Identifying groups of determinants in Bayesian model averaging using Dirichlet process clustering

Bettina Grün1 | Paul Hofmarcher2

1Institute for Statistics and Mathematics, Wirtschaftsuniversität Wien, Wien, Austria
2Salzburg Centre of European Union Studies (SCEUS), Department of Business, Economics and Social Theory, Paris Lodron University of Salzburg, Salzburg, Austria

Correspondence
Paul Hofmarcher, Salzburg Centre of European Union Studies (SCEUS), Department of Business, Economics and Social Theory, Paris Lodron University of Salzburg, Salzburg, Austria. Email: paul.hofmarcher@sbg.ac.at

Abstract
Model uncertainty is a pervasive problem in regression applications. Bayesian model averaging (BMA) takes model uncertainty into account and identifies robust determinants. However, it requires the specification of suitable model priors. Mixture model priors are appealing because they explicitly account for different groups of covariates as robust determinants. Specific Dirichlet process clustering (DPC) model priors are proposed; their correspondence to the binomial model prior derived and methods to perform the BMA analysis including a DPC postprocessing procedure to identify groups of determinants are outlined. The application of these model priors is demonstrated in a simulation exercise and in an empirical analysis of cross-country economic growth data. The BMA analysis is performed using the Markov chain Monte Carlo model composition sampler to obtain samples from the posterior of the model specifications. Results are compared with those obtained under a beta-binomial and a collinearity-adjusted dilution model prior.

KEYWORDS
Bayesian model averaging, Dirichlet process clustering, latent class analysis, mixture model, model prior, model uncertainty.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Scandinavian Journal of Statistics published by John Wiley & Sons Ltd on behalf of The Board of the Foundation of the Scandinavian Journal of Statistics.
INTRODUCTION

Inference under model uncertainty is a pervasive problem in many regression applications. Bayesian model averaging (henceforth BMA) in combination with suitable computational tools has become a standard method to account for model uncertainty (Hoeting et al., 1999; Hofmarcher & Grün, 2020; Raftery et al., 1997). The use of BMA provides better predictive performance (Madigan & Raftery, 1994) and identifies robust determinants (Fernández et al., 2001b). The method has been widely adopted for a range of different application areas including cross-country long-term economic growth (Fernández et al., 2001a; Ley & Steel, 2009, 2012; Masanjala & Papageorgiou, 2008) as well as other economic and finance applications (Moral-Benito, 2015; Ouysse & Kohn, 2010), wind speed prediction in weather forecasting (Baran, 2014; Sloughter et al., 2010), and the detection of differentially expressed genes in observational gene expression studies (Zhou et al., 2012).

A key aspect of any BMA application—as for any Bayesian estimation—is the specification of suitable priors. The priors in BMA consist of two parts: (1) the prior for covariate inclusion, so-called model priors, and (2) the prior for the regression coefficients and error variance (henceforth regression parameter priors). A standard default approach for the prior on the regression coefficients and the error variance is the use of (hyper-)g-priors for the regression coefficients of the covariates together with noninformative priors for the intercept and the error variance (see, e.g., Liang et al., 2008; Ley & Steel, 2009, 2012; Hofmarcher et al., 2015).

More controversy exists with respect to the specification of suitable model priors. The basic model prior is the binomial model prior which assigns an equal a priori inclusion probability \( \theta \) to all potential covariates. Sala-i-Martin et al. (2004) used \( \theta = \bar{k}/K \), with \( \bar{k} \) denoting the prior expected model size and \( K \) is the total number of potential covariates. The prior expected model size \( \bar{k} \) is fixed by the researcher a priori. To increase flexibility, and to reduce the dependence on the prior expected model size \( \bar{k} \), Ley and Steel (2009) proposed to put a beta hyperprior on the inclusion probability \( \theta \). This results in the beta-binomial (BB) model prior, which has become the standard workhorse model prior for most empirical BMA exercises. Indeed, the BB model prior is appealing if the model prior should cover the uncertainty regarding the true model size, that is, if researchers want to add one additional layer of uncertainty over the prior expected model size. One drawback of the BB model prior is that it does not take the correlation structure between covariates into account. To compensate for redundancy within the model space, Chipman (1996) and George (1999) introduced the concept of dilution priors. The aim of dilution priors is to account for similarities among the models (see also George, 2010) by, for example, down-weighting the prior probability of models which include highly correlated covariates and thus cover a similar model space. Moser and Hofmarcher (2014) studied the effect of dilution priors on BMA growth applications, in particular when interaction terms are present, to conclude that the different priors lead to similar predictive performance. Results presented in this article indicate that the BB prior should also be questioned as default model prior to be used if groups of covariates are suspected to be relevant and the aim of the BMA analysis is to identify these groups.

