IMF Working Paper

Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging

Martin Feldkircher and Stefan Zeugner

INTERNATIONAL MONETARY FUND

IMF Working Paper

Finance Department

Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging[*]

Prepared by Martin Feldkircher and Stefan Zeugner

Authorized for distribution by Craig Beaumont

September 2009

Abstract

Default prior choices fixing Zellner's g are predominant in the Bayesian Model Averaging literature, but tend to concentrate posterior mass on a tiny set of models. The paper demonstrates this *supermodel effect* and proposes to address it by a *hyper-g prior*, whose data-dependent shrinkage adapts posterior model distributions to data quality.

Analytically, existing work on the hyper-g-prior is complemented by posterior expressions essential to fully Bayesian analysis and to sound numerical implementation. A simulation experiment illustrates the implications for posterior inference. Furthermore, an application to determinants of economic growth identifies several covariates whose robustness differs considerably from previous results.

Contents

I Introduction

As a method of dealing with model uncertainty, Bayesian Model Averaging (BMA) has received considerable attention – ever since Raftery (1995) and Hoeting et al. (1999) demonstrated that inference neglecting model uncertainty leads to overstated confidence in statistical estimates. BMA, in contrast, tackles model uncertainty directly by basing inference on a weighted average of all potential covariate combinations. In the Bayesian framework, these weights arise naturally as posterior model probabilities (PMP), akin to the classical likelihood concept. The PMP for model M_s conditional on data (y, X) is proportional to its marginal likelihood $p(y|M_s, X)$ times a model prior $p(M_s)$:

$$p(M_s|y, X) \propto p(y|M_s, X)p(M_s)$$

Relying on this framework, numerous authors (e.g. Raftery 1995, Fernández et al. 2001a, Liang et al. 2008) have demonstrated that BMA outperforms other strategies in terms of predictive ability. Virtually all of them have so far concentrated on linear models with model-specific inference based on the natural-conjugate Normal-Gamma framework with *Zellner's g prior* (Zellner, 1986). This approach aims to represent the lack of prior knowledge by employing a conditional prior on model coefficients $\beta|\sigma^2 \sim N\left(0, g\sigma^2\left(X'X\right)^{-1}\right)$ that is partly determined by the scalar hyperparameter g.

The g prior structure has proven universally popular in BMA, since it leads to simple closed form expressions of posterior statistics and because it reduces prior elicitation to the choice of a single hyperparameter g. This applies in particular to the resulting marginal likelihood $p(y|M_s, X)$:[1]

$$p(y|M_s, X) \propto (1 + g)^{-\frac{k_s}{2}} \left(1 - \frac{g}{1+g} R_s^2\right)^{-\frac{N-1}{2}} \tag{1}$$

The elicitation of g is subject to intense debate (e.g. Liang et al. (2008), Hoeting et al. (1999), Fernández et al. (2001a), Eicher et al. (2009)) and constitutes the focus of the present paper. So far, literature has discussed the optimal choice of g practically by Monte Carlo simulations, and theoretically by tuning g according to two considerations: First, it focused on consistency, i.e. the choice of g such that BMA asymptotically uncovers 'the true model'. Second, the specification of g was studied in terms of its virtues as a model size penalty term to favor parsimonious models. In this respect, Fernández et al. (2001a) as well as Foster and George (1994) demonstrate how g can be calibrated to asymptotically mimic popular information criteria such as RIC or BIC by adjusting g to their respective model size penalties.

These studies were motivated by asymptotic consistency (as mentioned above), which focuses on a unique 'true' model and requires only a g increasing with N. However from a Bayesian viewpoint, many models might be 'true', in the sense that they are generating the data examined.[2] In this case, the quest for asymptotic consistency is analytically less clear-cut.

---

[1]Here, R_s^2 denotes the OLS R-squared of model M_s, and (N, k_s) the dimensions of its design matrix.

[2]Therefore, practitioners employing BMA focus much more on model-wise marginal distributions, such as posterior inclusion probabilities or posterior beta distributions.

The virtues of the different g elicitation mechanisms have been subject to debate – however the use of a constant hyperparameter g as such has been less frequently criticized.

Considered from the perspective of an applied researcher, the practical advantages of Zellner's g that render it ubiquitous in BMA come at a serious cost: g exerts non-negligible influence on posterior inference since it governs how posterior mass is spread over the models. Larger values of g incite posterior mass to concentrate on fewer models, whereas smaller values of g spread PMPs more evenly, irrespective of model sizes and their penalty terms. Posterior statistics, and in particular PMP and posterior inclusion probabilities (PIP) are thus notoriously sensitive to the value of the g prior. In other words, the researcher's prior on g plays a considerable role in determining how much posterior mass is attributed to a few, or the single best performing model – regardless whether these have been generating the data. Henceforth, we will refer to this concentration of posterior mass on a few models as the *supermodel effect*. While crucial in terms of prior sensitivity, this effect went more or less unnoticed in the focus on consistency as in Fernández et al. (2001a): Focusing on uncovering a single 'true' model in Monte Carlo simulations, previous studies longed for settings ascribing the bulk of posterior mass to this data-generating model – a job facilitated by choosing large values for g.

In our view, the earnest approach in tackling the issue is the introduction of a nondegenerate prior distribution on g, and thus 'let the data choose'. Only few papers have attempted this so far: among them are Cui and George (2008), Strachan and van Dijk (2004), Liang et al. (2008), with the latter being probably the most comprehensive contribution. Thus we propose using a *hyperprior distribution on g* in the vein of Liang et al. (2008). The advantages of a hyperprior are fivefold: First, it greatly reduces the g prior sensitivity of posterior mass. Second, it does so by adjusting the posterior distribution of g such that it reflects the data's signal-to-noise ratio, by shrinking the estimated coefficients more toward zero in noisier data sets.[3] Thus the hyper-g prior allows for data dependent shrinkage, in contrast to prior structures fixing g to some constant. Third, it still leaves ample space for the researcher to formulate her prior beliefs, but without incurring the risk of unwantedly affecting posterior statistics. Fourth it is computationally feasible and addresses the same theoretical considerations complying with asymptotic consistency as does the standard setting. Finally, the hyper-g prior is not exposed to the aforementioned supermodel effect a priori. It adjusts the distribution of posterior mass in dependence of the information provided by the data. Thus if noise dominates the data, PMPs under the hyper-g prior will be distributed more evenly, whereas in the case of minor noise, posterior mass will be concentrated even more as in fixed settings that impose large values for g.

In this sense, we complement the contribution of Liang et al. (2008) by providing further closed-form representations for important posterior quantities. Additionally, we present analytical expressions that allow for a sound numerical implementation in terms of accuracy. Furthermore we demonstrate the behavior of several prominent prior structures such as the benchmark prior put forward by Fernández et al. (2001a) under a typical situation where the data generating process cannot be traced back to one specific model. Our results show that under noisy data the hyper-g prior dilutes the posterior mass among models whereas the benchmark prior incorrectly favors one (wrong) model. By means of a simulation study

---

[3]i.e. by up- or downweighting the prior beliefs on coefficients β.

we examine the predictive properties of various settings for g indicating superior predictive ability for the hyper-g prior under varying signal-to-noise ratios.

The remainder of this study is organized as follows: the next section briefly summarizes Bayesian model averaging under the natural conjugate framework employing Zellner's g prior and reviews the prior settings that have resurfaced most often in the literature so far. Section III introduces the hyper-g prior and outlines derivations of further posterior quantities as well as an implementation strategy being of practical relevance. Section IV examines the supermodel effect inherent to traditional priors and highlights the predictive performance of flexible priors based on a simulation study. The following section demonstrates the sensitivity of posterior results to the choice of g by means of an empirical application to a prominent growth data set. Section VI concludes the paper.

II   Bayesian Model Averaging under Zellner's g Prior

Consider the canonical regression problem of sample size N with the dependent variable in the $N{\times}1$ vector y, X_s an $N{\times}k_s$ design matrix of covariates, and ε an N-dimensional vector of residuals in the following, linear model M_s:

$$y = \mathbf{1}\alpha_s + X_s\beta_s + \varepsilon$$

Here α denotes the (scalar) intercept, and the $k{\times}1$-vector β_s the nonzero regression coefficients. The residuals are assumed to be identically independently normally distributed with variance σ^2, i.e. $\varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Note that X_s can be assumed to be centered ($X_s'\mathbf{1} = \mathbf{0}$) without loss of generality, as this will only affect the posterior distribution of the constant α_s (Liang et al., 2008, p.412). Bayesian Model Averaging deals with uncertainty about the model M_s by drawing on the model-specific inference presented above. In the generic linear BMA problem, model uncertainty focuses on the choice of covariates X_s, which may be drawn from a set of K potential regressors. Thus there exist 2^K unique covariate combinations, as represented by the model candidate space $\mathcal{M} = \{M_1, M_2, \ldots, M_{2^K}\}$ (cf. Hoeting et al. 1999 for a more detailed account).

The Bayesian framework calls for specifying a prior distribution on the model's parameters α, β_s, and σ^2. The bulk of the BMA literature (Raftery, 1995; Chipman et al., 2001), favors the natural-conjugate approach, which puts a conditionally normal prior on coefficients $(\beta_s|\sigma^2, M_s) \sim N(\underline{\beta}_s, \sigma^2\underline{V}_s)$, where $\underline{\beta}_s$ and $\underline{V}_s$ represent hyperparameters.

With respect to the two other parameters, we depart from earlier tradition and follow Fernández et al. (2001a), who proposed improper priors on α and σ: Let $p(\alpha) \propto 1$, which corresponds best to a complete lack of prior information on the constant. Moreover, put an equally uninformative prior on $p(\sigma) \propto \sigma^{-1}$, which (in contrast to the traditional Gamma-priors) offers the additional advantage of being invariant under scale transformations (Fernández et al., 2001a, p. 391). The popularity of this prior structure is due to its closed-form solutions for posterior distributions, which allows for efficient coding with respect to large-scale model selection. Most notably, employing Bayes' Theorem via

$$p(\beta_s|y, X_s, M_s) = \int_0^\infty p(\beta_s|M_s, y, X_s, \sigma^2)dp(\sigma^2)$$

yields the posterior distribution of β as k_s-variate Student-t distribution whose variance structure is primarily shaped by the expression $\left(\underline{V}_s^{-1} + X_s'X_s\right)^{-1}$.

The above framework requires explicit formulation of $\underline{\beta}_s$ and $\underline{V}_s$, the prior hyperparameters on $\beta|\sigma^2$ – which is difficult to elicit given the many combinations possible in model selection problems. Virtually all linear BMA applications have thus opted for a common uninformative prior centered at zero, with the variance structure given by *Zellner's g prior*, i.e. $\underline{V}_s = g(X_s'X_s)^{-1}$ (Zellner, 1986). It thus assumes the prior covariance to be proportional to the posterior covariance expression $(X_s'X_s)^{-1}$ that arises from the sample, with the scalar g determining how much importance is attributed to the prior precision $\underline{V}_s^{-1}$. Apart from offering computational efficiency, Zellner's g thus reduces the elicitation of the covariance structure to simply choosing the scalar g. The conditional prior on the coefficients

$$\beta_s|\sigma^2, M_s, g \ \sim \ N\big(0, \sigma^2 g(X_s'X_s)^{-1}\big)$$

simplifies the posterior distribution of β such that it follows a $k-$variate student-t distribution with the following density function (for $N > 2$):

$$p(\beta_s|y, X_s, M_s, g) = \frac{\Gamma\left(\frac{N-1+k_s}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)\big((N-1)\pi\big)^{\frac{k_s}{2}}} \left|\Sigma\right|^{-\frac{1}{2}} \times$$

$$\times \left(1 + \tfrac{1}{N-1}(\beta_s - \tfrac{g}{1+g}\hat{\beta}_s)'\Sigma_s^{-1}(\beta_s - \tfrac{g}{1+g}\hat{\beta}_s)\right)^{-\frac{N-1+k_s}{2}} \tag{2}$$

$$\text{where} \quad \Sigma_s = \frac{\breve{y}'\breve{y}}{N-1}\left(1 - \tfrac{g}{1+g}R_s^2\right)\tfrac{g}{1+g}\big(X_s'X_s\big)^{-1}$$

Here, $\breve{y} = y - \mathbf{1}\bar{y}$ is the centered response vector, $\hat{\beta}_s$ the OLS estimator of the coefficient vector, and R_s^2 the OLS R-squared of model M_s. Apart from simplifying the posterior covariance of β_s, g also affects its expected value $E(\beta_s|y, X, M_s, g) = \frac{g}{1+g}\hat{\beta}_s$ which becomes a convex combination of its OLS estimator and the prior expected value (zero). Note that the value of the shrinkage factor $\frac{g}{1+g}$ thus determines the importance of the prior expected value $E(\beta_s) = 0$ with respect to the sample estimates.

Furthermore, Zellner's g yields a simple expression for the marginal likelihood of M_s:[4]

$$p(y|M_s, g) = \int_0^\infty \int_B p(y|\beta_s, \sigma^2, M_s)p(\beta_s, \sigma^2|g)d\beta d\sigma \ \propto (\breve{y}'\breve{y})^{-\frac{N-1}{2}} (1+g)^{-\frac{k_s}{2}} \left(1 - \tfrac{g}{1+g}R_s^2\right)^{-\frac{N-1}{2}} \tag{3}$$

with B denoting the parameter space of the β-coefficient vector. The Bayesian framework further calls for defining prior model probabilities $p(M_j)$ for all models contained in the model space $j \in \{1, 2, \ldots, 2^K\}$. While advocates of purism may call for subjective prior specification of $p(M_s)$, the sheer number of model candidates renders this virtue infeasible. Consequently, most authors have relied on the uniform model prior $p(M_s) = 2^{-K}$, whereas several (Ley and Steel, 2009; Brown et al., 1998; Sala-i-Martin et al., 2004) have proposed

---

[4]Note that although the term $(\breve{y}'\breve{y})^{-\frac{N-1}{2}}$ is constant over models, it is frequently included in the marginal likelihood expression, such as in Fernández et al. (2001a) – while others, such as Liang et al. (2008) omit it.

to specify model priors in dependence of average model size k_s, typically in such a way that prior elicitation is reduced to choosing the prior expected model size. The beta-binomial model prior put forward by Ley and Steel (2009) falls into this category,[5] and we will rely on it in the latter part of our study.

Given the prior model probabilities we can calculate the posterior model probabilities $p(M_s|y, X, g)$ conditional on (y, X, g) serving as the model weights in Bayesian model averaging:

$$p(M_s|y, X, g) = \frac{p(y|M_s, X, g)p(M_s)}{p(y|X)} = \frac{p(y|M_s, X, g)p(M_s)}{\sum_{j=1}^{2^K} p(y|M_j, X, g)p(M_j)}$$

The key component constituting the PMP is the marginal (or integrated) likelihood $p(y|M_s, X)$. The Bayes factor (i.e. the ratio of the marginal likelihoods for two competing models) allows for comparing any two models M_s and M_j by assessing their relative weights:

$$B(M_s : M_j) \equiv \frac{p(y|X, M_s, g)}{p(y|X, M_j, g)} = (1+g)^{\frac{k_j-k_s}{2}} \underbrace{\left(\frac{1 - \frac{g}{1+g}R_s^2}{1 - \frac{g}{1+g}R_j^2}\right)^{-\frac{N-1}{2}}}_{D_{sj}} \tag{4}$$

Multiplied by the prior odds, it yields the posterior odds $p(y|M_s, X, g)/p(y|M_j, X, g) = B(M_s : M_j)p(M_s)/p(M_j)$. Consequently, the posterior model probability given by the product of (3) and the model prior may be expressed as the (nested) Bayes factor with respect to the null model M_0 (times a model prior):

$$p(M_s|y, X) \propto (\breve{y}'\breve{y})^{-\frac{N-1}{2}} (1+g)^{-\frac{k_s}{2}} \left(1 - \frac{g}{1+g}R_s^2\right)^{-\frac{N-1}{2}} p(M_s) \propto B(M_s : M_0)p(M_s) \tag{5}$$

Finally, model averaging comes into play as the marginal posterior distribution of any statistic θ may be obtained as a mixture over posterior model probabilities:

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

This property is particularly useful in computing the posterior moments of the coefficient vector β as a weighted average over all models.[6] Likewise, posterior inclusion probabilities, popular for assessing the importance of single covariates, are obtained as the sum of probabilities for all models in which the covariate is included.

II.1   Popular Settings for Zellner's g

In view of equation (5), BMA inference thus hinges on posterior model probabilities and, in turn, on model priors $p(M_s)$ and marginal likelihoods $p(y|M_s, X)$ (and thus the hyperparameter g). Since the employed beta-binomial model prior already offers a sound statistical

---

[5]Please consult the technical appendix for further details.
[6]Note that we have retained the improper priors for α and σ as common to all models.

framework that aims at minimizing the impact of prior arbitrariness on posterior inference we consequently focus on the characteristics of the prior choice on g.

