Bernardo, J. Berger, A. Dawid, and A. Smith, pp. 157–177.

— (2010): "Dilution priors: Compensating for model space redundancy". In: *Borrowing Strength: Theory Powering Applications - A Festschrift for Lawrence D. Brown*. Ed. by J. Berger, T. Cai, and I. Johnstone. Institute of Mathematical Statistics, pp. 185–165.

Hoeting, J., D. Madigan, A. Raftery, and C. Volinsky (1999): "Bayesian model averaging: A tutorial". In: *Statistical Science* 14 (4), pp. 382–417.

Hofmarcher, P., J. Crespo Cuaresma, B. Grün, and K. Hornik (2014): "Last night a shrinkage saved my life: Economic growth, model uncertainty and correlated regressors". In: *Journal of Forecasting*, Forthcoming.

Ley, E. and M. Steel (2009): "On the effect of prior assumptions in Bayesian model averaging with applications to growth regressions". In: *Journal of Applied Econometrics* 24, pp. 651–674.

Ley, E. and M. F. J. Steel (2007): "Jointness in Bayesian variable selection with applications to growth regression". In: *Journal of Macroeconomics* 29 (3), pp. 476–493.

Liang, F. et al. (2008): "Mixtures of g priors for Bayesian variable selection". In: *Journal of the American Statistical Association* 103, pp. 410–423.

Lucas, R. J. (1988): "On the mechanics of economic development". In: *Journal of Monetary Economics* 22 (1), pp. 3–42.

Masanjala, W. and C. Papageorgiou (2008): "Rough and lonely road to prosperity: A reexamination of the sources of growth in Africa using Bayesian model averaging". In: *Journal of Applied Econometrics* 23, pp. 671–682.

Raftery, A. E. (1995): "Bayesian model selection in social research". In: *Sociological Methodology* 25, pp. 111–164.

Romer, P. M. (1994): "The origins of endogenous growth". In: *Journal of Economic Perspectives*, pp. 3–22.

Sala-i-Martin, X., G. Doppelhofer, and R. I. Miller (2004): "Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach". In: *American Economic Review* 94 (4), pp. 813–835.

Strachan, R. W. (2009): "Comment on 'Jointness of growth determinants' by Gernot Doppelhofer and Melvyn Weeks". In: *Journal of Applied Econometrics* 24 (2), pp. 245–247.

Zellner, A. (1986): "On assessing prior distributions and Bayesian regression analysis with g–prior distributions". In: *Bayesian Inference and Decision Techniques Essays in Honor of Bruno de Finetti* 6. Ed. by A. Zellner and P. Goel, pp. 233–243.

I Dilution Priors in BMA Growth Applications with Interactions

Later version published as: M. Moser and P. Hofmarcher (2014): "Model Priors Revisited: Interaction Terms In BMA Growth Applications". In: Journal of Applied Econometrics *29 (2), pp. 344–347*

Dilution Priors in BMA Growth Applications with Interactions*

Mathias Moser[a] and Paul Hofmarcher[b]

[a]Department of Economics, WU Vienna, Austria
[b]Department of Applied Statistics, Johannes Kepler University Linz

*Correspondence to: Mathias Moser, Department of Economics, WU Vienna, Augasse 2–6, 1090 Vienna, Austria. Phone: +43/1/31336-5481. E-mail: matmoser@wu.ac.at

Abstract

This paper provides a sensitivity analysis on the prior choice for interaction terms for the results of Masanjala and Papageorgiou (Rough and lonely road to prosperity: a reexamination of the sources of growth in Africa using Bayesian Model Averaging (BMA). Journal of Applied Econometrics 2008, **23**: 671–682.) which has been criticized for its implementation of interaction terms by Jesus Crespo Cuaresma (How different is Africa? Journal of Applied Econometrics 2011, **26**: 1041–1047). We perform inference based on different prior suggestions for model spaces with interaction terms and provide posterior inclusion probabilities, log predictive scores and continuous ranked probability scores. Our results show that the alternative priors deliver a similar degree of predictive performance compared to the original prior.

Keywords: Bayesian Model Averaging, Interaction, Dummy variable

1 Introduction

A recent discussion in the *Journal of Applied Econometrics* raised the question how to correctly implement interaction terms in Bayesian Model Averaging (BMA) applications. BMA has become very popular in growth applications where especially regional disparities are present.

Masanjala and Papageorgiou (2008, henceforth 'MP') use BMA techniques to target the problem of parameter heterogeneity for countries of Sub–Saharan Africa by adding regional interaction terms to their analysis. These interaction terms are treated as normal covariates in MP's approach.

In his comment on MP's work, Crespo Cuaresma (2011) points out that in such a setting models which include interaction terms and their corresponding parent variables are treated by the modeller as being *a priori* as likely as models that feature interaction terms but not (all) their corresponding parent covariates. He argues in favor of an alternative prior specification as opposed to the uniform model prior. In a reply Papageorgiou (2011) calls for a sensitivity analysis to shed more light on this issue.

The issue at hand, the treatment of interaction terms in model search algorithms, has also been raised by Chipman (1996) for the case of Bayesian Variable Selection methods. He points out that models with interaction terms should include all corresponding parent variables by 'convention', since effects may be hard to interpret otherwise. Chipman proposes to include such beliefs through model priors and introduces *heredity priors* which put a predefined penalty on models with missing main effects. Crespo Cuaresma (2011) adopts this approach to evaluate MP's results and furthermore argues that the problem of interaction terms is related to the treatment of correlated variables within BMA. He refers to George (1999) who proposes the idea of dilution priors to compensate for model space redundancy. Additionally, George (2010) presents several approaches for *dilution priors*. Inter alia he introduces a *tessellation defined dilution prior* which we include in our analysis (henceforth tessellation prior).

All these model prior suggestions of Chipman (1996) and George (2010) are reasonable alternatives to the uniform prior when interaction terms, and multicollinearity, are present in a dataset, as in MP's work. We will use the original cross-country dataset of MP which includes 24 covariates, one Sub-Saharan Africa dummy as well as 24 interactions for the mentioned dummy. In a first step we will replicate the BMA exercise of MP using the BMS software package for R[1] and show that MP's findings are consistent given their prior assumptions. Second, we carry out a prior sensitivity analysis based on the review of Crespo Cuaresma (2011) and alternative prior suggestions by Chipman (1996) as well as George (2010). The performance of the different priors are assessed both by changes in posterior inclusion probabilities as well as predictive power. We find that the inclusion of interaction terms is sensitive to the prior choice. Predictive abilities however remain almost unchanged compared to the default prior of MP.

2 Alternative Prior Choices

Following the work of MP and Crespo Cuaresma we consider a regression model of the form

$$y = \alpha \iota_n + Z_{1,j}\beta_{1,j} + I \times Z_{2,j}\beta_{2,j} + \sigma\epsilon, \tag{1}$$

where $Z_{1,j}$ is a matrix of K covariates, including the Sub–Saharan African dummy variable I and similar $Z_{2,j} \subseteq Z_{1,j} \setminus I$. The model space $\mathscr{M}$ includes all feasible combinations of $Z_{1,j}, Z_{2,j}$. The design matrix of an arbitrary model $M_j \in \mathscr{M}$ will be denoted by Z_j (instead of $(Z_{1,j}, I \times Z_{2,j})$) to simplify notation. Average per capita GDP–growth is denoted by the n dimensional vector y, the intercept is $\alpha\iota_n$ and $\sigma\epsilon$ are the regression errors with $\epsilon \sim N(0,1)$. The unconditional

[1] An adapted version of the BMS package for R is available at http://www.wu.ac.at/vw1/m/moser/mp-priors.

posterior distribution of any parameter Δ of interest, is the weighted average across all possible models $M_j \in \mathcal{M}$ so that

$$p(\Delta|y) = \sum_{j=1}^{2^{K'}} p(\Delta|y, M_j)p(M_j|y) \tag{2}$$

with $K' = 2K - 1$. We follow MP and Crespo Cuaresma (2011) who impose unit information priors (UIP) on the parameters Δ in each model. The posterior model probability which controls for model uncertainty in equation 2 is given by

$$p(M_j|y) \propto p(y|M_j)p(M_j). \tag{3}$$

In their analysis MP assume a uniform prior over models and accordingly let $p(M_j) = 2^{-K'}$. While this seems to be a natural choice for an ignorance–based model selection, it prefers mid–sized models as has been argued by Ley and Steel (2009). Similar to Chipman (1996) and George (2010), they also stress that, in a setup with interaction terms the '[...] assumption of prior independent inclusion of regressors can be contentious in some contexts.' In MP's view a change in the slope of a variable for African countries doesn't imply a level shift for African countries, while Crespo Cuaresma (2011) and Chipman (1996) would argue in favor of such an effect (by including both the interaction an the main effect). We will address these issues through two alternative prior specifications.

2.1 Weak and Strong Heredity Priors

Chipman (1996) proposes to relate the inclusion probability of an interaction term to the inclusion of the according parent variables in the model. In such a setting, a model featuring two covariates A and B as well as their interaction $A\#B$ is *a priori* more likely than a similar model which contains $A\#B$ but misses A, B or both main effects.

If we interpret this difference in prior beliefs as a penalty on models with missing main effects, a heredity prior can be expressed as a decreasing function of the number of missing parent variables in the specification. Through this penalty it is possible to add a heredity property to a number of different priors, e.g. to the uniform model prior,

$$p(M_j) \propto \begin{cases} 2^{-K'} & \text{if } (\eta = 0) \\ 2^{-K'} \times \frac{s}{\eta+1} & \text{if } (\eta > 0) \end{cases} \tag{4}$$

where η is a count variable for the number of missing parent variables and $s \geq 0$ a scale parameter. In our calibration exercise we will assume a grid of different scale parameters s between zero and one. Setting $s = 0$ results in the strong heredity prior used by Crespo Cuaresma (2011). Please note that for a weak heredity prior, i.e. $s > 0$ models with $\eta > 0$ are penalized but still admissible, so that $p(M_j|\eta > 0) > 0$, while such models are completely excluded from the sampling procedure in Crespo Cuaresma's approach.

2.2 Tessellation Prior

The tessellation prior belongs to the group of dilution priors (see George, 2010). Its purpose is to avoid placing little probability on unique models while assigning an excess mass of probability to regions of similar models (i.e. with highly correlated variables).

Tessellation priors achieve this property by projecting the single models Z_j to the surface of a unit sphere. They assign each point on the surface the one model whose spanned subspace minimizes the (euclidean) distance to a considered point. This results in model regions on the unit sphere's surface which form a Voronoi tessellation and deliver the desired diluted prior model probabilities.

Using MC^3 methods, and following George (2010), a Local–Spinner Process can be used to achieve the tessellation property:

1. Generate $Y^* \sim N_n(0, I)$

2. For design matrices $Z_{j'}$ in a neighborhood[2] of the current used Z_j select the nearest $Z_{j'}$ to the hypothetical data Y^* by minimizing $Y^*[I - Z_{j'}(Z_{j'}'Z_{j'})^{-1}Z_{j'}']Y^*/g(q_{j'})$.

$Z_{j'}(Z_{j'}'Z_{j'})^{-1}Z_{j'}'Y^*$ is the orthogonal projection of Y^* on the subspace spanned by the columns of $Z_{j'}$. $g(q_{j'})$ is a decreasing function, penalizing the dimensionality $q_{j'}$ of the matrix $Z_{j'}$. Natural choices for $g()$ would be the degrees of freedom correction $g(q_j) = (n - q_j)$ or its squared value for a stronger penalty on larger models (see George, 2010). In the following exercise we will use the latter option setting $g(q_j) = (n - q_j)^2$.

3 Which Prior to Choose?

Using the MP dataset we evaluate the set of considered priors by comparing posterior inclusion probabilities (PIP), log predictive scores (LPS) and continuous ranked probability scores (CRPS) based on an out–of–sample prediction exercise. To evaluate the model space $\mathscr{M} = 2^{49}$ we will utilize Markov Chain Monte Carlo Model Composition methods (MC^3, as used in e.g. Ley and Steel (2009)[3]). We use 4,000,000 drawings from which 2,000,000 are disregarded (burn-ins) and impose a unit information prior with Zellner's g equal to the number of observations. For predictive inference we randomly drop $\{70, 50, 20, 10, 5, 1\}$ observations from the dataset to create a training set and a hold–out set (see Eicher, Papageorgiou, and Raftery (2011)) for each mentioned prior. The reported mean values for LPS and CRPS result from 1000 single replications. The LPS is also used in Crespo Cuaresma (2011) and has been justified by Fernández, Ley, and Steel (2001) but has also been criticized as being too sensitive to e.g. outliers (see Eicher, Papageorgiou, and Raftery, 2011). An alternative score, the CPRS, is defined directly in terms of the predictive cumulative distribution functions and measures the squared error between the predicted and observed

[2]Here the neighborhood around model M_j is defined as models which differ only by one covariate.

[3]Also Eicher, Papageorgiou, and Raftery (2011) argue that this sampling procedure yields similar results compared to the one used by MP.

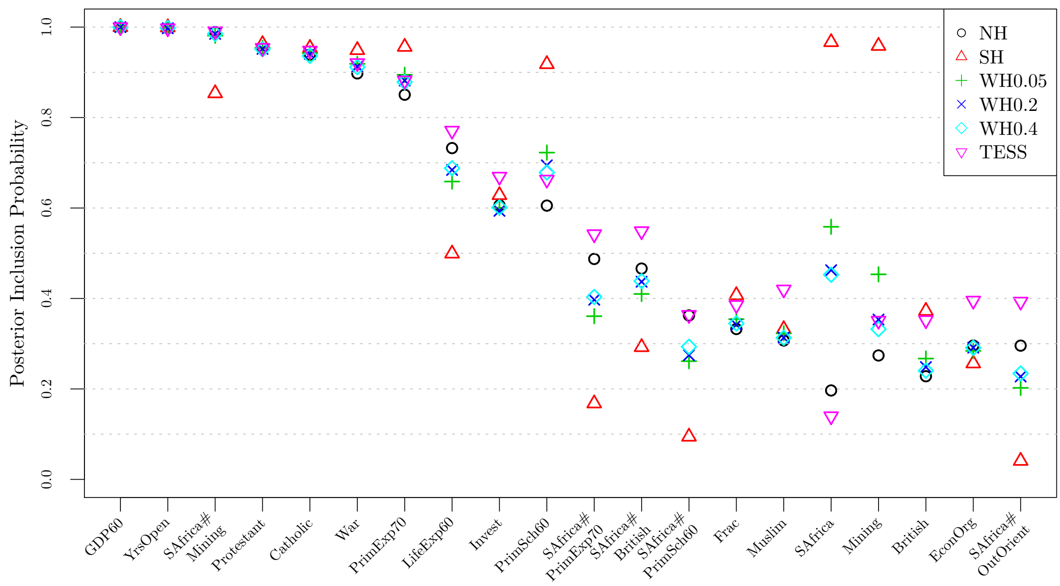


Figure I.1: Posterior Inclusion Probabilities for Top 15 Variables (all scenarios)

cumulative distributions. For a detailed discussion of CRPS see Gneiting and Raftery (2007).

Based on the original results of MP (*NH*, no heredity), we impose a strong heredity prior (*SH*), three weak heredity priors with different scale parameters *s* (*WH{0.01,0.1,0.5}*) and the tessellation prior (*TESS*) proposed by George (2010).

Figure I.1 reports PIP for the top 15 variables from the union of all models. We find that for important variables whose interaction with the Sub–Saharan Africa Dummy played a minor role in the original (*NH*) results, all priors produce PIP's within a narrow range compared to the original MP results (cf. GDP60, YRSOPEN, PROTESTANT or CATHOLIC). For interaction terms the heredity priors boost the importance of the parent effects, which is apparent for the SAFRICA#MINING term. Especially in the *SH* setup the inclusion probability of this interaction term decreases while the importance of the two parent variables SAFRICA and MINING rises close to one. The same effect applies to the weak heredity priors, however through the softer penalty results do not deviate as strongly from the *NH* scenario. We find similar effects for PRIMEXP70, PRIMSCH60 as well as the BRITISH colonialization dummy. Most of the interaction terms with low posterior inclusion probabilities behave in a similar fashion as can be seen in table I.1.

Table I.1: Posterior Inclusion Probabilities for different Model Priors

Regressors	NH	SH	WH0.05	WH0.2	WH0.4	TESS
GDP60	1.00	1.00	1.00	1.00	1.00	1.00
YrsOpen	1.00	1.00	1.00	1.00	1.00	1.00
SAfrica#Mining	0.99	0.85	0.98	0.99	0.99	0.99
Protestant	0.95	0.96	0.95	0.95	0.95	0.95
Catholic	0.94	0.95	0.94	0.94	0.94	0.95
War	0.90	0.95	0.92	0.91	0.91	0.92
PrimExp70	0.85	0.96	0.89	0.88	0.88	0.88
LifeExp60	0.73	**0.50**	0.66	0.68	0.69	0.77
Invest	0.61	0.63	0.60	0.59	0.60	0.67
PrimSch60	0.61	**0.92**	0.72	0.69	0.68	0.66
SAfrica#PrimExp70	0.49	**0.17**	**0.36**	0.40	0.40	0.54
SAfrica#British	0.47	**0.29**	0.41	0.44	0.44	0.55
SAfrica#PrimSch60	0.36	**0.09**	**0.26**	0.27	0.29	0.36
Frac	0.33	0.41	0.35	0.34	0.34	0.39
Muslim	0.31	0.33	0.32	0.31	0.31	**0.42**
EconOrg	0.30	0.26	0.28	0.29	0.29	**0.40**
SAfrica#OutOrient	0.30	**0.04**	**0.20**	0.23	0.23	**0.39**
Mining	0.27	**0.96**	**0.45**	**0.35**	0.33	**0.35**
SAfrica#Area	0.25	**0.04**	**0.18**	0.20	0.20	**0.34**
SAfrica#Frac	0.24	**0.04**	**0.16**	0.19	0.19	0.30
British	0.23	**0.37**	0.27	0.25	0.24	**0.35**
RERD	0.20	0.21	0.20	0.20	0.20	**0.31**
SAfrica	0.20	**0.97**	**0.56**	**0.46**	**0.45**	**0.14**
SAfrica#GDP60	0.19	**0.12**	0.17	0.17	0.17	0.17
SAfrica#AbslLat	0.17	**0.01**	**0.10**	**0.12**	**0.12**	**0.26**
SAfrica#LifeExp60	0.17	**0.05**	**0.12**	0.14	0.14	0.17
SAfrica#Invest	0.16	**0.06**	0.12	0.13	0.13	**0.20**

Table I.1: *(continued)*

Regressors	NH	SH	WH0.05	WH0.2	WH0.4	TESS
SAfrica#Rights	0.16	**0.02**	**0.11**	0.12	0.13	0.14
SAfrica#popGrowth	0.15	**0.02**	**0.10**	**0.11**	0.11	**0.23**
SAfrica#EconOrg	0.15	**0.02**	**0.10**	0.11	0.11	**0.20**
SAfrica#Muslim	0.15	**0.03**	**0.10**	**0.11**	0.11	**0.19**
SAfrica#CivilLib	0.14	**0.02**	**0.10**	0.11	0.11	0.14
Area	0.14	0.17	0.15	0.15	0.15	**0.25**
SAfrica#Catholic	0.14	0.16	0.15	0.14	0.14	**0.20**
SAfrica#French	0.14	**0.03**	**0.10**	0.11	0.11	0.17
SAfrica#Protestant	0.14	**0.09**	0.12	0.13	0.13	**0.22**
popGrowth	0.14	0.16	0.14	0.14	0.14	**0.22**
SAfrica#War	0.14	**0.10**	0.12	0.12	0.12	**0.20**
Other	0.13	0.13	0.13	0.13	0.13	**0.19**
SAfrica#RERD	0.13	**0.04**	**0.09**	0.10	0.11	**0.18**
SAfrica#YrsOpen	0.13	**0.17**	0.14	0.13	0.13	**0.16**
Rights	0.12	**0.18**	0.14	0.13	0.13	0.14
SAfrica#RevCoup	0.12	**0.01**	**0.08**	**0.09**	0.09	**0.15**
AbslLat	0.12	0.13	0.12	0.12	0.12	**0.18**
OutOrient	0.11	**0.16**	0.13	0.12	0.12	**0.17**
SAfrica#Other	0.11	**0.01**	**0.07**	**0.08**	0.09	**0.17**
CivilLib	0.11	**0.14**	0.12	0.12	0.12	0.14
French	0.11	0.14	0.12	0.12	0.11	**0.16**
RevCoup	0.11	0.12	0.11	0.11	0.11	**0.15**

While the weak heredity priors adjust the inclusion probability by design in the same direction as the strong heredity prior, this is not true for the *TESS* case. More specifically it promotes many of the effects found by MP. The tessellation prior reduces the importance of being a Sub–Saharan African country even further than in the original results while it focuses stronger on the interaction terms.