In the BMA framework, the model prior captures two potential veins of uncertainty: the model size and the inclusion of specific covariates or groups of covariates. To focus on model size uncertainty, that is, the number of included covariates, a vague prior on model size, which is, for example, induced by the BB model prior, needs to be imposed. The second vein of uncertainty might either only focus on the inclusion of single covariates or aim at assessing the joint
inclusion of several covariates. The latter is of particular importance if, for example, due to different proposed theories, competing groups of covariates are expected to equally well constitute robust determinants for the dependent variable.

To assess the joint inclusion of several covariates, Durlauf et al. (2008) proposed a model prior which explicitly accounts for the presence of competing theories and different groups of covariates being associated with each of the theories. Within the group of covariates for a given theory, Durlauf et al. (2008) employed dilution priors. This approach assesses the importance of a theory conditional on the associated covariates being selected. However a drawback of this approach is, that it requires the theories as well as the set of covariates associated with a specific theory to be a priori known.

To assess the uncertainty with respect to the covariate inclusion of single as well as pairs or groups of covariates, different postprocessing tools for the results of a BMA analysis were also investigated. In particular bivariate jointness measures were proposed to identify pairs of covariates which tend to appear together in the models (complements), or which tend to exclude each other (substitutes) (see Doppelhofer & Weeks, 2009; Ley & Steel, 2007, 2009; Strachan, 2009; Hofmarcher et al., 2018 as well as an empirical comparison by Man, 2018). Ley and Steel (2007) point out that these bivariate measures can easily be extended to more than two regressors and thus multivariate jointness measures obtained. In addition to these methods, Crespo Cuaresma et al. (2016) proposed an alternative postprocessing approach to account for the dependency in covariate inclusion. They analyzed the full multivariate posterior distribution of covariate inclusion using a Dirichlet process clustering (DPC) analysis (see, e.g., Kim et al., 2006; Molitor et al., 2010). This does not only account for pairwise jointness, but the full multivariate jointness between all covariates.

In this article we focus on a suitable BMA analysis for assessing model uncertainty with respect to the inclusion of groups of covariates where the specific group structure is not a priori known. We argue in favor of a DPC prior as model prior to facilitate the detection of a grouping structure in the covariates with varying group-specific joint inclusion probabilities. The reasons why DPC priors are suggested as model priors in BMA applications are twofold. First, mixture priors are a natural choice to a priori take into account that different groups of covariates may be relevant for the outcome variable. This is in line with the model formulation in Durlauf et al. (2008), where competing groups of explanatory variables emerging from different theories are assumed to be relevant for determining long-term per capita GDP growth. Second, mixture priors do not require to specify these groups of covariates a priori but they are determined in a data-driven way. In contrast to finite mixtures, Dirichlet process clustering priors have the advantage that they do not require the number of groups to be specified a priori. Compared with Durlauf et al. (2008), the proposed DPC model priors thus constitute a data-driven approach to identify theories and their associated covariates. Conditional on suitable model and regression parameter priors being specified, we outline how to proceed with the BMA analysis. We derive the correspondence between DPC model priors and the binomial model prior and describe how to obtain suitable posterior estimates given the suggested prior specifications. In particular we use the Markov chain Monte Carlo (MCMC) model composition (MC3) sampler to obtain samples from the posterior of the model specifications. We propose to use a DPC postprocessing approach to identify the groups of covariates and enable the assessment of joint inclusion and the identification of complements and substitutes based on these posterior samples. Overall the BMA approach presented is tailored for a situation where the focus is on model uncertainty with respect to specific groups of covariates being included and accounts for jointness of inclusion to provide insights on robust determinants for the dependent variable.
This article is organized as follows. Section 2 gives a short overview of the BMA methodology and describes the standard model priors used in the literature hitherto. The hierarchical specification of the DPC model prior and estimation and inference including a suitable Dirichlet process clustering postprocessing procedure to identify groups of determinants under this model prior are derived in Section 3. Section 4 presents a simulation exercise. The results based on an BMA analysis using the MCMC model composition sampler imply that the DPC model prior may lead to slightly better results than the BB prior and to considerably better results than the dilution prior when aiming at detecting the cluster structure present in the covariates for predicting the dependent variable when the size of the covariate groups is suitably specified and the state-of-the-art hyper-g-prior for the regression parameters is used. Section 5 presents the results of applying the DPC model prior to a cross-country growth data set and compares them to those obtained using the standard BB model prior as well as the dilution prior. Finally, Section 6 summarizes the results and relates them to previous work on post hoc jointness analysis of covariates in BMA.