With respect to the marginal likelihood $p(y|M_s, X_s)$, the discussion has concentrated on the elicitation of g, with two predominant considerations:

- Consistency: The choice of g such that posterior model probabilities asymptotically uncover 'the true model' M_T, i.e. $p(M_T|Y) \to 1$ as $N \to \infty$

- The importance of g as a penalty term enforcing parameter parsimony (the factor $(1+g)^{\frac{k_j-k_s}{2}}$ in (4))

Both issues have already been treated in the classical paper by Fernández et al. (2001a): With respect to consistency, they prove that a choice of $g = w(N)$ such that $\lim_{N\to\infty} w(N) = \infty$ and $\lim_{N\to\infty} \frac{w'(N)}{w(N)} = 0$ ensures consistency as it was mentioned above.

Still, consistency leaves open the exact specification of g, and over the course of more than a decade, various 'automatic' or 'default' specifications have been put forward that typically specify g in dependence of sample size. Mostly, their theoretical foundations build on the penalty term $(1+g)^{-\frac{k_s}{2}}$ in (5) and aim at asymptotically mimicking popular information criteria such as the BIC (cf. for instance Fernández et al. 2001a, p.424). Thus the specification of g was frequently debated in terms of its virtues as a model size penalty term. However from a Bayesian viewpoint, subjective and theoretical considerations on such a penalty should more properly be fused into the formulation of the model prior.[7] It is straightforward to neutralize the factor $(1+g)^{-\frac{k_s}{2}}$ in equation (5) by an appropriate model prior $p(M_j)$ and introduce more or less penalty as one pleases.[8]

In the remainder of this study, we thus try to appreciate g as what it was intended to be: a hyperparameter on the prior distribution of β. Consequently, focus on D_{sj}, the second factor in (4) (as the first factor might be adjusted by model priors): Cet.par., the larger g, the more D_{sj} tends away from unity[9] and the more tightly posterior mass is concentrated on a few 'super models'. The relative distribution of PMPs is therefore crucially affected by the choice of g. Thus the relative magnitude of the g hyperparameter for two competing prior structures will determine strength and direction of the supermodel effect.

With respect to the previously mentioned motivations, several studies have compared the performance of different specifications by means of Monte Carlo simulations, among them Fernández et al. (2001a), Eicher et al. (2009) and Liang et al. (2008). We briefly reiterate the most popular concepts in closely following Liang et al. (2008):

- *Risk Inflation Criterion Prior (g-RIC):* implies setting $g = K^2$. This calibrates the

---

[7]In fact, the discussion of model priors has largely focused on model size penalties, as for instance in Sala-i-Martin et al. (2004).

[8]Consider, for instance, the model prior $p(M_s) = \frac{(1+g)^{k_s/2}}{(\sqrt{1+g}+1)^K}$ that would completely neutralize the factor $(1 + g)^{-\frac{k_s}{2}}$ – and may be combined with other model priors adding more or less model size penalty, respectively.

[9]Note, however, that D_{ij} is bounded according to the values of R_j^2 and R_s^2. Nevertheless the exponent $(N-1)/2$ exacerbates any variations in g to quite a large extent.

posterior model probability to asymptotically match the risk inflation criterion proposed by Foster and George (1994).

- *Unit Information Prior (g-UIP):* In the linear case, it corresponds to $g = N$. It draws on the notion that the 'amount of information' contained in the prior equalize the amount of information in one observation (Kass and Wasserman, 1995). For this setting, Fernández et al. (2001a, p.424) demonstrate that as $N \to \infty$ the log of the Bayes factor in (4) approaches the ratio of the Bayesian information criterion (BIC) for the two models M_s and M_j.

- *Benchmark Prior (g-BRIC):* Corresponds to $g = \max(N, K^2)$. After an extensive study on various specifications for g involving different settings for N, K, and k_s, Fernández et al. (2001a) determined this combination of the g-UIP and g-RIC priors to perform best with respect to predictive performance.

- *Empirical Bayes – Local (EBL):* $g_s = \arg\max_g p(y|M_s, X, g)$. Authors such as George and Foster (2000) or Hansen and Yu (2001) advocate an 'Empirical Bayes' approach by using information contained in the data (y, X) to elicit g. The latter provide a theoretical underpinning for doing so locally, i.e. separately for each model. In the formulation given in Liang et al. (2008), this corresponds to $g_s = \max(0, F_s - 1)$ where F_s is the standard F-statistic for testing the OLS formulation of M_s, with $F_s = \frac{R_s^2(N-1-k_s)}{(1-R_s^2)k_s}$. Note that this formulation frequently raises objections as the data-dependency of g runs counter the intuition of a prior.

Several more 'automatic' specifications for g have been proposed (cf. Fernández et al. 2001a, George and Foster 2000 or Eicher et al. 2009), but the ones above have resurfaced most frequently and also serve as a benchmark reference for Liang et al. (2008).

III   The Hyper-g Prior: A Beta Prior on the Shrinkage Factor

Motivated on theoretical grounds,[10] Liang et al. (2008) introduce two prior distributions on the hyperparameter g. One of them, the so-called 'hyper-g' prior, is particularly interesting in that it provides closed form solutions for almost all posterior statistics of interest.

While Liang et al. (2008) ingeniously outline the basic features of the hyper-g prior, we derive further posterior quantities required for a fully Bayesian analysis. Equations (10)–

---

[10]Liang et al. (2008) motivate their paper with two 'paradoxes' that arise with constant g. First, they raise a BMA formulation of 'Bartlett's paradox' stating that if $g \to \infty$ for fixed N and K, the Bayes Factor $B(M_s : M_0)$ of any model with respect to the null model eventually goes to zero. Second, they refer to an 'information paradox' stating that for fixed N and K, if the R-squared of model M_s converges to unity, its Bayes factor with respect to any other fit-wise inferior model does not go to infinity. Note, however, the comment by Zellner 2008. Moreover, both arguments bite only in the case when N and K are kept constant: Bartlett's paradox in this case may be less relevant as typical specifications for g require it to rise in line with N. The 'information paradox' does not contradict the standard consistency argument that requires the respective Bayes Factor to converge to infinity only when N tends likewise to infinity.

(12) complement their article by providing the posterior distribution of $\beta|y, X$ and its second moments, as well as the second moment of the shrinkage factor.

In addition, equations focus (13)-(15) on facilitating the numerical implementation of the hyper-g prior. Most notably, posterior expressions given in Liang et al. (2008) involve ratios of Gaussian hypergeometric functions, which they propose to compute via Laplace approximations for reasons of computational feasibility. This approach gives rise to numerical inaccuracies, in particular with respect to the mentioned ratios. This section demonstrates how accurate statistics may be achieved in timely fashion by some algebraic transformations. We also show how the original hyper-g prior approach may be reconciled with consistency in the sense of Fernández et al. (2001a) with details given in the appendix.

The hyper-g prior for g translates into a Beta prior on the shrinkage factor $\frac{g}{1+g}$ that is common to all model candidates (Liang et al. (2008, p. 415)):

$$\frac{g}{1+g} \sim \text{Beta}\left(1, \frac{a}{2} - 1\right)$$

i.e. $\frac{g}{1+g}$ is Beta distributed with $E(\frac{g}{1+g}) = \frac{2}{a}$.[11] The elicitation of g is therefore supplanted by the choice of the hyperparameter $a \in (2, \infty)$: $a = 4$ renders the prior distribution of $\frac{g}{1+g}$ uniform, while moving a close to 2 concentrates the prior mass on the shrinkage factor close to 1. Conversely, any $a > 4$ tends to concentrate prior mass near 0. Liang et al. (2008) therefore omit those cases and concentrate on $a \in (2, 4]$ – a strategy we will follow in this study.

The authors derive the posterior distribution of $g|y, X, M_s$ and some posterior statistics by relying on an integral representation for the Gaussian hypergeometric function $_2F_1(a, b, c, z)$ (as, for instance, in Abramowitz and Stegun 1972, p. 563):

$$_2F_1(a, b, c, z) = \frac{\Gamma(c)}{\Gamma(c-b)\Gamma(b)} \int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-tz)^a} dt \quad c > b > 0$$

This integral representation is employed to derive the posterior distribution of g:

$$p(g|y, X_s, M_s) = \frac{k_s + a - 2}{2 \, _2F_1\left(\frac{N-1}{2}, 1, \frac{k_s+a}{2}, R_s^2\right)} (1+g)^{-\frac{k_s+a}{2}} (1 - \frac{g}{1+g} R_s^2)^{-\frac{N-1}{2}} \tag{6}$$

Moreover, the marginal likelihood may be expressed as (Liang et al., 2008, equation (17)):

$$p(y|X_s, M_s) \propto (\breve{y}'\breve{y})^{-\frac{N-1}{2}} \frac{a-2}{k_s + a - 2} \, _2F_1\left(\frac{N-1}{2}, 1, \frac{k_s+a}{2}, R_s^2\right) \tag{7}$$

The posterior expected value of the shrinkage factor is given by (Liang et al. (2008, equation (19))):[12]

$$E\left(\frac{g}{1+g}\bigg|y, X_s, M_s\right) = \frac{2}{k_s + a} \frac{_2F_1\left(\frac{N-1}{2}, 2, \frac{k_s+a}{2} + 1, R_s^2\right)}{_2F_1\left(\frac{N-1}{2}, 1, \frac{k_s+a}{2}, R_s^2\right)} \tag{8}$$

---

[11]Note that this is equivalent to putting the following prior on g: $p(g) = \frac{a-2}{2}(1+g)^{-\frac{a}{2}}$.

[12]Note that $E(\beta_s|y, X_s, M_s) = E\left(\frac{g}{1+g}\bigg|y, X_s, M_s\right)\hat{\beta}_s$.

Under this setting, the posterior expected value of the response under the hyper-g prior for model M_s is given by

$$E(y|X_s, M_s) = \mathbf{1} E(\alpha_s|X_s, M_s) + \mathrm{E}\left(\frac{g}{1+g}\bigg| y, X_s, M_s\right) X_s \hat{\beta}_s \tag{9}$$

with $\hat{\beta}_s$ denoting the OLS estimator for model M_s. From equation (9) the importance of the shrinkage factor becomes evident with the hyper-g prior allowing for model specific, data adaptive shrinkage as opposed to fixing the value for the shrinkage factor a priori.

The posterior statistics outlined so far suffice for the analysis in Liang et al. (2008). However, fully Bayesian inference requires several more expressions, notably with respect to second moments. Therefore, we provide the second moments of $\frac{g}{1+g}$, as well as those of $\beta|y, X$ and its posterior distribution below.[13] The posterior covariance of β_s is given by:

$$\mathrm{Cov}\left(\beta_s|y, X_s, M_s\right) = \frac{2}{k_s + a} \frac{\breve{y}'\breve{y}}{N-2} \frac{{}_2F_1(\frac{N-3}{2}, 2, \frac{k_s+a}{2}+1, R_s^2)}{{}_2F_1(\frac{N-1}{2}, 1, \frac{k_s+a}{2}, R_s^2)} (X_s'X_s)^{-1} \tag{10}$$

This covariance corresponds to the posterior distribution of β_s that may be represented as follows:

$$p(\beta_s|y, X_s, M_s) = \int_0^\infty p(\beta_s|y, X_s, M_s, g) p(g|y, X_s, M_s) dg =$$

$$= \frac{\Gamma\left(\frac{N-1+k_s}{2}\right) \Gamma\left(\frac{k_s+a}{2}\right)}{\Gamma\left(\frac{N-1+k+a}{2}\right)} \frac{N-1}{2} \frac{\sqrt{|X_s'X_s|}}{\pi^{\frac{k_s}{2}}} (\breve{y}'\breve{y})^{\frac{N-1}{2}} (\beta_s'X_s'X_s\beta_s)^{-\frac{N-1+k_s}{2}} \times$$

$$\times \frac{{}_2F_1\left(\frac{N-1+k_s}{2}, \frac{N-1}{2}+1, \frac{N-1+k_s+a}{2}, 1 - \frac{(y-X_s\beta_s)'(y-X_s\beta_s)}{\beta_s'X_s'X_s\beta_s}\right)}{{}_2F_1\left(\frac{N-1}{2}, 1, \frac{k_s+a}{2}, R_s^2\right)} \tag{11}$$

Note that this expression is of close, though not perfect resemblance to a hypergeometric function distribution of type II.[14]

Finally, the second posterior moment of the shrinkage factor is given by:[15]

$$\mathrm{E}\left(\left(\frac{g}{1+g}\right)^2\bigg| y, X_s, M_s\right) = \frac{8}{(k_s+a)(k_s+a+2)} \frac{{}_2F_1(\frac{N-1}{2}, 3, \frac{k_s+a}{2}+2, R_s^2)}{{}_2F_1(\frac{N-1}{2}, 1, \frac{k_s+a}{2}, R_s^2)} \tag{12}$$

The above posterior moments are all characterized by fractions of differing hypergeometric functions. As computing the value of hypergeometric functions is quite involved, this form

---

[13]For completeness, y's posterior predictive distribution is provided in the appendix.

[14]See Guptar and Nagar (2000) for the exact definition of the type II hypergeometric distribution.

[15]Note that with respect to equations (8) and (12) it is straightforward to derive the corresponding expressions for $E(g|y, X_s, M_s)$ and $E(g^2|y, X_s, M_s)$. However $E(g|y, X_s, M_s)$ will only be finite for $k_s + a > 4$ and $E(g^2|y, X_s, M_s)$ only for $k_s + a > 6$. We therefore concentrate on the posterior moments of the shrinkage factor.

renders numerical implementation extremely difficult in terms of computation time and accuracy. However, they may all be expressed in dependence of $F_s^* \equiv {}_2F_1(\frac{N-1}{2}, 1, \frac{k_s+a}{2}, R_s^2)$ using Gauss' relations for contiguous hypergeometric functions (Abramowitz and Stegun, 1972, p.563). Let $\bar{N} \equiv N - 3$ and $\bar{\theta}_s \equiv k_s + a - 2$ represent collected terms. Tedious, but straightforward algebra then yields the following results for (8),(10), and (12) (as long as $R_s^2 > 0$):[16]

$$\mathrm{E}\left(\frac{g}{1+g}\middle| y, X_s, M_s\right) = \frac{1}{R_s^2(\bar{N} - \bar{\theta}_s)}\left(\frac{\bar{\theta}_s}{F_s^*} - \bar{\theta}_s + \bar{N}R_s^2\right) \tag{13}$$

$$\mathrm{Cov}(\beta|y, X_s, M_s) = \frac{\breve{y}'\breve{y}}{N-2}(X'X)^{-1}\frac{\bar{N}}{(\bar{N}-\bar{\theta}_s-1)^2-1}\frac{1-R_s^2}{R_s^2}\times$$
$$\times\left(\left(1 + \frac{2}{\bar{N}}\frac{R_s^2}{1-R_s^2}\right)\frac{\bar{\theta}_s}{F_s^*} + ((\bar{N}-2)R_s^2 - \bar{\theta}_s)\right) \tag{14}$$

$$\mathrm{E}\left(\left(\frac{g}{1+g}\right)^2\middle| y, X_s, M_s\right) = \frac{1}{(R_s^2)^2(\bar{N}-\bar{\theta}_s)(\bar{N}-(\bar{\theta}_s+2))}\times$$
$$\times\left(((\bar{N}-2)R_s^2 - (\bar{\theta}_s+2))\frac{\bar{\theta}_s}{F_s^*} + \left(\bar{N}R_s^2 - \bar{\theta}_s\right)^2 - 2\left(\bar{N}(R_s^2)^2 - \bar{\theta}_s\right)\right) \tag{15}$$

Note that the equations above all contain the term $\bar{\theta}_s/F_s^*$ which is just $2/(a-2)$ times the integration constant of $p(g|y, X_s, M_s)$ or $\frac{a-2}{BF(M_s:M_0)}$ where $BF(M_s : M_0)$ is the null-based Bayes Factor for model M_s. So for each model's statistics, a hypergeometric function (or its Laplace approximation) has to be computed only once, which benefits numerical implementation in terms of computation speed.

Moreover, equations (13)-(15) reveal a certain resemblance to the respective posterior statistics under the 'Empirical Bayes - Local' (EBL) approach as outlined in section II.1: the main difference is the term $\bar{\theta}_s/F_s^*$, which guarantees non-negativity for the above expressions. Considering that the models associated with very low $\bar{\theta}_s/F_s^*$ (and thus high PMP) are disproportionally weighted into model averaging, this term thus virtually disappears from model-averaged statistics.[17] Moreover, the marginal model likelihood in (7) does not differ too far from its equivalent under EBL.[18] Section IV illustrates this effect in showing that hyper-g results differ far less from EBL than from settings under constant g.

One virtue of the hyper-g prior lies in the fact that the posterior distribution of the shrinkage factor $\frac{g}{1+g}$ can be interpreted in terms of goodness-of-fit. Equation (13) presents its model-specific expected value as close to $1 - 1/\hat{F}_s$, where $\hat{F}_s$ represents an adjusted OLS F-statistic

---

[16]In case $R_s^2 = 0$ (in particular for the null model), the respective quantities follow directly from (8),(10), and (12) since ${}_2F_1(a, b, c, 0) = 1$ for any (a, b, c).

[17]Note that ${}_2F_1(\frac{N-1}{2}, 1, \frac{k_s+a}{2}, R_s^2)$ goes quite rapidly towards infinity as R_s^2 increases. The term $\bar{\theta}_s/F_s^*$ could thus noticeably affect model-averaged posterior moments only in case the data examined offers a very low signal-to-noise ratio.

