

Dilution priors for interaction terms in BMA: An application to the financial crisis of 2008–09

Mathias Moser[†]

In this paper we investigate the usage of interaction terms in Bayesian Model Averaging. Standard model priors treat interactions like any other covariate and include them in models even if their parent variables have not been selected by the sampler. The literature proposes the usage of ‘Dilution Priors’ to address this problem. We apply two different heredity priors to a dataset of cumulated output loss in the financial crisis of 2008/09 and compare the effects of these priors on the inclusion of interaction terms. Furthermore predictive inference is carried out to evaluate the alternative priors.

Keywords: Bayesian Model Averaging

1 Introduction

If we consider regression models with a large number of potential covariates a critical question is how to determine good models (measured e.g. by their predictive performance). Since the number of possible models increases exponentially with the number of predictors K , automatic techniques for variable selection have become reasonably common, especially in growth econometrics. Moreover the treatment of model uncertainty has been addressed in numerous studies, specifically through Bayesian Model Averaging (BMA, cf. Fernández, Ley, and Steel 2001b).

For large model spaces (that is $> 2^{14}$ covariates) these methods use sampling algorithms to determine well performing models, such as Markov Chain Monte Carlo routines. Similarly different priors over the model space have received much attention mostly focusing on ‘objective priors’. One approach is an equal weighting of all possible models, which is referred to as a ‘(discrete) uniform model prior’. This prior choice assigns each model a prior inclusion probability of 2^{-K} and accordingly prefers mid-sized models.

[†]Department of Economics, Vienna University of Economics and Business (WU Vienna). Postal Address: Augasse 2-6, 1090 Vienna, Austria. Mail: matmoser@wu.ac.at

Even if such a weighting is commonly referred to as an ignorance prior (due to its uniform prior inclusion probabilities), such priors have been criticised for leaving out valuable information which may be crucial for the model search algorithm (cf. George 2010).

One case in which a uniform model prior might be inadequate is a model space that includes highly correlated variables. In this scenario the model space consists of partly redundant models which may include different variables that perform equally good or bad. An example: Consider some explanatory variable x_1 which is uncorrelated with a highly correlated set of variables $X_{2...K}$. Adding another correlated variable to the set $X_{2...(K+1)}$ under a uniform prior would c.p. reduce the prior weight on x_1 even though we added no new information. This effect becomes more severe if x_1 is in fact a key variable to explain some data y . In large model spaces this might be augmented by an MCMC algorithm that frequently visits reasonably good but redundant models and at the same time reduces the importance of well-performing unique models.

In growth empirics models have to deal with parameter heterogeneity, for example in a regional context. A well-known example are sub-saharan countries for which a regional dummy and/or interaction terms are considered in the model. In such a setting it is good practice in econometrics to include all parent variables in a model if an interaction term is included (cf. Chipman 1996). If we however use MCMC techniques with a uniform model prior the prior model probability of a well-specified model is as high as that of a model containing only an interaction without its parent variables (cf. Masanjala and Papageorgiou 2008).

In such a case two problems can arise: First, from a practical perspective, it can be hard to interpret interaction terms without parent variables (cf. Crespo Cuaresma 2011 and Papageorgiou 2011). Second, chances are high that the interaction term and some or all of its parents are highly correlated (depending on the structure of the variables). If we ignore the parent variables of an interaction term, it might catch some of the effect of the parent variable which leads in the best case to augmented standard errors and followingly insignificant coefficients for the interaction term.

Generally two solutions may be applied to this problem. What has been discussed is to specify priors that adjust or penalize models with missing parent variables and therefore include additional prior information. Another approach is to adjust the prior for the correlation between variables (cf. Garthwaite and Mubwandarikwa 2010).

In the following we will scrutinize different prior assumptions to correctly higher level interaction terms in Bayesian Model Averaging: Section 2 elaborates on the method and

different possible prior assumptions. In chapter 3 we introduce alternative prior concepts. As an application Section 4 applies the discussed priors on a dataset on output loss during the 2008–09 financial crisis and evaluates the results. Section 4 gives a short summary of the findings.