Considering the predictive performance of these different priors we find that all priors predict equally well. On average the LPS in table I.2 is in favor of

Table I.2: Out–of–sample CRPS/LPS Means for different Prior Choices (Std. Errors in parentheses)

Predicted Obs	NH	SH	WH0.05	WH0.2	WH0.4	TESS
70						
LPS	-3.4247	-3.4091	-3.4098	-3.4151	-3.4165	-3.4256
	(0.0407)	(0.0390)	(0.0391)	(0.0392)	(0.0397)	(0.0410)
CRPS	-0.0061	-0.0063	-0.0063	-0.0062	-0.0062	-0.0061
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
50						
LPS	-3.4283	-3.4131	-3.4138	-3.4189	-3.4203	-3.4292
	(0.0594)	(0.0585)	(0.0584)	(0.0579)	(0.0583)	(0.0599)
CRPS	-0.0061	-0.0063	-0.0063	-0.0062	-0.0062	-0.0061
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
20						
LPS	-3.4223	-3.4076	-3.4083	-3.4133	-3.4146	-3.4231
	(0.1203)	(0.1180)	(0.1179)	(0.1174)	(0.1183)	(0.1211)
CRPS	-0.0061	-0.0063	-0.0063	-0.0062	-0.0062	-0.0061
	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)
10						
LPS	-3.4298	-3.4172	-3.4178	-3.4219	-3.4227	-3.4306
	(0.1721)	(0.1695)	(0.1694)	(0.1683)	(0.1694)	(0.1732)
CRPS	-0.0060	-0.0063	-0.0062	-0.0062	-0.0062	-0.0060
	(0.0015)	(0.0015)	(0.0015)	(0.0015)	(0.0015)	(0.0015)
5						
LPS	-3.4420	-3.4248	-3.4257	-3.4318	-3.4338	-3.4428
	(0.2396)	(0.2344)	(0.2341)	(0.2328)	(0.2349)	(0.2411)
CRPS	-0.0060	-0.0062	-0.0062	-0.0061	-0.0061	-0.0060
	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0020)
1						
LPS	-3.4249	-3.4081	-3.4089	-3.4151	-3.4168	-3.4254
	(0.5681)	(0.5645)	(0.5635)	(0.5575)	(0.5605)	(0.5718)
CRPS	-0.0062	-0.0064	-0.0064	-0.0063	-0.0063	-0.0062
	(0.0046)	(0.0047)	(0.0047)	(0.0047)	(0.0047)	(0.0046)

George's Tessellation prior (*TESS*) which performs slightly better than the default prior (*NH*). In contrast the CRPS prefers the set of heredity priors over both the Tessellation and default scenario. The CRPS is slightly more volatile than the LPS which can be seen from the ratio of standard errors to the means which varies by roughly 5% and 1% respectively.

Based on these results, the prior choice for BMA setups with interaction terms is not a definite one. From a statistical point of view it seems advisable to impose a prior that enforces a 'conventional' model structure, especially if the forecasting performance remains almost unaltered. From a practical view one might argue that in this very case the use of a heredity or tessellation prior does not strongly affect the results of MP, since the effect of Sub–Saharan Africa remains in all setups.

The authors want to thank Jesus Crespo Cuaresma, Gernot Doppelhofer, Martin Feldkircher, Bettina Grün and Eduardo Ley for discussion and expert advice. The second author's research is supported by the Oesterreichische Nationalbank under the Jubiläumsfond grant 14663.

References

CHIPMAN, H. (1996): "Bayesian variable selection with related predictors". In: *Canadian Journal of Statistics* 24, pp. 17–36.

CRESPO CUARESMA, J. (2011): "How different is Africa? A comment on Masanjala and Papageorgiou". In: *Journal of Applied Econometrics* 26 (6), pp. 1041–1047.

EICHER, T., C. PAPAGEORGIOU, and A. RAFTERY (2011): "Default priors and predictive performance in Bayesian model averaging, with application to growth determinants". In: *Journal of Applied Econometrics* 26, pp. 30–55.

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

GEORGE, E. (1999): "Discussion of Bayesian model averaging and model search strategies by M.A. Clyde". In: *Bayesian Statistics 6*. Ed. by J. BERNARDO, J. BERGER, A. DAWID, and A. SMITH, pp. 157–177.

— (2010): "Dilution priors: Compensating for model space redundancy". In: *Borrowing Strength: Theory Powering Applications - A Festschrift for Lawrence D. Brown*. Ed. by J. BERGER, T. CAI, and I. JOHNSTONE. Institute of Mathematical Statistics, pp. 185–165.

GNEITING, T. and A. RAFTERY (2007): "Strictly proper scoring rules, prediction, and estimation". In: *Journal of the American Statistical Association* 102, pp. 359–378.

LEY, E. and M. STEEL (2009): "On the effect of prior assumptions in Bayesian model averaging with applications to growth regressions". In: *Journal of Applied Econometrics* 24, pp. 651–674.

MASANJALA, W. and C. PAPAGEORGIOU (2008): "Rough and lonely road to prosperity: A reexamination of the sources of growth in Africa using Bayesian model averaging". In: *Journal of Applied Econometrics* 23, pp. 671–682.

PAPAGEORGIOU, C. (2011): "How to use interaction terms in BMA: Reply to Crespo Cuaresma's comment on Masanjala and Papageorgiou (2008)". In: *Journal of Applied Econometrics* 26, pp. 1048–1050.

II Unveiling Covariate Inclusion Structures in Economic Growth Regressions Using Latent Class Analysis

European Economic Review, Revise and Resubmit

Unveiling Covariate Inclusion Structures In Economic Growth Regressions Using Latent Class Analysis*

Jesus Crespo Cuaresma[a,b,c,d], Bettina Grün[e], Paul Hofmarcher[a], Stefan Humer[a], and Mathias Moser[a]

[a]Department of Economics, Vienna University of Economics and Business (WU)

[b]Austrian Institute of Economic Research (WIFO)

[c]International Institute of Applied System Analysis (IIASA)

[d]Wittgenstein Centre for Demography and Global Human Capital (WIC)

[e]Department of Applied Statistics, Johannes Kepler University Linz (JKU)

*The authors would like to thank Sylvia Frühwirth-Schnatter, two anonymous referees and the participants at the *Second Bayesian Young Statisticians Meeting* for helpful comments. Corresponding Author: Jesus Crespo Cuaresma. Address: Welthandelsplatz 1, 1020 Vienna, Austria, Tel: +43(0)131336-4530, Fax: +43(0)131336-728, Email: jcrespo@wu.ac.at. Paul Hofmarcher's research is supported by funds of the Oesterreichische Nationalbank (Oesterreichische Nationalbank, Anniversary Fund, project number: 14663).

Abstract

We propose the use of Latent Class Analysis methods to analyze the covariate inclusion patterns across specifications resulting from Bayesian Model Averaging exercises. Using Dirichlet Process clustering, we are able to identify and describe dependency structures among variables in terms of inclusion in the specifications that compose the model space. We apply the method to two datasets of potential determinants of economic growth. Clustering the posterior covariate inclusion structure of the model space formed by linear regression models reveals interesting patterns of complementarity and substitutability across economic growth determinants.

JEL Classification: C11, C21, O47.
Keywords: Economic Growth Determinants, Bayesian Model Averaging, Latent Class Analysis, Dirichlet Processes.

1 Introduction

Bayesian Model Averaging (BMA) has become a popular tool for economic growth applications in economics (for a comprehensive introduction to BMA, see Hoeting, Madigan, Raftery, and Volinsky, 1999). The so-called *open-endedness* of economic theory concerning the factors driving income per capita differences across countries (Brock and Durlauf, 2001) allows the empirical researcher to specify a large number of models to quantify the effect of potential drivers on economic growth. The use of techniques that explicitly assess model uncertainty (mostly within the class of linear regression models) has thus become widespread in econometric research dealing with the empirical determinants of income growth differences across countries (for some seminal contributions to this literature, see e.g. Fernández, Ley, and Steel, 2001; Sala-i-Martin, Doppelhofer, and Miller, 2004; Masanjala and Papageorgiou, 2008; Durlauf, Kourtellos, and Tan, 2008; Ley and Steel, 2009b).

Economic growth applications of BMA tend to quantify the relative importance of a given covariate by calculating its so-called posterior inclusion probability (PIP), which is defined as the sum of posterior probabilities of specifications which contain that particular variable. Such a statistic has become a standard tool in econometric applications of BMA and is routinely used to measure the relative importance of different potential drivers of income growth differences across economies. While standard PIPs are intuitive measures that provide valuable insights into the overall importance of individual covariates as economic growth determinants, they face a number of shortcomings. The PIP neglects the heterogeneity across typical model specifications and accordingly does not inform about whether the degree of *importance* of the variable is evenly spread across potential specifications (that is, it is relatively independent of whether other covariates are part of the model) or, on the contrary, it is particular to specific combinations of explanatory variables.

Previous work assessing joint covariate inclusion in BMA applications has focused on capturing relevant dependency structures using bivariate measures, that is, concentrating on the analysis of the joint posterior distribution of the inclusion of pairs of variables over the model space. Such a concept has been quantified in the form of bivariate *jointness* measures in the context of BMA, put forward first by G. Doppelhofer and M. Weeks in a working paper of 2005 (Doppelhofer and Weeks, 2005), which got published in a slightly different version as Doppelhofer and Weeks (2009a). Ley and Steel (2007), Strachan (2009) and Ley and Steel (2009a) offer alternative measures of jointness. In particular, Ley and Steel (2007) formulate a set of properties for jointness measures and show that Doppelhofer and Weeks's statistics do not fulfill them. Additionally, they propose two other indices which satisfy all of their suggested properties. Strachan (2009) shows that the interpretability of the jointness measure of Doppelhofer and Weeks (2009a) may be limited in contexts where one or both of the analyzed variables have a negligible PIP and offers yet another measure in order to tackle this shortcoming. Doppelhofer and Weeks (2009b), on the other hand, argue that another desirable property of jointness measures happens to be fulfilled by their indicator but not accounted for in the indices of Ley and Steel (2007) or Strachan (2009).[1]

In this paper we propose an alternative approach aimed at succinctly and comprehensibly describing the dependency structure across variables in the model space using *latent class analysis* (LCA, see, e.g., Vermunt and Magidson, 2002) and apply it to economic growth regressions. This method was first introduced by Lazarsfeld (1950) to describe dependency structures between observed discrete variables based on latent traits and has gained widespread popularity in such research fields as psychometrics or political science (see, e.g., Breen, 2000; Blaydes and Linzer, 2008). The main idea behind LCA is to relate the realizations of observed variables to an unobserved, categorical latent variable which captures the dependency structure between the observed variables. This latent variable groups observations in such a way that the dependency between

[1] Interestingly, the measures proposed by Doppelhofer and Weeks (2009a), Ley and Steel (2007) and Strachan (2009) were independently developed earlier in the context of data mining. The statistic of Doppelhofer and Weeks (2005) is known as *log-ratio*, the measures of Ley and Steel (2007) are related to the *Jaccard index*. The index of Doppelhofer and Weeks (2009a) is known as *odds-ratio* and Strachan (2009)'s measure is closely related to the so-called *two-way support* (see Tan, Kumar, and Srivastava, 2004; Glass, 2013).

variables is reduced to a minimum within groups. By applying LCA methods to the covariate inclusion structure of best models identified by BMA, we are able to capture the dependency patterns across included covariates through a (unobserved) latent variable which induces classes with independent covariates conditional on class membership. Such a setting implies that PIPs within clusters constitute sufficient information to describe the importance of the variables and the differences of PIPs between clusters are representative of the dependencies in the inclusion of a covariate with respect to (all) other variables.

The method proposed in this paper provides a tool for applied econometricians that goes beyond the identification of individual robust determinants of socioeconomic variables by distilling the joint covariate structures that underlie the distribution of the posterior model probability across specifications. Suitable theoretical frameworks based on the results of the clustering can then be inferred based on the identity of the corresponding groups of variables. In the spirit of Durlauf, Kourtellos, and Tan (2008), the applied researcher may be interested in incorporating prior beliefs about the relative importance of some theoretical frameworks (defined over the joint prior inclusion probability of certain covariate groups) in order to assess the evidence for or against them. The modeling tool provided by our method is able to incorporate this information in a straightforward manner.

We apply this approach to the two datasets that have been most widely used for assessing the robustness of economic growth determinants (those in Fernández, Ley, and Steel, 2001, and Sala-i-Martin, Doppelhofer, and Miller, 2004, henceforth FLS and SDM, respectively). Our results for the FLS dataset reveal patterns of complementarity and substitutability across geographical, institutional and religious variables. For the SDM dataset, we find that the importance of the variable related to malaria prevalence is highly dependent on the inclusion of other covariates in the specification. The insights gained from the clustering exercise for the SDM dataset partly reconcile some of the contradictory results found in the literature concerning the importance of malaria prevalence as a determinant of income growth differences across countries (see for example

Sala-i-Martin, Doppelhofer, and Miller, 2004; Schneider and Wagner, 2012; Hofmarcher, Crespo Cuaresma, Grün, and Hornik, 2014).

The remainder of this paper is structured as follows. In Section 2, we present the econometric setting used to analyze the anatomy of covariate inclusion over the model space within BMA applications and outline the LCA approach. Section 3 presents the results of the LCA analysis applied to the set of best models identified for the FLS and SDM datasets. Section 4 concludes and proposes further paths of research.

2 Evaluating Covariate Inclusion Dependency Using Latent Class Analysis

2.1 Model Uncertainty and Economic Growth Determinants: The Econometric Framework

The standard setting for BMA analysis in the framework of cross-country growth regressions assumes that the growth rate of income per capita (y) can be linearly related to a group of covariates (X_j) chosen from a set of potential growth determinants (X). Assuming that n observations are available, a typical linear regression model (M_j) is given by

$$y|\alpha, \beta_j, \sigma \sim N(\alpha\iota + X_j\beta_j, \sigma^2 I), \tag{1}$$

where ι is a column vector of ones of dimension n. Assuming that a total of K variables are available, inference on a quantity of interest (Δ) is given by

$$p(\Delta|y) = \sum_{j=1}^{2^K} p(\Delta|y, M_j)p(M_j|y), \tag{2}$$

where $p(M_j|y)$ is the posterior model probability, which in turn is proportional to the product of the prior model probability $p(M_j)$ and its marginal likelihood $p(y|M_j)$. After eliciting priors over model-specific parameters ($p(\beta_j|M_j)$ and $p(\sigma|M_j)$), as well as over models ($p(M_j)$), posterior model probabilities and thus the posterior distributions given by equation (2) can be computed. The problems caused by the exorbitantly large number of summands in equation (2) when K is not small can be overcome in a straightforward manner by sampling from the model space using Markov Chain Monte Carlo (MCMC) methods (Madigan and York, 1995).

In the spirit of the literature on jointness in BMA applications, we propose to analyze the anatomy of the set of models sampled by the Markov chain in order to carry out inference about the covariate inclusion structures existing in the model space. While existing jointness measures tend to concentrate on the analysis of the $K \times K$ matrix of bivariate inclusion frequencies in the Markov chain, we aim at gaining knowledge about the overall structure of covariate inclusion by analyzing the full $M \times K$ matrix of inclusion profiles of the specifications sampled by the Markov chain, where M is the number of sampled models. A model profile γ_i, for $i = 1, \dots, M$ (that is, one of the rows of the matrix), is a K-dimensional vector of ones and zeros indicating the variables which are included in model i, with typical element $\gamma_{ik} = 1$ if variable k is part of model i and $\gamma_{ik} = 0$ otherwise. We propose to perform the analysis of the inclusion patterns over the model space assuming the existence of implicit latent groups to which model specifications are assigned depending on their covariate inclusion pattern.

2.2 Latent Classes and Covariate Inclusion: A Bayesian Approach Using Dirichlet Processes

We propose to use a method that resembles existing BMA applications dealing with the computation of jointness measures among covariates. It takes a two-step approach in terms of analyzing the posterior probability distribution over model specifications obtained using standard BMA methods. Using clustering methods based on LCA, it aims at unveiling clusters of model profiles among the specifications sampled in the Markov chain Monte Carlo model composite procedure.

Following the methods put forward by Molitor, Papathomas, Jerrett, and Richardson (2010), we apply Dirichlet Process Clustering (DPC) in order to carry out inference about the latent classes governing covariance inclusion structures in economic growth regressions. Compared to other methods in the literature (Forgy, 1965; Hartigan and Wong, 1979; Patterson, Dayton, and Graubard, 2002), DPC eliminates the need to set the number of latent classes a priori. While selecting a suitable number of clusters has been a widely discussed problem in the LCA and finite mixture literatures (McLachlan and Peel, 2000, Chap. 6), the nature of Bayesian inference using DPC allows for the automatic selection of an optimal number of clusters for given prior settings.

We assume that γ_i, the K-dimensional vector summarizing the variable inclusion profile for model i, has elements that arise from a mixture of infinitely, but countably many distributions,

$$p(\gamma_i) = \sum_{c=1}^{\infty} p(g_i = c) \prod_{k=1}^{K} p(\gamma_{ik} | g_i = c), \tag{3}$$

where $p(g_i = c)$ denotes the probability that model i is assigned to cluster c and $p(\gamma_{ik} \mid g_i = c)$ governs the inclusion probability of the k-th covariate in cluster c. In turn, for our application we use

$$\begin{aligned} p(\gamma_{ik} \mid g_i = c) &\sim \text{Bern}(\pi_{ck}), \\ \pi_{ck} &\sim \text{Beta}(\delta, \delta), \\ p(g_i = c) &= V_c \prod_{j=1}^{c-1} (1 - V_j), \\ V_c &\sim \text{Beta}(1, \alpha). \end{aligned}$$

Such a mixture model implies, that given assignment to a cluster, the inclusion of covariate k resembles the probabilistic process proposed, for example, in

Ley and Steel (2009b). The inclusion probability of covariate k in a given cluster c is thus governed by a Bernoulli distribution whose parameter follows a Beta distribution. The probabilistic structure that governs assignment to the different clusters, $p(g_i = c)$, on the other hand, corresponds to the so-called stick-breaking process formulation of the Dirichlet process (see Sethuraman, 1994; Papaspiliopoulos, 2008; Liverani, Hastie, Papathomas, and Richardson, 2013). This representation can be interpreted as determining the mixing proportions $p(g_i = c)$ by successive divisions of the unit interval whose relative sizes are determined by independent draws from the Beta$(1, \alpha)$ distribution.

Posterior inference for DPC can be carried out using MCMC methods. Papaspiliopoulos and Roberts (2008), for instance, present an approach using retrospective sampling. However, identifying a DPC model is difficult due to label switching (Redner and Walker, 1984). We follow Molitor, Papathomas, Jerrett, and Richardson (2010) and derive a suitable partitioning of the set of sampled model profiles using the information on co-assignment to the same clusters during sampling. This information is collapsed into an association matrix that can be interpreted as a similarity matrix between model profiles when assuming that model specifications often assigned to the same cluster are *similar*. A clustering technique relying only on similarity measures between specification profiles can then be used to find the final clustering, for instance Partitioning Around Medoids (PAM, Kaufman and Rousseeuw, 1990), which is the approach used in our empirical application.