[18]Please refer to section A.2 in the appendix for a theoretical underpinning of this claim.

for the model M_s: $\hat{F}_s = \frac{R_s^2(\bar{N}-\bar{\theta}_s)}{(1-R_s^2)\bar{\theta}_s}$. Larger values of the shrinkage factor hence correspond to more variance explained by the model M_s.

The model-averaged expected value of the shrinkage factor $E(\frac{g}{1+g}|y,X)$ may be interpreted likewise. As long as there are some Bayes factors considerably larger than one, the following inequality holds:[19]

$$\frac{1}{1-E(\frac{g}{1+g}|y,X)} \leq \frac{R_F^2}{(1-R_F^2)}\frac{(N-E(k|y,X)-a-1)}{(E(k|y,X)+a-2)} \tag{16}$$

where R_F^2 is the OLS R-squared of the 'full model' with K regressors, and $E(k|y,X)$ is the expected posterior model size. The right-hand side thus constitutes a pseudo F-statistic that relates R_F^2 with the 'number of parameters' $E(k|y,X)+a-2$. It thus forms an upper bound for the 'goodness-of-fit' that can be achieved by BMA. Additionally, (16) gives rise to the following inequality, which may serve to establish a relationship to the classical interpretation of R-squared:

$$R_F^2 \geq \frac{E(k|y,X)+a-2}{N-E\left(\frac{g}{1+g}|y,X\right)\left(N-E(k|y,X)-a-1\right)}$$

Finally, the hyperparameter a can still be trimmed to represent prior beliefs on the shrinkage factor. It is straightforward, for instance, to specify the prior beliefs such that the expected shrinkage factor matches the expressions laid out in section II.1. In general, most popular settings for g can thus be emulated by $a = 2 + 2/w(N)$, with $w(N) > 0$ and $\lim_{n\to\infty} w(N) = \infty$, thus positioning the prior expected value at $E(\frac{g}{1+g}) = \frac{w(N)}{1+w(N)}$. Note that in this case, 'consistency' in the sense of Fernández et al. (2001a, p.6) is ensured for the 'hyper-g' prior[20] (cf. section A.1 in the appendix).

In this light, we propose the following specifications for the prior beliefs on the shrinkage factor:

- *HG-UIP*: $a = 2 + \frac{2}{N}$ corresponds to the 'g-UIP'-shrinkage factor with $E\left(\frac{g}{1+g}\right) = \frac{N}{1+N}$. Then 95% of the prior mass on the shrinkage factor is contained in the interval $[1 - 0.95^N, 1]$

- *HG-RIC*: $a = 2 + \frac{2}{K^2}$ corresponds to 'g-RIC'-shrinkage with $E\left(\frac{g}{1+g}\right) = \frac{K^2}{1+K^2}$. In this case 95% of the prior mass is contained in the interval $[1 - 0.95^{K^2}, 1]$

Similarly, other specifications akin to 'classic' g formulations could be implemented – as long as they depend on N as defined above, in order to retain asymptotic consistency. However, as posterior expressions are quite insensitive to the value of a, and most of these formulations will lead to a close to 2, the resulting posterior statistics will be virtually identical. We therefore limit our attention to the two specifications above.

---

[19]Even though this inequality will hold in virtually all relevant cases, it may not hold in case of less dependent-covariate correlation than expected under a null hypothesis of no relation. As a rule of thumb, $R_F^2 > \frac{K+a-2}{/}N - 3$ is sufficient for (16) to hold in any case. Please refer to section A.3 in the appendix for further details.

[20]Consistency does not directly apply to the g-RIC prior outlined below. However, throughout the following sections, g-RIC is in practice identical with the g-BRIC prior (as always $K^2 > N$). Since the latter qualifies for consistency, the notion may be extended to g-RIC, at least in our case.

IV A Simulation Study

In this section we carry out a simulation study empirically investigating the supermodel effect as well as assessing predictive performance of selected prior structures. We can broadly distinguish two classes of prior settings, the degenerate ones fixing g values (*fixed* prior settings) as opposed to model specific and data dependent g prior structures (*flexible* prior settings). In the following, we concentrate on the 8 prior structures given in Table 1.

Fixed Prior Settings	
g-RIC	Risk inflation criterion, $g = K^2$.
g-UIP	Unit information prior, $g = N$.
g-E$(\frac{g}{1+g}|Y)$	$\frac{g}{1+g}$ is set to the posterior mean under the HG-4 prior (i.e. $E(\frac{g}{1+g}|Y)$).

Flexible Prior Settings	
EBL	Local empirical Bayes estimate of g.
HG-3	Hyper-g prior with $a = 3$.
HG-4	Hyper-g prior with $a = 4$.
HG-RIC	Hyper-g prior with $a = 2 + 2/K^2$.
HG-UIP	Hyper-g prior with $a = 2 + 2/N$.

Table 1: Definition of Prior Settings.

The first two fixed settings correspond to what Fernández et al. (2001a) coined the "benchmark" prior and is widely used in applied work.[21] In these the suggestion made by Fernández et al. (2001a) often results in the g-RIC prior and will hence serve as our reference prior. The implied (large) value for g under g-RIC is expected to have two consequences: first g-RIC will favor parsimonious models, and second posterior mass will be concentrated on a small set of models.[22] The unit information prior and the g-E$(\frac{g}{1+g}|Y)$ complete the set up for fixed prior structures on g. For the latter we impose $\frac{g}{1+g}$ a-priori to equate the (model weighted) posterior mean of $(\frac{g}{1+g}|Y)$ under the HG-4 setting (the hyperprior with $a = 4$). We have chosen this particular prior structure to exemplify the impact of adaptive shrinkage: a prior both priors are expected to be very similar in posterior mass distribution. However, posterior results are expected to seriously differ regarding the assignment of PMPs. In principle, the g-E$(\frac{g}{1+g}|Y)$ setting will favor more strongly parsimonious models and those with comparably small posterior support under the HG-4 due to keeping g constant.

Secondly, we propose more flexible prior structures that embody model specific g values and data dependent shrinkage. In particular these settings are the local empirical Bayes estimates and the hyper-g prior corresponding to a fully Bayesian approach. One strength in placing a prior on g lies in the fact that we can incorporate our prior beliefs following the rules of Bayesian statistics[23] via the hyperparameter a. For the simulation study, we devise

---

[21]See for example Fernández et al. (2001b), Masanjala and Papageorgiou (2008) and Koop and Potter (2003) among others.

[22]This facilitates quick convergence of stochastic search algorithms such as the MC^3 to the target distribution.

[23]See Laud and Ibrahim (1995) for a model selection approach designing information criteria that allow

four different values for a: HG-3 (a=3) corresponds to a prior expected shrinkage factor of $\frac{2}{3}$, whereas HG-4 (a=4) corresponds closely to a uniform prior over the shrinkage factor. We contrast these two settings with two settings calibrated to match the g-RIC and g-UIP prior structure (HG-RIC, HG-UIP). That is, the prior expected value of the shrinkage factor $E(\frac{g}{1+g})$ conforms to the shrinkage factors induced by g-RIC (K^2) or g-UIP (N).

Data-wise, we employ two different settings, where the first set up "A" is as in Fernández et al. (2001a). Each Monte Carlo run draws 10 potential explanatory variables $(\mathbf{x}_1, \ldots, \mathbf{x}_{10})$ with $N =100$ observations from a standard normal distribution for each covariate. Additional 5 variables are generated by multiplying the first five regressors by $[0.3, 0.5, 0.7, 0.9, 1.1]$ inducing a correlation structure among the covariates. Note that this correlation structure impedes uncovering the data generating model.

The second set-up "B" is more demanding since the data generating process cannot be traced back to a single model. This is more in line with Bayesian thinking whose question is not whether the preferred model is perfectly true (to which the answer is no), but whether under the assumed model(s) the observed data is a plausible outcome.[24] The data generating process is composed of 5 partially nested models with unequal model weights imposed. This creates a "hierarchy" of models with y_4 and y_5 relatively dominating the remaining models in terms of explained variation.

Setup "A": $\quad y = 4 + 2\mathbf{x}_1 - \mathbf{x}_5 + 1.5\mathbf{x}_7 + \mathbf{x}_{11} + 0.5\mathbf{x}_{13} + \sigma\varepsilon$

Setup "B": $\quad y = 0.2y_1 + 0.1y_2 + 0.1y_3 + 0.3y_4 + 0.3y_5$
$\qquad\qquad\quad y_1 = 4 + 2\mathbf{x}_1 - \mathbf{x}_5 + 1.5\mathbf{x}_7 + \mathbf{x}_{11} + 0.5\mathbf{x}_{13} + \sigma\varepsilon$
$\qquad\qquad\quad y_2 = 4 + 4\mathbf{x}_1 - \mathbf{x}_5 + 1.5\mathbf{x}_2 + \mathbf{x}_8 + 0.5\mathbf{x}_{11} + \sigma\varepsilon$
$\qquad\qquad\quad y_3 = 4 + 1\mathbf{x}_5 - \mathbf{x}_7 + 1.5\mathbf{x}_3 + \mathbf{x}_9 + 0.5\mathbf{x}_6 + \sigma\varepsilon$
$\qquad\qquad\quad y_4 = 4 + 2\mathbf{x}_1 - \mathbf{x}_2 + 1.5\mathbf{x}_4 + \mathbf{x}_7 + 0.5\mathbf{x}_6 + \sigma\varepsilon$
$\qquad\qquad\quad y_5 = 4 + 2\mathbf{x}_1 - \mathbf{x}_{10} + 1.5\mathbf{x}_{11} + \mathbf{x}_{12} + 0.5\mathbf{x}_{13} - 2\mathbf{x}_{14} + \sigma\varepsilon$

Posterior inference under the different prior structures will be examined with varying signal-to-noise ratios. In particular we conduct the simulation study for four increasing levels of noise: $\sigma = 1/2, \sigma = 1, \sigma = 2.5, \sigma = 5$. The relatively low number $K = 15$ allows for enumerating the model space and basing posterior inference on the results of the full set of 2^K models. This guarantees that differences of results for the competing priors are not influenced by additional variation due to stochastic search.

Applied research often focuses on the posterior inclusion probabilities (PIPs) of the variables entering the analysis and the posterior moments of the related coefficients. Table 2 and 3 show PIPs for setting "A", averaged over 50 Monte Carlo draws (standard deviations in parentheses): Under situations characterized by small degrees of noise ($\sigma = 1/2$ and $\sigma = 1$) results do not differ considerably between fixed and data dependent priors for g. Under the $\sigma = 2.5$ setting, PIPs corresponding to coefficients of the data generating model exhibit differences in magnitude but still lead to the same interpretation. Results change when looking at the $\sigma = 5$ setting. Posterior mass under the flexible priors is spread more evenly as with fixed priors employed. The g-RIC prior shows strong support for the first variable,

---

for the input of prior knowledge.

[24]See for example Gelman et al. (1995).

with a large PIP for β_1 of approximately 80%. The remaining variables receive negligible posterior support tempting the researcher to believe that the data generating process is solely driven by the first variable. In contrast, flexible priors still 'identify' all variables. As expected, mass is spread more evenly, and over larger models, resulting in a high share of covariates with PIP close to 0.5 - which reflect the serious degree of noise in the data.

Besides the PIP, the posterior model probability of the model that was used in generating the data can be of interest to examine consistency properties in the sense of Fernández et al. (2001a). Tables 6 and 7 show summary statistics for the posterior model probabilities based on the 50 Monte Carlo draws. In line with the asymptotic results, more information in the data leads the hyperprior to uncover the data generating process with highest precision, whereas increasing noise deteriorates the selection ability of BMA for all settings. The ratio of the posterior model probability for the data generating process to the one with highest PMP is given in Table 7. The results show that in situations described by higher degrees of noise in the data all specifications favor a model different from the one generating the data. However, the deterioration of PMPs coupled with a surge in the PMP ratio of true to best model for flexible priors when noise increases indicated that mass is more diluted. In other words, while flexible priors fail to uncover the data generating model (as do the fixed ones) the assigned PMP for the best model is small as compared to fixed priors. Hence the surge of uncertainty is reflected in posterior mass distribution. Figures 1 and 2 exemplify the differences in PMP ascription for the 8 priors. The first figure shows the cumulative posterior mass of the 100 best models under the four signal-to-noise settings, averaged over 50 Monte Carlo draws. From the picture and the figures from Table 6 it becomes evident that flexible priors uncover the data generating model with highest precision *and* concentrate most mass on this model(s) in situations characterized by a high degree of information provided by the data. This means that the posterior mean $E(\frac{g}{1+g}|y, X)$ is larger for the flexible priors than the constant values $\frac{g}{1+g}$ under the fixed priors. As noise increases, the flexible priors distribute mass more evenly among explanatory variables reflecting the surge of uncertainty. In contrast, approaches fixing g are not capable of adjusting posterior mass distribution to uncertainty inherent in the data. The merits of Bayesian model averaging regarding handling of model uncertainty and predictive abilities are thus limited in these settings. Figure 2 shows a QQ-plot for the prior settings with the g-RIC specification as the reference prior. For all data dependent priors we see that differences increase with noise as is expected.[25]

Under setting "B" we view the employed models rather as approximations and uncovering a "true" model is of minor importance. The results exemplify once again the supermodel effect behavior of fixed prior settings illustrated in Figures 3 and 4. Small degrees of noise trigger a concentration of posterior mass under the hyper-g prior and the empirical Bayes approach. A surge in noise is reflected in a wider spread of posterior mass among models under flexible priors, whereas fixed priors still concentrate on a small number of models.

Finally we draw attention to the shrinkage factor's role in weighting prior versus posterior coefficients and its implications for prediction. Results from a prediction exercise are expected to vary considerably between fixed and flexible prior settings, since the latter incorporate data adaptive shrinkage. Akin to Liang et al. (2008) we calculate the root mean squared error (rmse) based on 30 out of sample observations, averaged over 50 Monte Carlo steps.

---

[25]We have omitted results from the HG-3 setting, since results are very similar to that of HG-4.

The rmse statistics shown in Table 8 are normalized with respect to forecasting results under the g-RIC prior. Thus values below 1 indicate better predictive performance (in terms of accuracy) of the respective prior structure as compared to the g-RIC prior. The top panel of Table 8 shows mixed results for setting "A". As expected, the g-RIC prior excels in nearly all signal-to-noise settings, concentrating on a single (and luckily the correct data-generating) model. In the $\sigma = 1/2$ test bed, however, the flexible priors concentrate mass even more tightly than does the g-RIC and consequently yield comparably better predictions in terms of rmse. As noise increases the g-RIC outperforms the other priors by greater margins exploiting the comparative advantage that the data generating process is composed of a single model.

In contrast, predictive merits of flexible priors are more pronounced in the complex data generating process structure of setting "B": Predictive abilities differ by a greater margin, with flexible priors nearly dominating throughout all signal-to-noise settings. Especially the HG-3 prior and the empirical Bayes approach demonstrate superior predictive abilities with the latter one outperforming the g-RIC prior for all signal-to-noise setups. This contrasts with simulation exercises in the literature, as their data generating processes can be traced back to single models, which plays in favor of (large) fixed priors because of the supermodel effect.

V   Growth Determinants Revisited

We now apply the different prior settings in order to examine growth determinants in a cross country growth data set. There is a dense empirical growth literature that has employed model averaging techniques.[26] We use the data set given in Fernández et al. (2001b) and described in Sala-i-Martin (1997). Following Fernández et al. (2001b) we use 41 potential growth determinants for 72 countries. The data comprises proxy variables for human capital, institutional quality indicators, investment variables and regional dummy variables. For the sake of comparison we employ uniform model priors instead of the beta-binomial model prior. Results on the posterior inclusion probabilities are shown in Table 9, where we have used 3,000,000 posterior draws after a burn-in phase of 2,000,000 draws.[27] Variables with posterior inclusion probabilities greater than 0.5 are often identified as 'robustly related' to the dependent variable.[28] However, note that such a threshold should be allowed to vary with the information content of the data, as well as with the implied model size penalty of model priors. As with our model prior structure, the model size penalty term in the marginal likelihoods has not been adjusted for: a larger posterior shrinkage factor thus corresponds to more weight being put on a small subset of parsimonious models. A PIP that only

---

[26]See Sala-i-Martin et al. (2004), Crespo Cuaresma and Doppelhofer (2007), Sala-i-Martin (1997) for traditional approaches to model averaging and Fernández et al. (2001b), Eicher et al. (2009) for Bayesian strategies.

[27]The computer program is coded in R (R Development Core Team, 2008) and is available from the authors upon request.

[28]Eicher et al. (2009) translate the scale of evidence put forward originally by Kass and Raftery (1995) into four categories: weak (50-75% PIP), substantial (75-95%), strong (95-99%) and decisive (99%+) evidence. Barbieri and Berger (2003), on the other hand, highlight the predictive merits of the median model, consisting of those regressors whose PIP exceeds 0.5.

slightly exceeds the 0.5 threshold coupled with a rather small posterior mean of $\frac{g}{1+g}$ cannot be interpreted in the same way as in the case when $E(\frac{g}{1+g}|y, X)$ is large. In particular this applies to comparing results with the g-RIC prior, under which the (posterior) shrinkage factor is much larger than otherwise.