2 BMA AND PRIOR SPECIFICATIONS

The standard BMA setting assumes that there is a linear relationship between $n$ observations $y$ and a group of covariates $X_l$ chosen from the complete set of potential covariates $X$. The linear model assumes additive noise and that the dependent variable follows a normal distribution with the mean given by the linear predictor:

$$y \sim N(\mathbf{1}\beta_0 + X_l\beta, \sigma^2 I),$$

where $N(\mu, \Sigma)$ is the multivariate normal distribution with mean $\mu$ and variance–covariance matrix $\Sigma$, $\mathbf{1}$ is a vector of ones of length $n$ and $X \in \mathbb{R}^{n \times K}$. $X_l$ is obtained from $X$ by selecting a subset of covariates. This subset selection can also be indicated by a 0–1 vector $\gamma_l$ of length $K$. Thus the vector $\gamma_l$ characterizes the set of covariates included in a specific regression model. The vector $\gamma_l$ can take $2^K$ different values, corresponds to the covariate inclusion vector for a given model and represents a particular model specification.

Taking model uncertainty into account the posterior distribution of the parameters given the data corresponds to

$$p(\beta_0, \beta, \sigma^2 | y) = \sum_{l=1}^{2^K} p(\beta_0, \beta, \sigma^2 | y, \gamma_l) p(\gamma_l | y).$$

A Bayesian approach requires specifying the regression parameter prior $p(\beta_0, \beta, \sigma^2 | \gamma_l)$ and the model prior $p(\gamma_l)$. In general both priors jointly impact on the posterior of the regression parameters and on the posterior of the model specifications.

2.1 Regression parameter priors

A usual choice for the regression parameter prior, conditional on a specific model specification $\gamma_l$, is a constant flat prior for $\beta_0$, an uninformative prior $p(\sigma^2) \propto 1/\sigma^2$ for the error variance and the so-called Zellner’s $g$-prior for the regression coefficients $\beta$ (Zellner, 1986):
\[ \beta | \sigma^2, \gamma_l, g \sim N \left( 0, \sigma^2 g (X_l'X_l)^{-1} \right) . \]

The g-prior has the advantage that only one parameter, namely g, needs to be specified and that the marginal likelihood can be evaluated analytically. The influence of g has also been investigated in combination with different model priors (see, e.g., Fernández et al., 2001a; Ley & Steel, 2009). In particular Eicher et al. (2011) give an explicit formula for a specific model prior setting to tradeoff the specification of g in the regression parameter prior with the model prior. Specific choices for g are discussed and compared in Fernández et al. (2001a). To alleviate the influence of g, hyperpriors for g were proposed with a benchmark prior given by Ley and Steel (2012). In the following the hyper-g-prior is considered using a beta distribution for \( g/(1 + g) \) such that

\[ \frac{g}{1 + g} \sim \text{Beta} \left( 1, \frac{1}{n} \right) . \]

This implies that \( g/(1 + g) \) has mean \( n/(n + 1) \). This corresponds to the shrinkage imposed on the regression coefficients by the unit information prior where \( g = n \) (Zeugner & Feldkircher, 2015).

Alternative regression parameter priors for BMA include, among others, independent Gaussian priors, lasso and elastic net priors, and robust g-priors. In the BMA context Malsiner-Walli et al. (2019) compare independent Gaussian priors to the g-prior and Hofmarcher et al. (2015) compare lasso and elastic net priors to other parameter priors. Bayarri et al. (2012) provide explicit results on the impact of the robust g-prior on the posterior of the model specifications. The horseshoe prior has also been proposed as regression parameter prior (Carvalho et al., 2010). However, this prior already constitutes an alternative to the spike-and-slab prior used in BMA to perform variable selection and thus is in general not used within BMA applications (Piironen & Vehtari, 2017).

### 2.2 Model priors

Several priors were proposed hitherto in the literature for the covariate inclusion vectors \( \gamma_l \). In the following we present the binomial model prior proposed by Sala-i-Martin et al. (2004) and the extension to the BB model prior put forward by Ley and Steel (2009). These two model priors treat the covariates exchangeably. In fact, for these priors, the prior weight of a specific covariate inclusion vector only depends on the number of covariates being included and not which specific covariates are included. The priors differ in the prior weight assigned to models of different size, with the binomial prior having more weight concentrated close to the prior expected model size and the BB prior distributing the prior weight in a more dispersed way.

Finally, the collinearity-adjusted dilution (DIL) model prior proposed by George (2010) is presented. This prior takes the correlation between covariates into account and reduces the joint inclusion probability of highly correlated covariates thus not only taking the number of covariates included into account.

#### 2.2.1 Binomial model prior

Sala-i-Martin et al. (2004) specified their prior model probabilities by choosing a prior expected model size \( \tilde{k} \), with each covariate having a prior inclusion probability of \( \theta = \frac{\tilde{k}}{K} \), and \( K \) denoting
the number of potential covariates. The marginal probability of the model induced by $\gamma_i$ is a priori given by

$$p(\gamma_i | k) = \left( \frac{k}{K} \right)^{k_i} \left( 1 - \frac{k}{K} \right)^{K-k_i} \propto \left( \frac{k}{K - k} \right)^{k_i},$$

where $\gamma_i$ contains $k_i$ covariates, that is, $1^\prime \gamma_i = k_i$. The prior probability of a specific model $\gamma_i$ thus is the same for all models with the same number of covariates included.