Once a partition has been chosen, several statistics can be used to assess the goodness of fit of the clustering. In our application we rely on measures based on the likelihood ratio chi-squared test statistic (G^2), which measures goodness-of-fit by relating the observed counts of specification profiles in each cluster to the counts predicted by the estimated model. The test statistic is given by $G^2 = 2\sum_j^{2^K} q_j \ln \dfrac{q_j}{Q_j}$, where q_j refers to the observed number of counts of specification profile γ_j and Q_j is the expected number of counts assuming independency of the explanatory variables (see for example Brier, 1980). We calculate this G^2 statistic separately for each cluster and the aggregated BMA results.

In addition, in order to identify substitutability/complementarity of variables based on the cluster solution, we compute a simple measure of *interestingness* of a variable (IM) in the spirit of the literature on association rules. The interestingness measure IM is determined as the square root of the mean squared deviation of PIPs with respect to the unclustered case across clusters, weighted by the cluster-specific mass of posterior model probabilities. Thus, this measure reflects the stability of the relative importance of the variable across model structures and is able to give an indication of the existence of substitutability/complementarity inclusion patterns across covariates in the model space.

2.3 A Simulation Exercise

We assess the performance of the method by making use of a small-scale simulation exercise. We consider a set of ten potential covariates, x_k, $k = 1, \dots, 10$ and two settings based on different data generating processes. In the first setting, the dependent variable is a linear combination of the first five covariates and a random error term, $y_i = \sum_{k=1}^{5} x_{ik} + \varepsilon_i$, where $\varepsilon_i \sim N(0, 0.01)$ and all covariates are drawn from standard normal distributions. In the second setting, the dependent variable can be represented by two different linear combinations of predictors, so that $y_i = \sum_{k=1}^{5} x_{ik} + \varepsilon_i = -\sum_{j=6}^{10} x_{ik} + \varepsilon_i$.[2]

Using simulated datasets with 50 observations for each one of the settings, we perform standard BMA (assuming a single cluster of model specifications) as well as the clustering procedure proposed over the sampled model profiles. We use a Beta-Binomial prior for covariate inclusion (Ley and Steel, 2009b) and a unit information prior for the parameters in the BMA application. For this small example with $K = 10$ a complete enumeration of all models is performed. For the clustering procedure, we use a Gamma(2, 1) prior over α, elicit $\delta = 90$ and retain the top 500 models. The posterior inference is based on 1,500 MCMC iterations, after 1,000 burn-in runs. The results for the first (single cluster) setting are presented in Table II.1, where we report

[2]Technically, we implement this setting by defining $x_{i,10} \approx -\sum_{k=1}^{9} x_{ik}$.

the posterior inclusion probabilities and the mean of the posterior distribution of the parameters associated to each one of the covariates, averaged over 100 simulated datasets.

The standard BMA method (see results in panel (a) of Table II.1) correctly identifies the covariates included in the true model and the mean of the posterior distribution of the relevant parameters appear very close to the true value of unity. The clustering approach identifies two clusters, with the first one covering over 99% of the models in the BMA procedure and reproducing the same results as those in the non-clustered case in terms of PIP and means of the posterior distribution of the associated parameters (see panels (b) and (c) in Table II.1). In the second setting, whose results are presented in Table II.2, the standard BMA procedure *averages out* the effects of the two alternative data generating processes. The PIP values obtained using BMA are around 0.6 for all variables and the mean of the posterior distribution over the parameters is approximately 0.5 for the first five covariates and -0.5 for the rest of the variables. DPC is able to disentangle the two competing data generating processes, assigning roughly the same posterior mass to each one of the two clusters found. The mean of the posterior distribution of the parameters are in line with the actual values in the true model(s) and the covariates which are not included in the alternative specification have a relatively low PIP and an expected effect which is very close to zero.

3 Covariate Inclusion Clustering in Economic Growth Regressions

The clustering method presented in Section 2 is applied to the datasets compiled by Fernández, Ley, and Steel (2001) and Sala-i-Martin, Doppelhofer, and Miller (2004) (henceforth, FLS and SDM datasets). These two datasets comprise cross-country information on a large number of potential determinants of income growth and have been extensively used to assess empirically the role played by model uncertainty in economic growth regressions. In addition to GDP per

Table II.1: Simulation Results: Single cluster

(a) Standard BMA			**(b)** DPC: Cluster 1 (>99%)			**(c)** DPC: Cluster 2 (<1%)		
	PIP	P. Mean		PIP	P. Mean		PIP	P. Mean
β_1	1.0000	0.9799	β_1	1.0000	0.9799	β_1	1.0000	0.9799
β_2	1.0000	0.9822	β_2	1.0000	0.9822	β_2	1.0000	0.9824
β_3	1.0000	0.9810	β_3	1.0000	0.9810	β_3	1.0000	0.9811
β_4	1.0000	0.9828	β_4	1.0000	0.9828	β_4	1.0000	0.9825
β_5	1.0000	0.9796	β_5	1.0000	0.9796	β_5	1.0000	0.9804
β_6	0.2030	−0.0001	β_6	0.2009	−0.0001	β_6	0.9975	−0.0004
β_7	0.2053	−0.0007	β_7	0.2033	−0.0007	β_7	1.0000	−0.0025
β_8	0.2032	0.0003	β_8	0.2011	0.0003	β_8	1.0000	0.0010
β_9	0.2063	−0.0002	β_9	0.2042	−0.0002	β_9	1.0000	−0.0013
β_{10}	0.2053	−0.0001	β_{10}	0.2034	−0.0001	β_{10}	1.0000	−0.0010

Simulation results averaged over 100 simulated datasets. Data generating process: $y_i = \sum_{k=1}^{5} x_{ik} + \varepsilon_i$. Column labelled *PIP* reports the posterior inclusion probability, column labelled *P. Mean* reports the mean of the posterior distribution of the corresponding parameter. See text for details on the setting of the simulation.

capita growth figures, the FLS dataset is composed by 41 covariates and spans information for 72 countries, while the SDM dataset includes information on 67 different determinants for 88 economies. The variables in both datasets are presented in the Appendix A.

The BMA analysis of both datasets is carried out using a Beta-Binomial prior on covariate inclusion probabilities with a prior average model size of $K/2$ (20.5 for the FLS dataset and 33.5 for the SDM dataset) and the hyper g-prior proposed in Liang et al. (2008) for the regression coefficients. We base our inference concerning the inclusion probability of covariates on five million MCMC model draws, whereby the first two million draws were discarded. Alternatively, we also implemented dilution priors over the model space following George (1999) (see also Durlauf, Kourtellos, and Tan, 2008). Such a model prior assigns relatively lower prior probability to specifications with highly correlated covariates by weighting the prior model probability using the determinant of the correlation matrix of the explanatory variables. The results obtained using

Table II.2: Simulation Results: Two clusters

	(a) Standard BMA			(b) DPC: Cluster 1 (49%)			(c) DPC: Cluster 2 (51%)	
	PIP	P. Mean		PIP	P. Mean		PIP	P. Mean
β_1	0.6030	0.4951	β_1	0.1995	0.0024	β_1	1.0000	0.9817
β_2	0.6038	0.4949	β_2	0.2009	0.0023	β_2	1.0000	0.9814
β_3	0.6032	0.4931	β_3	0.2000	0.0012	β_3	1.0000	0.9789
β_4	0.6039	0.4945	β_4	0.2014	0.0018	β_4	1.0000	0.9809
β_5	0.6028	0.4941	β_5	0.1991	0.0020	β_5	1.0000	0.9801
β_6	0.5983	−0.4854	β_6	1.0000	−0.9774	β_6	0.2009	0.0006
β_7	0.5989	−0.4874	β_7	1.0000	−0.9808	β_7	0.2025	0.0000
β_8	0.5985	−0.4853	β_8	1.0000	−0.9772	β_8	0.2018	0.0006
β_9	0.5983	−0.4851	β_9	1.0000	−0.9771	β_9	0.2011	0.0008
β_{10}	0.5983	−0.4857	β_{10}	1.0000	−0.9782	β_{10}	0.2009	0.0006

Simulation results averaged over 100 simulated datasets. Data generating process: $y_i = \sum_{k=1}^{5} x_{ik} + \varepsilon_i = -\sum_{k=6}^{10} x_{ik} + \varepsilon_i$. Column labelled *PIP* reports the posterior inclusion probability, column labelled *P. Mean* reports the mean of the posterior distribution of the corresponding parameter. See text for details on the setting of the simulation.

such a dilution prior are not qualitatively different from those with the standard Beta-Binomial prior which are presented below.[3]

Using the top 500 unique models visited by the Markov chain (weighted by their posterior model probability), we apply the clustering procedure described in Section 2 in order to unveil clusters of inclusion patterns across specifications. Technically, we create an auxiliary dataset composed by the 500 top model profiles drawn where the number of observations of each model profile is proportional to its posterior probability. We normalize this auxiliary dataset so that the profile corresponding to the 500^{th} top model is included exactly once and the relative importance of the rest of the models is preserved. For the FLS

[3]For the FLS dataset, for instance, the correlation between the posterior inclusion probabilities obtained with the dilution prior and the standard Beta-Binomial prior, as well as between the means and standard deviations of the posterior parameter distributions, tend to be above 0.8. Detailed results of the BMA exercise using George (1999)'s dilution prior are available from the authors upon request.

and SDM dataset the weighted top 500 model profiles in the auxiliary datasets span 33,480 and 28,800 model profile observations, respectively.[4]

Concerning prior elicitation for DPC, we use a setting that implies a relative preference for a smaller number of broad clusters over a multitude of clusters populated by few model structures, which may eventually lack interpretability. We use a Gamma(2, 1) prior over α and $\delta = 90$. Posterior inference is based on 1,500 MCMC iterations, after 1,000 burn-in runs. This choice of priors is relatively standard in LCA applications (see e.g. Liverani, Hastie, Papathomas, and Richardson, 2013).[5]

3.1 Results for the FLS Dataset

DPC identifies an optimal partition of seven clusters of models by inclusion structure in the FLS dataset. Table II.3 provides an overview of the main characteristics of these different model clusters regarding the number of model specifications in the cluster, as well as the mean model size and the average adjusted R^2 for specifications within the cluster. These statistics are also presented for the unclustered model space considered. Although the top 500 models used for the analysis only cover approximately 8% of the posterior model probability in the space of potential specifications, the overall unclustered results are very similar to those in Fernández, Ley, and Steel (2001) concerning the PIP of individual variables.[6]

[4]Expanding the set of top models to cover a larger part of the posterior model probability leads to significant computational complications. For the case of the FLS dataset, which contains less covariates, we also implemented the method for the top 1,000 models, leading to similar results as those presented for the top 500 specifications. Such a result is not very surprising given the fact that the increase in the posterior model probability covered by the top models is very modest when moving from the top 500 to the top 1,000 specifications.

[5]We have carried out several robustness checks changing the elicitation of the priors which did not lead to any significant differences in the inference results as long as the prior setting implies a preference for a relatively small number of clusters.

[6]It should be noted that, in contrast to Fernández, Ley, and Steel (2001) and Ley and Steel (2007), we employ a hyperprior for prior inclusion probabilities and model-specific parameters, following Ley and Steel (2009b) and Liang et al. (2008), respectively.

The first two clusters capture more than half of the posterior mass covered by the set of specifications considered, while clusters 6 and 7 cover a negligible part of the model space in terms of posterior model probability. Cluster 7 is composed by very large models and due to its minimal importance in terms of posterior probability does not appear particularly relevant in terms of interpretation. The cluster-specific G^2 statistics imply an improvement in fit as compared to the unclustered results once the covariate inclusion structures are assigned to the classes identified. The reduction in the G^2 statistic is very sizable and widespread across the clusters.

Table II.3: Summary of FLS clusters

	Overall	1	2	3	4	5	6	7
$\sum$ PMP	0.08	0.03	0.02	0.01	0.01	0.01	0.00	0.00
Avg model size	10.46	10.46	8.68	8.44	11.59	10.95	18.15	41.00
Avg adj. R^2	0.83	0.84	0.81	0.80	0.85	0.84	0.90	0.91
G^2 stat. ($\times 10^5$)	3.52	0.24	0.24	0.13	0.15	0.09	0.19	0.00

Figure II.1 offers a graphical representation of the differences in PIPs for individual covariates across the identified clusters. The covariates are sorted by their PIPs in the standard (unclustered) BMA exercise, which are plotted as a solid line together with their corresponding within-cluster PIPs, depicted as bars. It should be noted that the patterns of PIP across variables in all clusters differ structurally from the unclustered BMA results, so that no individual cluster mimics the PIPs obtained by the standard BMA exercise closely. The color of the bars in Figure II.1 corresponds to the value of the IM statistic.

The PIPs of the four most robust variables of the FLS dataset (`Confucian`, `GDP60`, `LifeExp` and `Equipinv`) appear stable across clusters. The variables with a higher degree of variability in PIPs across clusters tend to be related

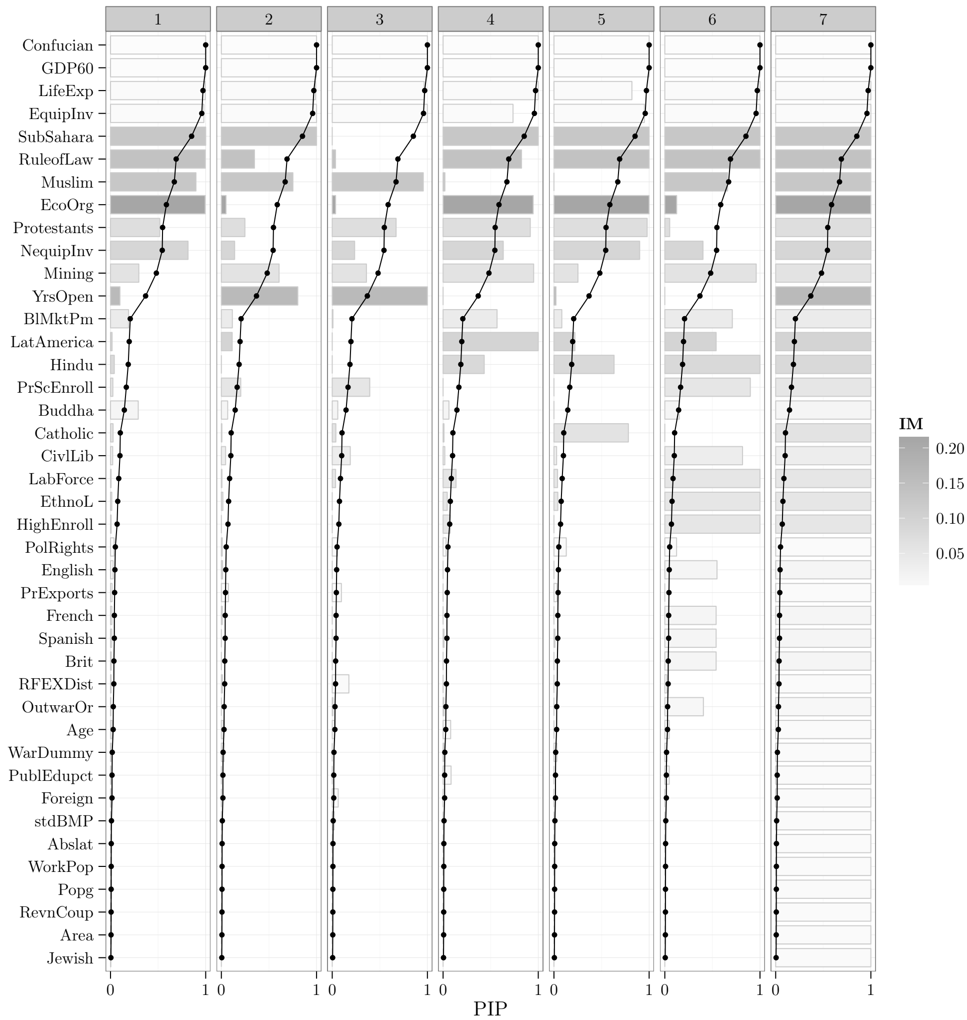


Figure II.1: FLS dataset: PIPs in unclustered BMA (solid line) and by identified cluster (bars)

Table II.4: FLS dataset: Weighted correlation of cluster-specific PIPs for variables with $IM > 0.5\max(IM)$

	SubSahara	EcoOrg	YrsOpen	Muslim	RuleofLaw
SubSahara	1.00	0.50	−0.65	−0.34	0.73
EcoOrg		1.00	−0.87	−0.33	0.87
YrsOpen			1.00	0.45	−0.96
Muslim				1.00	−0.31
RuleofLaw					1.00

to geography (`SubSahara`), institutions (`EcoOrg`, `RuleofLaw` and `YrsOpen`[7]) and religion (`Muslim`). The characteristics of the inclusion structure of these variables across clusters can be best understood by computing the weighted correlation matrix of cluster-specific PIPs, which is presented in Table II.4. The correlation among covariate inclusion variables reveals that `SubSahara`, `EcoOrg` and `RuleofLaw` tend to contain complementary information in the sense of appearing together in specifications. The same is true for the group of variables formed by `YrsOpen` and `Muslim`, while the inclusion of these two sets of variables presents sizable negative correlation. This result indicates that some of the effects of institutions and geographical variables on economic growth can be alternatively modeled using these two groups of covariates in a robust manner, but that once that they are controlled for, the inclusion of variables of the other group appears redundant.

The interplay of changes in PIPs across clusters presented in Figure II.1 indicates that the set of religious, institutional and geographical variables used in cross-country growth regressions often contain redundant information which can be replicated using different subgroups of them. An example of such a phenomenon is observed when comparing clusters 1 and 3. The importance of `SubSahara` and `RuleofLaw` as growth determinants which can be inferred from the results

[7]The variable `YrsOpen` is based on the Sachs-Warner index of openness, which has a strong institutional component. For example, socialist economies are automatically considered closed to trade by this indicator.

in cluster 1 disappears in cluster 3 and their fall in PIPs occurs in parallel to a strong increase in PIP for `YrsOpen`. The set of religious variables (`Muslim`, `Catholic`, `Protestants`, `Hindu` and, to a lesser extent, `Buddha`) also presents large variation in PIPs across clusters.

3.2 Results for the SDM Dataset

Ley and Steel (2009a) found very weak (bivariate and/or trivariate) jointness in the group of covariates included in the SDM dataset. Our procedure splits the model space into three different model clusters by covariate inclusion patterns. Table II.5 presents the summary statistics for the identified clusters. The top 500 unique specifications cover 40% of the posterior model probability, a much larger proportion than in the case of the FLS dataset. The structure of variable inclusion for the SDM dataset appears to have a different nature as compared to the results for the FLS dataset. In addition to the lower number of identified clusters, the first two classes of inclusion structures identified exhibit relatively similar characteristics in terms of the posterior model probability covered. As in the case of the FLS dataset, the cluster specific G^2 statistics are lower than the corresponding value for the model without clustering, thus supporting the method employed.

Figure II.2 depicts the PIPs of the variables in the SDM dataset computed using the top 500 models, as well as those derived from the models in the single clusters.[8] The results show a large degree of variability in PIPs across clusters for many of the covariates, including those presenting the highest PIPs in the unclustered case.

Given the large posterior probability mass over models covered by the first two clusters, we concentrate on the differences in PIPs observed between these two. Remarkable differences in PIPs across these two clusters can be observed for the `MALFAL66` variable, which presents a much higher PIP in the second

[8]Variables with PIP lower than 5% have been excluded in order to improve the readability of the graph. For these variables no remarkable changes could be detected when comparing the BMA results with the cluster-specific PIPs.