The flexible prior structures identify a range of additional growth determinants as compared to the g-RIC prior setting manifested in the differences of the posterior mean model sizes given at the bottom of the table. Figure 5 top right panel illustrates the behavior under the different prior structures. Due to the degree of noise in the date set, flexible priors distribute posterior mass more evenly than fixed priors, in particular the g-RIC setting. The following figures exemplify the variation of posterior inference for the 8 prior settings with variables on the X-axis ordered according to the posterior inclusion probabilities under the g-RIC setting. The figures reveal the remarkably small differences within the class of flexible priors, again illustrating the close inter-relatedness of this group. By virtue of their smaller g-values, the g-E$(\frac{g}{1+g}|Y)$ and the g-UIP are the fixed priors coming closest to the results of the flexible priors, whereas the g-RIC is far off.

Moreover, the graphs show that inference under the hyper-g priors is insensitive to the prior choice on the hyperparameter a. Due to the supermodel effect and model size penalty, the g-RIC prior identifies a smaller subset of growth determinants. However, this holds not true for all variables: the number of years an economy has been open (YrsOpen) looses significance under the flexible priors with the g-RIC setting being the only prior structure identifying the variable as an important growth determinant. The distribution of posterior mass under the employed prior structures is shown in the top left panel of Figure 5. Due to the degree of noise in the data, the flexible priors distribute posterior mass more evenly than the fixed ones. Besides the (robust) identification of growth drivers we are interested in the posterior means and the standardized coefficients (i.e. posterior mean / posterior standard deviation) as a further significance indicator given in the bottom panel of Figure 5.[29] Disparities in posterior means emanate from differences in the posterior inclusion probabilities *and* magnitudes of the shrinkage factor among the prior settings. As expected for some variables (a regional dummy (Hindu) and two proxy variables for human capital (HighEnroll and PublicEducpt)) posterior means vary considerably as compared to the g-RIC setting.

Variations of posterior inference are solely due to differences in the value of the g hyperparameter, thus once again emphasizing the importance of this parameter for Bayesian model averaging inference. For the growth data exercise, the (posterior) shrinkage factor varies from 0.999 (g-RIC) to 0.951 (HG-4). Fixing g - in our view - bears the danger of ignoring the information in the data and exerts non-negligible influence on posterior results.

VI   Concluding Remarks

The ubiquity of Zellner's g prior in linear BMA rests on two reasons: it provides closed-form solutions and reduces the complexity of prior elicitation to, in practice, one scalar g.

---

[29]Table 10 lists results on the posterior means for the whole set of prior structures. Note that we have fully (response variable and covariates) standardized the coefficients implying that the slopes refer to changes in terms of standard deviations.

Consequently, theoretical considerations have mostly focused on the choice of g, in particular its virtues as a penalty term for model size.

This study deviates in bringing forward two arguments that have been overlooked so far: First, model size considerations should be decoupled from the prime feature of g (scaling coefficient covariance) and more properly be fused into the formulation of model priors. The elicitation of g should thus not interfere with prior desiderata on model size.

Second, we demonstrated that fixing g to arbitrary values may have unintended consequences on posterior model probabilities: The higher g, the more tightly posterior mass will concentrate on the few best-performing 'super models' – regardless of model sizes, number of observations or signal-to-noise ratios. Ultimately, a large value for g will favor a single model, thereby acting in a model-selection fashion. As previous studies predominantly have assessed BMA performance on simulated data generated by a single model, they tended to favor g-specifications ascribing larger values to g, that effectively *select* the right model. We demonstrate in section IV that the g-RIC prior suggested by Fernández et al. (2001a) is particularly prone to this supermodel behavior.

In order to overcome problems, we propose to put a prior distribution on the g parameter: Such a hyperprior allows for data-dependent shrinkage, thus adjusting the weight of prior beliefs more properly according to data quality. In discriminating models only as far as data quality allows for, a prior on g thus offers a remedy for the supermodel effect. In this manner, we focus on the *hyper-g* prior introduced by Liang et al. (2008), whose formulation offers three main advantages: First, it admits closed form solutions for almost any quantity of interest, thereby facilitating implementation. Second, it allows for BMA consistency. Third, its hyperparameter allows for formulating prior beliefs on coefficient variance, but without incurring the risk of unintended consequences on posterior model mass. We complement the existing literature on the hyper-g prior by providing additional posterior expressions that allow for fully Bayesian inference, as well as for sound numerical implementation.

Section IV contrasts various formulations of fixed and hyper-g priors in simulations, concentrating on predictive performance under varying signal-to-noise ratios. As expected the fixed (especially the g-RIC) priors perform considerably well when the data generating process rests on a single model that is part of the candidate model space. However, in more complex settings, the virtues of flexible prior structures become pronounced: Flexible priors outperform fixed g settings (in particular g-RIC) in terms of forecasting accuracy and exhibit a more stable structure of posterior model and inclusion probabilities as noise varies.

The final section illustrates these considerations by applying the same priors to a prominent growth data set. The results demonstrate that fixing g runs the risk of grossly over- or understating the importance of some variables – the degree of openness, for instance, is not as important to growth as one may think under the g-RIC prior. In this data set, fixing g to values larger than implied by flexible priors leads to stronger discrimination among posterior inclusion probabilities, which may incite overconfidence in BMA results. Finally, the magnitudes of several coefficients differ markedly between fixed and hyper-g priors, but are negligible among the hyper-g prior structures.

Concluding, the hyper-g prior offers a sound, fully Bayesian approach that features the virtues of prior input and predictive gains without incurring the risk of misspecification.

References

Abramowitz, M. and Stegun, I. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* National Bureau of Standards Applied Mathematics Series 55. Tenth Printing.

Barbieri, M. M. and Berger, J. O. (2003). Optimal Predictive Model Selection. *Ann. Statist.*, 32:870–897.

Bernardo, J. and Smith, A. (1994). *Bayesian Theory.* John Wiley and Sons, New York.

Brown, P., Vannucci, M., and Fearn, T. (1998). Multivariate Bayesian Variable Selection and Prediction. *Journal of the Royal Statistical Society B*, 60:627–641.

Chipman, H., George, E., and McCulloch, R. (2001). *The Practical Implementation of Bayesian Model Selection.* Institute of Mathematical Statistics Lecture Notes-Monograph Series Vol. 38. Beachwood, Ohio.

Crespo Cuaresma, J. and Doppelhofer, G. (2007). Nonlinearities in Cross-Country Growth Regressions: A Bayesian Averaging of Thresholds (BAT) Approach. *Journal of Macroeconomics*, 29:541–554.

Cui, W. and George, E. (2008). Empirical Bayes vs. fully Bayes variable selection. *Journal of Statistical Planning and Inference*, 138:4:888–900.

Eicher, T., Papageorgiou, C., and Raftery, A. (2009). Determining growth determinants: default priors and predictive performance in Bayesian model averaging. *Journal of Applied Econometrics, forthcoming.*

Eklund, J. and Karlsson, S. (2007). Forecast Combination and Model Averaging using Predictive Measures. *Econometric Reviews*, 26:329–362.

Fernández, C., Ley, E., and Steel, M. F. (2001a). Benchmark Priors for Bayesian Model Averaging. *Journal of Econometrics*, 100:381–427.

Fernández, C., Ley, E., and Steel, M. F. (2001b). Model Uncertainty in Cross-Country Growth Regressions. *Journal of Applied Econometrics*, 16:563–576.

Foster, D. P. and George, E. I. (1994). The Risk Inflation Criterion for Multiple Regression. *The Annals of Statistics*, 22:1947–1975.

Gelman, A., Carlin, J. B., Stern, S. H., and Rubin, B. D. (1995). *Bayesian Data Analysis.* Chapman & Hall.

George, E. and Foster, D. (2000). Calibration and empirical Bayes variable selection. *Biometrika*, 87(4):731–747.

Guptar, A. K. and Nagar, D. K. (2000). *Matrix Variate Distributions.* Chapman & Hall /CRC, Monographs and Surveys in Pure and Applied Mathematics, 104.

Hansen, M. and Yu, B. (2001). Model selection and the principle of minimum description length. *Journal of the American Statistical Association*, 96(454):746–774.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science*, 14, No. 4:382–417.

Kass, R. and Raftery, A. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90:773–795.

Kass, R. and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, pages 928–934.

Koop, G. and Potter, S. (2003). Forecasting in Large Macroeconomic Panels Using Bayesian Model Averaging . *FRB NY Staff Report*, 163.

Laud, P. and Ibrahim, J. (1995). Predictive model selection. *Journal of the Royal Statistical Society*, Series B 57:247–262.

Ley, E. and Steel, M. F. (2009). On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regressions. *Journal of Applied Econometrics*, 24:4:651–674.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g Priors for Bayesian Variable Selection. *Journal of the American Statistical Association*, 103:410–423.

Masanjala, W. and Papageorgiou, C. (2008). Rough and Lonely Road to Prosperity: A Re-examination of the Sources of Growth in Africa Using Bayesian Model Averaging. *Journal of Applied Econometrics*, 23:671–682.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0 edition. http://www.R-project.org.

Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, 25:111–163.

Sala-i-Martin, X. (1997). I Just Ran 2 Million Regressions. *American Economic Review*, 87:178–183.

Sala-i-Martin, X., Doppelhofer, G., and Miller, R. I. (2004). Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. *American Economic Review*, 94:813–835.

Strachan, R. and van Dijk, H. (2004). Exceptions to Bartletts Paradox. *Keele Economic Research Papers*.

Zellner, A. (1986). *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, chapter On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions. North-Holland: Amsterdam.

Zellner, A. (2008). Comments on Mixtures of g-priors for Bayesian Variable Selection by F. Liang, R. Paulo, G. Molina, MA Clyde and JO Berger.

A   Technical Appendix

A.1   Consistency of the Hyper-g Prior

Fernández et al. (2001a) define asymptotic 'consistency' as follows: Consider that only Model M_s is true, while all other models $M_j \neq M_s$ are not true. Consistency then requires:

$$\plim_{n\to\infty} p(M_s|y, X_s) = 1 \quad \text{and} \quad \plim_{n\to\infty} p(M_j|y, X_s) = 0 \ \forall M_j \neq M_s$$

Liang et al. (2008, Appendix B) have proven the above for the hyper-g prior except for the case where the true model M_s is the null model M_0. They stop short their proof because in this case the Bayes factor $B(M_j : M_0)$ is (Liang et al., 2008, p.423):

$$\frac{p(M_j|y, X_s)}{p(M_0|y)} \geq \int_0^\infty (1+g)^{-\frac{k_j}{2}} \, p(g) dg \tag{A.1}$$

Moreover they state that if the above integral vanishes as $N \to \infty$, then consistency is ensured.

Applying the hyper-g setting transforms the right-hand side in (A.1) into the following (by $a > 2$):

$$\int_0^\infty (1+g)^{-\frac{k_j}{2}} \, p(g) dg = \frac{a-2}{2} \int_0^1 (1+g)^{-\frac{k_j+a}{2}} dg = \frac{a-2}{k_j + a - 2}$$

If $a = 2 + w(N)$ with $w(N) > 0$ and $\lim_{N\to\infty} w(N) = 0$, then the integral vanishes and thus concludes the proof.

A.2   Relationship between Hyper-g Prior and EBL

Due to perceived numerical difficulties, Liang et al. (2008) propose the use of an Laplace approximation for the posterior model likelihood under the hyper-g distribution (Liang et al. (2008, equation (17))). Depending on the data, Laplace approximations can be prone to substantial numerical inaccuracies. However, they may be useful for the purpose of this section, namely a tentative approach to establishing a rough equivalence between particular forms of the hyper-g prior and Empirical Bayes with respect to posterior statistics.
Consider the familiar form of the Laplace approximation

$$\int_\Theta \exp(h(\theta)) d\theta \approx \sqrt{\frac{2\pi}{-h''(\hat{\theta})}} \exp h(\hat{\theta})$$

. . . where $\hat{\theta}$ is the maximizer of the integrand' logarithm $h(\theta)$. Consider in turn the null-based model Bayes Factor for the hyper-g prior formulation as in (7):

$$BF_h = \frac{a-2}{2} \int_0^\infty (1+g)^{\frac{N-1-k-a}{2}} (1+g(1-R^2))^{-\frac{N-1}{2}} dg$$

Letting

$$h(g) = \frac{1}{2} \left((N-1-k-a)\log(1+g) - (N-1)\log(1+(1-R^2)g) \right)$$

yields the maximizer:

$$\hat{g} = \max\left(\frac{R^2(N-1-k-a)}{(1-R^2)(k+a)} - 1, 0\right)$$

where $\hat{g} = 0$ if and only if $k + a \geq R^2(N-1)$. Liang et al. (2008, p.421) note the similarity to the local Empirical Bayes (EBL) estimator of g, but abstain from further investigating the issue.

The second derivative of $h(g)$ is given as

$$h''(g) = \frac{1}{2}\left(-\frac{N-1-k-a}{(1+g)^2} + \frac{(N-1)(1-R^2)^2}{(1+(1-R^2)g)^2}\right)$$

The Bayes factor is thus approximately equal to:

$$BF_h \approx (a-2)\sqrt{\frac{\pi}{\frac{N-1-k-a}{(1+\hat{g})^2} - \frac{(N-1)(1-z)^2}{(1+(1-z)\hat{g})^2}}}(1+\hat{g})^{\frac{N-1-k-a}{2}}(1+\hat{g}(1-R^2))^{-\frac{N-1}{2}}$$

In case we have $\hat{g} > 0$, then algebraic manipulation of the expression above yields:

$$BF_h \approx (a-2)\sqrt{\pi}\sqrt{\frac{N-1}{(N-1-k-a)(k+a)}}\left(\frac{R^2}{(1-R^2)}\frac{N-1-k-a}{k+a}\right)^{-\frac{k+a-2}{2}}\left(\frac{(1-R^2)(N-1)}{N-1-k-a}\right)^{-\frac{N-1}{2}}$$

Now consider the equivalent null-based model Bayes Factor for the EBL approach which is:

$$BF_{EBL} = \left(\frac{R^2}{1-R^2}\frac{N-1-k}{k}\right)^{-\frac{k}{2}}\left(\frac{(1-R^2)(N-1)}{(N-1-k)}\right)^{-\frac{N-1}{2}}$$

in case if $k \leq R^2(N-1)$

Therefore, if $k + a \leq R^2(N-1)$:

$$BF_h \approx (a-2)\sqrt{\pi}\sqrt{\frac{N-1}{(k+a)(N-1-k-a)}}\left(\frac{(1-z)}{z}\frac{k+a}{N-1-k-a}\right)^{\frac{a-2}{2}}\left(\frac{k+a}{k}\right)^{\frac{k}{2}}\left(\frac{N-1-k-a}{N-1-k}\right)^{\frac{N-1-k}{2}}BF_{EBL}$$

So if $a \to 2$, the hyper-g Bayes Factor is approximatively equivalent to an EBL Bayes factor times a k-based model prior (that does not depend on z). Moreover, this model prior is bounded in a relatively narrow range: Note that

$$(1+a)^{-a/2} \leq \left(\frac{k+a}{k}\right)^{\frac{k}{2}}\left(\frac{N-1-k-a}{N-1-k}\right)^{\frac{N-1-k}{2}} < 1$$

The upper bound follows from the fact that $((k+a)/k)^{k/2} = (1+\frac{a}{2}/\frac{k}{2})^{\frac{k}{2}} < \exp(\frac{a}{2})$. Similarly $((N-1-k-a)/(N-1-k))^{\frac{N-1-k}{2}} < \exp(-\frac{a}{2})$. Setting $k = 1$ and letting $N-1 \geq k+a+1$ performs the lower bound.[30] The effect of the term in square roots actually counters the impact of the latter term, as $\sqrt{\frac{4}{N-1}} \leq \sqrt{\frac{N-1}{(N-1-k-a)(k+a)}} \leq 1$ for $k + a < N - 1$. The 'model prior' thus results in an upweighting of models with few or many coefficients, while intermediate model sizes are downweighted.

---

[30]Note that if $k = 0$, $BF_{EBL} = BF_h = 1$.

The impact of the k-based 'model prior', however, is virtually negligible with respect to the size of BF_{EBL}. Thus, at least as long as $R^2(N-1) > k+a$, BF_h is quite close to BF_{EBL}. And as long as the signal-to-noise ratio in the data is not too small, BMA posterior statistics will be disproportionally based on models with large PMPs (and thus $(N-1)R^2 \gg k+a$). Models with large differences between BF_h and BF_{EBL} will thus hardly affect posterior statistics.

A.3   The Shrinkage Factor and Goodness-of-Fit

In order to demonstrate equation (16) consider a reformulation of the posterior expected value of the shrinkage factor (13)

$$\mathrm{E}\left(\frac{g}{1+g}\Big|y,X\right) = \epsilon + \sum_{j=1}^{2^K} p(M_s|y,X)\frac{\bar{N}R_s^2 - \bar{\theta}_s}{R_s^2(\bar{N}-\bar{\theta}_s)} \tag{A.2}$$

$$\text{where}\quad \epsilon = \sum_{j=1}^{2^K} p(M_s|y,X)\frac{\bar{\theta}_s}{F_s^* R_s^2(\bar{N}-\bar{\theta}_s)}$$

The ϵ term is based on the expression $\bar{\theta}_s/F_s^*$ in (13), whose only role is to keep $E(\frac{g}{1+g}|X,y)$ non-negative in case of a 'bad' model, whereas it rapidly vanishes for models with higher signal-to-noise ratios. As long as the null model is not the single 'true' model, ϵ vanishes as $N \to \infty$ for fixed K – but even in small samples, ϵ tends rapidly towards zero as data quality increases. Moreover, in BMA sampling with any viable signal-to-noise ratio, any models with very low PMP will hardly affect posterior results, and hence the expression ϵ will vanish as soon as there exist some models with considerable null-based Bayes factors (which therefore must have their $F_s^* \gg 1$).