### 2.2.2 BB model prior

To increase the flexibility of the model prior and to reduce the influence of the prior expected model size $k$, Ley and Steel (2009) proposed a hyperprior on the inclusion probability $\theta$. This changes the prior for $\gamma_i$ to the BB prior, that is, in the hierarchical specification:

$$p(\gamma_i | \theta) = \theta^{k_i} (1-\theta)^{K-k_i},$$

$$\theta \sim \text{Beta}(a, b),$$

which gives

$$p(\gamma_i | a, b) = \frac{\Gamma(a + b)}{\Gamma(a) \Gamma(b)} \int_{\Omega_0} \theta^{k_i} (1-\theta)^{(K-k_i)} \theta^{a-1} (1-\theta)^{b-1} d\theta$$

$$= \frac{\Gamma(a + b)}{\Gamma(a) \Gamma(b) \Gamma(a + b + K)} \Gamma(a + k_i) \Gamma(b + K - k_i).$$

In the BB model prior, the probability of a specific model $\gamma_i$ with $k_i$ included covariates conditional on a specific value of $\theta$ equals the probability of the binomial model prior with $\theta$ corresponding to $\tilde{k}/K$. Integrating over $\theta$ gives that the marginal probability of a specific model $\gamma_i$ is proportional to the product of two Gamma functions, with arguments depending on the choice of the beta parameters $a, b$ and the model size $k_i$ induced by $\gamma_i$. Again the prior probability of a specific model does only depend on the number of covariates included $k_i$, but not the specific covariates. A detailed discussion of this model prior can be found in Ley and Steel (2009). They proposed to set $a = 1$ and $b = (K - k)/k$ to induce a prior expected model size of $\tilde{k}$.

### 2.2.3 Collinearity-adjusted dilution model prior

The model priors presented in the previous sections do not take the collinearity between the potential covariates into account, but assume two covariates are a priori equally likely to be included regardless of if they are uncorrelated or highly correlated. George (2010) criticized this approach as putting disproportionate mass on correlated covariates and proposed to adjust the prior model weight of a model $\gamma_i$ by taking the value of the determinant of the correlation matrix $C_{\gamma_i}$ of $X_i$ into account. Note that $|C_{\gamma_i}| = 1$ with $| \cdot |$ denoting the determinant, if the included covariates are uncorrelated and $|C_{\gamma_i}| = 0$ if they are perfectly collinear. In the DIL model prior proposed by George (2010) the prior probability of a specific model $\gamma_i$ is given by
for some monotonic function \( f \) satisfying \( f(1) = 1 \) and \( f(0) = 0 \). In the following we use the identity function for \( f \). The function \( f \) controls the down-weighting applied to groups of covariates in dependence of their correlation structure compared with the binomial model prior with \( \theta = k/K \).

If a set of covariates is collinear the a priori probability of including that particular set is zero for this prior. Each noncollinear subset has a positive a priori probability to be included. If covariates are orthogonal, \( \theta = k/K \) also induces a prior model size of \( k \) as the binomial prior.

The prior proposed by George (2010) is based on capturing the global dependency structure between the covariates in a model by only considering the determinant. Alternatively a prior could be considered which takes a more local (e.g., pairwise) correlation structure into account to adjust the model prior for redundancy in the covariates. This would result in an alternative dilution prior in addition to the DIL model prior.

### 3 THE DPC MODEL PRIOR

BMA is in general used to identify covariates which are robust determinants regardless of other covariates being included. If for a specific regression problem different groups of covariates are assumed to be present which are associated with the outcome variable, BMA may also be employed to identify robust determinants as part of a group of covariates. To be in line with such an assumption, using a mixture prior based on latent class analysis as a model prior explicitly accounts for different groups of covariates being relevant for the outcome. Lazarsfeld (1950) developed latent class analysis to account for dependencies in categorical multivariate data. These dependencies are assumed to be caused by latent groups where the categorical variables are independent within groups. If the categorical multivariate data consists of binary variables only, the latent class analysis model corresponds to a mixture of independent Bernoulli distributions.

Finite or infinite mixture models could be used as mixture priors. However, infinite mixtures, such as Dirichlet process clustering models, have the advantage that they do not require to a priori fix the number of groups and thus facilitate the determination of a suitable number of groups for a given data set (see, e.g., McAuliffe et al., 2006). We thus propose to use a DPC model prior which also corresponds to the limiting case of a finite mixture prior where the number of components grows to infinity and the parameter used for the Dirichlet weights times the number of components converges to a fixed positive finite value (Green & Richardson, 2001). We would expect that results for finite mixtures are similar, in particular if the sparse finite mixture approach suggested by Malsiner-Walli et al. (2016) is implemented. Frühwirth-Schnatter and Malsiner-Walli (2019), for example, illustrated that sparse finite mixtures may give similar results to infinite mixtures based on DPC if the prior parameter settings are suitably matched.