2 Bayesian Model Averaging

A model M_j with a set of regressors $Z_j \in Z$ may be written as

$$y = \alpha \iota_n + Z_j \beta_j + \epsilon \quad (1)$$

where α is the intercept and ι_n is an n -dimensional vector of ones. To carry out Bayesian inference we need to multiply the likelihood with a suitable prior on the parameters θ to derive the posterior distribution.

$$p(\theta|y, M_j) = p(\theta|M_j)p(y|\theta, M_j) \quad (2)$$

Following numerous studies (cf. Ley and Steel 2009) we will use non-informative priors for α and σ

$$p(\alpha, \sigma) = \sigma^{-1}$$

and a g -prior for all β (Zellner 1986), where g is $1/\max(n, K^2)$ the so-called BRIC prior (Fernández, Ley, and Steel 2001a).

$$p(\beta_j|\alpha, \sigma, M_j) = f_N^{k_j}(\beta_j|0, \sigma^2(gZ_j'Z_j)^{-1}) \quad (3)$$

To address model uncertainty we can take the average over all models, that is the posterior of each model times its posterior model probability:

$$p(\theta|y) = \sum_{j=1}^{2^K} p(\theta|y, M_j)p(M_j|y) \quad (4)$$

the latter is given as the marginal likelihood multiplied with the prior model probability

$$p(M_j|y) = p(y|M_j)p(M_j)$$

The prior model probability $p(M_j)$ can be chosen in various ways. One way is a uniform

prior that assigns the same *ex ante* probability to each model. Another solution by Ley and Steel 2009 introduces a hierarchical Binomial–Beta prior which will be used in the remainder of this paper.

Prediction can be based upon the predictive density (cf. Zellner 1971)

$$p(\tilde{y}|y, Z, \tilde{Z}) \sim t(\tilde{Z}\beta, \frac{\nu s^2}{\nu - 2} \underbrace{[I + \tilde{Z}(Z'Z)^{-1}\tilde{Z}']}_{H^{-1}}, \nu) \quad (5)$$

The marginal posterior density function is then

$$p(\tilde{y}_i) = \frac{\tilde{y}_i - \tilde{Z}(i)\hat{\beta}}{\sqrt{h_{ii}}} \sim t_\nu \quad (6)$$

which can be used to calculate the Log Predictive Score

$$LPS_{BMA} = - \sum_{i=1}^n \log \sum_{j=1}^{2^K} p(\tilde{y}_i) \quad (7)$$

3 Dilution: Weak and Strong Heredity Priors

Heredity priors apply to models with interaction terms and disfavour models with interactions but missing parents. More specifically they can adjust the prior model probability with regard to the chosen covariates. In a three variable model (A, B and their interaction AB) the following cases can be distinguished if an interaction term AB is included ($\delta_{AB} = 1$):

$$Pr(\delta_{AB} = 1 | \delta_A, \delta_B) = \begin{cases} p_{00} & \text{if } (\delta_A, \delta_B) = (0, 0) \\ p_{01} & \text{if } (\delta_A, \delta_B) = (0, 1) \\ p_{10} & \text{if } (\delta_A, \delta_B) = (1, 0) \\ p_{11} & \text{if } (\delta_A, \delta_B) = (1, 1) \end{cases}$$

A prior which does not follow the heredity principle would simply let $p_{00} = p_{01} = p_{10} = p_{11}$ and therefore treat all models as *ex ante* equally important. A more restrictive prior may set $(p_{00}, p_{01}, p_{10}, p_{11}) = (0, 0, 0, q)$, which would ignore all models with interaction terms and missing parent variables. This has been referred to as a strong-heredity prior (Chipman 1996; Crespo Cuaresma 2011). If we relax these assumption by setting $(p_{00}, p_{01}, p_{10}, p_{11}) = (q_1, q_2, q_3, q_4)$ this leads to a prior according to the weak heredity principle.