In the following, suppose that $K+a < N$. Now proceed to demonstrating the inequality in (16) by considering that as long as $\bar{\theta}_s \leq \bar{N}$, the following holds:[31]

$$E_{\mathcal{M}}\left(\tfrac{1}{\bar{N}-\bar{\theta}_s}\right) \geq \frac{1}{E_{\mathcal{M}}\left(\bar{N}-\bar{\theta}_s\right)}$$

where $E_{\mathcal{M}}(x) = \sum_{j=1}^{2^K} x p(M_j|y,X)$ denotes the expected value over model probabilities. Multiply with $\bar{N}$ and subtract unity to obtain

$$E_{\mathcal{M}}\left(\tfrac{\bar{\theta}_s}{\bar{N}-\bar{\theta}_s}\right) \geq \frac{E_{\mathcal{M}}\left(\bar{\theta}_s\right)}{E_{\mathcal{M}}\left(\bar{N}-\bar{\theta}_s\right)}$$

Moreover, since any nested model's R-squared R_s^2 cannot exceed the R-squared of the full model R_F^2, we have that $\frac{1-R_s^2}{R_s^2} \geq \frac{1-R_F^2}{R_F^2}$ and therefore:

---

[31]This stems from the fact that for a random variable $x \geq 0$, by the definition of covariance $E(\frac{1}{x})E(x) = 1 - Cov(x,\frac{1}{x})$, and since $x \geq 0$ we have $Cov(x,\frac{1}{x}) \leq 0$. Therefore $E(\frac{1}{x}) \geq \frac{1}{E(x)}$.

$$E_{\mathcal{M}}\left(\frac{1-R_s^2}{R_s^2}\frac{\theta_s}{\bar{N}-\bar{\theta}_s}\right) \geq \frac{1-R_F^2}{R_F^2}\frac{E_{\mathcal{M}}(\bar{\theta}_s)}{E_{\mathcal{M}}(\bar{N}-\bar{\theta}_s)}$$

Retransforming and integrating in (A.2) yields another representation of (16) (recall that $\bar{N} \equiv N - 3$ and $E_{\mathcal{M}}(\bar{\theta}_s) \equiv E_{\mathcal{M}}(k_s) + a - 2$):

$$E_{\mathcal{M}}\left(\frac{g}{1+g}\big|y,X\right) - \epsilon = 1 - E_{\mathcal{M}}\left(\frac{1-R_s^2}{R_s^2}\frac{\theta_s}{\bar{N}-\bar{\theta}_s}\right) \leq \frac{R_F^2\bar{N} - E_{\mathcal{M}}(\bar{\theta}_s)}{R_F^2(\bar{N} - E_{\mathcal{M}}(\bar{\theta}_s))}$$

How close $E_{\mathcal{M}}\left(\frac{g}{1+g}\big|y,X\right)$ comes to this upper bound is mainly determined by the posterior variance of model size (the less variance, the closer), and by parsimoniousness of the model priors.Note that the term ϵ on the left-hand side might break the inequality in rare instances. However this term tends to be very small: Numerical simulations of a null hypothesis with varying N, K, a and standard deviations have yielded no single instance in which $R_F^2 > \frac{K+a-2}{\bar{N}}$ and $E_{\mathcal{M}}(\frac{g}{1+g}|y,X)$ larger than the right-hand side above. Therefore, if $R_F^2 > \frac{K+a}{N}$ ($> \frac{K+a-2}{N-3}$)[32], then the ϵ term can be safely omitted from the inequality above.

A.4 The Posterior Predictive Distribution and the Hyper-g Prior

Consider using the data (X, y) to forecast the dependent variable ($\hat{y}$ conditional on 'prediction' covariates $\hat{X}$. Let X be $N \times k$ matrix, y be $N \times 1$, while $\hat{y}$ is $l \times y$ and $\hat{X}$ $l \times k$. The posterior predictive distribution of $\hat{y}$ is then given as a a multivariate t-distribution of dimension l (Eklund and Karlsson, 2007, equation (A.15))[33]

$$\hat{y}|\hat{X}, X, y, g \sim t_l(\bar{y} + s\hat{X}\hat{\beta}, \Sigma, N - 1)$$
$$\text{where } \Sigma = \left(I_l + s\hat{X}(X'X)^{-1}\hat{X}'\right)\frac{y'y}{N-1}(1 - sR^2)$$

Here, s denotes the shrinkage factor $s = \frac{g}{1+g}$, and R^2 the (centered) R-squared of y on X, where $\bar{y}$ is an N-dimensional vector whose elements are the arithmetic mean of y.
Integrating the density function of $\hat{y}|\hat{X}, X, y, g$ with respect to the shrinkage factor yields the integrand of the following equation (after some rearrangement):

---

[32]Note that this threshold is just slightly higher than the expected value of R_F^2 under the classic null hypothesis of no significant variance explanation by a regression model. As a rule of thumb, if the standard F-statistic for the full model is 'significant' by at least 20%, then the inequality above is guaranteed to hold.

[33]The slight differences with respect to Eklund and Karlsson (2007) are due to the fact that we employ an improper prior on beta variance *sigma* and the constant.

$$f(\hat{y}|\hat{X}, X, y) = \frac{\Gamma\left(\frac{N-1+l}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)\pi^{\frac{l}{2}}} \overbrace{\frac{(k+a-2)((\breve{y}'\breve{y})^{\frac{N-1}{2}})}{{}_2 2F_1\left(\frac{N-1}{2}, 1, \frac{k+a}{2}, R^2\right)}}^{\frac{a-2}{2}\frac{1}{p(y|X)}} \times$$

$$\times \int_0^1 \left|I_l + s\hat{X}(X'X)^{-1}\hat{X}'\right|^{-\frac{1}{2}} (1-s)^{\frac{k+a-4}{2}} \times$$

$$\times \left(\breve{y}'\breve{y}(1-sR^2) + (\hat{y}-\bar{y}-s\hat{X}\hat{\beta})'\left(I_l + s\hat{X}(X'X)^{-1}\hat{X}'\right)^{-1}(\hat{y}-\bar{y}-s\hat{X}\hat{\beta})\right)^{-\frac{N-1+l}{2}} ds$$

... where $\breve{y} \equiv y - \bar{y}$. To our knowledge, there is no closed-form solution to to the integral above, neither to its Laplace approximation. We therefore recommend to resort to numerical integration.

A.5   The Beta-binomial Prior over the Model Space

BMA calls for eliciting a prior distribution of models in $\mathcal{M}$. Two typical prior specifications have been usually imposed in the literature: a) an uninformative flat prior over all models, which implies that the posterior odds ratio resembles solely the Bayes factor and comparison of models is governed by their relative marginal likelihoods, and b) a prior that discriminates among models according to the number of regressors they include, so that a larger prior probability mass falls over models of a given size (see Sala-i-Martin et al. (2004)). This second alternative assumes that each covariate enters a model with probability ϑ, which implies that the prior mass for model j which includes k_j variables amounts to $\overline{p}(M_j) = \vartheta^{k_j}(1-\vartheta)^{K-k_j}$. The uninformative prior in a) is nested in b) by imposing $\vartheta = 1/2$, which results into equal model probabilities of 2^{-K} for all models for each matrix.

Ley and Steel (2009) show that fixing $\vartheta = 1/2$ puts most mass on models with $K/2$ regressors, since they are dominant in number. Their recommendation is thus to treat ϑ as random and placing a (hyper)prior on it. The proposal of Ley and Steel (2009) is to impose that the model size follows a beta-binomial(a, b) distribution (Bernardo and Smith (1994)) with $a = 1$, so that

$$P(k = k_j) = \frac{\Gamma(1+b)}{\Gamma(1) + \Gamma(b) + \Gamma(1+b+K)}\binom{K}{k_j}\Gamma(1+k_j)\Gamma(b+K-k_j) \quad k_j = 0, \dots, K. \quad \text{(A.3)}$$

The prior can be elicited by anchoring the prior expected model size, m.[34] Ley and Steel (2009) quantify the influence that a poorly specified prior exerts on posterior results when ϑ is fixed, which leads to the relative merits of BMA being less pronounced and its predictive power deteriorating. In contrast, the results in Ley and Steel (2009) indicate that the choice of m has no influential impact on posterior inference and the prior over models is purely non-informative.

---

[34]Note that b is then implicitly defined through $b = (K-m)/m$.

A.6 Charts and Tables

Figure 1: Cumulated posterior model probabilities for Setting A. Top panel corresponds to a signal-to-noise ratio of $\sigma = 1/2$ (left) and $\sigma = 1$ (right). Bottom panel to a ratio of $\sigma = 2.5$ (left) and $\sigma = 5$ (right).

Figure 2: QQ-plot of cumulated posterior mass for different choices of g against that of the g-RIC setting (Setting A based on 50 Monte Carlo draws).

	signal-to-noise ratio of $\sigma = 1/2$								signal-to-noise ratio of $\sigma = 1$							
	g-RIC	g-UIP	g-E$(\frac{g}{1+g}|Y)$	EBL	HG-3	HG-4	HG-RIC	HG-UIP	g-RIC	g-UIP	g-E$(\frac{g}{1+g}|Y)$	EBL	HG-3	HG-4	HG-RIC	HG-UIP
β_1	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_2	0.085	0.114	0.068	0.055	0.055	0.059	0.050	0.050	0.094	0.132	0.109	0.115	0.116	0.124	0.107	0.107
	(0.072)	(0.069)	(0.067)	(0.058)	(0.055)	(0.058)	(0.051)	(0.051)	(0.125)	(0.128)	(0.127)	(0.128)	(0.127)	(0.128)	(0.127)	(0.127)
β_3	0.083	0.112	0.066	0.054	0.052	0.056	0.047	0.047	0.085	0.125	0.101	0.108	0.109	0.117	0.099	0.099
	(0.067)	(0.061)	(0.064)	(0.055)	(0.046)	(0.049)	(0.043)	(0.043)	(0.073)	(0.084)	(0.078)	(0.081)	(0.080)	(0.083)	(0.077)	(0.077)
β_4	0.077	0.106	0.061	0.050	0.050	0.054	0.046	0.046	0.078	0.117	0.093	0.100	0.101	0.110	0.092	0.092
	(0.065)	(0.064)	(0.062)	(0.060)	(0.059)	(0.060)	(0.057)	(0.057)	(0.057)	(0.068)	(0.062)	(0.063)	(0.063)	(0.065)	(0.060)	(0.060)
β_5	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_6	0.086	0.116	0.068	0.055	0.055	0.060	0.050	0.050	0.067	0.105	0.082	0.089	0.090	0.098	0.081	0.081
	(0.059)	(0.061)	(0.052)	(0.046)	(0.045)	(0.048)	(0.042)	(0.042)	(0.021)	(0.031)	(0.025)	(0.029)	(0.029)	(0.032)	(0.027)	(0.027)
β_7	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_8	0.113	0.136	0.098	0.085	0.082	0.086	0.077	0.077	0.087	0.129	0.104	0.111	0.111	0.121	0.101	0.102
	(0.153)	(0.132)	(0.156)	(0.149)	(0.138)	(0.141)	(0.135)	(0.135)	(0.061)	(0.075)	(0.068)	(0.069)	(0.068)	(0.071)	(0.065)	(0.065)
β_9	0.084	0.114	0.066	0.054	0.054	0.058	0.049	0.049	0.083	0.125	0.099	0.107	0.108	0.117	0.098	0.098
	(0.066)	(0.070)	(0.058)	(0.055)	(0.060)	(0.057)	(0.051)	(0.051)	(0.049)	(0.066)	(0.057)	(0.065)	(0.064)	(0.068)	(0.060)	(0.060)
β_{10}	0.086	0.115	0.068	0.056	0.056	0.060	0.051	0.051	0.100	0.145	0.118	0.126	0.127	0.136	0.116	0.116
	(0.082)	(0.078)	(0.077)	(0.069)	(0.066)	(0.069)	(0.062)	(0.062)	(0.086)	(0.106)	(0.095)	(0.105)	(0.103)	(0.107)	(0.099)	(0.099)
β_{11}	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_{12}	0.085	0.115	0.066	0.054	0.054	0.059	0.049	0.050	0.109	0.151	0.126	0.133	0.133	0.142	0.123	0.123
	(0.054)	(0.056)	(0.048)	(0.043)	(0.042)	(0.044)	(0.040)	(0.040)	(0.122)	(0.134)	(0.128)	(0.131)	(0.129)	(0.131)	(0.126)	(0.126)
β_{13}	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.992**	**0.992**	**0.992**	**0.992**	**0.992**	**0.992**	**0.992**	**0.992**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.049)	(0.045)	(0.047)	(0.047)	(0.048)	(0.047)	(0.049)	(0.049)
β_{14}	0.079	0.109	0.063	0.052	0.052	0.055	0.047	0.047	0.097	0.136	0.112	0.120	0.120	0.129	0.111	0.111
	(0.069)	(0.065)	(0.067)	(0.064)	(0.063)	(0.064)	(0.061)	(0.061)	(0.124)	(0.130)	(0.127)	(0.129)	(0.127)	(0.129)	(0.125)	(0.125)
β_{15}	0.093	0.122	0.075	0.062	0.062	0.066	0.057	0.057	0.101	0.144	0.119	0.125	0.126	0.135	0.115	0.115
	(0.096)	(0.094)	(0.090)	(0.085)	(0.084)	(0.087)	(0.080)	(0.080)	(0.096)	(0.109)	(0.103)	(0.101)	(0.100)	(0.102)	(0.096)	(0.096)
$E(\frac{g}{1+g}|Y)$	0.996	0.990	0.998	0.999	0.998	0.998	0.999	0.999	0.996	0.990	0.994	0.993	0.992	0.991	0.993	0.993

Table 2: Posterior Inclusion Probabilities for Setting A with standard deviations in parenthesis. Left Panel corresponds to a signal-to-noise ratio of $\sigma = 1/2$, right panel to a ratio of $\sigma = 1$. Coefficients corresponding to variables of the data generating model are in **bold**. PIP values exceeding 0.5 in bold. Results are averaged over 50 Monte Carlo Steps.

Table 3 (landscape). Two panels.

Left panel — signal-to-noise ratio of $\sigma = 2.5$

	g-RIC	g-UIP	g-E$(\frac{g}{1+g}\mid Y)$	EBL	HG-3	HG-4	HG-RIC	HG-UIP
β_1	**1.000** (0.002)	**1.000** (0.001)	**1.000** (0.000)	**1.000** (0.000)	**1.000** (0.000)	**1.000** (0.000)	**1.000** (0.000)	**1.000** (0.000)
β_2	0.086 (0.106)	0.131 (0.126)	0.317 (0.149)	0.304 (0.153)	0.304 (0.151)	0.330 (0.151)	0.276 (0.150)	0.276 (0.150)
β_3	0.094 (0.064)	0.141 (0.077)	0.331 (0.109)	0.316 (0.115)	0.317 (0.113)	0.343 (0.114)	0.289 (0.111)	0.289 (0.111)
β_4	0.108 (0.119)	0.158 (0.144)	0.339 (0.165)	0.327 (0.172)	0.327 (0.169)	0.351 (0.168)	0.300 (0.169)	0.300 (0.169)
$\boldsymbol{\beta_5}$	**0.796** (0.244)	**0.826** (0.215)	**0.871** (0.155)	**0.873** (0.155)	**0.870** (0.156)	**0.875** (0.150)	**0.865** (0.164)	**0.865** (0.163)
β_6	0.062 (0.044)	0.103 (0.068)	0.282 (0.115)	0.272 (0.118)	0.273 (0.117)	0.298 (0.119)	0.245 (0.113)	0.245 (0.113)
$\boldsymbol{\beta_7}$	**0.997** (0.014)	**0.999** (0.005)	**1.000** (0.001)	**1.000** (0.001)	**1.000** (0.001)	**1.000** (0.001)	**1.000** (0.001)	**1.000** (0.001)
β_8	0.075 (0.065)	0.121 (0.095)	0.305 (0.136)	0.292 (0.151)	0.292 (0.149)	0.317 (0.149)	0.265 (0.148)	0.265 (0.148)
β_9	0.131 (0.161)	0.187 (0.199)	0.369 (0.217)	0.356 (0.219)	0.356 (0.216)	0.380 (0.213)	0.329 (0.218)	0.330 (0.218)
β_{10}	0.082 (0.089)	0.127 (0.113)	0.313 (0.144)	0.302 (0.151)	0.303 (0.148)	0.328 (0.149)	0.275 (0.146)	0.275 (0.146)
$\boldsymbol{\beta_{11}}$	**0.828** (0.240)	**0.869** (0.205)	**0.925** (0.141)	**0.922** (0.146)	**0.919** (0.147)	**0.924** (0.141)	**0.913** (0.156)	**0.913** (0.156)
β_{12}	0.065 (0.036)	0.109 (0.053)	0.301 (0.105)	0.288 (0.109)	0.289 (0.107)	0.315 (0.111)	0.259 (0.102)	0.260 (0.102)
$\boldsymbol{\beta_{13}}$	0.437 (0.332)	**0.504** (0.322)	**0.651** (0.261)	**0.640** (0.268)	**0.637** (0.267)	**0.653** (0.258)	**0.617** (0.277)	**0.617** (0.277)
β_{14}	0.098 (0.129)	0.143 (0.146)	0.321 (0.157)	0.309 (0.163)	0.310 (0.161)	0.334 (0.159)	0.283 (0.162)	0.283 (0.162)
β_{15}	0.165 (0.147)	0.212 (0.156)	0.386 (0.148)	0.372 (0.160)	0.373 (0.158)	0.396 (0.156)	0.347 (0.161)	0.347 (0.161)
$E(k\mid Y)$	5.027	5.629	7.711	7.572	7.568	7.844	7.260	7.264
$E(\frac{g}{1+g}\mid Y)$	0.996	0.990	0.949	0.955	0.947	0.939	0.956	0.956