Table II.5: Summary of SDM clusters

	Overall	1	2	3
$\sum$ Posterior model prob.	0.40	0.21	0.17	0.03
Average model size	5.47	6.54	3.95	6.77
Average adjusted R^2	0.67	0.71	0.62	0.70
G^2 statistic ($\times 10^5$)	10.25	1.33	2.36	0.58

cluster, making it the second most important variable for models within that cluster. Such a phenomenon is accompanied by a sizable decrease in PIP for `P60`, `IPRICE1`, `TROPICAR`, `GDPCH60L` and `DENS65C`. The empirical literature on model uncertainty in cross-country growth regressions which analyzes the SDM dataset often reports on the effect that the use of different approaches to parameter shrinkage has on the importance of `MALFAL66`. Schneider and Wagner (2012) as well as Hofmarcher, Crespo Cuaresma, Grün, and Hornik (2014), for instance, find that the robustness of `MALFAL66` as a determinant of income growth differences across countries improves when estimation methods based on LASSO and elastic nets are used. In addition, the results in Schneider and Wagner (2012) and Hofmarcher, Crespo Cuaresma, Grün, and Hornik (2014) also indicate a loss of importance of `DENS65C` when methods implying a more stringent shrinkage are used in the estimation. These are precisely two of the variables which present the highest values of IM in our results, hinting to the fact that their relative importance depends on the *type* of model (as represented by the variable inclusion structure cluster) considered.

Such a pattern of substitutability across covariates is easily recognizable from the weighted correlation matrix of cluster-specific PIPs for the group of variables with the highest IM values, which is presented in Table II.6. The correlation patterns present in the model space indicate that `MALFAL66` tends to act as a substitute of the group of variables composed by `IPRICE1`, `TROPICAR`, `GDPCH60L` and `DENS65C`. The difference in average model size across these two important clusters in the space of posterior inclusion probability structures is

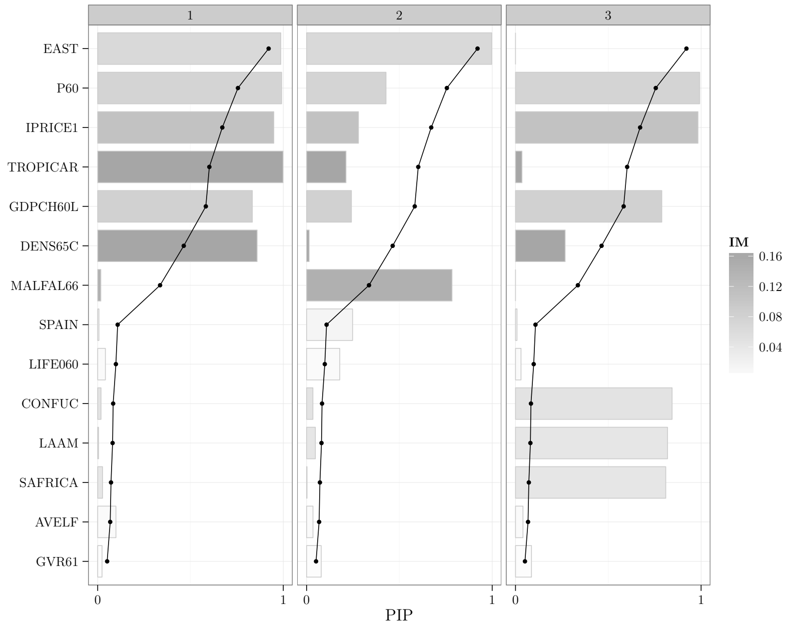


Figure II.2: SDM dataset: PIPs in unclustered BMA (solid line) and by identified cluster (bars)

Table II.6: SDM dataset: Weighted correlation of cluster-specific PIPs for variables with $IM > 0.5\max(IM)$

	DENS65C	GDPCH60L	IPRICE1	MALFAL66	TROPICAR
DENS65C	1.00	0.95	0.92	−0.93	0.97
GDPCH60L		1.00	1.00	−1.00	0.83
IPRICE1			1.00	−1.00	0.79
MALFAL66				1.00	−0.80
TROPICAR					1.00

in line with the strong impact of different parameter shrinkage approaches on the relative importance of the variables which is highlighted in previous literature. In addition, in their study of pairwise jointness measures, Doppelhofer and Weeks (2009a) report that `P60`, `IPRICE1`, `DENS65C` and `TROPICAR` exhibit significant negative bivariate jointness with `MALFAL66`, a result that can be easily reconciled with the output of our analysis. While Ley and Steel (2007) find very limited evidence for jointness structures in the SDM dataset, the only triplets of important variables for which disjointness is reported also involve `TROPICAR` and `MALFAL66`.

In spite of the fact that the third cluster that DPC identifies covers a very small part of the posterior mass over models, its PIP structure also reveals interesting patterns as compared to the other two clusters. In this group of models, two of the most relevant variables in terms of (unclustered) PIP, `EAST` and `TROPICAR`, lose their importance and their information is captured by a different set of geographical and religious variables (`CONFUC`, `LAAM` and `SAFRICA`). The results in Doppelhofer and Weeks (2009a) concerning the complementarity of `EAST` and `TROPICAR` and the substitutability of `EAST` with respect to `CONFUC`, `LAAM` and `SAFRICA` are perfectly in line with these results. In addition, Doppelhofer and Weeks (2009a) find the latter to be complements, which is also supported by the comparison of the PIPs in our third cluster with those in the other two.

4 Conclusions and Future Paths of Research

In this contribution we are concerned with covariate inclusion patterns of BMA exercises with large model spaces. Recent research on such *jointness* structures tends to choose a low-dimensional approach to such an analysis and thus concentrates on bivariate or trivariate approaches, by calculating the inclusion relationships of few explaining factors at a time. We propose a novel approach by utilizing LCA techniques and apply DPC to two well known datasets in the BMA growth literature. The clustering method put forward in our contribution aims at unveiling commonalities in the joint inclusion of variables and thus offering the applied econometrician evidence about the competing structures (as formed by *groups* of variables that appear together) that are covered by the posterior over the model space.

Our results indicate that within the set of models sampled by the Markov chain in the BMA analysis of determinants of economic growth, several distinct clusters of models by covariate inclusion can be identified. For the FLS data, we identify seven clusters of models which differ in the inclusion structure for geographic, institutional and religious covariates. In contrast, the SDM dataset only reveals three latent classes with very different dependency structures. The inclusion of the variable measuring malaria prevalence is shown to vary strongly across clusters, with its effect on economic growth being captured often by other factors such as the fraction of tropical area and coastal population density.

We show that the study of dependency structures in covariate inclusion for large model spaces appears particularly relevant in order to understand the nature of the factors affecting global patterns of income growth. The proposed method lends itself to further straightforward expansions such as the use of low-dimensional jointness measures for the analysis of within-cluster inclusion patterns for small groups of covariates. The assessment of covariate inclusion clusters in the model space under different shrinkage priors can also shed light on the effects of multicollinearity on the robustness of economic growth determinants to model uncertainty.

In order to make our method and results comparable to those in the literature on *jointness* measures, we decided to follow a two-step procedure and use the clustering method on the model profiles visited by the Markov chain of the BMA procedure. The LCA and DPC methods proposed in this contribution would also lend themselves to create priors over suitable covariate combinations in the specifications that compose the model space. This path of further research, which we are pursuing at the moment, appears particularly promising in order to unify the literature on jointness and dilution priors in BMA applications.

Appendix

A Datasets

Table A.1: Variable names and descriptive statistics — FLS

	Abbreviation	Variable	Mean	Std. Dev.
1	Abslat	Absolute latitude	25.73	17.250
2	Age	Age	23.71	37.307
3	Area	Area (Scale Effect)	972.92	2051.976
4	BlMktPm	Black Market Premium	0.16	0.291
5	Brit	British Colony dummy	0.32	0.470
6	Buddha	Fraction Buddhist	0.06	0.184
7	Catholic	Fraction Catholic	0.42	0.397
8	CivlLib	Civil Liberties	3.47	1.712
9	Confucian	Fraction Confucian	0.02	0.087
10	EcoOrg	Degree of Capitalism	3.54	1.266
11	English	Fraction of Pop. Speaking English	0.08	0.239
12	EquipInv	Equipment investment	0.04	0.035
13	EthnoL	Ethnolinguistic fractionalization	0.37	0.296
14	Foreign	Fraction speaking foreign language	0.37	0.422
15	French	French Colony dummy	0.12	0.333
16	GDP60	GDP level in 1960	7.49	0.885
17	HighEnroll	Higher education enrollment	0.04	0.052
18	Hindu	Fraction Hindu	0.02	0.101
19	Jewish	Fraction Jewish	0.01	0.097
20	LabForce	Size labor force	9305.38	24906.056
21	LatAmerica	Latin American dummy	0.28	0.451
22	LifeExp	Life expectancy	56.58	11.448
23	Mining	Fraction GDP in mining	0.04	0.077
24	Muslim	Fraction Muslim	0.15	0.295
25	NequipInv	Non-Equipment Investment	0.15	0.055
26	OutwarOr	Outward Orientation	0.39	0.491
27	PolRights	Political Rights	3.45	1.896
28	Popg	Population Growth	0.02	0.010
29	PrExports	Primary exports, 1970	0.67	0.299
30	Protestants	Fraction Protestant	0.17	0.252
31	PrScEnroll	Primary School Enrollment, 1960	0.80	0.246
32	PublEdupct	Public Education Share	0.02	0.009
33	RevnCoup	Revolutions and coups	0.18	0.238
34	RFEXDist	Exchange rate distortions	121.71	41.001
35	RuleofLaw	Rule of law	0.55	0.335
36	Spanish	Spanish Colony dummy	0.22	0.419
37	stdBMP	SD of black-market premium	45.60	95.802
38	SubSahara	Sub-Saharan dummy	0.21	0.409
39	WarDummy	War dummy	0.40	0.494
40	WorkPop	Ratio workers to population	−0.95	0.189
41	y	GDP per capita growth	0.02	0.018
42	YrsOpen	Number of Years open economy	0.44	0.355

Table A.2: Variable names and descriptive statistics — SDM

	Abbreviation	Variable	Mean	Std. Dev.
1	ABSLATIT	Absolute latitude	23.21	16.843
2	AIRDIST	Air distance to big cities	4324.17	2613.763
3	AVELF	Ethnolinguistic fractionalization	0.35	0.302
4	BRIT	British colony	0.32	0.468
5	BUDDHA	Fraction Buddhist	0.05	0.168
6	CATH00	Fraction Catholic	0.33	0.415
7	CIV72	Civil liberties	0.51	0.326
8	COLONY	Colony dummy	0.75	0.435
9	CONFUC	Fraction Confucian	0.02	0.079
10	DENS60	Population density costal 1960's	108.07	201.445
11	DENS65C	Population density 1960	146.87	509.828
12	DENS65I	Interior density	43.37	88.063
13	DPOP6090	Population growth rate 1960-1990	0.02	0.009
14	EAST	East Asian dummy	0.11	0.319
15	ECORG	Capitalism	3.47	1.381
16	ENGFRAC	English-speaking population	0.08	0.252
17	EUROPE	European dummy	0.22	0.414
18	FERTLDC1	Fertility in 1960's	1.56	0.419
19	GDE1	Defense spending share	0.03	0.025
20	GDPCH60L	GDP 1960 (log)	7.35	0.901
21	GEEREC1	Public education spending share in GDP in 1960's	0.02	0.010
22	GGCFD3	Government consumption share deflated with GDP prices	0.05	0.039
23	GOVNOM1	Nominal government GDP share 1960's	0.15	0.058
24	GOVSH61	Government share of GDP	0.17	0.071
25	GR6096	Average growth rate of GDP per capita 1960-1996	0.02	0.019
26	GVR61	Government consumption share 1960's	0.12	0.075
27	H60	Higher education in 1960	0.04	0.050
28	HERF00	Religous intensity	0.78	0.193
29	HINDU00	Fraction Hindu	0.03	0.125
30	IPRICE1	Investment price	92.47	53.678
31	LAAM	Latin American dummy	0.23	0.421
32	LANDAREA	Land area	867188.52	1814688.290
33	LANDLOCK	Landlocked country dummy	0.17	0.378
34	LHCPC	Hydrocarbon deposits in 1993	0.42	4.351
35	LIFE060	Life expectancy in 1960	53.72	12.062
36	LT100CR	Fraction of land area near navigable water	0.47	0.380
37	MALFAL66	Malaria prevalence in 1960's	0.34	0.431
38	MINING	Fraction GDP in mining	0.05	0.077
39	MUSLIM00	Fraction Muslim	0.15	0.296
40	NEWSTATE	Time of independence	1.01	0.977

Table A.2: *(continued)*

	Abbreviation	Variable	Mean	Std. Dev.
41	OIL	Oil-producing country dummy	0.06	0.233
42	OPENDEC1	(Imports+exports)/GDP	0.52	0.336
43	ORTH00	Fraction Orthodox	0.02	0.098
44	OTHFRAC	Fraction speaking foreign language	0.32	0.414
45	P60	Primary schooling 1960	0.73	0.293
46	PI6090	Average inflation 1960-1990	13.13	14.990
47	POP1560	Fraction population less than 15	0.39	0.075
48	POP60	Population in 1960	20308.08	52538.387
49	POP6560	Fraction population over 65	0.05	0.029
50	PRIEXP70	Primary exports in 1970	0.72	0.283
51	PRIGHTS	Political rights	3.82	1.997
52	PROT00	Fraction Protestant	0.14	0.285
53	RERD	Real exchange rate distortions	125.03	41.706
54	REVCOUP	Revolution and coups	0.18	0.232
55	SAFRICA	African dummy	0.31	0.464
56	SCOUT	Outward orientation	0.40	0.492
57	SIZE60	Size of the economy	16.15	1.820
58	SOCIALIST	Socialist dummy	0.07	0.254
59	SPAIN	Spanish colony	0.17	0.378
60	SQPI6090	Square of inflation 1960-1990	394.54	1119.699
61	TOT1DEC1	Terms of trade growth in 1960's	0.00	0.035
62	TOTIND	Terms of trade ranking	0.28	0.190
63	TROPICAR	Fraction of tropical area	0.57	0.472
64	TROPPOP	Fraction population in tropics	0.30	0.373
65	WARTIME	Fraction spent in war 1960-1990	0.07	0.152
66	WARTORN	War participation 1960-1990	0.40	0.492
67	YRSOPEN	Years open	0.36	0.344
68	ZTROPICS	Tropical climate zone	0.19	0.269

B Posterior Inclusion Probabilities by Cluster

Table B.1: PIPs within detected clusters — FLS

	Overall	1	2	3	4	5	6	7	IM
GDP level in 1960	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
Fraction Confucian	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
Life expectancy	0.97	0.99	0.99	0.97	1.00	0.82	1.00	1.00	0.00
Equipment investment	0.96	0.98	1.00	1.00	0.73	0.95	1.00	1.00	0.01
Sub-Saharan dummy	0.85	1.00	1.00	0.00	1.00	1.00	1.00	1.00	0.13
Rule of law	0.69	1.00	0.35	0.03	0.82	1.00	1.00	1.00	0.14
Fraction Muslim	0.67	0.90	0.75	0.96	0.02	0.00	0.66	1.00	0.13
Degree of Capitalism	0.59	0.99	0.05	0.04	0.95	1.00	0.12	1.00	0.21
Fraction Protestant	0.55	0.52	0.25	0.67	0.92	0.98	0.05	1.00	0.07
Non-Equipment Investment	0.54	0.81	0.14	0.24	0.63	0.90	0.40	1.00	0.09
Fraction GDP in mining	0.48	0.30	0.61	0.36	0.95	0.25	0.96	1.00	0.06
Number of Years open economy	0.37	0.10	0.80	1.00	0.00	0.02	0.00	1.00	0.16
Black Market Premium	0.21	0.19	0.12	0.01	0.57	0.08	0.71	1.00	0.04
Latin American dummy	0.20	0.02	0.11	0.00	1.00	0.22	0.54	1.00	0.10
Fraction Hindu	0.19	0.04	0.00	0.00	0.43	0.63	1.00	1.00	0.09
Primary School Enrollment, 1960	0.17	0.02	0.20	0.39	0.00	0.00	0.90	1.00	0.05
Fraction Buddhist	0.15	0.29	0.07	0.06	0.06	0.00	0.15	1.00	0.02
Fraction Catholic	0.10	0.03	0.01	0.04	0.01	0.78	0.00	1.00	0.06
Civil Liberties	0.10	0.02	0.04	0.19	0.02	0.03	0.82	1.00	0.04
Size labor force	0.09	0.01	0.01	0.04	0.14	0.04	1.00	1.00	0.05
Ethnolinguistic fractionalization	0.08	0.01	0.02	0.00	0.04	0.04	1.00	1.00	0.05
Higher education enrollment	0.07	0.01	0.00	0.00	0.08	0.00	1.00	1.00	0.05
Political Rights	0.05	0.04	0.01	0.06	0.03	0.13	0.12	1.00	0.00
Fraction of Pop. Speaking English	0.05	0.03	0.01	0.00	0.00	0.00	0.55	1.00	0.02
Primary exports, 1970	0.04	0.02	0.07	0.10	0.00	0.03	0.00	1.00	0.00
French Colony dummy	0.04	0.01	0.01	0.00	0.00	0.00	0.54	1.00	0.02
Spanish Colony dummy	0.04	0.01	0.00	0.00	0.01	0.01	0.54	1.00	0.02
British Colony dummy	0.04	0.01	0.00	0.00	0.00	0.00	0.54	1.00	0.02
Exchange rate distortions	0.03	0.01	0.01	0.17	0.00	0.00	0.02	1.00	0.01
Outward Orientation	0.03	0.00	0.00	0.01	0.01	0.00	0.41	1.00	0.01
Age	0.03	0.02	0.02	0.02	0.08	0.00	0.05	1.00	0.00
War dummy	0.02	0.01	0.03	0.00	0.01	0.03	0.02	1.00	0.00
Public Education Share	0.02	0.00	0.00	0.00	0.08	0.00	0.05	1.00	0.00
Fraction speaking foreign language	0.02	0.01	0.00	0.06	0.00	0.00	0.00	1.00	0.00
SD of black-market premium	0.01	0.00	0.01	0.02	0.00	0.00	0.00	1.00	0.00
Absolute latitude	0.01	0.01	0.00	0.01	0.00	0.00	0.00	1.00	0.00
Ratio workers to population	0.01	0.01	0.00	0.01	0.00	0.00	0.00	1.00	0.00
Population Growth	0.01	0.01	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Revolutions and coups	0.01	0.01	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Area (Scale Effect)	0.01	0.01	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Fraction Jewish	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00