#### 3.1 Hierarchical specification

The DPC model has been developed in Bayesian nonparametrics (Antoniak, 1974; Ferguson, 1973) and has various applications in mixture modeling (see, e.g., Kim et al., 2006). The hierarchical relationships of the DPC model imply the following model prior specification on \( \gamma \):

\[
p(\gamma | \theta) \propto f(|C_{\gamma}|) \theta^k (1 - \theta)^{K-k},
\]
\[
p(y|\theta_l) = \prod_{j=1}^{K} \theta_{lj}^{y_{lj}} (1 - \theta_{lj})^{1-y_{lj}},
\]

\[
\theta_{l|G} \sim G(\cdot),
\]

\[
G|G_0, \alpha \sim DP(\cdot|G_0, \alpha).
\]

The Dirichlet process is denoted by \(DP(\cdot|G_0, \alpha)\), where \(G_0\) is the base measure and \(\alpha\) is the scaling parameter.

In the case of latent class analysis with a product of Bernoulli distributions as class distributions, the base measure \(G_0\) corresponds to

\[
\prod_{j=1}^{K} \text{Beta}(a, b).
\]

The set of hyperparameters for the DPC model prior thus consists of \((\alpha, a, b)\), the scaling parameter for the Dirichlet process \(\alpha\) and the parameters \(a\) and \(b\) for the base measure.

Alternatively the DPC model prior can be specified using the stick-breaking representation proposed by Sethurman (1994) (see also Liverani et al., 2015). The probability distribution for the covariate inclusion parameters \(\theta_l\) is then given by:

\[
G(\cdot) = \sum_{c=1}^{\infty} \pi_c \delta_{\theta_c}(\cdot),
\]

where \(\delta_{\theta}(\cdot)\) denotes the point mass at \(\theta\). This distribution is determined by first drawing the sticks

\[
V_c \sim \text{Beta}(1, \alpha), \quad c = 1, \ldots, \infty.
\]

This corresponds to

\[
f(v_c|\alpha) = \alpha(1 - v_c)^{\alpha-1},
\]

and induces the weights

\[
\pi_1 = v_1,
\]

\[
\pi_c = v_c \prod_{l=1}^{c-1} (1 - v_l), \quad c = 2, \ldots, \infty.
\]

The cluster parameters \(\theta_c\) are drawn from the base measure:

\[
\theta_c \sim \prod_{j=1}^{K} \text{Beta}(a, b), \quad c = 1, \ldots, \infty.
\]

The scaling parameter \(\alpha\) influences how many different values for the covariate inclusion parameters \(\theta\) are a priori expected. If \(\alpha\) is large, \(v_c\) tends to be close to zero and therefore small cluster sizes and many clusters can be expected. On the other hand small values of \(\alpha\) imply large cluster sizes and few clusters can be expected.
sizes and only few clusters, that is, different values for \( \theta \). To attenuate the influence of a specific value of \( a \) a Gamma hyperprior is added, that is, \( a \sim \text{Gamma}(2, 1) \).

### 3.2 Marginal prior

The marginal prior \( p(\gamma | a, b) \) is obtained by

\[
p(\gamma | a, b) = \int_{\Omega_{\Theta, V}} p(\gamma | \Theta, V)p(\Theta | a, b)p(V | a, b)d\Theta dV
\]

\[
= \int_{\Omega_{\Theta, V}} \left\{ \sum_{c=1}^{\infty} \left[ \prod_{i=1}^{c-1} (1 - V_i) \right] \left[ \prod_{j=1}^{K} \theta_{cj}^{\gamma_j}(1 - \theta_{cj})^{1-\gamma_j} \right] \right\} d\Theta dV
\]

\[
= \int_{\Omega_V} \sum_{c=1}^{\infty} \left\{ \left[ \prod_{i=1}^{c-1} (1 - V_i) \right] \left[ \prod_{u=1}^{\infty} (1 - V_u)^{a-1} \right] \right\} dV
\]

\[
= \int_{\Omega_V} \left\{ \sum_{c=1}^{\infty} \left[ \prod_{i=1}^{c-1} (1 - V_i) \right] \left[ \prod_{u=1}^{\infty} (1 - V_u)^{a-1} \right] \frac{\alpha^k b^{K-k}}{(a+b)^K} \right\} dV
\]

\[
= \frac{\alpha^k b^{K-k}}{(a+b)^K} \left( \frac{a}{a+b} \right)^k \left( 1 - \frac{a}{a+b} \right)^{K-k},
\]

using \( \Theta = (\theta_{c})_{c=1, \ldots, \infty} \) and \( V = (V_c)_{c=1, \ldots, \infty} \). This implies that \( p(\gamma | a, b) \) is independent of \( a \) and the same as obtained for the binomial model prior for a given prior expected model size.