Right panel — signal-to-noise ratio of $\sigma = 5$

	g-RIC	g-UIP	g-E$(\frac{g}{1+g}\mid Y)$	EBL	HG-3	HG-4	HG-RIC	HG-UIP
β_1	**0.792** (0.257)	**0.842** (0.208)	**0.955** (0.066)	**0.953** (0.068)	**0.946** (0.077)	**0.946** (0.075)	**0.884** (0.207)	**0.905** (0.171)
β_2	0.038 (0.037)	0.067 (0.058)	**0.512** (0.149)	0.455 (0.171)	0.445 (0.166)	0.481 (0.159)	0.370 (0.185)	0.380 (0.180)
β_3	0.050 (0.052)	0.081 (0.074)	**0.514** (0.160)	0.460 (0.180)	0.450 (0.176)	0.486 (0.167)	0.376 (0.197)	0.386 (0.192)
β_4	0.044 (0.072)	0.074 (0.104)	**0.503** (0.158)	0.449 (0.176)	0.439 (0.172)	0.476 (0.164)	0.366 (0.195)	0.376 (0.190)
$\boldsymbol{\beta_5}$	0.184 (0.267)	0.236 (0.286)	**0.617** (0.215)	**0.579** (0.246)	**0.567** (0.244)	**0.594** (0.228)	0.497 (0.279)	**0.509** (0.273)
β_6	0.048 (0.070)	0.080 (0.100)	**0.515** (0.185)	0.463 (0.208)	0.453 (0.204)	0.488 (0.194)	0.380 (0.226)	0.390 (0.221)
$\boldsymbol{\beta_7}$	**0.506** (0.342)	**0.585** (0.340)	**0.845** (0.183)	**0.829** (0.208)	**0.817** (0.212)	**0.829** (0.195)	**0.750** (0.284)	**0.766** (0.268)
β_8	0.049 (0.075)	0.080 (0.098)	**0.509** (0.157)	0.460 (0.172)	0.450 (0.169)	0.486 (0.160)	0.376 (0.194)	0.386 (0.188)
β_9	0.045 (0.061)	0.077 (0.094)	**0.508** (0.165)	0.455 (0.176)	0.445 (0.173)	0.482 (0.165)	0.372 (0.198)	0.382 (0.192)
β_{10}	0.042 (0.040)	0.072 (0.059)	**0.520** (0.147)	0.465 (0.163)	0.454 (0.159)	0.491 (0.152)	0.377 (0.181)	0.388 (0.175)
$\boldsymbol{\beta_{11}}$	0.305 (0.316)	0.369 (0.323)	**0.717** (0.222)	**0.689** (0.246)	**0.677** (0.246)	**0.699** (0.229)	**0.609** (0.294)	**0.622** (0.285)
β_{12}	0.054 (0.067)	0.093 (0.109)	**0.533** (0.171)	0.480 (0.196)	0.469 (0.192)	**0.504** (0.182)	0.393 (0.217)	0.404 (0.211)
$\boldsymbol{\beta_{13}}$	0.083 (0.119)	0.129 (0.151)	**0.558** (0.189)	**0.511** (0.216)	0.500 (0.212)	**0.532** (0.200)	0.426 (0.239)	0.437 (0.233)
β_{14}	0.043 (0.058)	0.072 (0.083)	**0.504** (0.160)	0.449 (0.176)	0.439 (0.172)	0.476 (0.164)	0.366 (0.191)	0.375 (0.187)
β_{15}	0.074 (0.114)	0.109 (0.131)	**0.520** (0.155)	0.470 (0.175)	0.461 (0.171)	0.496 (0.162)	0.387 (0.195)	0.398 (0.191)
$E(k\mid Y)$	2.359	2.965	8.830	8.167	8.011	8.467	6.928	7.103
$E(\frac{g}{1+g}\mid Y)$	0.996	0.990	0.783	0.817	0.795	0.760	0.856	0.849

Table 3: Posterior Inclusion Probabilities for Setting "A" with standard deviations in parenthesis. Left Panel corresponds to a signal-to-noise ratio of $\sigma = 2.5$, right panel to a ratio of $\sigma = 5$ noise. Coefficients corresponding to variables of the data generating model are in **bold**. PIP values exceeding 0.5 in bold. Results correspond to 50 Monte Carlo Steps.
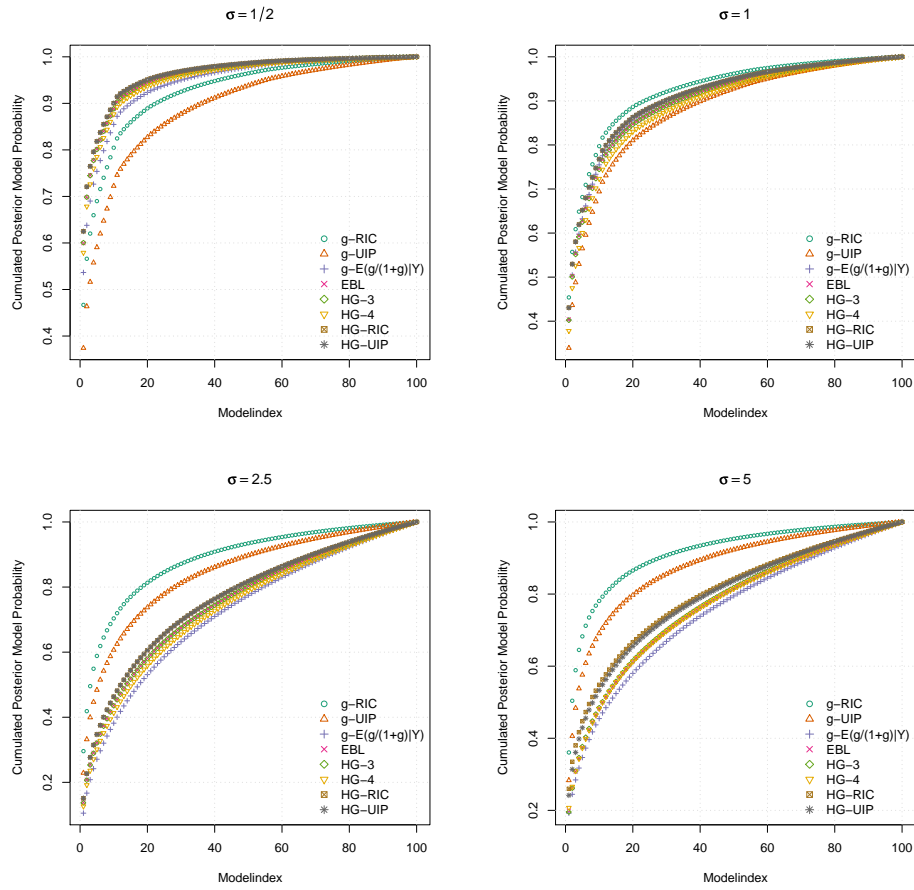
Figure 3: Cumulated posterior model probabilities for Setting B. Top panel corresponds to a signal-to-noise ratio of $\sigma = 1/2$ (left) and $\sigma = 1$ (right). Bottom panel to a ratio of $\sigma = 2.5$ (left) and $\sigma = 5$ (right)
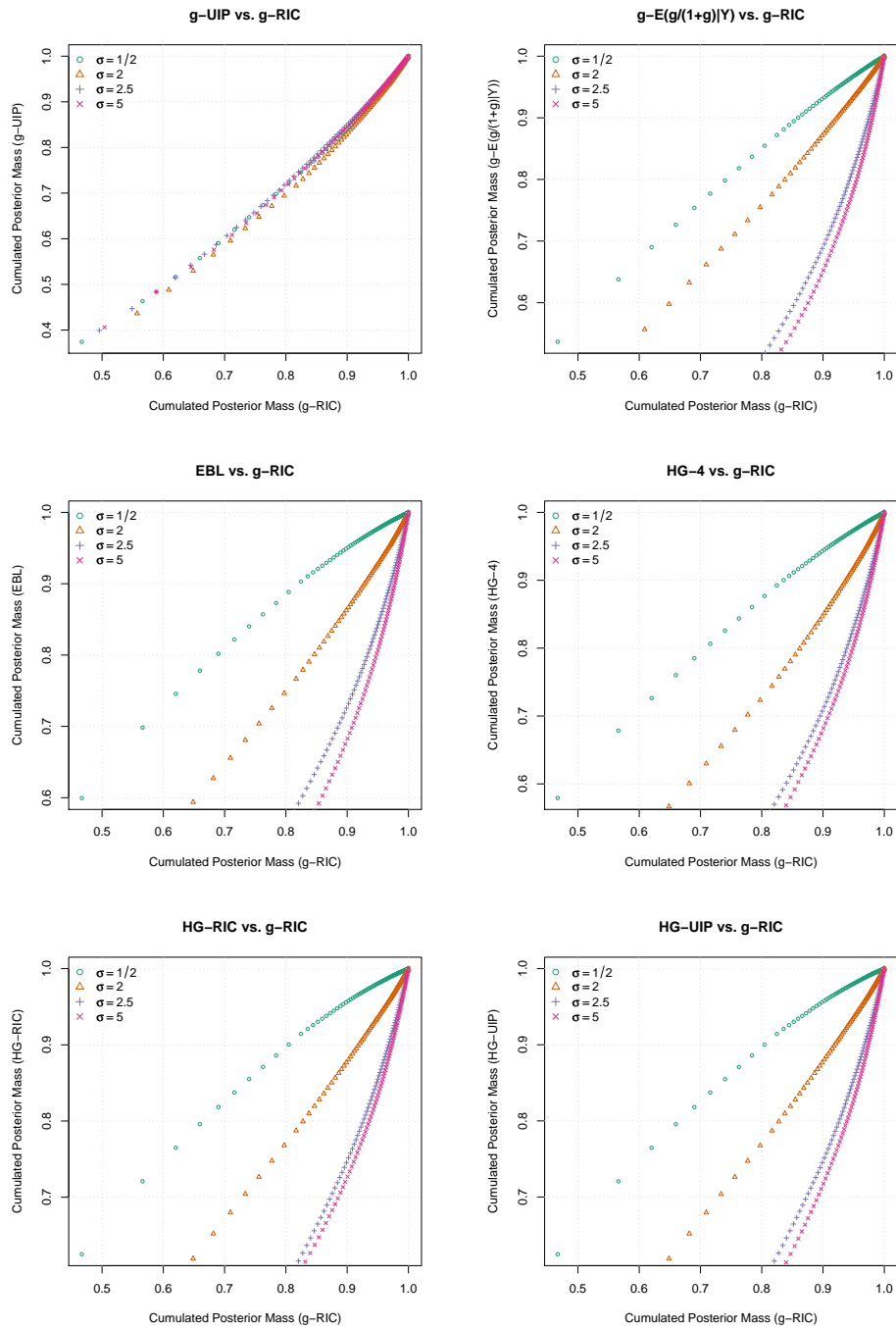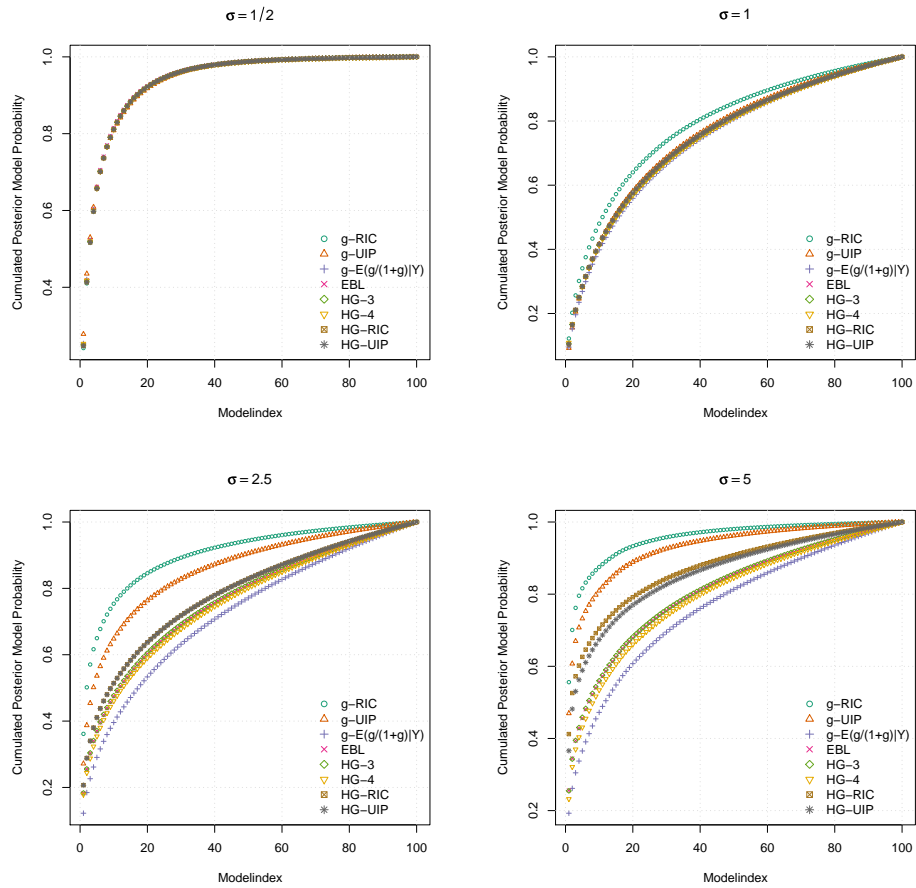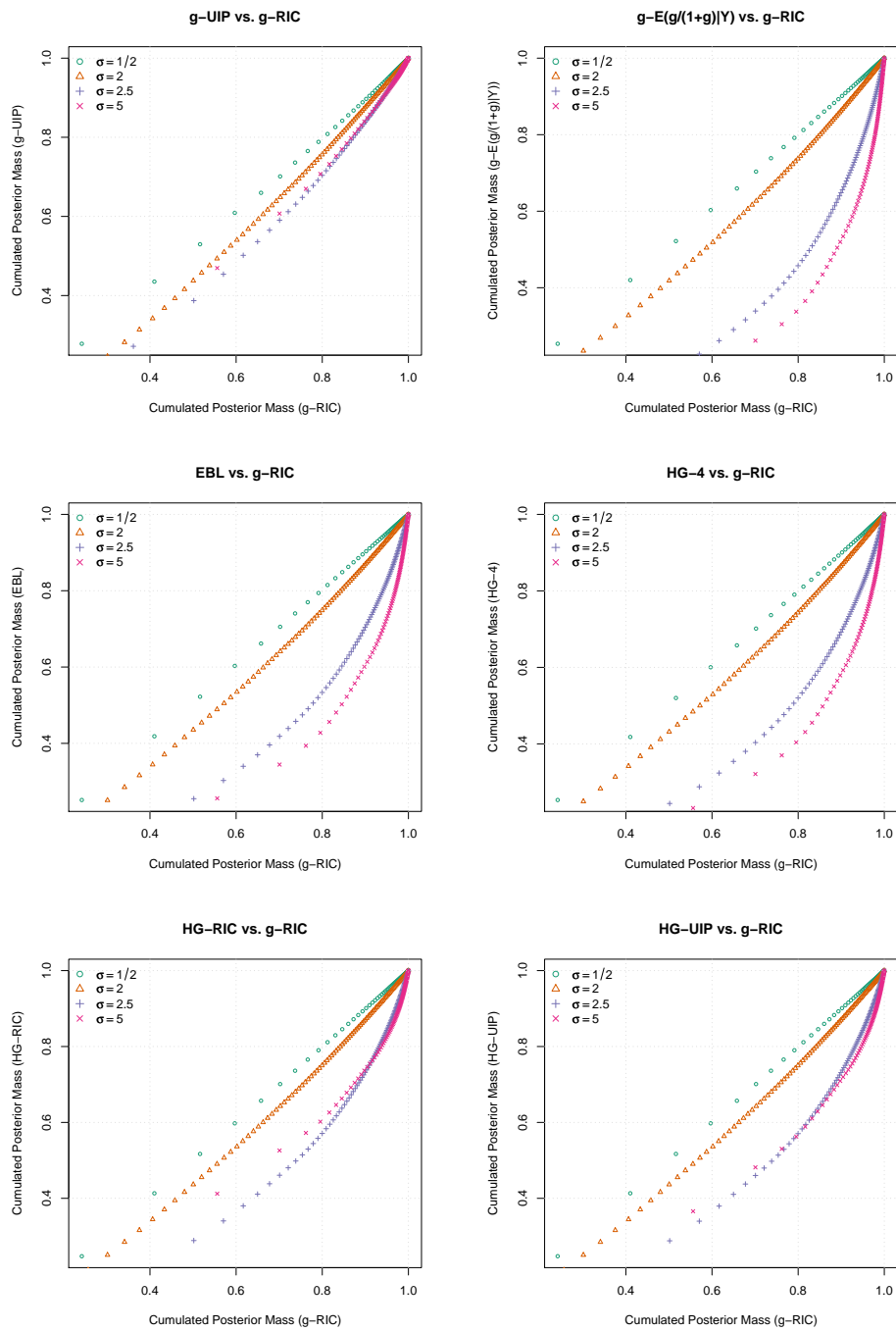
Figure 4: QQ-plot of cumulated posterior mass for different choices of g against that of the g-RIC setting (Setting B based on 50 Monte Carlo draws).