Table B.2: PIPs within detected clusters — SDM

	Overall	1	2	3	IM
East Asian dummy	0.92	0.99	1.00	0.00	0.06
Primary schooling 1960	0.76	0.99	0.43	0.99	0.08
Investment price	0.67	0.95	0.28	0.98	0.11
Fraction of tropical area	0.60	1.00	0.21	0.03	0.17
GDP 1960 (log)	0.58	0.83	0.24	0.79	0.08
Population density 1960	0.46	0.86	0.01	0.27	0.17
Malaria prevalence in 1960's	0.34	0.02	0.78	0.00	0.14
Spanish colony	0.11	0.01	0.25	0.01	0.01
Life expectancy in 1960	0.10	0.04	0.18	0.03	0.00
Fraction Confucian	0.08	0.02	0.03	0.84	0.04
Latin American dummy	0.08	0.00	0.05	0.82	0.04
African dummy	0.07	0.03	0.00	0.81	0.04
Ethnolinguistic fractionalization	0.07	0.10	0.03	0.04	0.00
Government consumption share 1960's	0.05	0.02	0.08	0.09	0.00
Political rights	0.05	0.09	0.00	0.00	0.00
Years open	0.05	0.04	0.05	0.04	0.00
Fraction Muslim	0.04	0.03	0.03	0.15	0.00
Fraction Buddhist	0.04	0.04	0.00	0.28	0.00
Fraction GDP in mining	0.04	0.03	0.02	0.15	0.00
Population density costal 1960's	0.03	0.06	0.00	0.08	0.00
Higher education in 1960	0.03	0.03	0.03	0.00	0.00
(Imports+exports)/GDP	0.03	0.03	0.02	0.02	0.00
Government share of GDP	0.02	0.01	0.02	0.14	0.00
Fraction speaking foreign language	0.02	0.02	0.02	0.03	0.00
Primary exports in 1970	0.02	0.00	0.04	0.00	0.00
Air distance to big cities	0.02	0.04	0.00	0.00	0.00
Real exchange rate distortions	0.02	0.02	0.01	0.02	0.00
Fraction population less than 15	0.02	0.03	0.01	0.00	0.00
Government consumption share deflated with GDP prices	0.01	0.01	0.00	0.10	0.00
Fraction Protestant	0.01	0.01	0.02	0.01	0.00
Fraction population in tropics	0.01	0.01	0.01	0.03	0.00
Absolute latitude	0.01	0.01	0.01	0.00	0.00
Civil liberties	0.01	0.01	0.00	0.00	0.00
Colony dummy	0.01	0.01	0.01	0.00	0.00
Revolution and coups	0.01	0.01	0.01	0.00	0.00
Outward orientation	0.01	0.01	0.00	0.00	0.00
Fraction Hindu	0.01	0.01	0.00	0.00	0.00
Average inflation 1960-1990	0.01	0.01	0.00	0.00	0.00
European dummy	0.00	0.00	0.00	0.01	0.00
Size of the economy	0.00	0.01	0.00	0.00	0.00

Table B.2: *(continued)*

	Overall	1	2	3	IM
Hydrocarbon deposits in 1993	0.00	0.01	0.00	0.00	0.00
Fertility in 1960's	0.00	0.01	0.00	0.00	0.00
Fraction population over 65	0.00	0.00	0.00	0.00	0.00
British colony	0.00	0.01	0.00	0.00	0.00
English-speaking population	0.00	0.00	0.01	0.00	0.00
Square of inflation 1960-1990	0.00	0.01	0.00	0.00	0.00
Defense spending share	0.00	0.01	0.00	0.00	0.00
Landlocked country dummy	0.00	0.01	0.00	0.00	0.00
Religous intensity	0.00	0.01	0.00	0.00	0.00
Oil-producing country dummy	0.00	0.01	0.00	0.00	0.00
Time of independence	0.00	0.01	0.00	0.00	0.00
Socialist dummy	0.00	0.01	0.00	0.00	0.00
Fraction Catholic	0.00	0.00	0.00	0.01	0.00
Population growth rate 1960-1990	0.00	0.00	0.00	0.00	0.00
Nominal government GDP share 1960's	0.00	0.01	0.00	0.00	0.00
Public education spending share in GDP in 1960's	0.00	0.00	0.00	0.00	0.00
Capitalism	0.00	0.01	0.00	0.00	0.00
Terms of trade growth in 1960's	0.00	0.00	0.00	0.00	0.00
Tropical climate zone	0.00	0.00	0.00	0.00	0.00
Fraction spent in war 1960-1990	0.00	0.00	0.00	0.00	0.00
War participation 1960-1990	0.00	0.00	0.00	0.00	0.00
Land area	0.00	0.00	0.00	0.00	0.00
Population in 1960	0.00	0.00	0.00	0.00	0.00
Fraction Orthodox	0.00	0.00	0.00	0.00	0.00
Fraction of land area near navigable water	0.00	0.00	0.00	0.00	0.00
Interior density	0.00	0.00	0.00	0.00	0.00
Terms of trade ranking	0.00	0.00	0.00	0.00	0.00

References

BLAYDES, L. and D. A. LINZER (2008): "The political economy of women's support for fundamentalist Islam". In: *World Politics* 60 (4), pp. 576–609.

BREEN, R. (2000): "Why is support for extreme parties underestimated by surveys? A latent class analysis". In: *British Journal of Political Science* 30 (2), pp. 375–382.

BRIER, S. S. (1980): "Analysis of contingency tables under cluster sampling". In: *Biometrika* 67 (3), pp. 591–596.

BROCK, W. A. and S. N. DURLAUF (2001): "What have we learned from a decade of empirical research on growth? Growth empirics and reality". In: *World Bank Economic Review* 15 (2), pp. 229–272.

DOPPELHOFER, G. and M. WEEKS (2005): *Jointness of growth determinants*. Cambridge Working Papers in Economics.

DOPPELHOFER, G. and M. WEEKS (2009a): "Jointness of growth determinants". In: *Journal of Applied Econometrics* 24 (2), pp. 209–244.

— (2009b): "Jointness of growth determinants: Reply to comments by Rodney Strachan, Eduardo Ley and Mark FJ Steel". In: *Journal of Applied Econometrics* 24 (2), pp. 252–256.

DURLAUF, S., A. KOURTELLOS, and C. TAN (2008): "Are any growth theories robust?" In: *Economic Journal* 118 (527), pp. 329–346.

FERNÁNDEZ, C., E. LEY, and M. STEEL (2001): "Model uncertainty in cross–country growth regressions". In: *Journal of Applied Econometrics* 16 (5), pp. 563–576.

FORGY, E. W. (1965): "Cluster analysis of multivariate data: Efficiency versus interpretability of classifications". In: *Biometrics* 21 (3), pp. 768–769.

GEORGE, E. (1999): "Discussion of Bayesian model averaging and model search strategies by M.A. Clyde". In: *Bayesian Statistics 6*. Ed. by J. BERNARDO, J. BERGER, A. DAWID, and A. SMITH, pp. 157–177.

GLASS, D. H. (2013): "Confirmation measures of association rule interestingness". In: *Knowledge-Based Systems* 44, pp. 65–77.

HARTIGAN, J. A. and M. A. WONG (1979): "Algorithm AS 136: A k-means clustering algorithm". In: *Applied Statistics* 28 (1), pp. 100–108.

HOETING, J., D. MADIGAN, A. RAFTERY, and C. VOLINSKY (1999): "Bayesian model averaging: A tutorial". In: *Statistical Science* 14 (4), pp. 382–417.

HOFMARCHER, P., J. CRESPO CUARESMA, B. GRÜN, and K. HORNIK (2014): "Last night a shrinkage saved my life: Economic growth, model uncertainty and correlated regressors". In: *Journal of Forecasting*, Forthcoming.

KAUFMAN, L. and P. J. ROUSSEEUW (1990): *Finding groups in data*. New York, USA: John Wiley & Sons, Inc.

LAZARSFELD, P. F. (1950): "The logical and mathematical foundations of latent structure analysis". In: *Measurement and Prediction*. Ed. by S. A. STOUFFER. New York: John Wiley & Sons, pp. 362–412.

LEY, E. and M. STEEL (2009a): "On the effect of prior assumptions in Bayesian model averaging with applications to growth regressions". In: *Journal of Applied Econometrics* 24, pp. 651–674.

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

— (2009b): "On the effect of prior assumptions in Bayesian model averaging with applications to growth regression". In: *Journal of Applied Econometrics* 24 (4), pp. 651–674.

LIANG, F. et al. (2008): "Mixtures of g priors for Bayesian variable selection". In: *Journal of the American Statistical Association* 103, pp. 410–423.

LIVERANI, S., D. I. HASTIE, M. PAPATHOMAS, and S. RICHARDSON (2013): "PReMiuM: an R package for profile regression mixture models using Dirichlet processes". In: *arXiv preprint arXiv:1303.2836*.

MADIGAN, D. and J. YORK (1995): "Bayesian graphical models for discrete data". In: *International Statistical Review* 63 (2), pp. 215–232.

MASANJALA, W. and C. PAPAGEORGIOU (2008): "Rough and lonely road to prosperity: A reexamination of the sources of growth in Africa using Bayesian model averaging". In: *Journal of Applied Econometrics* 23, pp. 671–682.

MCLACHLAN, G. J. and D. PEEL (2000): *Finite mixture models*. Wiley.

MOLITOR, J., M. PAPATHOMAS, M. JERRETT, and S. RICHARDSON (2010): "Bayesian profile regression with an application to the national survey of children's health". In: *Biostatistics* 11 (3), pp. 484–498.

PAPASPILIOPOULOS, O. and G. ROBERTS (2008): "Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models". In: *Biometrika* 95 (1), pp. 169–186.

PAPASPILIOPOULOS, O. (2008): *A note on posterior sampling from Dirichlet mixture models*. Technical Report 8, CRISM Paper. University of Warwick. Centre for Research in Statistical Methodology.

Patterson, B. H., C. M. Dayton, and B. I. Graubard (2002): "Latent class analysis of complex sample survey data: application to dietary data". In: *Journal of the American Statistical Association* 97 (459), pp. 721–741.

Redner, R. A. and H. F. Walker (1984): "Mixture densities, maximum likelihood and the EM algorithm". In: *SIAM Review* 26 (2), pp. 195–239.

Sala-i-Martin, X., G. Doppelhofer, and R. I. Miller (2004): "Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach". In: *American Economic Review* 94 (4), pp. 813–835.

Schneider, U. and M. Wagner (2012): "Catching growth determinants with the adaptive LASSO". In: *German Economic Review* 13, pp. 71–85.

Sethuraman, J. (1994): "A constructive definition of Dirichlet priors". In: *Statistica Sinica* 4 (2), pp. 639–650.

Strachan, R. W. (2009): "Comment on 'Jointness of growth determinants' by Gernot Doppelhofer and Melvyn Weeks". In: *Journal of Applied Econometrics* 24 (2), pp. 245–247.

Tan, P.-N., V. Kumar, and J. Srivastava (2004): "Selecting the right objective measure for association analysis". In: *Information Systems* 29 (4), pp. 293–313.

Vermunt, J. K. and J. Magidson (2002): "Latent class cluster analysis". In: *Applied Latent Class Analysis*. Ed. by J. Hagenaars and A. McCutcheon. Cambridge, UK: Cambridge University Press, pp. 89–106.

III A Comprehensive Approach to Posterior Jointness Analysis in Bayesian Model Averaging Applications

Department of Economics Working Paper #193, WU Vienna

A Comprehensive Approach to Posterior Jointness Analysis in Bayesian Model Averaging Applications*

Jesus Crespo Cuaresma[a,b,c,d], Bettina Grün[e], Paul Hofmarcher[a], Stefan Humer[a], and Mathias Moser[a]

[a]Department of Economics, Vienna University of Economics and Business (WU)

[b]Austrian Institute of Economic Research (WIFO)

[c]International Institute of Applied System Analysis (IIASA)

[d]Wittgenstein Centre for Demography and Global Human Capital (WIC)

[e]Department of Applied Statistics, Johannes Kepler University Linz (JKU)

*Corresponding Author: Mathias Moser. Address: Welthandelsplatz 1, 1020 Vienna, Austria, Tel: +43(0)131336-5481, Email: matmoser@wu.ac.at.

Abstract

Posterior analysis in Bayesian model averaging (BMA) applications often includes the assessment of measures of *jointness* (joint inclusion) across covariates. We link the discussion of jointness measures in the econometric literature to the literature on association rules in data mining exercises. We analyze a group of alternative jointness measures that include those proposed in the BMA literature and several others put forward in the field of data mining. The way these measures address the joint exclusion of covariates appears particularly important in terms of the conclusions that can be drawn from them. Using a dataset of economic growth determinants, we assess how the measurement of jointness in BMA can affect inference about the structure of bivariate inclusion patterns across covariates.

JEL Classification: C11, C55, O40.
Keywords: Bayesian Model Averaging, Jointness, Robust Growth Determinants, Machine Learning, Association Rules.

1 Introduction

Addressing model uncertainty concerns in econometric applications through the use of Bayesian model averaging (BMA) is becoming a standard practice in empirical studies where no unique theoretical guidelines exist. One of such areas in economics where BMA has established itself as a useful tool of analysis is economic growth. A growing number of studies aims at identifying robust determinants of income per capita growth differences across countries without having to rely on specific theoretical frameworks (see for example Fernández, Ley, and Steel, 2001b; Brock and Durlauf, 2001; Sala-i-Martin, Doppelhofer, and Miller, 2004; Moral-Benito, 2012; Eicher, Helfman, and Lenkoski, 2012; Moral-Benito, 2014). In these studies, the robustness of individual covariates as determinants of income growth differences is routinely measured through *posterior inclusion probabilities* (PIP), i.e., the posterior probability covered by all models that contain that particular variable. This represents an average over a (possibly) large number of very different models.

Moving beyond the development of robustness measures based on individual covariates, some contributions in the literature aim at identifying particular structures in the posterior distribution of joint covariate inclusion. The literature tends to concentrate on the assessment of measures based on bivariate inclusion structures and uses the term *jointness* to refer to the dependence in the inclusion of groups (most often, pairs) of variables. Doppelhofer and Weeks (2005), Ley and Steel (2007, henceforth LS), Doppelhofer and Weeks (2009a, henceforth DW) and Strachan (2009) are the most relevant references dealing with measuring posterior inclusion dependence of regressors in economic growth applications. Using a different approach from these studies, Crespo Cuaresma et al. (2015) employ clustering methods to identify covariate inclusion patterns over the structure revealed by the posterior model probabilities of BMA exercises.

To quantify the association of covariate inclusion, the BMA literature has proposed several measures of *jointness*. These measures and the properties that define them have been studied in a strand of independent literature in the field of data mining, which aims at evaluating the quality of so-called association rules.

A common example for such a problem in data mining is finding sets of products that tend to be purchased together in a shopping basket. The development of rules that define the inclusion patterns existing between two or more items is conceptually very similar to finding jointness structures for a given set of covariates in the model space after the posterior model probabilities have been computed. However, the choice of measures to quantify these associations has generated a vivid discussion in the machine learning literature. Several studies provide comparisons of a large number of concepts and try to identify suitable measures through the kind of properties they fulfill (Geng and Hamilton, 2006; Glass, 2013). Besides these attempts to select measures based on objective criteria, some authors also adopt a subjective approach, in which the researcher tries to quantify a priori expectations (Tan, Kumar, and Srivastava, 2004). Some studies also show that many of the proposed measures produce similar rankings and therefore can be used exchangeably in many applications (Vaillant, Lenca, and Lallich, 2004; Tew, Giraud-Carrier, Tanner, and Burton, 2014).

The controversy around measuring jointness in BMA applications was born from the contributions by Ley and Steel (2009a), Strachan (2009) and Doppelhofer and Weeks (2009b). In their exchange of ideas the different authors raised concerns about how the different measures in the BMA context were defined. These discussions especially revolved around cases were several measures are undefined, or give contradictory results. Especially the question of whether the probability that two variables are not included in a model should influence the value of a jointness measure or not was debated vividly. We bring insights from the literature on association measures used in data mining and provide a thorough analysis of the differential characteristics of a larger set of jointness measures which nests those proposed hitherto in BMA applications. More specifically, we review properties of jointness measures, which have been proposed in the machine learning literature and focus on the property of null-invariance. We show that, while most measures in the BMA literature have this property, it is not favorable in BMA applications. Based on this discussion, we select a subset of measures that fulfill the afore discussed properties and use them to investigate jointness in the data set of Fernández, Ley, and Steel (2001b).

The paper is structured as follows. In section 2 we briefly review the standard implementation of jointness measures in the context of BMA. We present a short summary of relevant concepts from the literature on association rule analysis and how these are related to jointness in section 3. The empirical application based on the cross-country growth regression dataset in Fernández, Ley, and Steel (2001a), is carried out in section 4. Section 5 concludes and puts forward avenues of further research.

2 BMA and Jointness Measures: A Review

BMA methods aim at obtaining posterior distributions of the quantities of interest in a regression model which incorporate the uncertainty concerning model specification. Let our quantity of interest be related to the parameters of a linear regression model of the form

$$y|\alpha,\beta_j,\sigma \sim N(\alpha\iota + X_j\beta_j, \sigma^2 I), \tag{1}$$

where y is an $n \times 1$ vector whose elements are the observations of the dependent variable of interest, ι a vector of ones of the same length and the $n \times k$ matrix X_j is composed by the observations of k variables out of a total set of K covariates. Model uncertainty can be explicitly addressed by basing our inference on the parameters of interest on the posterior distribution

$$p(\alpha,\beta,\sigma|y) = \sum_{j=1}^{2^K} p(\alpha,\beta,\sigma|y,M_j)p(M_j|y), \tag{2}$$

where each specification-specific posterior distribution $p(\alpha,\beta,\sigma|y,M_j)$ is weighted by the corresponding posterior model probability $p(M_j|y)$. The posterior model probability is in turn proportional to the marginal likelihood of the model multiplied with the prior model probability,

$$p(M_j|y) \propto p(y|M_j)p(M_j). \tag{3}$$

It is standard in BMA applications to elicit improper non-informative priors on α and σ, so that $p(\alpha) \propto 1$ and $p(\sigma) \propto \sigma^{-1}$. A common choice for the prior of the slope coefficients β is Zellner's g–prior (Zellner, 1986),

$$p(\beta|M_j, \sigma) \sim N(0, \sigma^2(\frac{1}{g_0}X_j'X_j)^{-1}), \tag{4}$$

so that the prior variance matrix is scaled by the parameter g_0 and has the structure of the covariance matrix of the OLS estimator. Several fixed values for the g parameter have been proposed (see e.g. Foster and George, 1994; Fernández, Ley, and Steel, 2001a). To allow for more flexibility, hyperpriors on g have also been put forward in the literature by Liang et al. (2008), Feldkircher and Zeugner (2009), and Ley and Steel (2012).

For the prior model probabilities, a straightforward approach is to elicit a flat prior over all specifications entertained, so that $p(M_j) = 2^{-K}$ for all j. Given that this prior embodies a preference for models of size around $K/2$, Ley and Steel (2009b) argue for a binomial-beta prior on covariate inclusion, a setting which is able to achieve a very flexible prior structure over model size and includes a purely uninformative distribution over the number of included covariates.

Since analyzing the whole model space of 2^K models is often computationally infeasible, the relevant parts of the model space can be explored via Markov Chain Monte Carlo Model Composition (MC^3) methods (Madigan and York, 1995) in order to compute the relevant posterior distributions.

Among the many interesting features of the posterior over model specifications, the joint distribution of covariate inclusion constitutes the basis to create measures of jointness. Following Doppelhofer and Weeks (2009a), let model specifications be represented by a 0-1 vector of covariate inclusion profiles (as defined by the inclusion variables γ_k, $k = 1, \ldots, K$), so that

$$p(M_j|y) = p(\gamma_1 = c_1, \gamma_2 = c_2, ..., \gamma_K = c_K|y) , \tag{5}$$

where c_k is the binary variable representing the inclusion of covariate k in the model. Given these inclusion profiles, jointness quantifies to which degree two variables A and B tend to appear jointly across models $\big(p(A \cap B|y) \equiv p(AB|y)\big)$ as opposed to the posterior probability to appear without the respective other variable $\big(p(A \cap \bar{B}|y) \equiv p(A\bar{B}|y)$ and $p(\bar{A} \cap B|y) \equiv p(\bar{A}B|y)\big)$.

The comparison of these probabilities allows to consider two covariates as complements, substitutes or independent a posteriori, given their relative (common) appearance. The group of jointness measures that have been proposed in the BMA context uses these probabilities to generate a single statistic which allows a categorization of such pairs (or eventually, triplets) of variables. Positive values for these indicators typically refer to joint appearance (and therefore a certain degree of complementarity between them), while negative values are related to the fact that the two covariates act as substitutes in specifications. So far, five different measures of jointness have been proposed in the econometric literature dealing with BMA, which differ in the way they incorporate the different marginal and joint inclusion probabilities.

The earliest jointness measure in the BMA context is attributed to Doppelhofer and Weeks (2005), who propose to use

$$J = \ln\left(\frac{p(AB)}{p(A) \times p(B)}\right), \tag{6}$$

which resembles the logarithm of the posterior odds ratio. The use of posterior odds ratios as jointness indicator was criticized by Ley and Steel (2007), who note that the measure may be misleading for variables with high PIP and that the measure hardly allows for comparisons across different pairs of variables.