The proposed DPC model prior is developed to be suitable for BMA applications where a group structure of covariate inclusion is suspected. It also has the advantage to marginally correspond to a model prior already previously used in the literature, thus providing also guidance for choosing a suitable model prior among already known model priors if the group structure in covariate inclusion is to be analyzed a posteriori.

### 3.3 Inference

In BMA analysis using a DPC model prior the posterior of interest does not only consist of the regression parameters \( (\beta_0, \beta, \sigma^2) \), but also the group-specific parameters \( \Theta \) and group sizes induced by \( V \) which characterize the posterior distribution of \( \gamma \). The posterior of all parameters can be decomposed in the following way:

\[
p(\beta_0, \beta, \sigma^2, \gamma, \Theta, V | y, a, b) = p(\beta_0, \beta, \sigma^2 | \gamma, y)p(\Theta, V | \gamma, a, b)p(\gamma | y, a, b).
\]

For Bayesian inference the marginal posterior of the regression parameters consisting of \( \beta_0, \beta \) and \( \sigma^2 \) can be determined separately from the marginal posterior of the Dirichlet process clustering...
parameters:

\[ p(\beta_0, \beta, \sigma^2 | y, \alpha, a, b) = \sum_{l=1}^{2^K} p(\beta_0, \beta, \sigma^2 | \gamma_l, y) p(\gamma_l | y, a, b), \]

\[ p(\Theta, V | y, \alpha, a, b) = \sum_{l=1}^{2^K} p(\Theta, V | \gamma_l, \alpha, a, b) p(\gamma_l | y, a, b). \]

If the usual choice for the regression parameter prior is made, the posterior of the regression parameters is in general approximated using samples from \( p(\gamma | y, a, b) \) as the conditional probabilities of the regression parameters given a particular model specification are available in closed form. Similar the posterior of the DPC parameters could also be approximated using these samples. However, this is complicated because the conditional probabilities are not directly available. We will propose an alternative approach to determine point estimates for the Dirichlet process clustering parameters.

### 3.3.1 Covariate inclusion and regression parameters

The results in Section 3.2 indicate that the binomial and DPC model priors are marginally identical. This implies that the posteriors \( p(\beta_0, \beta, \sigma^2 | y, \alpha, a, b) \) and \( p(\gamma | y, a, b) \) are also identical for both model priors. Conditional on \( \gamma \) the posterior for the regression parameters are independent of the model prior. The posterior for \( \gamma \) is proportional to the marginal likelihood times marginal prior and thus also identical for both model priors. Thus the same techniques can be used to approximate them.

To determine the posterior \( p(\gamma | y, a, b) \), all possible models can be evaluated for a small number of covariates \( K \). This is not computational feasible for larger values of \( K \) and the posterior is approximated using MCMC methods to obtain a sample from \( p(\gamma | y, a, b) \). The MCMC methods allow to explore the model space in an efficient way and switch between models based on the ratio of the posterior model weights of the current and the proposed model. If the usual regression parameter prior is used, the Bayes factors for two models can be determined in closed form (Liang et al., 2008) and thus the posterior \( p(\gamma | y, a, b) \) can be directly approximated with the regression parameters integrated out. This allows to use the MCMC model composition (MC\(^3\)) algorithm developed for BMA by Madigan et al. (1995) and available in the R (R Core Team, 2020) package BMS (Zeugner & Feldkircher, 2015). The MC\(^3\) algorithm implemented uses the marginal likelihood and prior probabilities of the model prior to switch between models using birth and death as well as swap moves, that is, single covariates are either added or dropped or a pair of covariates change their roles from being included to dropped. Using this sampling scheme one can either use the empirical frequencies of the visited models or the posterior probabilities implied by the marginal likelihoods times the prior probabilities to approximate the posterior probabilities.

If mixing of the MC\(^3\) algorithm is poor using the specific model prior, one could alternatively also consider using a different sampling scheme which ensures that the sampling scheme explores the model space well, for example, by using suitably adapted values for the different models instead of the model prior values. The posterior model weights for the visited models may then only be determined based on the marginal likelihood and the prior probabilities of the specified model prior, while the empirical frequencies of the visited models are not suitable measures. Further alternative sampling schemes with better mixing have also been proposed, see, for
example, Clyde et al. (2011) and Hubin and Storvik (2018). In the following empirical analysis, only the MC\(^3\) algorithm is employed.

In the following it is assumed that \(M\) model vectors are available which are drawn using \(M\) MCMC iterations to approximate \(p(\gamma | y, a, b)\) and that these are combined in \(\Gamma\), that is,

\[
\Gamma = \{ \gamma^1, \ldots, \gamma^M \}.
\]

Given \(\Gamma\) the posterior of the regression parameters in Equation (1) can be approximated using

\[
p(\beta_0, \beta, \sigma^2 | y, a, b) \approx \frac{1}{M} \sum_{m=1}^{M} p(\beta_0, \beta, \sigma^2 | y, \gamma^m).
\]

Alternatively the sum can be taken over the unique set of covariate inclusion vectors in \(\Gamma\) using as weights the theoretic weights based on the marginal likelihoods, rescaled to sum to one over the set of considered covariate inclusion vectors (for the derivation of the marginal likelihoods see, e.g., Clyde et al., 2011).