For this approach it could be useful to relate the prior model probability to the number of missing parent variables, i.e. a model with an interaction term which lacks both parent variables is more unlikely than a model which includes at least one parent. Therefore we will propose a structure like $p_{00} < (p_{01} = p_{10}) < p_{11}$.

The process of choosing the probabilities p can be simplified by expressing them as a function of the number of missing parent variables. Through this approach one could imply e.g. a linear, logarithmic or exponential penalty for additional missing parent variables. Let the number of missing parents be denominated by η then we could express the penalty for the simple case of a uniform prior as

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

(8)

By premultiplying the numerator or denominator of the penalty term $\frac{1}{\eta}$ by a scalar s (which we may refer to as a shape parameter), the form of this logarithmic function can be easily adjusted. Equation 3 is an extension to the standard uniform prior, as it downweights models with missing parents but keeps the prior model probability of other models unaltered. In the following analysis we will use a standard logarithmic penalty, that is a shape parameter s equal to one.

4 Estimating output loss for the recent financial crisis

Crespo Cuaresma and Feldkircher 2012 examined output loss during the 2008 financial crisis for 153 countries around the world. The dataset includes a large number of variables and higher level interaction terms.¹ The inclusion of such ‘triple-interactions’ (e.g. $cese \times real_gdp_growth \times net_fdi$) leaves room for a large number of models with missing parent variables if we apply BMA techniques. In most cases were the sampler selects such a three-component interaction term not all three parent variables will be included in the model. We repeat the BMA procedure of Crespo Cuaresma and Feldkircher 2012 and

¹A register of the included variables can be found in the Appendix of their paper.

introduce two forms of heredity priors to adjust for these effects. The first model weighs models through a standard binomial–beta prior. The second approach utilizes the concept of strong heredity and sets the prior model probability of a model with interactions but (partly) missing parents to zero. In a third setting we implement a weak heredity prior with a penalty according to chapter 3 on the Binomial–Beta prior.

Table 1: Results for Cumulated Output Loss between 2007 and 2009
(Top 15 variables from each model)

	No Heredity			Strong Heredity			Weak Heredity		
	PIP	PMean	PSD	PIP	PMean	PSD	PIP	PMean	PSD
rgdpcap	0.84	-1.7767	1.0096	0.86	-1.9007	1.0107	0.83	-1.7517	1.0164
UA	0.77	-12.3058	8.3511	0.88	-15.2819	7.4963	0.80	-13.0868	8.2636
EU.15	0.74	-4.3212	3.1919	0.71	-4.2836	3.2754	0.75	-4.4602	3.2132
real.gdp.gr	0.67	0.4242	0.3471	0.80	0.5168	0.3254	0.73	0.4677	0.3393
transition#FinOpenn	0.65	-7.4634	6.5124	0.37	-5.0304	7.0072	0.64	-7.4073	6.5122
transition#real.gdp.gr									
#ext.debt.gdp	0.64	-0.0160	0.0141	0.09	-0.0020	0.0067	0.53	-0.0132	0.0141
imp.from.US.gdp	0.51	-0.1557	0.7396	0.49	-0.1630	0.8209	0.55	-0.1706	0.7562
tradeExp.US.gdp	0.50	-0.0732	0.7286	0.44	-0.0611	0.8099	0.47	-0.0649	0.7456
baltics	0.47	-5.7729	7.0831	0.92	-14.1800	5.7719	0.55	-7.1530	7.4096
transition#net.fdi.infl	0.44	0.2962	0.4287	0.07	0.0207	0.1740	0.39	0.2560	0.4022
dGapExo	0.35	-0.0225	0.0352	0.25	-0.0150	0.0302	0.32	-0.0201	0.0338
reerm	0.35	-0.0050	0.0080	0.45	-0.0073	0.0094	0.35	-0.0051	0.0081
transition#real.gdp.gr									
#net.fdi.infl	0.35	0.0180	0.0353	0.08	0.0054	0.0224	0.29	0.0160	0.0333
pop	0.33	0.2269	0.3827	0.26	0.1658	0.3357	0.30	0.2066	0.3688
dom.credit	0.33	-0.0104	0.0177	0.27	-0.0085	0.0166	0.33	-0.0109	0.0184
transition	0.11	0.4726	2.6997	0.92	-0.5852	5.9498	0.19	0.7761	3.4477
FinOpenn	0.11	0.1660	0.6705	0.40	0.6575	1.2280	0.12	0.1803	0.6990
net.fdi.infl	0.11	0.0145	0.0594	0.31	0.0426	0.0969	0.13	0.0167	0.0635
cpi_corruption	0.30	-0.2173	0.3925	0.26	-0.1970	0.3882	0.30	-0.2174	0.3924
Log Pred Score		2.8490			2.9893			2.8614	