In a later study Doppelhofer and Weeks (2009a) propose a cross-product ratio of inclusion probabilities as another measure,

$$\mathscr{J} = \ln\left(\frac{p(AB) \times p(\bar{A}\bar{B})}{p(A\bar{B}) \times p(\bar{A}B)}\right). \tag{7}$$

In a reply Ley and Steel (2009a) are again not in favor of this approach, since the DW measure is not defined in cases where a variable has a PIP of 1 or 0. Instead LS highlight two alternative measures (Ley and Steel, 2007):

$$\mathscr{J}^* = \frac{p(AB)}{p(A) + p(B) - p(AB)} \tag{8}$$

$$\mathscr{J}' = \frac{p(AB)}{p(A\bar{B}) + p(\bar{A}B)} \quad . \tag{9}$$

While $\mathscr{J}'$ relates the joint inclusion to the probability of including either one of the two variables, $\mathscr{J}^*$ uses the probability of including either one but not both variables in the denominator.

Another measure was introduced by Strachan (2009), who proposes to only look at relevant variables in terms of PIP. This is accomplished by adapting DW's cross-product ratio in such a way, that it includes the marginal probabilities of both variables,

$$\tilde{\mathscr{J}} = p(A)p(B)\ln\left(\frac{p(AB)}{p(A\bar{B}) \times p(\bar{A}B)}\right). \tag{10}$$

A major discussion in the jointness literature also involves the treatment of $p(\bar{A} \cap \bar{B}|y) \equiv p(\bar{A}\bar{B}|y)$. This *exclusion margin* indicates to which extent both variables do not tend to appear together in specifications and therefore may be considered as representing a measure of (un)importance of bivariate jointness. While DW stress the importance of this aspect in the discussion (Doppelhofer and Weeks, 2009b), this property is not included in the jointness measures proposed by Strachan (2009) and Ley and Steel (2009a). The treatment of the information concerning joint exclusion of covariates constitutes a differential characteristic across association measures known as *null-invariance* in the data mining literature (Glass, 2013).

3 From Association Rules to Jointness Measures

The measures used in the literature on jointness of covariates in BMA analysis are often applied in data mining when describing association rules, although the linkages between the two strands of literature has not been explicitly acknowledged hitherto. Data mining is often concerned with the exploration of huge datasets of so-called transactions, which may for example each represent shopping baskets with different sets of items (products). Association analysis aims at finding patterns in these data structures to learn about consumer behavior and the interrelation across purchased items. The major tool used are association rules of the form $A \longrightarrow B$ (if A is included in the basket, B tends to be included), where A and B can include either individual items or disjoint itemsets.

For a large number of items, the count of rules can potentially grow very large. The number of itemsets is $2^K - 1$ for K items (variables) which implies $3^K - 2^{K+1} + 1$ possible association rules (excluding empty sets) between itemsets of all sizes. Therefore association rules are routinely *mined* to only include such rules which are “interesting” for the application. This refers on the one hand to associations which are frequent, as measured by the *support*. On the other hand, rules should be strong as measured by the *confidence*, which relates the occurrence of a pattern to the number of counterexamples in the data.

The most common strategy to extract such rules is the *apriori algorithm* (Aggarwal and Yu, 1998; Hahsler, Grün, and Hornik, 2005), which reduces the complexity of the problem by reasoning that all item subsets of a frequent itemset must also be frequent and vice versa. This approach is also related to *support-based pruning* and has been applied by a large number of studies in the data mining literature (Tan, Kumar, and Srivastava, 2004).

In addition to support and confidence — which are relevant to achieve computational feasibility — the interestingness of these rules can be quantified using several measures. Similar to the jointness literature, a number of such indicators has been proposed in the data mining context. Recent surveys in this field collect as many as 40 different measures and try to provide a structural overview of the alternative measures available (Glass, 2013; Geng and Hamilton, 2006; Tan, Kumar, and Srivastava, 2004).[1] Some of these measures resemble the ones proposed in the BMA jointness literature. The first jointness measure of Doppelhofer and Weeks (2005) is equivalent to the Log-Ratio or equivalently, the log of the Interest (Lift) measure (Geng and Hamilton, 2006). Ley and Steel (2007)'s $\mathscr{J}^*$ is identical to the long-used Jaccard index and their $\mathscr{J}'$ measure is a derivation thereof. As another alternative, Strachan (2009) introduces a measure ($\tilde{\mathscr{J}}$) that has been known as the Two-Way support (Geng and Hamilton, 2006). Finally, the statistic proposed by Doppelhofer and Weeks (2009a) has also been known as the Odds-Ratio in the field of data mining (Tew, Giraud-Carrier, Tanner, and Burton, 2014).

Another similarity between the two strands of literature is the debate on which measure is the most appropriate for a given application. Tan, Kumar, and Srivastava (2004) propose the use of subjective measures, which depend on the user to rank a small predefined set of associations for a specific application. Using this approach, an appropriate measure can be selected, which reproduces the user's ranking. More generally, objective measures have been analyzed based on certain properties they are expected to fulfill. Ley and Steel (2007) propose four properties, that BMA jointness measures should fulfill: An indicator should be *interpretable* in such a sense, that it has a "clear intuitive meaning" and is well *calibrated* against a clearly defined scale. Furthermore, the property of *extreme jointness* states that a measure should reach its maximum when both variables always appear together. Also, a measure should always be defined (*definition*) when either variable is included with positive probability. In contrast, the association analysis literature tends to impose a larger number of characteristics that are expected to be fulfilled. In the following section we shortly review the

[1] A detailed overview of interestingness measures can be found in Appendix A.1.

most important properties proposed in the literature, discuss their implications for jointness and relate them to the measures in the BMA literature where applicable.

3.1 Desirable Properties of Interestingness Measures for BMA

The properties that have been independently discussed in the BMA context, partly reflect those which are used in the machine learning (ML) literature. Finding a suitable measure clearly depends on the properties that are required for a certain application. For example, while machine learning problems are often concerned with positive association, BMA results additionally need to reflect negative association in the form of variable substitutes. Furthermore the type of assertion that is being made, needs to be considered and especially the question whether two variables are considered exchangeable, so that $A \to B \equiv B \to A$. In the following we select four properties, which can be considered crucially relevant for jointness based on the insights from the BMA and ML discussions.

Interestingness vs. Confirmation A confirmation measure is an interestingness measure m that, for a given threshold τ, satisfies that

$$\begin{aligned} m(A,B) > \tau &\iff Pr(A|B) > Pr(A), \\ m(A,B) = \tau &\iff Pr(A|B) = Pr(A), \\ m(A,B) < \tau &\iff Pr(A|B) < Pr(A). \end{aligned}$$

The indicator is thus anchored at some threshold value τ that defines statistical independence (e.g. 0 for DW's Odds-Ratio). For the case of jointness indicators discussed in the BMA literature, this property is implicitly given for all proposed measures and seems to be a reasonable characteristic to be fulfilled. We therefore limit our empirical analysis to the set of confirmation measures that have been proposed in the data mining context (Glass, 2013).

Symmetry Implication rules that imply that the proposition $A \to B$ differs from $B \to A$ are asymmetric. Since jointness measures are interested in measuring the common appearance (or lack thereof) of two explanatory variables, a suitable measure should therefore be symmetric with regard to the ordering of variables. The assertion that certain covariates are "substitutes" or "complements" implies thus commutativity.[2] All jointness measures proposed in the BMA literature fulfill this requirement. A number of measures from the data mining literature are however asymmetric and thus excluded from the empirical analysis carried out in the following sections.[3]

Monotonicity and Maximality The range of interestingness measures should be bounded and monotonically increasing between the two extreme cases. This property is partly reflected in the more restrictive Piatesky-Shapiro conditions: $m = 0$ if $p(AB) = 0$, m monotonically increases with $p(AB)$ and m monotonically decreases with $p(A)$ or $p(B)$ (Piatetsky-Shapiro, 1991; Tan, Kumar, and Srivastava, 2004). Maximality corresponds to *extreme jointness*, the property introduced by Ley and Steel (2007) in the jointness literature. This property defines that a measure should reach its maximum when both variables always appear together.

Table III.1: Interestingness Measures for Jointness

	Value	Range	k
Non null-invariant			
Collective Strength	$\ln\left[\frac{\Pr(AB)+\Pr(\bar{A}\bar{B})}{\Pr(A)\Pr(B)+\Pr(\bar{A})\Pr(\bar{B})} \times \frac{1-\Pr(A)\Pr(B)-\Pr(\bar{A})\Pr(\bar{B})}{1-\Pr(AB)-\Pr(\bar{A}\bar{B})}\right]$	$]-\infty,\infty[$	
Relative Risk	$\ln\left[\frac{\Pr(B\|A)}{\Pr(B\|\bar{A})}\right]$	$]-\infty,\infty[$	
Yule's Q	$\frac{\Pr(AB)\Pr(\bar{A}\bar{B})-\Pr(A\bar{B})\Pr(\bar{A}B)}{\Pr(AB)\Pr(\bar{A}\bar{B})+\Pr(A\bar{B})\Pr(\bar{A}B)}$	$[-1,1]$	
Normalized Difference	$\Pr(B\|A)-\Pr(B\|\bar{A})$	$[-1,1]$	
ϕ-Coefficient	$\frac{\Pr(AB)-\Pr(A)\Pr(B)}{\sqrt{\Pr(A)\Pr(B)\Pr(\bar{A})\Pr(\bar{B})}}$	$[-1,1]$	

[2] This property is often called commutative symmetry (Glass, 2013).

[3] Tan, Kumar, and Srivastava (2004) suggest to symmetrize measures by using $\max(p(A|B), p(B|A))$.

Table III.1: *(continued)*

	Value	Range	k
Null-invariant			
Two-Way Support	$\Pr(AB)\ln\left[\frac{\Pr(AB)}{\Pr(A)\Pr(B)}\right]$	$[0,\infty[$	
AllConf	$\min(\Pr(B\vert A),\Pr(A\vert B))$	$[0,1]$	$-\infty$
Coherence	$(\Pr(A\vert B)^{-1}+\Pr(B\vert A)^{-1}-1)^{-1}$	$[0,1]$	-1
Cosine	$\frac{\Pr(AB)}{\sqrt{\Pr(A)\Pr(B)}}$	$[0,1]$	0
Kulczynski	$(\Pr(A\vert B)+\Pr(B\vert A))/2$	$[0,1]$	1
MaxConf	$\max(\Pr(B\vert A),\Pr(A\vert B))$	$[0,1]$	$+\infty$

Null-invariance Measures that are null-invariant ignore so-called null transactions, in which neither A nor B occur. Whether null-invariance is a desirable property for an association measure depends on the nature of the empirical application under scrutiny. For the case of jointness measures in BMA analysis, different views concerning the desirability of null-invariance have been voiced in the literature. Doppelhofer and Weeks (2009b) criticize null-invariance, since "*[...] jointness can manifest itself in both the inclusion and exclusion margin of the joint posterior distribution*". In contrast, Strachan (2009) and Ley and Steel (2009a) stress the effect of low-probability models, which are represented only sparsely in the model matrix and which would be "uninteresting" for most non null-invariant measures where the common exclusion probability is respected.

3.2 Confirmation Measures for Jointness Analysis

Based on the extensive surveys of interestingness measures in the data mining literature (Tan, Kumar, and Srivastava, 2004; Geng and Hamilton, 2006; Glass, 2013; Tew, Giraud-Carrier, Tanner, and Burton, 2014), we select a subset of indicators which fulfill the properties put forward above and that are therefore potentially suitable to analyze jointness in BMA applications. More specifically, all interestingness measures analyzed here are (a) confirmation measures, (b) symmetric around a threshold that implies inclusion independence and (c) reach their maxima when both variables are highly complementary. We group these measures by whether they fulfill null-invariance or not. Table III.1 provides an overview of these indicators.[4]

Table III.2: Comparison of Interestingness Measures: Independency

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Probabilities								
$p(A)$	0.10	0.50	0.90	0.70	0.50	0.60	0.50	0.90
$p(B)$	0.10	0.10	0.10	0.20	0.50	0.40	0.90	0.90
$p(A\|B)$	0.10	0.50	0.90	0.70	0.50	0.60	0.50	0.90
$p(B\|A)$	0.10	0.10	0.10	0.20	0.50	0.40	0.90	0.90
$p(AB)$	0.01	0.05	0.09	0.14	0.25	0.24	0.45	0.81
$p(A\bar{B})$	0.09	0.05	0.01	0.06	0.25	0.16	0.45	0.09
$p(\bar{A}B)$	0.09	0.45	0.81	0.56	0.25	0.36	0.05	0.09
$p(\bar{A}\bar{B})$	0.81	0.45	0.09	0.24	0.25	0.24	0.05	0.01
Non null-invariant								
Collective Strength	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Relative Risk	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Yule's Q	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Normalized Difference	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ϕ-Coefficient	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Null-invariant								
Two-Way Support	0.00	0.01	0.03	0.05	0.12	0.12	0.30	0.73
AllConf	0.10	0.10	0.10	0.20	0.50	0.40	0.50	0.90
Coherence	0.05	0.09	0.10	0.18	0.33	0.32	0.47	0.82
Cosine	0.10	0.22	0.30	0.37	0.50	0.49	0.67	0.90
Kulczynski	0.10	0.30	0.50	0.45	0.50	0.50	0.70	0.90
MaxConf	0.10	0.50	0.90	0.70	0.50	0.60	0.90	0.90

Note: Independency defined as $p(AB) = p(A)p(B)$

This choice of measures subsumes all the indicators used in the BMA jointness literature, while we adhere to the naming conventions used in data mining. We replace the *Odds Ratio* with its projection on the $[-1, 1]$ interval, which is known as *Yule's Q*.[5] The *Collective Strength* measure was introduced by Aggarwal and Yu (1998) and compares the violation rate of an itemset to its expected value

[4] A full list of the interestingness measures used in the literature and that have been considered to select the particular indicators considered here is presented in Appendix A.1.

[5] The *Odds Ratio*, *Yule's Q*, and the log transformation of *Yule's Q*, *Yule's Y*, produce the same rankings of association rules and are therefore considered equivalent (Tew, Giraud-Carrier, Tanner, and Burton, 2014).

under statistical independence. It is defined between zero and ∞, where a value of unity signals statistical independence, a lower value indicates substitutability and a larger value complementarity. We use the log transformed measure which is defined around 0 as the independence threshold. *Relative Risk* is a measure widely used in case studies, where an exposed group (numerator) is compared to a non-exposed group (denominator). Log-transforming this measure, we define independence at a value of zero and substitutes (complements) below (above) this value. *Normalized Difference* is simply the difference between two probabilities and hence defined in $[-1, 1]$. The *ϕ-Coefficient* is basically a correlation measure and closely related to the χ^2 statistic, bounded in the interval $[-1, 1]$.

Table III.3: Comparison of Interestingness Measures: Complementarity

	Substitutes					Complements				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Probabilities										
$p(A)$	0.10	0.90	0.70	0.50	0.70	0.10	0.50	0.90	0.30	0.40
$p(B)$	0.10	0.10	0.20	0.50	0.30	0.10	0.40	0.10	0.30	0.20
$\boldsymbol{p(A\vert B)}$	**0.01**	**0.09**	**0.07**	**0.05**	**0.07**	**0.90**	**1.00**	**1.00**	**1.00**	**1.00**
$p(B\vert A)$	0.01	0.01	0.02	0.05	0.03	0.90	0.80	0.11	1.00	0.50
$p(AB)$	0.00	0.01	0.01	0.02	0.02	0.09	0.40	0.10	0.30	0.20
$p(A\bar{B})$	0.10	0.09	0.19	0.48	0.28	0.01	0.00	0.00	0.00	0.00
$p(\bar{A}B)$	0.10	0.89	0.69	0.48	0.68	0.01	0.10	0.80	0.00	0.20
$p(\bar{A}\bar{B})$	0.80	0.01	0.11	0.03	0.02	0.89	0.50	0.10	0.70	0.60
Non null-invariant										
Collective Strength	-0.12	-2.48	-1.43	-2.94	-2.80	2.38	2.20	0.13	Inf	1.15
Relative Risk	-2.40	-4.51	-3.43	-2.94	-3.43	4.39	Inf	Inf	Inf	Inf
Yule's Q	-0.85	-1.00	-0.98	-0.99	-1.00	1.00	1.00	1.00	1.00	1.00
Normalized Difference	-0.10	-0.90	-0.60	-0.90	-0.90	0.89	0.80	0.11	1.00	0.50
ϕ-Coefficient	-0.10	-0.90	-0.69	-0.90	-0.90	0.89	0.82	0.11	1.00	0.61
Null-invariant										
Two-Way Support	0.00	0.00	0.00	0.00	0.00	0.08	0.36	0.03	0.30	0.14
AllConf	0.01	0.01	0.02	0.05	0.03	0.90	0.80	0.11	1.00	0.50
Coherence	0.01	0.01	0.02	0.03	0.02	0.82	0.80	0.11	1.00	0.50
Cosine	0.10	0.30	0.37	0.50	0.46	0.10	0.45	0.30	0.30	0.28
Kulczynski	0.01	0.05	0.04	0.05	0.05	0.90	0.90	0.56	1.00	0.75
MaxConf	0.01	0.09	0.07	0.05	0.07	0.90	1.00	1.00	1.00	1.00

Notes: Substitutes defined as $p(AB) = 0.1 \times p(A)p(B)$
Complements defined as $p(AB) = \min(1, 9 \times p(A)p(B))$

As described by Wu, Chen, and Han (2010), five common null-invariant measures can be represented by the generalized mean of the two conditional probabilities $p(A|B)$ an $p(B|A)$ with parameter k. This representation nests the

AllConf measure (*Confidence*), *Coherence* (*Jaccard*, Ley and Steel (2007)), *Cosine* (similar to Doppelhofer and Weeks (2005)), *Kulczynski* and *MaxConf*, which we employ as examples of alternative null-invariant measures. These measures present themselves as differently weighted means, so that *Coherence* describes the harmonic mean, *Cosine* the geometric and *Kulczynski* the arithmetic mean of the two probabilities (Wu, Chen, and Han, 2010). Additionally, we include Strachan (2009)'s measure, which is known as *Two-Way-Support* in the data mining literature. This measure is a combination of two basic interestingness indicators and can be reproduced by scaling the log *Lift* with the *Support* of a rule (Yao and Zhong, 1999).

Based on the reasoning by Doppelhofer and Weeks (2009b) concerning the fact that a sensible jointness measure should equal zero for independence, we provide a synthetic example for different measures in Table III.2. The eight columns provide scenarios where A and B are statistically independent, so that $p(AB) = p(A)p(B)$, but differ in the values for $P(A)$ and $P(B)$. Based on this assumption, we calculate the different jointness measures for each scenario. Column 5 depicts the scenario described in Doppelhofer and Weeks (2009b), which is the special case of $p(A) = p(B) = 0.5$. While Doppelhofer and Weeks (2009b) only argued based on an example with equal posterior inclusion probability across covariates ($p(A) = p(B)$), we also consider differing individual posterior probabilities of inclusion in Table III.2.

As expected, the non null-invariant measures regard all eight scenarios presented in Table III.2 as independent, since they explicitly take care of the exclusion margin $p(\bar{A}\bar{B})$. In contrast, the null-invariant measures only agree in terms of the absolute size of the indicator for cases where the posterior inclusion of both variables is equally likely (see columns 1, 5 and 8). Even in these scenarios, the measures do not provide a clear independence threshold. The value defining independence varies with $p(A)$ and $p(B)$, the posterior inclusion probabilities of both variables. *AllConf* and *MaxConf*, which only consider the minimum or maximum of the two conditional probabilities, $P(A|B)$ and $P(B|A)$, are exceptions. It has been argued that null-invariant measures are hardly able to correctly quantify positive and negative association, since they do not account for varying sizes of the exclusion margin (Glass, 2013). This can be considered less of a problem for certain applications in data mining, where only a small set of positive relationships out of a large set of transactions containing many zeros is of interest. However for the application to jointness the researcher mostly faces “balanced” datasets, where variable inclusion and exclusion are both similarly frequent.