### 3.3.2 DPC parameters

The posterior \(p(\Theta, V | y, a, a, b)\) could also be approximated using

\[
p(\Theta, V | y, a, a, b) \approx \frac{1}{M} \sum_{m=1}^{M} p(\Theta, V | \gamma^m, a, a, b),
\]

where \(\gamma^m\) are drawn from \(p(\gamma | y, a, b)\). However, the conditional probabilities of the Dirichlet process clustering parameters given the particular model specification are not readily available. In the following we thus proceed in a different way to obtain point estimates for \(\Theta\) and \(V\). These point estimates of \(\Theta\) and \(V\) help to understand any group structure present in covariate inclusion. The distribution \(p(\gamma | y, a, b)\), for \(l = 1, \ldots, 2^k\) is approximated by \(\Gamma\). Assuming a mixture distribution for \(\gamma\), a suitable partition into different groups is determined and point estimates for \(\Theta\) and \(V\) derived based on the partition.

The approach follows the procedure suggested in Molitor et al. (2010) and Liverani et al. (2015) for employing the DPC model and consists of the following three steps:

1. Use MCMC sampling to obtain a sample of partitions which group covariate inclusion vectors together in the Dirichlet process clustering model.
2. Postprocess the sample of partitions to obtain a final grouping.
3. Determine point estimates for the group-specific parameters conditional on the final grouping.

Regarding Step 1 different MCMC samplers have been proposed to obtain the sampled partitions. In the following the slice sampler proposed by Kalli et al. (2011) is used as implemented in the R package \textit{PReMiU}M (Liverani et al., 2015). If \(J\) MCMC sampling iterations are recorded after

\[1\]Package \textit{PReMiU}M performs DPC using a range of different base measures including a product of Bernoulli distributions for binary data.
discarding the burn-in iterations, Step 1 results in $J$ sampled partitions $P_1, \ldots, P_J$. Each partition $P_j$ groups the $M$ covariate inclusion vectors $\gamma^m$, $m = 1, \ldots, M$ contained in $\Gamma$ into nonempty subsets such that every vector $\gamma^m$ is included in one and only one subset.

For Step 2, Molitor et al. (2010) proposed the following procedure to determine a consensus partition. First, an estimate of the posterior coassignment probabilities to the same cluster for pairs of observations using the sampled partitions is determined. This matrix of posterior coassignment probabilities represents a similarity matrix between observations and thus can be converted to a dissimilarity matrix and used as input to a standard clustering method for grouping observations based on a dissimilarity matrix. Each entry of the dissimilarity matrix contains the average number of partitions where the two corresponding covariate inclusion vectors are assigned to different groups of the partition, that is, the dissimilarity between $\gamma_k$ and $\gamma_l$ is determined by

$$d_{kl} = \frac{1}{J} \sum_{j=1}^{C(j)} \sum_{c=1}^{C(j)} \left(1 - 1(\gamma_k \in P_{jc})1(\gamma_l \in P_{jc})\right),$$

where $1(\cdot)$ is the indicator function and the partition $P_j$ is given by the set $\{P_{jc}\}_{c=1, \ldots, C(j)}$, with $C(j)$ the number of groups in the $j$th partition. Each $P_{jc}$ contains the indices of the covariate inclusion vectors assigned to the $c$th group in the partition. The complete dissimilarity matrix is then given by $D = (d_{kl})_{k,l=1,\ldots,M}$.

Given a dissimilarity matrix, a standard clustering method which may be employed is partitioning around medoids (PAM; Kaufman & Rousseeuw, 1990). PAM determines an optimal partition based on a dissimilarity matrix for a given number of subsets. The PAM algorithm is similar to the $k$-means algorithm and groups observations into $k$ groups, with $k$ being prespecified. The standard $k$-means algorithm uses the squared Euclidean distance and employs an iterative scheme where (1) cluster centers are determined which have minimal distance to the observations currently assigned to the cluster and (2) observations are assigned to their closest cluster centers. This iterative scheme is varied in PAM by enforcing that cluster centers correspond to actual observations in the data set and are thus referred to as “medoids” and by enabling also other distance and dissimilarity measures, that is, the dissimilarity measure used to determine the dissimilarity matrix. In contrast to $k$-means, PAM does not require the data matrix, but can also be applied if only the dissimilarity matrix between observations is available.

