Variable Jointness in BMA via Dirichlet Clustering

Jesus Crespo Cuaresma[1], Paul Hofmarcher[2], Stefan Humer[3], Mathias Moser[3]

Second Bayesian Young Statisticians Meeting (BAYSM 2014)
Vienna, September 18–19, 2014

[1] Department of Economics, Vienna University of Economics and Business (WU)
Austrian Institute of Economic Research (WIFO)
International Institute of Applied System Analysis (IIASA)
Wittgenstein Centre for Demography and Global Human Capital (WIC)
jcrespo@wu.ac.at

[2] Department of Applied Statistics, Johannes Kepler University, Linz
paul.hofmarcher@jku.at

[3] Department of Economics, Vienna University of Economics and Business (WU)
<shumer,matmoser>@wu.ac.at

Abstract

In the economic literature, a vivid discussion concerning the analysis of jointness of variables within Bayesian Model Averaging (BMA) regressions took place. Several bivariate measures, like the *odds ratios* (Doppelhofer and Weeks, 2009a) or the *Jaccard index* (Ley and Steel, 2007) have been proposed to analyze the joint inclusion of covariates in such a setting. This paper uses Dirichlet Process priors to examine the jointness of sampled models within a BMA framework and tries to identify *common* model specifications. Furthermore we show how sensitive prior choices in the BMA procedure are with respect to this clustering of sampled models.

Keywords: Bayesian Model Averaging, Variable Jointness, Clustering

1 Introduction

In the economic literature (see Doppelhofer and Weeks, 2009a; Strachan, 2009; Doppelhofer and Weeks, 2009b) a vivid discussion took place how to correctly analyze the (dis)jointness of variables in Bayesian Model Averaging (BMA).

Common measures of importance focus on the frequency at which a variable is included in the sampled models (Posterior Inclusion Probabilities). However, this approach neglects the interdependencies of variables and the fact that different clusters of models can prefer varying sets of covariates which are frequently included. Such an analysis would provide a deeper understanding of the visited models, in order to provide guidance on questions concerning supplementary or complementary effects between variable sets.

Still there is no consent in the literature regarding the right measure of jointness. To the best of our knowledge Doppelhofer and Weeks (2009a) (henceforth DW) were the first who introduced a new measure which allows to analyze the jointness among explanatory variables in such a setting.

To formalize a measure of dependency between the explanatories x_i and x_j (henceforth i and j) DW defined their *jointness* statistic as the cross product ratio, i.e.,

$$J_{ij} = ln\left[\frac{p(i \cap j)}{p(\bar{i} \cap j)} \frac{p(\bar{i} \cap \bar{j})}{p(i \cap \bar{j})}\right] \tag{1}$$

In the data mining literature this measure is know as Odds ratio α (Tan et al., 2004).

Ley and Steel (2009) criticized this measure and proposed four criteria an appropriate measure of jointness should fulfill:

Interpretability Any jointness measure should have either a clear meaning or should be statistically well founded.

Calibration Values of jointness should be calibrated against some predefined scale.

Extreme Jointness Situations were two variables appear always together should result in maximum jointness.

Support A jointness measure should always be defined whenever at least one variable is included.

More recently Glass (2013) discussed properties and advantages of different association rules such as null–invariance, monotonicity and symmetry requirements. However the literature on association rules disagrees on the importance of these different properties for association rules (and jointness measures). Therefore these attributes have to be evaluated in a systematic way and their applicability for model comparisons has to be discussed in detail.

In general, inspecting the c_i estimates for the visited modles allows to gather a more general picture of the sampled model space. Moreover it allows us to identify similarities between the sampled models. Therefore we are able to analyze jointness more extensively compared to the bivariate case.

For such an analysis the three key properties defined by Piatetsky-Shapiro (1991) should be satisfied, yet there are numerous extensions to these basic characteristics. An overview is provided by Tan et al. (2004).

2 Jointness of Dirichlet Clustered Models

In order to discuss the jointness of variables and determine a more detailed relationship between the variables and the sampling behavior of the MCMC methods, we make use of Dirichlet Process Clustering (DPC).

The idea of utilizing clustering to profile patterns in data (here similar models) is not new (see Molitor et al., 2010). Compared to the methods used in literature, the advantages of the Dirichlet clustering approach are manifold: First, the DPC extends the bivariate comparison to a more general picture of similar models. Furthermore the Bayesian clustering approach allows the number of groups to vary and accordingly can uncover subgroups and examines their association with an outcome of interest.

Most importantly, clustering facilitates interpretation of jointness effects, which can be ambiguous for the whole model space. In this setting jointness measures can not only be applied to single clusters but also to the global model space. The estimated clusters will however capture most of the jointness effect between groups of models, so that their inherent structure can be addressed more directly using the proposed jointness measures.

Formally, we consider models of the form

$$\begin{aligned} y_i|c,\phi &\sim F(\phi(c_i)) \\ c_i|p &\sim Discrete(p_1,\ldots,p_K) \\ \phi(c) &\sim G_0 \\ p_1,\ldots,p_K &\sim Dirichlet(\alpha/K,\ldots,\alpha/K). \end{aligned}$$

c_i indicates latent class of observation y_i and K the number of components (clusters) which is estimated within the MCMC procedure.

References

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

Doppelhofer, G. and Weeks, M. (2009b). Jointness of growth determinants: Reply to comments by rodney strachan, eduardo ley and mark fj steel. *Journal of Applied Econometrics*, 24(2):252–256.

Geng, L. and Hamilton, H. J. (2006). Interestingness measures for data mining: A survey. *ACM Computing Surveys (CSUR)*, 38(3):9.

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

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

Ley, E. and Steel, M. F. (2009). Comments on 'jointness of growth determinants'. *Journal of Applied Econometrics*, 24(2):248–251.

Molitor, J., Papathomas, M., Jerrett, M., and Richardson, S. (2010). Bayesian profile regression with an application to the national survey of children's health. *Biostatistics*, 11(3):484–498.

Piatetsky-Shapiro, G. (1991). Discovery, analysis and presentation of strong rules. In Piatetsky-Shapiro, G. and Frawley, W., editors, *Knowledge Discovery in Databases*, pages 229–248. MIT Press, Cambridge, MA.

Strachan, R. W. (2009). Comment on jointness of growth determinants by gernot doppelhofer and melvyn weeks. *Journal of Applied Econometrics*, 24(2):245–247.

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

Wu, T., Chen, Y., and Han, J. (2010). Re-examination of interestingness measures in pattern mining: a unified framework. *Data Mining and Knowledge Discovery*, 21(3):371–397.