Table III.3 provides insights to the reaction of these different measures to substitutes and complements. We choose substitution relations in joint inclusion (columns 1 to 5) in such a way that the probability of common occurrence is one tenth of the independence threshold, or $P(AB) = 0.1 \times p(A)p(B)$. We find that non null-invariant measures regard all these scenarios as substitutes, leading to jointness values below zero. *Yule’s Q* is in this regard very consistent, as it finds values close to its absolute minimum of -1 for all five cases. As a counterexample, *Normalized Difference* and the *ϕ-Coefficient* agree in the scenarios entertained where the exclusion margin is low (columns 2, 4 and 5) by regarding the pair as highly substitutes, but gain in value (towards independence) when this margin increases (columns 1 and 3). For the extreme case of $p(A) = p(B) = 0.1$ this results in a large exclusion margin of 0.8, while at the same time both indicators approach zero (-0.1).

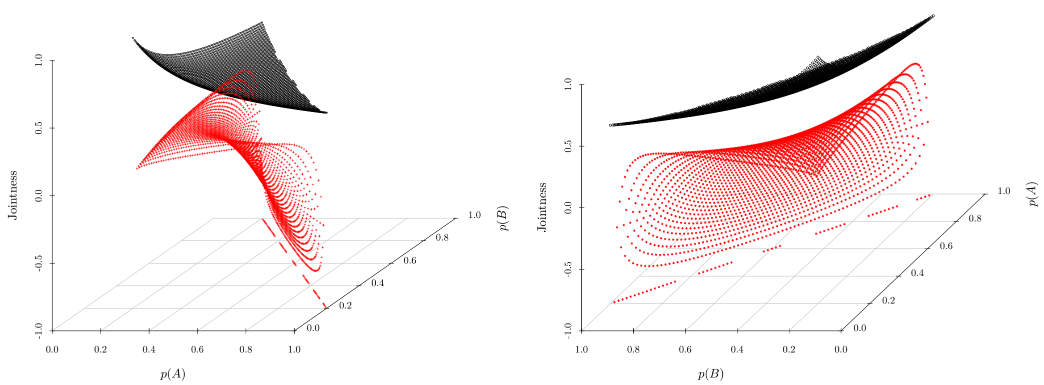


Figure III.1: Cosine (black) and Yule’s Q (red) for $p(AB) = 0.2$

While the independence threshold is not uniquely defined for null-invariant measures, most of these present very low values, close to their common lower bound of zero. *Two-way-support* is exactly at its lower bound of zero for all our scenarios involving substitutes. Still, it is hard to gauge substitutability with this measure, since its independence level is also very close to zero $(0, 0.03, 0.05, 0.12, 0.07)$ for most of the five scenarios. Another example is *Cosine*, whose value is in every scenario identical to the independence threshold.

Columns 6–10 in Table III.3 present five examples of bivariate complements, where p(AB) is set to be a multiple of the independence threshold, $\Pr(AB) = 9 \times \Pr(A)\Pr(B)$. Our findings for non null-invariant measures suggest that we correctly identify complements in all of these five scenarios. For $p(A|B) = 1$ the *Relative Risk* measure is infinite by construction. As before, the *Normalized Difference* and the *ϕ-Coefficient* are more ambiguous in their assertion of the complementarity relationships between pairs. Especially for cases which have a high level of mutual exclusion (column 8) — $p(\bar{A}B) = 0.8$ in this case — both measures shift in value towards independence. A similar case can be made for *Collective Strength* and for the scenario depicted in column 10.

Identifying complements via null-invariant measures seems to be a harder task, since we need to interpret these values relative to the (non unique) independence point. *MaxConf* always represents $p(A|B)$, which was chosen to be large, and therefore also ranges at its upper border of unity. *AllConf, Coherence* and *Kulczynski* all represent similar patterns to *Normalized Difference* and the *ϕ-Coefficient*, that is, high values when mutual exclusion is low and a drop in the indicator level as soon as either of these probabilities rise.

The effect of extreme values for the exclusion margin can also be grasped by assessing the jointness measures graphically. Figure III.1 depicts the sensitivity of two measures, Cosine (null-invariant) and Yule's Q (Odds Ratio, non null-invariant) for a given level of joint occurrence $p(AB) = 0.2$ and varying values of $p(A)$ and $p(B)$ (X-, Y-axes). The Cosine measure is represented by a slightly convex plane, varying between 0.33 and 1, whereas mean and median lie close to 0.45. The maximum of the measure is found at the minimum values of $p(A)$

and $p(B)$, which correspond to the joint probability of 0.2. The measure then decreases towards the extreme values $\{1, 0.2\}$ and $\{0.2, 1\}$. In both cases the exclusion margin $p(\bar{A}\bar{B})$ is zero, however the probabilities of $p(A\bar{B})$ and $p(\bar{A}B)$ vary and cause the measure to react. The non null-invariant measure, Yule's Q, varies in a stronger fashion in an interval between 0.2 and its absolute minimum of -1. We find that this indicator is rather stable for individual inclusion values up to 0.4, which is twice the value of $p(AB)$, and for cases where $p(A) \gg p(B)$ or vice versa. For cases where the inclusion probabilities of both variables become large, the measure drops sharply indicating substitutability instead of complementarity. In our opinion this is a desirable indication. If the joint probability of occurrence if far below the marginal inclusion probabilities of the two variable, a measure should not classify them as complements.

4 Jointness of Economic Growth Determinants Revisited

Table III.1: Results of the BMA routine for the FLS data set

	PIP	Post Mean	Post SD
GDP60	1.000	-0.016	0.003
Confucian	0.993	0.060	0.016
LifeExp	0.971	0.001	0.000
EquipInv	0.907	0.124	0.062
SubSahara	0.885	-0.016	0.008
Mining	0.815	0.031	0.020
Hindu	0.717	-0.050	0.042
NequipInv	0.696	0.034	0.029
RuleofLaw	0.666	0.008	0.007
LabForce	0.655	0.000	0.000
EcoOrg	0.614	0.001	0.001
Muslim	0.598	0.007	0.008
BlMktPm	0.566	-0.004	0.004
LatAmerica	0.563	-0.006	0.007
EthnoL	0.561	0.006	0.007
Protestants	0.559	-0.005	0.006
HighEnroll	0.554	-0.049	0.055
PrScEnroll	0.495	0.008	0.011
CivlLib	0.430	-0.001	0.002
Spanish	0.427	0.004	0.006

In our empirical application we apply alternative jointness measures to the dataset used in Fernández, Ley, and Steel (2001b, henceforth FLS), which includes information on income per capita growth and 41 potential determinants of economic growth differences for 72 countries.[6] In a first step, we apply BMA methods to obtain the posterior inclusion probabilities for all variables, as well as the mean and standard deviation of the posterior distribution of the parameters associated with each covariate. For this application we employ a hyper-g prior over the parameters (Liang et al., 2008) and a Binomial-Beta model prior following Ley and Steel (2009b). The BMA results are obtained using five million Markov Chain Monte Carlo iterations over the model space, where the first two million are disregarded as burn-in. Out of the three million visited models, approximately two thirds are unique, with a mean number of 19.8 included explanatory variables. Table III.1 presents the posterior inclusion probabilities for the top 20 variables, together with the mean and standard deviation of the posterior distribution of their respective parameters. The BMA results confirm the robustness of several economic growth determinants such as *GDP60, Confucian, LifeExp* or *EquipInv*, which have a PIP above 0.9.

Using the top $10,000$ unique models weighted by posterior model probabilities, we construct the binary matrix of model profiles, defined by the inclusion binary variables, γ_j. Since the top $10,000$ models have been included approximately $130,000$ times in the three million MCMC draws, this matrix has dimensions $130,000 \times 41$, where each cell describes whether covariate 1–41 is included (1) or not (0) in a given model. From this model profile matrix we can construct rules based on joint inclusion of variables.[7] A common further step in association analysis involves support-based pruning, where the rules are reduced given a minimum and/or maximum value for support, i.e., the frequency of a rule, and confidence, which measures the occurrences of a rule relative to the number of counterexamples. Pruning with respect to support eliminates infrequent rules, which only appear very rarely in the data. Table III.2 shows the number of

[6]See Appendix B.1 for a description of the variables, as well as some descriptive statistics.

[7]We concentrate on bivariate jointness. A straightforward extension would be to analyze jointness based on triplets, for which tools such as the *apriori algorithm* can be used.

bivariate rules found for the FLS data set given different thresholds for support and confidence. We find a total of 1,640 bivariate rules if we impose no restrictions, which is no lower bounds for support or confidence. Twelve rules satisfy the most rigorous pruning, implied by only keeping highly frequent pairs which have a support value larger than 0.9. Following the association rules analysis literature, we use a low level of support pruning (0.1), so that we end up with a set of 582 distinct rules to analyze. In addition, we prune rules with extremely high support, namely the twelve resulting from a support level of 0.9. These may be of interest in the data mining context, but do not provide enough variation to analyze whether the covariates involved are substitutes or complements in the jointness context. The rules selected involve 29 of the 41 covariates.

Table III.2: Number of rules by minimum confidence and support

Support/Confidence	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	1640	1002	750	611	530	467	403	347	291	217
0.1	304	304	264	227	215	201	168	145	128	101
0.2	154	154	154	140	133	124	99	83	75	62
0.3	110	110	110	110	106	102	81	66	59	48
0.4	88	88	88	88	88	86	70	60	55	44
0.5	74	74	74	74	74	74	61	53	48	41
0.6	38	38	38	38	38	38	38	34	30	25
0.7	28	28	28	28	28	28	28	28	25	20
0.8	20	20	20	20	20	20	20	20	20	16
0.9	12	12	12	12	12	12	12	12	12	12

For the overall set of identified joint variable inclusions and the pruned subset we obtain the interestingness measures described in Table III.1 and calculate Spearman rank correlations, to quantify the concordance of the orderings implied by the different measures. Table III.3 presents the results from this exercise. In the lower triangle, the results for the total of 1,640 rules are presented, while in the upper-right triangle we show the correlations for the pruned subset. The rank correlations within the group of non null-invariant measures imply highly congruent rankings by these indicators. These measures provide rankings that are only loosely correlated with those delivered by their null-invariant counterparts. Comparing rank correlations for the full and pruned sets of associations, we find that the agreement increases above the diagonal in Table III.3, which indicates that the exclusion of extreme cases causes the rankings implied by the measures to converge.

Within the set of null-invariant measures we find significantly less within-group correlation. While the measures *Coherence*, *Cosine* and *Kulczynski* tend to agree in terms of ranking bivariate inclusion relationships, this is not the case for *AllConf* and *MaxConf*. Since these two measures actually represent minima and maxima functions over the conditional inclusion probabilities $p(A|B)$ and $p(B|A)$, they frequently take extreme values at 0 or 1 and therefore produce rankings with a large number of ties around these values. Similarly, the rank correlations for the pruned set of bivariate inclusions are higher than for the full set.

Given these results, we restrict our subsequent analysis to four distinct measures. On the one hand, we select the *Yule's Q* (an Odds-Ratio transformation) and the *ϕ-Coefficient*, which have been shown to react differently to the exclusion margin in the simulations. On the other hand, we concentrate on the null-invariant measures *Cosine* (Jaccard) and *Kulczynski*.

Table III.3: Spearman Correlation of Ranked Measures

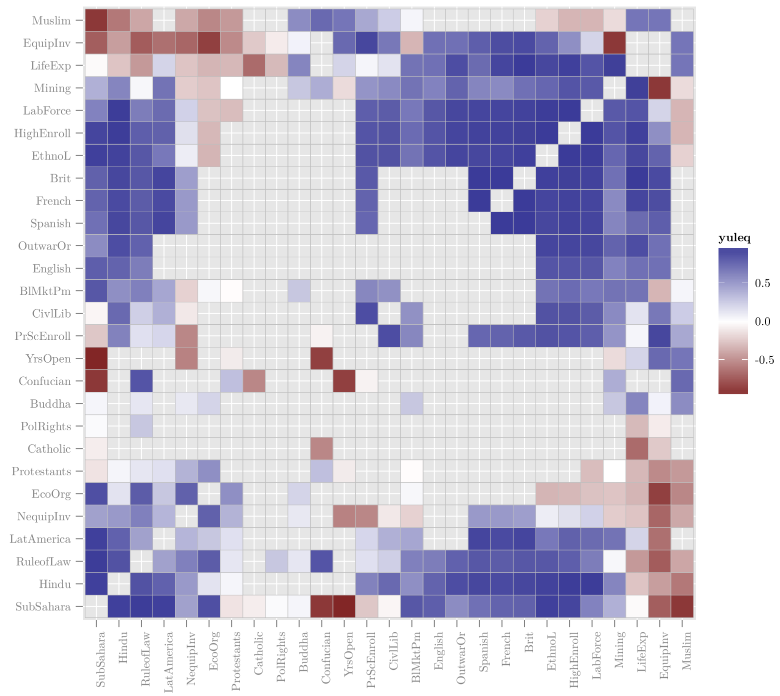
	Non null-invariant					Null-invariant					
	1	2	3	4	5	6	7	8	9	10	11
Non null-invariant											
1 Collective Strength		0.89	0.91	0.93	0.99	0.19	0.38	0.39	0.36	0.28	0.14
2 Relative Risk	0.83		0.91	0.86	0.90	0.03	0.24	0.24	0.22	0.19	0.10
3 Yule's Q	0.81	0.94		0.87	0.93	0.20	0.34	0.36	0.38	0.38	0.34
4 Normalized Difference	0.95	0.86	0.87		0.94	0.18	0.33	0.35	0.34	0.30	0.19
5 ϕ-Coefficient	0.99	0.82	0.86	0.94		0.26	0.43	0.44	0.43	0.37	0.26
Null-invariant											
6 Two-Way Support	-0.15	-0.20	-0.11	-0.01	-0.04		0.82	0.87	0.97	0.92	0.78
7 AllConf	0.04	-0.15	-0.08	0.13	0.13	0.86		0.99	0.86	0.64	0.38
8 Coherence	0.03	-0.12	-0.03	0.13	0.14	0.92	0.98		0.91	0.71	0.47
9 Cosine	0.05	-0.06	0.04	0.16	0.17	0.94	0.94	0.99		0.93	0.76
10 Kulczynski	-0.02	-0.01	0.11	0.13	0.13	0.93	0.81	0.91	0.95		0.93
11 MaxConf	-0.22	0.12	0.25	-0.03	-0.06	0.54	0.17	0.32	0.43	0.64	

Notes: Lower-left triangle: Rank correlations for all 1640 rules
Upper-right triangle: Rank correlations for 568 pruned rules (support min 0.1/max 0.9)

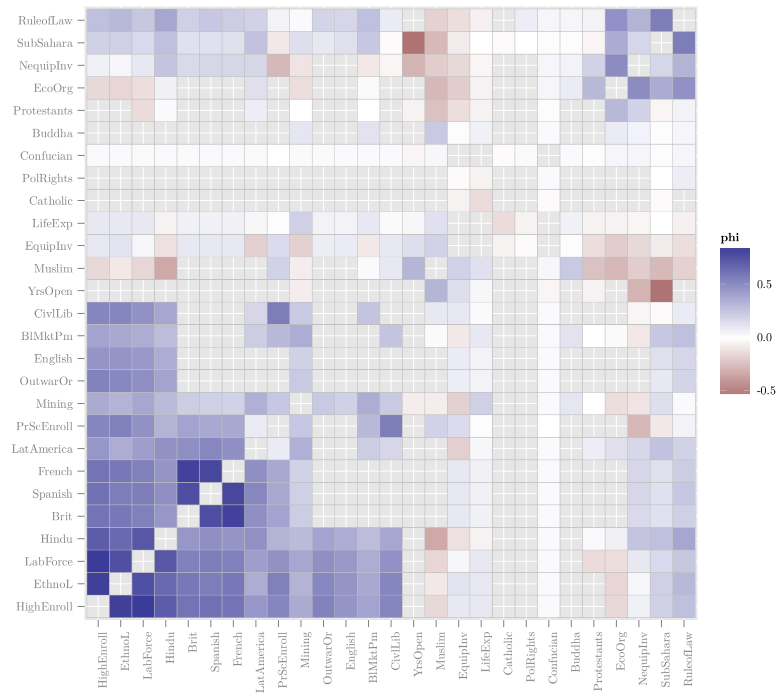
Figures III.1 and III.2 represent graphically the degree of jointness implied by these four measures. The pairs of variables in these figures are ordered in such a way that high jointness patterns can be found along the diagonal of the matrix depicted in them (Hahsler, Hornik, and Buchta, 2008; Tan, Kumar, and Srivastava, 2004). For Yule's Q in Figure III.1a, we find a number of strong complementary relationships, represented by the blue shaded tiles. These clusters are primarily composed of the colonial dummies (*Brit, English, Spanish* and *French*) as well as geographical factors (*Latin America, SubSahara, EthnoL*). We also find a number of complements in the set of economic system-related variables, *OutwardOrientation, RuleOfLaw, LabForce* and *BlackMarketPremium*. In contrast, Yule's Q unveil very few substitutability relationships between pairs of variables. These are mainly related to religious variables (*Muslim, Confucian*) and their relation to the Sub-Saharan African dummy.

The ϕ-Coefficient (see Figure III.1b) presents a similar picture with respect to colonial variables and *RuleOfLaw* or *OutwardOrientation*. However, it highlights even less substitutability relationships than Yule's Q besides the connection between *SubSahara* and *YrsOpen*.

The two null-invariant measures in Figure III.2 show very similar patterns for complementarity of colonial and geographical variables. However, they tend to emphasize bivariate relationships of variables that have very high PIPs in the BMA exercise. For these covariates there are hardly any models where they do not appear together, so that these types of measures consider them to be strongly related in a complementarity sense. This applies to all the variables that present very high PIPs: *GDP60, Confucian, EquipInv* or *LifeExp*. The Cosine and the Kulczynski measures also find a number of substitutes, with *YearsOpenEconomy* and *NequipInv* being an example of these.