By eliciting a strong heredity prior it can be seen from table 4 that the posterior inclusion probabilities of interaction terms drop sharply compared to a scenario with a standard model prior. Furthermore the inclusion probabilities of the single parent variables are pushed upwards, since they now could capture some of the effects of the missing interaction terms. The transition variable and its interactions with other variables show for example that the PIP of the interaction terms drop compared to the no heredity scenario (see e.g. transition#FinOpenn). However the inclusion probability of the single parent variables Transition and FinOpenn rises sharply since they have to be included in any model featuring their interaction.

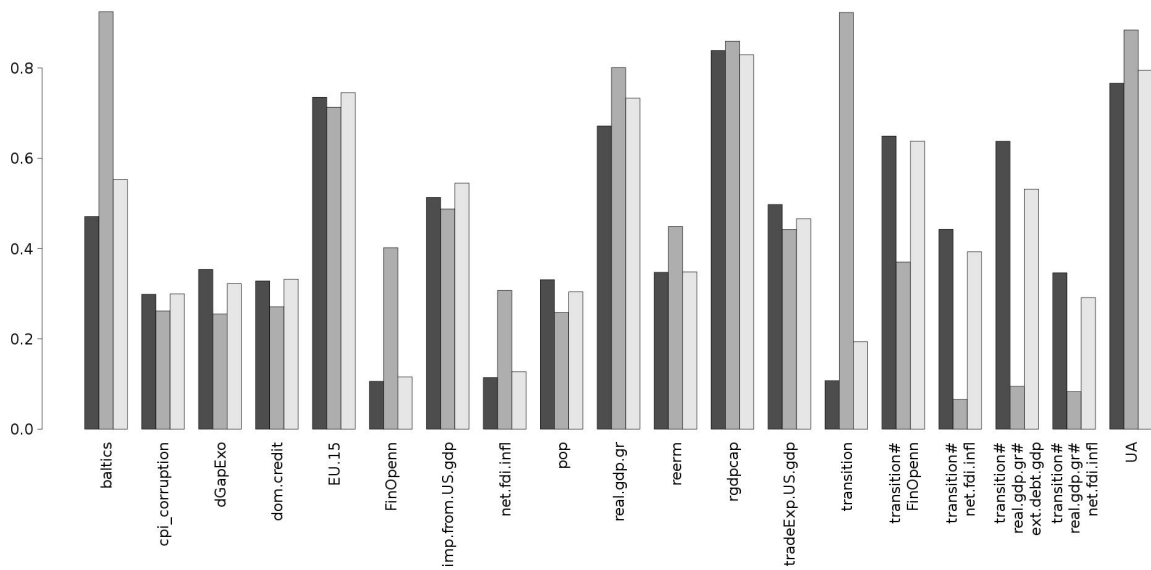


Figure 1: Posterior Inclusion Probabilites for the top 15 variables from each model

In contrast the weak heredity prior penalizes models with missing parent variables but still includes them if they perform excessively well. Accordingly inclusion probabilities for interaction terms are between those of the no heredity and strong heredity scenarios. While the weak heredity prior does not overvalue parent variables, like in the strong heredity setting, it emphasizes the importance of some not interacted variables like the EU15 dummy or import from US relative to GDP ('imp.from.US.gdp').