Table 4: Posterior Inclusion Probabilities for Setting B with standard deviations in parenthesis. Left Panel corresponds to a signal-to-noise ratio of $\sigma = 1/2$, right panel to a ratio of $\sigma = 1$. PIP values exceeding 0.5 in bold. Results are averaged over 50 Monte Carlo Steps.

signal-to-noise ratio of $\sigma = 1/2$

	g-RIC	g-UIP	g-E($\frac{g}{1+g}|Y$)	EBL	HG-3	HG-4	HG-RIC	HG-UIP
β_1	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_2	**0.801**	**0.835**	**0.818**	**0.810**	**0.808**	**0.813**	**0.804**	**0.804**
	(0.223)	(0.182)	(0.205)	(0.218)	(0.217)	(0.213)	(0.223)	(0.222)
β_3	**0.762**	**0.802**	**0.781**	**0.773**	**0.772**	**0.777**	**0.767**	**0.767**
	(0.239)	(0.195)	(0.220)	(0.228)	(0.227)	(0.222)	(0.232)	(0.232)
β_4	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
β_5	**0.885**	**0.897**	**0.891**	**0.889**	**0.888**	**0.889**	**0.886**	**0.886**
	(0.147)	(0.119)	(0.135)	(0.140)	(0.140)	(0.136)	(0.143)	(0.143)
β_6	**0.943**	**0.952**	**0.948**	**0.947**	**0.946**	**0.947**	**0.944**	**0.944**
	(0.116)	(0.093)	(0.106)	(0.107)	(0.108)	(0.105)	(0.111)	(0.111)
β_7	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_8	**0.696**	**0.749**	**0.720**	**0.712**	**0.711**	**0.718**	**0.704**	**0.704**
	(0.234)	(0.190)	(0.216)	(0.221)	(0.220)	(0.215)	(0.225)	(0.225)
β_9	**0.679**	**0.734**	**0.704**	**0.694**	**0.693**	**0.700**	**0.686**	**0.686**
	(0.256)	(0.211)	(0.238)	(0.247)	(0.246)	(0.241)	(0.252)	(0.252)
β_{10}	**0.998**	**0.998**	**0.999**	**0.999**	**0.998**	**0.998**	**0.998**	**0.998**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
β_{11}	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_{12}	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
β_{13}	**0.989**	**0.990**	**0.989**	**0.989**	**0.989**	**0.989**	**0.989**	**0.989**
	(0.053)	(0.043)	(0.048)	(0.050)	(0.050)	(0.049)	(0.052)	(0.052)
β_{14}	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_{15}	0.486	**0.575**	**0.524**	**0.508**	**0.509**	**0.519**	0.499	0.499
	(0.191)	(0.155)	(0.177)	(0.185)	(0.183)	(0.179)	(0.187)	(0.187)
$E(k|Y)$	13.238	13.530	13.373	13.318	13.313	13.348	13.275	13.275
$E(\frac{g}{1+g}|Y)$	0.996	0.990	0.994	0.995	0.994	0.993	0.994	0.994

signal-to-noise ratio of $\sigma = 1$

	g-RIC	g-UIP	g-E($\frac{g}{1+g}|Y$)	EBL	HG-3	HG-4	HG-RIC	HG-UIP
β_1	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_2	0.374	0.487	**0.567**	**0.562**	**0.560**	**0.575**	**0.544**	**0.544**
	(0.266)	(0.262)	(0.242)	(0.247)	(0.245)	(0.240)	(0.250)	(0.250)
β_3	0.393	0.480	**0.550**	**0.548**	**0.548**	**0.560**	**0.533**	**0.534**
	(0.295)	(0.270)	(0.243)	(0.248)	(0.247)	(0.241)	(0.253)	(0.253)
β_4	**0.747**	**0.806**	**0.840**	**0.840**	**0.838**	**0.845**	**0.831**	**0.831**
	(0.282)	(0.242)	(0.207)	(0.206)	(0.206)	(0.199)	(0.215)	(0.215)
β_5	0.437	**0.516**	**0.577**	**0.573**	**0.572**	**0.584**	**0.559**	**0.560**
	(0.293)	(0.271)	(0.245)	(0.249)	(0.247)	(0.241)	(0.254)	(0.254)
β_6	0.470	**0.560**	**0.625**	**0.621**	**0.620**	**0.632**	**0.606**	**0.606**
	(0.332)	(0.303)	(0.272)	(0.277)	(0.276)	(0.269)	(0.283)	(0.283)
β_7	**0.980**	**0.988**	**0.991**	**0.990**	**0.990**	**0.990**	**0.989**	**0.989**
	(0.039)	(0.023)	(0.018)	(0.020)	(0.021)	(0.019)	(0.022)	(0.022)
β_8	0.325	0.430	**0.509**	**0.506**	**0.506**	**0.520**	0.489	0.490
	(0.237)	(0.242)	(0.227)	(0.233)	(0.231)	(0.227)	(0.236)	(0.236)
β_9	0.255	0.364	0.452	0.449	0.449	0.465	0.431	0.431
	(0.189)	(0.196)	(0.188)	(0.197)	(0.195)	(0.192)	(0.198)	(0.198)
β_{10}	**0.737**	**0.804**	**0.841**	**0.838**	**0.836**	**0.843**	**0.828**	**0.828**
	(0.271)	(0.225)	(0.190)	(0.192)	(0.193)	(0.186)	(0.200)	(0.200)
β_{11}	**0.997**	**0.998**	**0.999**	**0.999**	**0.998**	**0.999**	**0.998**	**0.998**
	(0.022)	(0.013)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)
β_{12}	**0.671**	**0.754**	**0.804**	**0.795**	**0.793**	**0.802**	**0.782**	**0.782**
	(0.300)	(0.248)	(0.207)	(0.217)	(0.218)	(0.210)	(0.227)	(0.227)
β_{13}	**0.728**	**0.773**	**0.804**	**0.801**	**0.800**	**0.806**	**0.793**	**0.794**
	(0.303)	(0.265)	(0.233)	(0.236)	(0.235)	(0.229)	(0.243)	(0.243)
β_{14}	**0.993**	**0.996**	**0.997**	**0.996**	**0.996**	**0.996**	**0.996**	**0.996**
	(0.030)	(0.021)	(0.016)	(0.018)	(0.018)	(0.017)	(0.020)	(0.020)
β_{15}	0.285	0.380	0.460	0.457	0.457	0.472	0.441	0.441
	(0.201)	(0.194)	(0.180)	(0.185)	(0.184)	(0.180)	(0.188)	(0.188)
$E(k|Y)$	9.391	10.335	11.015	10.976	10.964	11.089	10.822	10.824
$E(\frac{g}{1+g}|Y)$	0.996	0.990	0.982	0.983	0.980	0.978	0.982	0.982

Table header: signal-to-noise ratio of $\sigma = 2.5$ (left panel) | signal-to-noise ratio of $\sigma = 5$ (right panel)

	g-RIC	g-UIP	g-E$(\frac{g}{1+g}\|Y)$	EBL	HG-3	HG-4	HG-RIC	HG-UIP	g-RIC	g-UIP	g-E$(\frac{g}{1+g}\|Y)$	EBL	HG-3	HG-4	HG-RIC	HG-UIP
β_1	**0.999**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.819**	**0.851**	**0.944**	**0.940**	**0.930**	**0.927**	**0.804**	**0.857**
	(0.004)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.274)	(0.244)	(0.113)	(0.109)	(0.115)	(0.107)	(0.239)	(0.192)
β_2	0.060	0.105	0.355	0.377	0.373	0.409	0.327	0.328	0.030	0.048	0.412	0.392	0.376	0.423	0.261	0.282
	(0.059)	(0.096)	(0.179)	(0.196)	(0.193)	(0.188)	(0.197)	(0.197)	(0.050)	(0.068)	(0.163)	(0.183)	(0.174)	(0.168)	(0.171)	(0.172)
β_3	0.112	0.165	0.414	0.432	0.426	0.460	0.382	0.382	0.040	0.063	0.422	0.399	0.384	0.428	0.264	0.288
	(0.150)	(0.170)	(0.200)	(0.211)	(0.209)	(0.203)	(0.213)	(0.213)	(0.069)	(0.094)	(0.184)	(0.206)	(0.197)	(0.187)	(0.177)	(0.184)
β_4	0.055	0.105	0.376	0.394	0.389	0.426	0.342	0.342	0.033	0.053	0.436	0.412	0.395	0.442	0.276	0.298
	(0.060)	(0.114)	(0.208)	(0.225)	(0.221)	(0.216)	(0.224)	(0.224)	(0.049)	(0.066)	(0.168)	(0.186)	(0.177)	(0.170)	(0.178)	(0.178)
β_5	0.056	0.093	0.332	0.355	0.351	0.387	0.306	0.307	0.028	0.047	0.420	0.397	0.382	0.428	0.263	0.286
	(0.066)	(0.083)	(0.159)	(0.185)	(0.183)	(0.179)	(0.186)	(0.186)	(0.025)	(0.039)	(0.168)	(0.187)	(0.178)	(0.173)	(0.169)	(0.171)
β_6	0.083	0.128	0.374	0.396	0.391	0.428	0.344	0.345	0.024	0.040	0.395	0.372	0.357	0.404	0.239	0.262
	(0.154)	(0.174)	(0.210)	(0.216)	(0.213)	(0.207)	(0.218)	(0.218)	(0.034)	(0.051)	(0.149)	(0.180)	(0.170)	(0.163)	(0.148)	(0.157)
β_7	0.230	0.304	**0.547**	**0.555**	**0.549**	**0.579**	**0.508**	**0.508**	0.062	0.086	0.446	0.423	0.407	0.451	0.288	0.312
	(0.262)	(0.298)	(0.292)	(0.299)	(0.296)	(0.283)	(0.311)	(0.311)	(0.144)	(0.165)	(0.193)	(0.216)	(0.209)	(0.199)	(0.209)	(0.210)
β_8	0.057	0.095	0.335	0.359	0.355	0.392	0.308	0.309	0.024	0.040	0.396	0.375	0.360	0.407	0.245	0.267
	(0.072)	(0.103)	(0.172)	(0.188)	(0.185)	(0.183)	(0.185)	(0.185)	(0.027)	(0.042)	(0.157)	(0.183)	(0.174)	(0.168)	(0.166)	(0.168)
β_9	0.056	0.097	0.332	0.356	0.353	0.388	0.309	0.309	0.025	0.042	0.397	0.379	0.364	0.411	0.250	0.271
	(0.063)	(0.098)	(0.179)	(0.193)	(0.190)	(0.186)	(0.193)	(0.193)	(0.029)	(0.046)	(0.158)	(0.175)	(0.166)	(0.162)	(0.160)	(0.161)
β_{10}	0.107	0.154	0.402	0.424	0.419	0.453	0.376	0.376	0.046	0.071	0.435	0.417	0.401	0.446	0.286	0.308
	(0.193)	(0.210)	(0.235)	(0.252)	(0.249)	(0.241)	(0.255)	(0.255)	(0.072)	(0.102)	(0.197)	(0.209)	(0.202)	(0.193)	(0.206)	(0.205)
β_{11}	**0.503**	**0.580**	**0.764**	**0.768**	**0.761**	**0.781**	**0.727**	**0.728**	0.138	0.170	**0.524**	0.501	0.484	**0.525**	0.360	0.387
	(0.352)	(0.344)	(0.269)	(0.260)	(0.261)	(0.245)	(0.283)	(0.282)	(0.243)	(0.257)	(0.236)	(0.245)	(0.241)	(0.226)	(0.249)	(0.250)
β_{12}	0.115	0.173	0.433	0.451	0.445	0.480	0.399	0.399	0.040	0.065	0.433	0.410	0.394	0.440	0.274	0.297
	(0.164)	(0.197)	(0.241)	(0.243)	(0.239)	(0.234)	(0.243)	(0.243)	(0.067)	(0.098)	(0.173)	(0.196)	(0.188)	(0.179)	(0.187)	(0.187)
β_{13}	0.126	0.178	0.411	0.430	0.426	0.458	0.384	0.384	0.073	0.094	0.436	0.415	0.400	0.444	0.284	0.307
	(0.175)	(0.201)	(0.227)	(0.243)	(0.240)	(0.232)	(0.246)	(0.246)	(0.178)	(0.188)	(0.187)	(0.211)	(0.205)	(0.193)	(0.212)	(0.212)
β_{14}	0.234	0.318	**0.600**	**0.600**	**0.592**	**0.625**	**0.546**	**0.547**	0.051	0.080	0.475	0.448	0.431	0.475	0.308	0.332
	(0.273)	(0.291)	(0.276)	(0.286)	(0.285)	(0.269)	(0.304)	(0.304)	(0.072)	(0.097)	(0.198)	(0.216)	(0.208)	(0.198)	(0.208)	(0.209)
β_{15}	0.045	0.083	0.327	0.350	0.346	0.383	0.300	0.300	0.033	0.051	0.410	0.391	0.376	0.422	0.260	0.282
	(0.039)	(0.066)	(0.151)	(0.169)	(0.167)	(0.163)	(0.170)	(0.170)	(0.073)	(0.092)	(0.159)	(0.181)	(0.173)	(0.168)	(0.173)	(0.173)
$E(k\|Y)$	2.837	3.577	7.001	7.244	7.173	7.649	6.556	6.564	1.465	1.803	6.981	6.670	6.444	7.073	4.662	5.036
$E(\frac{g}{1+g}\|Y)$	0.996	0.990	0.933	0.926	0.913	0.897	0.931	0.931	0.996	0.990	0.782	0.796	0.767	0.716	0.866	0.850

Table 5: Posterior Inclusion Probabilities for Setting B with standard deviations in parenthesis. Left Panel corresponds to a signal-to-noise ratio of $\sigma = 2.5$, right panel to a ratio of $\sigma = 5$ noise. PIP values exceeding 0.5 in bold. Results correspond to 50 Monte Carlo Steps.

	g-RIC	g-UIP	g-E($\frac{g}{1+g}$|Y)	EBL	HG-3	HG-4	HG-RIC	HG-UIP
Min.	0.1349	0.1279	0.1573	0.1888	0.1934	0.1809	0.2096	0.2094
Mean	0.4618	0.3725	0.5306	0.5951	0.5973	0.5758	0.6220	0.6217
Max.	0.6297	0.5106	0.7037	0.7669	0.7644	0.7461	0.7845	0.7843
St.Dev.	0.1342	0.1019	0.1482	0.1551	0.1490	0.1487	0.1484	0.1484
Min.	0.0539	0.0317	0.0426	0.0308	0.0320	0.0289	0.0362	0.0362
Mean	0.4433	0.3290	0.3922	0.3944	0.3932	0.3690	0.4219	0.4215
Max.	0.6115	0.4849	0.5578	0.5954	0.5922	0.5658	0.6220	0.6216
St.Dev.	0.1373	0.1138	0.1283	0.1358	0.1342	0.1293	0.1392	0.1392
Min.	0.0021	0.0022	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Mean	0.1201	0.1048	0.0556	0.0665	0.0660	0.0606	0.0721	0.0720
Max.	0.4609	0.3392	0.1493	0.1978	0.1968	0.1768	0.2216	0.2213
St.Dev.	0.1133	0.0834	0.0387	0.0487	0.0478	0.0441	0.0524	0.0524
Min.	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Mean	0.0012	0.0023	0.0010	0.0025	0.0024	0.0021	0.0027	0.0028
Max.	0.0215	0.0324	0.0114	0.0323	0.0308	0.0254	0.0347	0.0347
St.Dev.	0.0035	0.0057	0.0026	0.0062	0.0060	0.0051	0.0067	0.0067