(a) Yule's Q (Odds Ratio)



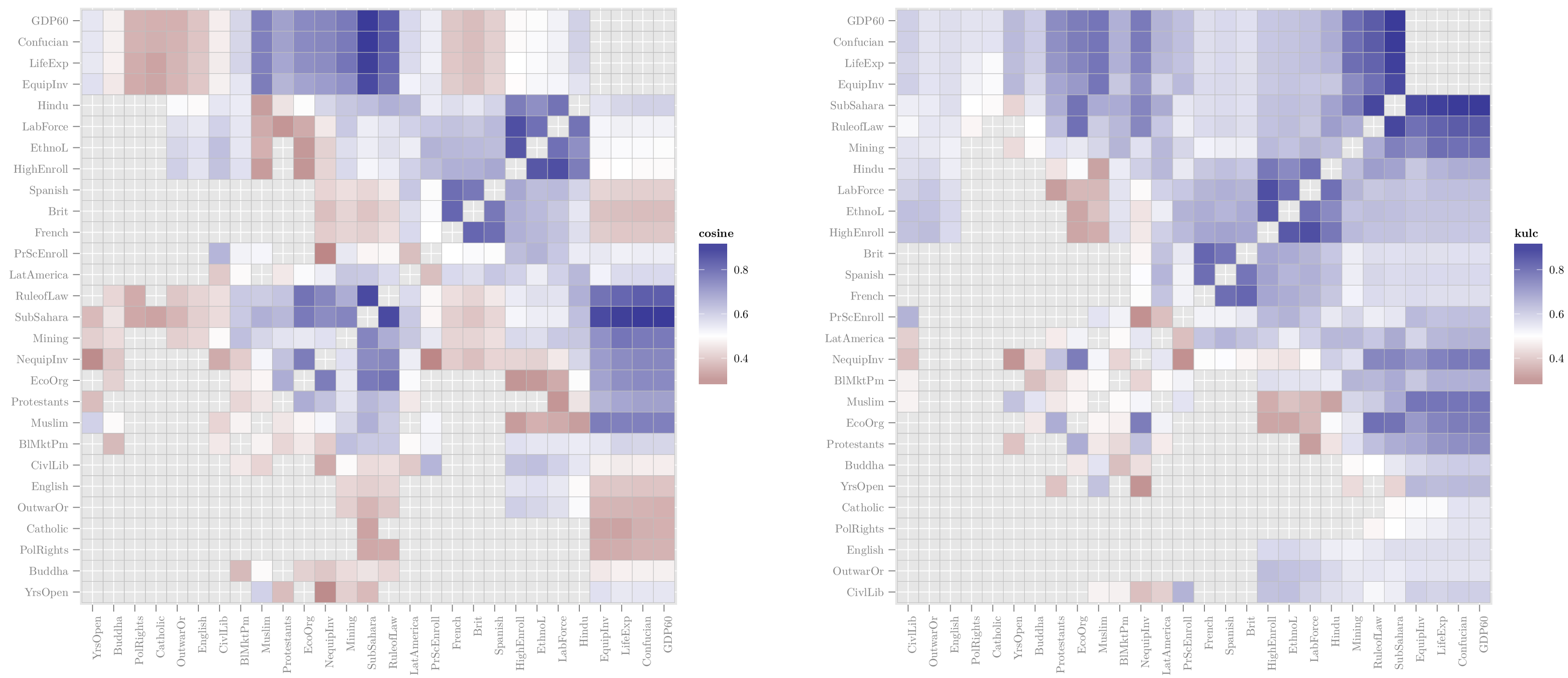
(b) ϕ-Coefficient

Figure III.1: non null-invariant Jointness Measures, FLS Dataset

To sum up, both types of measure provide similar insights into the bivariate covariate inclusion structure in the model space. On the one hand, the additional weighting for the probability of joint exclusion in the non null-invariant measures causes relationships with high individual PIPs to lose importance as compared to the bivariate jointness of variables with average PIP. On the other hand, null-invariant measures ignore this exclusion margin and stress the importance of variable relationships where both variables have a high individual PIP.

In contrast to the results of LS for this data set, the jointness results found here are not exclusively related to variables with high PIP. For the measures introduced by LS, high jointness is concentrated among the top 5 regressors (*GDP60, Confucian, LifeExp, EquipInvest* and *Sub-Sahara*). This can be reproduced by restricting the analysis to the two null-invariant measures considered here. If however, the exclusion margin is included into the analysis, other jointness relationships are discovered. One example for these are colonial variables, which are less frequent, but still exhibit complementary behavior.

In their analysis, DW employ the dataset of Sala-i-Martin, Doppelhofer, and Miller (2004, SDM data set), for which PIPs tend to be more concentrated on a few variables. Accordingly LS also find less jointness in this data set, using their null-invariant measures.



(a) Cosine

(b) Kulczynski

Figure III.2: Null-invariant Jointness Measures, FLS Dataset

5 Conclusion

In this paper we investigate the issue of measuring jointness of robust growth determinants as raised by Ley and Steel (2007), Doppelhofer and Weeks (2009a) and others in the BMA literature. We link the measurement of joint inclusion of covariates to the field of assessing association in data mining, where similar problems are studied. We argue that the search for substitutes and complements in model profiles is similar to the data mining issue of finding “interesting” combinations of e.g. products in a shopping basket.

We link the properties that have been introduced for jointness to the concepts that are used for categorizing interestingness measures for association rules analysis. In particular, the jointness literature in BMA is concerned with a subset of these interestingness measures, referred to as confirmation measures. Furthermore, we highlight the role of null-invariance, that is, the effect of cases were both variables in a bivariate inclusion relationship are excluded. Based on these properties we select a set of interestingness measures and show how they relate to the jointness indicators proposed in the literature.

We show that null-invariant measures fail to give a comprehensive view on jointness since they cannot gauge the effect of statistical independence consistently across different dependence structures. We examine further how sensitive different measures are with regard to varying dependence structures across included covariates. Finally, we provide an empirical application of these measures to the well known dataset of economic growth determinants used by Fernández, Ley, and Steel (2001b) and discuss the complementarity and substitutability inclusion structures found.

Using non null-invariant measures, such as Yule’s Q, we find a large number of complementary relationships but only few substitutes among bivariate pairs of variables. The latter are primarily related to the combination of socioeconomic specifics (*Confucian, Muslim*) and geographical variables (*SubSahara*). Complementary relationships are manifold and can be found for example between different colonial variables, such as *Brit, English, Spanish* or *French*. Furthermore the quality of institutions (*RuleOfLaw*) and economic variables (*OutwardOrientation, BlackMarketPremium*) seem to exhibit such relationships.

As highlighted by Doppelhofer and Weeks (2009b), the treatment of the *exclusion margin* is highly relevant for an analysis of jointness. Null-invariance may lead to ambiguous results since these measures cannot quantify substitutes and complements in an appropriate fashion (Glass, 2013). Given this theoretical justification, we do find differences in the rank correlations between the two types of measures, but these only partly influence the general picture of complementary and substitute covariates found in the FLS dataset.

Appendix

A Review of measures of interestingness and confirmation

Table A.1: Definition of Jointness measures

#		Measure	Value
1	ϕ	ϕ-Coefficient	$\frac{\Pr(AB)-\Pr(A)\Pr(B)}{\sqrt{\Pr(A)\Pr(B)\Pr(\bar{A})\Pr(\bar{B})}}$
2	AV	Added Value	$\Pr(B\mid A)-\Pr(B)$
3	AC	AllConf	$\min(\Pr(B\mid A),\Pr(A\mid B))$
4	b	Carnap	$\Pr(AB)-\Pr(A)\Pr(B)$
5	cf	Certainty Factor	$\frac{\Pr(B\mid A)-\Pr(B)}{1-\Pr(B)}$ if $\Pr(B\mid A)>\Pr(B)$
6	χ^2	Chi-square (χ^2)	$\frac{(\Pr(AB)-\Pr(A)\Pr(B))^2 N}{\Pr(A)\Pr(\bar{A})\Pr(B)\Pr(\bar{B})}$
7	κ	Coehen's Kappa (κ)	$\frac{\Pr(B\mid A)\Pr(A)+\Pr(\bar{B}\mid\bar{A})\Pr(\bar{A})-\Pr(A)\Pr(B)-\Pr(\bar{A})\Pr(\bar{B})}{1-\Pr(A)\Pr(B)-\Pr(\bar{A})\Pr(\bar{B})}$
8	coh	Coherence	$(\Pr(A\mid B)^{-1}+\Pr(B\mid A)^{-1}-1)^{-1}$
9	cs	Collective Strength	$\ln\left[\frac{\Pr(AB)+\Pr(\bar{A}\bar{B})}{\Pr(A)\Pr(B)+\Pr(\bar{A})\Pr(\bar{B})}\times\frac{1-\Pr(A)\Pr(B)-\Pr(\bar{A})\Pr(\bar{B})}{1-\Pr(AB)-\Pr(\bar{A}\bar{B})}\right]$
10	conf	Confidence	$\Pr(B\mid A)$
11	conv	Conviction	$\ln\left[\frac{\Pr(A)\Pr(B)}{\Pr(A,\bar{B})}\right]$
12	IS	Cosine	$\frac{\Pr(AB)}{\sqrt{\Pr(A)\Pr(B)}}$
13	G	Gini index	$\Pr(A)(\Pr(B\mid A)^2+\Pr(\bar{B}\mid A)^2)+\Pr(\bar{A})(\Pr(B\mid\bar{A})^2+\Pr(\bar{B}\mid\bar{A}))-\Pr(B)^2-\Pr(\bar{B})^2$
14	IR	Imbalance Ratio	$\frac{\mid\Pr(A\mid B-\Pr(B\mid A\mid}{Pr(A\mid B)+\Pr(B\mid A)-\Pr(A\mid B)\Pr(B\mid A)}$
15	I	Interest	$\frac{\Pr(AB)}{\Pr(A)\Pr(B)}$
16	J	J-Measure	$\Pr(AB)\log\frac{\Pr(B\mid A)}{\Pr(B)}+\Pr(A\bar{B})\log\frac{\Pr(\bar{B}\mid A)}{\Pr(\bar{B})}$
17	ζ	Jaccard (ζ)	$\frac{\Pr(AB)}{\Pr(A)+\Pr(B)-\Pr(AB)}$
18	k	Kemeny-Oppenheim	$\frac{\Pr(A\mid B)-\Pr(A\mid\bar{B})}{\Pr(A\mid B)+\Pr(A\mid\bar{B})}$
19	kl	Klosgen	$\sqrt{\Pr(AB)}\times\max(\Pr(B\mid A)-\Pr(B),\Pr(A\mid B)-\Pr(A))$
20	kulc	Kulczynski	$(\Pr(A\mid B)+\Pr(B\mid A))/2$
21	L	Laplace	$\frac{N\times\Pr(AB)+1}{N\times\Pr(A)+2}$
22	l	Lift	$\frac{\Pr(B\mid A)}{\Pr(B)}$
23	ll	Log-Likelihood	$\ln\left[\frac{\Pr(A\mid B)}{\Pr(A\mid\bar{B})}\right]$
24	r	Log-Ratio	$\ln\left[\frac{\Pr(B\mid A)}{\Pr(B)}\right]$
25	MC	MaxConf	$\max(\Pr(B\mid A),\Pr(A\mid B))$
26	M	Mutual Information	$\Pr(AB)\log\frac{\Pr(AB)}{\Pr(A)\Pr(B)}+\Pr(A\bar{B})\log\frac{A\bar{B}}{\Pr(A)\Pr(\bar{B})}$ $+\ \Pr(\bar{A}B)\log\frac{\Pr(\bar{A}B)}{\Pr(\bar{A})\Pr(B)}+\Pr(\bar{A}\bar{B})\log\frac{\Pr(\bar{A}\bar{B})}{\Pr(\bar{A})\Pr(\bar{B})}$
27	s	Normalized Difference	$\Pr(B\mid A)-\Pr(B\mid\bar{A})$
28	α	Odds Ratio	$\ln\left[\frac{\Pr(AB)\Pr(\bar{A}\bar{B})}{\Pr(A,\bar{B})\Pr(\bar{A}B)}\right]$
29	ows	One-Way Support	$\Pr(B\mid A)\ln\left[\frac{\Pr(AB}{\Pr(A)\Pr(B}\right]$
30	PS	Piatetsky-Shapiro's	$N\times(\Pr(AB)-\Pr(A)\Pr(B))$
31	rr	Relative Risk	$\ln\left[\frac{\Pr(B\mid A)}{\Pr(B\mid\bar{A})}\right]$
32	sup	Support	$\Pr(AB)$
33	tws	Two-Way Support	$\Pr(AB)\ln\left[\frac{\Pr(AB)}{\Pr(A)\Pr(B)}\right]$
34	yq	Yule's Q	$\frac{\Pr(AB)\Pr(\bar{A}\bar{B})-\Pr(A\bar{B})\Pr(\bar{A}B)}{\Pr(AB)\Pr(\bar{A}\bar{B})+\Pr(A\bar{B})\Pr(\bar{A}B)}$
35	yy	Yule's Y	$\frac{\sqrt{\Pr(AB)\Pr(\bar{A}\bar{B})}-\sqrt{\Pr(A\bar{B})\Pr(\bar{A}B)}}{\sqrt{\Pr(AB)\Pr(\bar{A}\bar{B})}+\sqrt{\Pr(A\bar{B})\Pr(\bar{A}B)}}$

B Data description

Table B.1: Variable Names and Descriptive Statistics — FLS

	Abbreviation	Variable	μ	σ
1	Age	Age	23.71	37.307
2	Area	Area (Scale Effect)	972.92	2051.976
3	BlMktPm	Black Market Premium	0.16	0.291
4	Brit	British Colony dummy	0.32	0.470
5	Buddha	Fraction Buddhist	0.06	0.184
6	Catholic	Fraction Catholic	0.42	0.397
7	CivlLib	Civil Liberties	3.47	1.712
8	Confucian	Fraction Confucian	0.02	0.087
9	EcoOrg	Degree of Capitalism	3.54	1.266
10	English	Fraction of Pop. Speaking English	0.08	0.239
11	EquipInv	Equipment investment	0.04	0.035
12	EthnoL	Ethnolinguistic fractionalization	0.37	0.296
13	Foreign	Fraction speaking foreign language	0.37	0.422
14	French	French Colony dummy	0.12	0.333
15	GDP60	GDP level in 1960	7.49	0.885
16	HighEnroll	Higher education enrollment	0.04	0.052
17	Hindu	Fraction Hindu	0.02	0.101
18	Jewish	Fraction Jewish	0.01	0.097
19	LabForce	Size labor force	9305.38	24906.056
20	LatAmerica	Latin American dummy	0.28	0.451
21	LifeExp	Life expectancy	56.58	11.448
22	Mining	Fraction GDP in mining	0.04	0.077
23	Muslim	Fraction Muslim	0.15	0.295
24	NequipInv	Non-Equipment Investment	0.15	0.055
25	OutwarOr	Outward Orientation	0.39	0.491
26	PolRights	Political Rights	3.45	1.896
27	Popg	Population Growth	0.02	0.010
28	PrExports	Primary exports, 1970	0.67	0.299
29	Protestants	Fraction Protestant	0.17	0.252
30	PrScEnroll	Primary School Enrollment, 1960	0.80	0.246
31	PublEdupct	Public Education Share	0.02	0.009
32	RevnCoup	Revolutions and coups	0.18	0.238
33	RFEXDist	Exchange rate distortions	121.71	41.001
34	RuleofLaw	Rule of law	0.55	0.335
35	stdBMP	SD of black-market premium	45.60	95.802
36	SubSahara	Sub-Saharan dummy	0.21	0.409
37	WarDummy	War dummy	0.40	0.494
38	WorkPop	Ratio workers to population	−0.95	0.189
39	YrsOpen	Number of Years open economy	0.44	0.355

References

AGGARWAL, C. C. and P. S. YU (1998): "A new framework for itemset generation". In: *PODS 98, Symposium on Principles of Database Systems*. Seattle, WA, USA, pp. 18–24.

BROCK, W. A. and S. N. DURLAUF (2001): "What have we learned from a decade of empirical research on growth? Growth empirics and reality". In: *World Bank Economic Review* 15 (2), pp. 229–272.

CRESPO CUARESMA, J. et al. (2015): *Unveiling covariate inclusion structures in economic growth regressions using latent class analysis*.

DOPPELHOFER, G. and M. WEEKS (2005): *Jointness of growth determinants*. Cambridge Working Papers in Economics.

DOPPELHOFER, G. and M. WEEKS (2009a): "Jointness of growth determinants". In: *Journal of Applied Econometrics* 24 (2), pp. 209–244.

— (2009b): "Jointness of growth determinants: Reply to comments by Rodney Strachan, Eduardo Ley and Mark FJ Steel". In: *Journal of Applied Econometrics* 24 (2), pp. 252–256.

EICHER, T. S., L. HELFMAN, and A. LENKOSKI (2012): "Robust fdi determinants: bayesian model averaging in the presence of selection bias". In: *Journal of Macroeconomics* 34 (3), pp. 637–651.

FELDKIRCHER, M. and S. ZEUGNER (2009): "Benchmark priors revisited: On adaptive shrinkage and the supermodel effect in Bayesian model averaging". In: *IMF Working Papers*.

FERNÁNDEZ, C., E. LEY, and M. STEEL (2001a): "Benchmark priors for Bayesian model averaging". In: *Journal of Econometrics* 100, pp. 381–427.

— (2001b): "Model uncertainty in cross–country growth regressions". In: *Journal of Applied Econometrics* 16 (5), pp. 563–576.

FOSTER, D. P. and E. I. GEORGE (1994): "The risk inflation criterion for multiple regression". In: *The Annals of Statistics* 22 (4), pp. 1947–1975.

GENG, L. and H. J. HAMILTON (2006): "Interestingness measures for data mining: A survey". In: *ACM Computing Surveys (CSUR)* 38 (3), p. 9.

GLASS, D. H. (2013): "Confirmation measures of association rule interestingness". In: *Knowledge-Based Systems* 44, pp. 65–77.

HAHSLER, M., B. GRÜN, and K. HORNIK (2005): "Arules – A computational environment for mining association rules and frequent item sets". In: *Journal of Statistical Software* 14 (15), pp. 1–25. ISSN: 1548-7660. URL: http://www.jstatsoft.org/v14/i15/.

Hahsler, M., K. Hornik, and C. Buchta (2008): “Getting things in order: An introduction to the r package seriation”. In: *Journal of Statistical Software* 25 (3), pp. 1–34. issn: 1548-7660. url: http://www.jstatsoft.org/v25/i03/.

Ley, E. and M. Steel (2012): “Mixtures of g-priors for Bayesian model averaging with economic applications”. In: *Journal of Econometrics* 171 (2), pp. 251–266.

Ley, E. and M. Steel (2009a): “On the effect of prior assumptions in Bayesian model averaging with applications to growth regressions”. In: *Journal of Applied Econometrics* 24, pp. 651–674.

Ley, E. and M. F. J. Steel (2007): “Jointness in Bayesian variable selection with applications to growth regression”. In: *Journal of Macroeconomics* 29 (3), pp. 476–493.

— (2009b): “On the effect of prior assumptions in Bayesian model averaging with applications to growth regression”. In: *Journal of Applied Econometrics* 24 (4), pp. 651–674.

Liang, F. et al. (2008): “Mixtures of g priors for Bayesian variable selection”. In: *Journal of the American Statistical Association* 103, pp. 410–423.

Madigan, D. and J. York (1995): “Bayesian graphical models for discrete data”. In: *International Statistical Review* 63 (2), pp. 215–232.

Moral-Benito, E. (2012): “Panel growth regressions with general predetermined variables”. In: *CEMFI Working Paper* 1006.

Moral-Benito, E. (2014): “Growth empirics in panel data under model uncertainty and weak exogeneity”. In: *Journal of Applied Econometrics*.

Piatetsky-Shapiro, G. (1991): “Discovery, analysis and presentation of strong rules”. In: *Knowledge Discovery in Databases*. Ed. by G. Piatetsky-Shapiro and W. Frawley. MIT Press, Cambridge, MA, pp. 229–248.

Sala-i-Martin, X., G. Doppelhofer, and R. I. Miller (2004): “Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach”. In: *American Economic Review* 94 (4), pp. 813–835.

Strachan, R. W. (2009): “Comment on 'Jointness of growth determinants' by Gernot Doppelhofer and Melvyn Weeks”. In: *Journal of Applied Econometrics* 24 (2), pp. 245–247.

Tan, P.-N., V. Kumar, and J. Srivastava (2004): “Selecting the right objective measure for association analysis”. In: *Information Systems* 29 (4), pp. 293–313.

Tew, C., C. Giraud-Carrier, K. Tanner, and S. Burton (2014): “Behavior-based clustering and analysis of interestingness measures for association rule mining”. In: *Data Mining and Knowledge Discovery* 28 (4), pp. 1004–1045.

VAILLANT, B., P. LENCA, and S. LALLICH (2004): “A clustering of interestingness measures”. In: *Discovery Science*. Springer, pp. 290–297.

WU, T., Y. CHEN, and J. HAN (2010): “Re-examination of interestingness measures in pattern mining: a unified framework”. In: *Data Mining and Knowledge Discovery* 21 (3), pp. 371–397.

YAO, Y. and N. ZHONG (1999): “An analysis of quantitative measures associated with rules”. In: *Methodologies for Knowledge Discovery and Data Mining*. Springer, pp. 479–488.

ZELLNER, A. (1986): “On assessing prior distributions and Bayesian regression analysis with g–prior distributions”. In: *Bayesian Inference and Decision Techniques Essays in Honor of Bruno de Finetti* 6. Ed. by A. ZELLNER and P. GOEL, pp. 233–243.