Prediction was carried out by randomly dropping 15% of the sample and calculating the LPS of the predicted values compared to the actual realizations of cumulated output loss. Especially the strong heredity prior performs distinctively worse than the default setup with a LPS score of 2.99 (compared to 2.85). As can be seen from the PIP the weak heredity prior also favours models with interaction terms with not all of their parents included. Accordingly its LPS is close to the no heredity prior.

5 Discussion

The treatment of interaction terms in Bayesian Model Averaging has only been addressed theoretically in the literature. A number of authors propose that such interactions should be dealt with in a special way rather than treating them as normal covariates. Since a standard sampler selects models with no regard to their specification, it has been suggested to impose a suitable model prior to address the problem of missing parent variables in

regressions with interaction terms.

In this paper we implement different variations of a so-called heredity prior and evaluate their performance compared to a standard BMA procedure with a default model prior. The strong heredity scenario puts a severe penalty on uncommon models, that include interaction terms but not their parents, and effectively dismisses them as improper. Another alternative to introduce prior information on model specification is a prior according to the weak heredity principle. We propose a prior specification that puts a linear penalty on models with missing interaction terms and downweighs their model probability dependent on the number of missing parent variables.

This study uses a dataset on cumulated output loss during the financial crisis of 2008/09 to assess the effects of these different priors. Compared to an ignorance prior the strong heredity prior forces a correct model specification. Followingly the model space is effectively reduced to models with a) no interaction terms or b) interaction terms and all their parent variables included. Results show that for this dataset, the inclusion probabilities for interaction terms drop sharply if we apply a strong heredity prior. At the same time the parent variables become more important, since they are now necessarily included if the interaction term is in the model. Predictive inference shows that the strong heredity principle performs worse than a prior without dilution property.

As an alternative we introduced a weak heredity prior, which in our dataset, stresses the importance of the interaction terms. Accordingly it allows more often for an uncommon model, that is a model with interaction terms and not all of their parents included. Predictive inference shows that the weak heredity prior in this example performs equally well as a no heredity prior.

The weak heredity prior seems to be a reasonable choice for models that feature interaction terms. On the one hand it includes prior information on how to handle these kind of variables. On the other it allows the model search algorithm to deviate from this prior recommendation if uncommon models perform extremely well over others.

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* (forthcoming).
- Crespo Cuaresma, J.** and **M. Feldkircher** (2012): “Drivers of output loss during the 2008-09 crisis: A focus on emerging Europe”.
- Fernández, C., E. Ley,** and **M.J.F. 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.
- Garthwaite, P.H.** and **E. Mubwandarikwa** (2010): “Selection of Prior Weights for Weighted Model Averaging”. In: *Australian & New Zealand Journal of Statistics* 52.4, pp. 363–382.
- George, E.I.** (2010): “Dilution priors: Compensating for model space redundancy”. In: *Borrowing Strength: Theoriz Powering Applications - A Festschrift for Lawrence D. Brown*. Ed. by **J.O. Berger, T.T. Cai,** and **I.M. Johnstone**. Institute of Mathematical Statistics, pp. 185–165.
- Ley, E.** and **M.J.F. Steel** (2009): “On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regressions”. In: *Journal of Applied Econometrics* 24.4, pp. 651–674.
- Masanjala, W.H.** 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* 23.5, pp. 671–682.
- Papageorgiou, C.** (2011): “How to use Interaction Termin in BMA: Reply to Crespo Cuaresma’s Comment on Masanjala and Papageorgiou (2008)”. In: *Journal of Applied Econometrics* 26, pp. 1048–1050.
- Zellner, A.** (1971): *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley.
- (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.K. Goel**, pp. 233–243.