Table 6: Summary statistics of posterior model probabilities for true model based on setting "A" and 50 Monte Carlo Steps. Top panel corresponds to $\sigma = 1/2$, second panel to $\sigma = 1$, third panel to $\sigma = 2.5$, fourth panel to $\sigma = 5$

	g-RIC	g-UIP	g-E($\frac{g}{1+g}$|Y)	EBL	HG-3	HG-4	HG-UIP	HG-RIC
Min.	0.4752	0.6133	0.4704	0.5192	0.5386	0.5175	0.5695	0.5691
Mean	0.9806	0.9919	0.9807	0.9872	0.9908	0.9902	0.9914	0.9914
Max.	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
St.Dev.	0.1342	0.1019	0.1482	0.1551	0.1490	0.1487	0.1484	0.1484
Min.	0.1363	0.1115	0.1226	0.1131	0.1188	0.1158	0.1239	0.1238
Mean	0.9650	0.9552	0.9604	0.9612	0.9626	0.9604	0.9657	0.9656
Max.	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
St.Dev.	0.1373	0.1138	0.1283	0.1358	0.1342	0.1293	0.1392	0.1392
Min.	0.0071	0.0076	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Mean	0.4683	0.5202	0.5516	0.5325	0.5331	0.5274	0.5382	0.5383
Max.	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
St.Dev.	0.1133	0.0834	0.0387	0.0487	0.0478	0.0441	0.0524	0.0524
Min.	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Mean	0.0067	0.0173	0.0070	0.0203	0.0200	0.0147	0.0284	0.0285
Max.	0.1643	0.3744	0.0842	0.2835	0.2737	0.1685	0.4719	0.4688
St.Dev.	0.0035	0.0057	0.0026	0.0062	0.0060	0.0051	0.0067	0.0067

Table 7: Summary statistics of the ratio of posterior model probabilities of true model and best model for setting "A" based on 50 Monte Carlo Steps. Top panel corresponds to $\sigma = 1/2$, second panel to $\sigma = 1$, third panel to $\sigma = 2.5$, fourth panel to $\sigma = 5$

	g-RIC	g-UIP	g-E($\frac{g}{1+g}$|Y)	EBL	HG-3	HG-4	HG-RIC	HG-UIP
$\sigma = 1/2$	-	1.00877	0.99754	0.99798	0.99793	0.99817	0.99771	0.99771
$\sigma = 1$	-	1.00347	1.00200	1.00128	1.00219	1.00315	1.00126	1.00127
$\sigma = 2.5$	-	0.99501	1.00079	1.00556	1.00320	1.00699	1.00039	1.00042
$\sigma = 5$	-	0.99034	1.00697	1.00594	1.00720	1.01958	1.01256	1.00692
$\sigma = 1/2$	-	0.99754	0.99926	0.99794	0.99948	0.99910	0.99998	0.99998
$\sigma = 1$	-	0.98166	0.97396	0.97316	0.97501	0.97398	0.97648	0.97647
$\sigma = 2.5$	-	0.98875	0.97284	0.96760	0.97580	0.97847	0.97631	0.97627
$\sigma = 5$	-	0.99578	1.00427	0.99968	1.00747	1.02216	1.01853	1.00976

Table 8: Relative Root Mean Squared Error based on 30 out of sample forecasts averaged over 50 Monte Carlo Steps. Values below 1 indicate superior predictive performance as compared to the g-RIC setting. Top panel corresponds to setting "A" and bottom panel to setting "B".

	g-RIC	g-UIP	g-E($\frac{g}{1+g}|Y$)	EBL	HG-3	HG-4	HG-UIP	HG-RIC
GDP60	**0.9989**	**1.0000**	**0.9999**	**0.9999**	**0.9998**	**0.9998**	**0.9998**	**0.9999**
Confucian	**0.9880**	**0.9990**	**0.9953**	**0.9959**	**0.9955**	**0.9952**	**0.9964**	**0.9961**
LifeExp	**0.9314**	**0.9890**	**0.9751**	**0.9787**	**0.9748**	**0.9746**	**0.9792**	**0.9764**
EquipInv	**0.9271**	**0.9245**	**0.8949**	**0.9031**	**0.8958**	**0.8895**	**0.8996**	**0.8974**
SubSahara	**0.7288**	**0.9480**	**0.9087**	**0.9129**	**0.9096**	**0.9042**	**0.9127**	**0.9151**
Muslim	**0.6524**	**0.5196**	**0.5695**	**0.5673**	**0.5646**	**0.5644**	**0.5596**	**0.5592**
RuleofLaw	0.4855	**0.7569**	**0.6761**	**0.6822**	**0.6697**	**0.6611**	**0.6797**	**0.6826**
YrsOpen	**0.5172**	0.2452	0.3530	0.3433	0.3546	0.3630	0.3384	0.3373
EcoOrg	0.4536	**0.6328**	**0.6400**	**0.6320**	**0.6402**	**0.6370**	**0.6371**	**0.6447**
Protestants	0.4488	**0.5665**	**0.5795**	**0.5759**	**0.5772**	**0.5825**	**0.5782**	**0.5820**
Mining	0.4572	**0.8975**	**0.8633**	**0.8698**	**0.8638**	**0.8587**	**0.8693**	**0.8656**
NequipInv	0.4297	**0.7159**	**0.7207**	**0.7209**	**0.7216**	**0.7185**	**0.7218**	**0.7264**
LatAmerica	0.2084	**0.6297**	**0.5842**	**0.5962**	**0.5870**	**0.5861**	**0.5946**	**0.5956**
PrScEnroll	0.2144	**0.5114**	0.4987	**0.5118**	0.4964	0.4928	**0.5032**	0.4960
Buddha	0.1986	0.3526	0.4232	0.4190	0.4223	0.4225	0.4183	0.4135
BlMktPm	0.1796	**0.6471**	**0.6145**	**0.6197**	**0.6104**	**0.6103**	**0.6176**	**0.6176**
Catholic	0.1277	0.2388	0.3138	0.3066	0.3103	0.3157	0.3019	0.3040
CivlLib	0.1281	0.4559	0.4501	0.4536	0.4484	0.4445	0.4492	0.4485
Hindu	0.1213	**0.8613**	**0.8198**	**0.8358**	**0.8259**	**0.8176**	**0.8345**	**0.8331**
PrExports	0.1039	0.1828	0.2784	0.2684	0.2762	0.2846	0.2661	0.2679
PolRights	0.0959	0.3354	0.3820	0.3810	0.3821	0.3821	0.3787	0.3819
RFEXDist	0.0836	0.2072	0.2740	0.2742	0.2769	0.2744	0.2700	0.2672
Age	0.0817	0.3064	0.3590	0.3544	0.3583	0.3639	0.3533	0.3553
WarDummy	0.0776	0.2593	0.3380	0.3336	0.3421	0.3421	0.3351	0.3347
LabForce	0.0771	**0.8017**	**0.7414**	**0.7697**	**0.7519**	**0.7392**	**0.7608**	**0.7587**
Foreign	0.0697	0.1501	0.2248	0.2206	0.2271	0.2346	0.2196	0.2222
English	0.0689	0.3687	0.3822	0.3868	0.3863	0.3835	0.3854	0.3860
EthnoL	0.0584	**0.6826**	**0.6206**	**0.6461**	**0.6277**	**0.6078**	**0.6299**	**0.6321**
Spanish	0.0563	0.4485	0.4341	0.4496	0.4367	0.4293	0.4386	0.4374
stdBMP	0.0495	0.1264	0.2049	0.2002	0.2030	0.2073	0.1971	0.1971
French	0.0510	0.4092	0.3992	0.4190	0.4055	0.3956	0.4078	0.4004
Abslat	0.0434	0.1529	0.2273	0.2247	0.2243	0.2377	0.2261	0.2236
WorkPop	0.0427	0.1333	0.2064	0.1994	0.2066	0.2135	0.2014	0.2012
HighEnroll	0.0448	**0.6947**	**0.6165**	**0.6415**	**0.6169**	**0.6049**	**0.6308**	**0.6261**
Popg	0.0368	0.1485	0.2182	0.2135	0.2179	0.2218	0.2093	0.2087
Brit	0.0386	0.3304	0.3298	0.3382	0.3316	0.3260	0.3345	0.3289
OutwarOr	0.0383	0.3387	0.3507	0.3613	0.3489	0.3463	0.3517	0.3501
Jewish	0.0355	0.1289	0.1966	0.1927	0.1995	0.2096	0.1949	0.1922
RevnCoup	0.0300	0.1252	0.1946	0.1924	0.1925	0.1962	0.1877	0.1842
PublEdupct	0.0315	0.2951	0.3152	0.3167	0.3148	0.3182	0.3151	0.3133
Area	0.0298	0.1391	0.2152	0.2097	0.2164	0.2180	0.2101	0.2072
$E(k|Y)$	10.442	19.657	20.389	20.518	20.411	20.377	20.395	20.368
$E(\frac{g}{1+g}|Y)$	0.999	0.986	0.955	0.960	0.955	0.951	0.958	0.958

Table 9: Posterior Inclusion Probabilities for different prior settings. Values larger than 0.5 in bold.

	g-RIC	g-UIP	g-E($\frac{g}{1+g}|Y$)	EBL	HG-3	HG-4	HG-UIP	HG-RIC
GDP60	**-0.7810**	**-0.8019**	**-0.7743**	**-0.7785**	**-0.7740**	**-0.7704**	**-0.7768**	**-0.7769**
Confucian	**0.2703**	**0.3055**	**0.2881**	**0.2921**	**0.2884**	**0.2858**	**0.2905**	**0.2901**
LifeExp	**0.5267**	**0.5546**	**0.5348**	**0.5411**	**0.5358**	**0.5331**	**0.5386**	**0.5366**
EquipInv	**0.3040**	**0.2361**	**0.2222**	**0.2259**	**0.2227**	**0.2202**	**0.2240**	**0.2230**
SubSahara	**-0.2545**	**-0.4094**	**-0.3619**	**-0.3686**	**-0.3632**	**-0.3586**	**-0.3683**	**-0.3693**
Muslim	**0.1448**	**0.1017**	**0.1096**	**0.1089**	**0.1083**	**0.1084**	**0.1073**	**0.1067**
RuleofLaw	0.1322	**0.1658**	**0.1406**	**0.1422**	**0.1391**	**0.1368**	**0.1418**	**0.1429**
YrsOpen	**0.1439**	0.0325	0.0482	0.0465	0.0482	0.0502	0.0462	0.0455
EcoOrg	0.0816	**0.0888**	**0.0866**	**0.0850**	**0.0861**	**0.0860**	**0.0862**	**0.0875**
Protestants	-0.0772	**-0.0804**	**-0.0799**	**-0.0785**	**-0.0794**	**-0.0800**	**-0.0796**	**-0.0806**
Mining	0.0800	**0.1486**	**0.1419**	**0.1433**	**0.1421**	**0.1415**	**0.1432**	**0.1426**
NequipInv	0.0737	**0.1066**	**0.1066**	**0.1065**	**0.1070**	**0.1064**	**0.1068**	**0.1079**
LatAmerica	-0.0429	**-0.1759**	**-0.1464**	**-0.1528**	**-0.1478**	**-0.1455**	**-0.1515**	**-0.1510**
PrScEnroll	0.0598	**0.1262**	0.1099	**0.1142**	0.1090	0.1066	**0.1116**	0.1101
Buddha	0.0264	0.0335	0.0410	0.0404	0.0408	0.0410	0.0400	0.0398
BlMktPm	-0.0219	**-0.0783**	**-0.0702**	**-0.0713**	**-0.0697**	**-0.0694**	**-0.0711**	**-0.0709**
Catholic	-0.0050	-0.0112	-0.0111	-0.0100	-0.0111	-0.0110	-0.0111	-0.0121
CivlLib	-0.0271	-0.0944	-0.0853	-0.0877	-0.0852	-0.0832	-0.0866	-0.0859
Hindu	-0.0184	**-0.3591**	**-0.3003**	**-0.3169**	**-0.3041**	**-0.2949**	**-0.3123**	**-0.3104**
PrExports	-0.0168	-0.0162	-0.0271	-0.0252	-0.0264	-0.0273	-0.0250	-0.0252
PolRights	-0.0157	-0.0454	-0.0490	-0.0490	-0.0494	-0.0493	-0.0487	-0.0494
RFEXDist	-0.0095	-0.0148	-0.0187	-0.0190	-0.0190	-0.0186	-0.0186	-0.0183
Age	-0.0078	-0.0252	-0.0282	-0.0280	-0.0283	-0.0287	-0.0279	-0.0282
WarDummy	-0.0081	-0.0222	-0.0288	-0.0284	-0.0293	-0.0296	-0.0286	-0.0288
LabForce	0.0102	**0.2997**	**0.2487**	**0.2646**	**0.2525**	**0.2445**	**0.2599**	**0.2572**
Foreign	0.0068	0.0038	0.0071	0.0068	0.0074	0.0078	0.0069	0.0069
English	-0.0057	-0.0350	-0.0337	-0.0346	-0.0342	-0.0336	-0.0344	-0.0344
EthnoL	0.0055	**0.1331**	**0.1084**	**0.1155**	**0.1094**	**0.1045**	**0.1122**	**0.1116**
Spanish	0.0053	0.1066	0.0821	0.0888	0.0833	0.0792	0.0857	0.0852
stdBMP	-0.0034	-0.0021	-0.0037	-0.0035	-0.0036	-0.0039	-0.0035	-0.0035
French	0.0037	0.0585	0.0448	0.0492	0.0458	0.0432	0.0471	0.0464
Abslat	0.0006	-0.0071	-0.0066	-0.0065	-0.0065	-0.0069	-0.0067	-0.0068
WorkPop	-0.0030	-0.0035	-0.0062	-0.0060	-0.0062	-0.0064	-0.0062	-0.0061
HighEnroll	-0.0044	**-0.1871**	**-0.1441**	**-0.1551**	**-0.1453**	**-0.1393**	**-0.1513**	**-0.1492**
Popg	0.0027	0.0072	0.0110	0.0104	0.0106	0.0113	0.0105	0.0103
Brit	-0.0018	0.0413	0.0263	0.0303	0.0271	0.0249	0.0288	0.0281
OutwarOr	-0.0020	-0.0286	-0.0266	-0.0280	-0.0265	-0.0258	-0.0270	-0.0267
Jewish	-0.0012	-0.0009	-0.0019	-0.0018	-0.0020	-0.0021	-0.0020	-0.0018
RevnCoup	-0.0001	-0.0003	-0.0002	-0.0002	-0.0003	-0.0003	-0.0002	-0.0003
PublEdupct	0.0004	0.0266	0.0229	0.0238	0.0228	0.0226	0.0237	0.0235
Area	-0.0005	-0.0026	-0.0030	-0.0031	-0.0031	-0.0030	-0.0031	-0.0031
$E(k|Y)$	10.442	19.657	20.389	20.518	20.411	20.377	20.395	20.368
$E(\frac{g}{1+g}|Y)$	0.999	0.986	0.955	0.960	0.955	0.951	0.958	0.958

Table 10: Fully standardized posterior means for different prior settings. Coefficients corresponding to covariates with PIP exceeding 0.5 in bold.
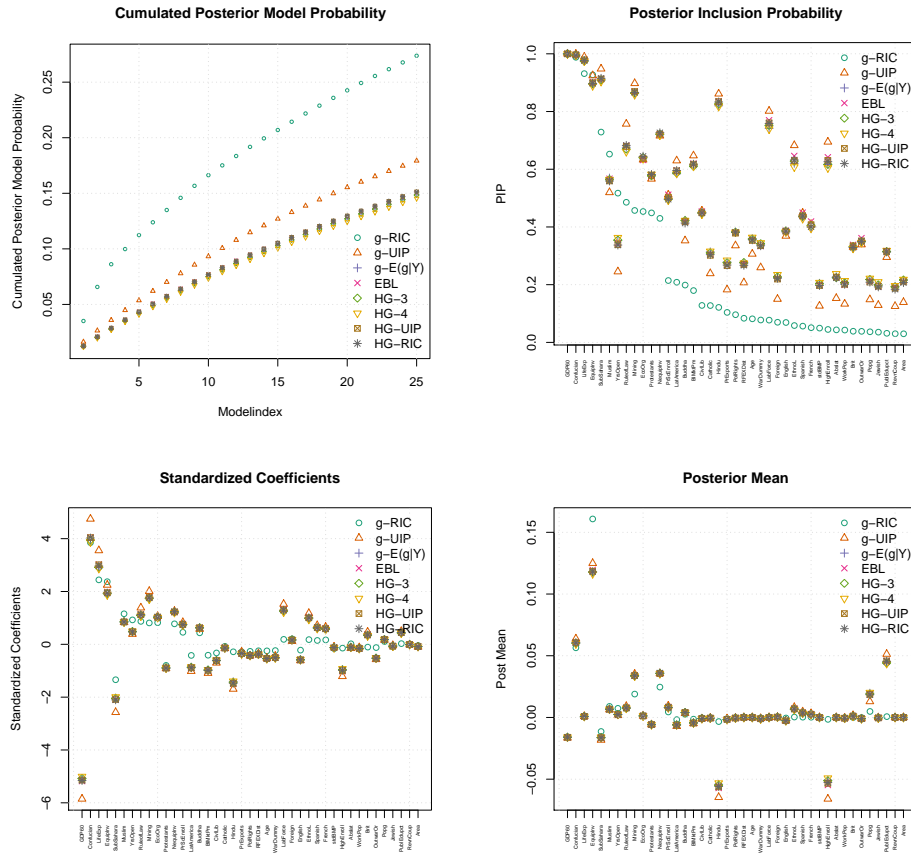
Figure 5: Top left panel shows the cumulative posterior mass, left panel the posterior inclusion probabilities, bottom left panel the standardized coefficients and bottom right panel the posterior mean for the growth determinants exercise.