Kaufman and Rousseeuw (1990) proposed to use the average silhouette width as criterion to select the optimal number of clusters. The average silhouette width is determined as the mean of the silhouette values of all observations. Rousseeuw (1987) defines the silhouette value for an observation as the difference between the average dissimilarity of this observation with all observations in its next closest, that is, neighboring, cluster and with the observations in its own cluster divided by the maximum of the two values to obtain a score in $[-1, 1]$. Negative values indicate that the average dissimilarity of this observation is smaller for the neighboring cluster than for its own cluster. The higher the average silhouette width the better is a clustering solution. Using average silhouette width as criterion only selects a minimum number of two clusters. In order to also consider one cluster as a suitable clustering solution the $C$ index as implemented in Charrad et al. (2015) is used as an additional criterion to assess if forming only one cluster would be preferable (see also Charrad et al., 2014). In Charrad et al. (2015), the $C$ index is determined by subtracting from the average within-cluster dissimilarities the minimum within-cluster dissimilarity and dividing by the difference between maximum and minimum within-cluster dissimilarity. If
the C index does not suggest to prefer the one cluster solution, the best clustering solution (with at least two clusters) is selected according to the average silhouette width.

Step 2 results in a consensus partition of the covariate inclusion vectors in \( \Gamma \). In Step 3 this consensus partition is the basis for determining cluster sizes and cluster-specific parameters. In the following only point estimates based on the best consensus partition are considered. Point estimates for \( V \) are determined based on the relative sizes of the groups in the best consensus partition. The best consensus partition splits the model vectors \( y \) into different groups. The point estimates for \( \Theta \) are obtained by calculating the mean covariate inclusion separately for each group of model vectors \( y \).

The DPC postprocessing procedure might be applied regardless of model prior used in the BMA analysis to obtain the model vector posterior samples \( \Gamma \). However, if not the DPC model prior is specified this corresponds to a post hoc analysis which aims at identifying a structure which has not been a priori supported by the specified model priors. In the following empirical analysis, we also combine the Dirichlet process clustering postprocessing step with BMA results obtained with other model priors to assess the impact of a suitable model prior specification.

4 | SIMULATION STUDY

The simulation study focuses on investigating the differences in results obtained for a BMA analysis with different model priors and using the DPC postprocessing procedure to identify groups of determinants with artificial data. The model priors compared are the DPC, the BB, and the DIL model prior. The data generation process assumed for the artificial data is in line with a Dirichlet process clustering model prior. Two competing groups of covariates predict the dependent variable equally well and these two competing groups of covariates are of equal size. Of primary interest are the results obtained with model priors where the prior expected model size corresponds to the size of the true competing groups of covariates and using the hyper-g-prior for the regression parameters. In addition the differences in results obtained are discussed if the prior expected model size is increased and different regression parameter priors are considered.

The artificial data are generated using a set of 20 potential covariates, \( x_k \), \( k = 1, \ldots, 20 \). The dependent variable \( y \) is generated such that either the first three covariates \( (k = 1, 2, 3) \) or the next three covariates \( (k = 4, 5, 6) \) are suitable regressors in the linear model. This means

\[
y = \sum_{k=1}^{3} x_k + \epsilon \approx -\sum_{k=4}^{6} x_k + \epsilon,
\]

where \( \epsilon \sim N(0, \sigma^2) \). To ensure that the dependent variable can be predicted using either group of covariates, the covariates are generated in the following way: First, 20 independent, standard normally distributed covariates \( x_k \) are generated and the expected value \( y \) is assumed to be equal to \( \sum_{k=1}^{3} x_k \). Then, \( x_6 \) is redefined to ensure that \( -\sum_{k=4}^{6} x_k \approx \sum_{k=1}^{3} x_k \), that is,

\[
x_6 = -\left[ \sum_{k=1}^{3} x_k + \sum_{k=4}^{5} x_k \right] + \epsilon,
\]

where \( \epsilon \sim N(0, 0.001^2) \). The normally distributed noise is added in order to avoid multicollinearity problems. This implies a negative correlation between \( x_6 \) and the covariates \( x_j \), \( j = 1, \ldots, 5 \), while
all other covariates are uncorrelated. The first three covariates as well as \( x_6 \) are directly correlated to \( y \), whereas the covariates \( x_4 \) and \( x_5 \) have substantial partial correlation with \( y \) conditional on \( x_6 \). Predictions of \( y \) based on the column space spanned by the first three covariates perform equally well as those based on the column space spanned by the next three covariates.

In the simulation study the number of observations is set to 40. This corresponds to a usual scenario where BMA methods are employed to account for model uncertainty in the context of long-term economic growth. The standard deviation \( \sigma \) of the error term \( \epsilon \) is varied to assess how results differ depending on the strength of the linear relationship. The influence of the specified model prior on the results is assumed to be smaller in case \( \sigma \) is smaller and the information contained in the data is stronger. The five different values considered for \( \sigma \) are \( \{0.01, 0.1, 0.5, 1, 2\} \).

Note that the number of observations and \( \sigma \) are related and a similar signal-to-noise ratio can be achieved by increasing both, the number of observations and \( \sigma \). For each setting 100 artificial data sets are generated.

The analysis is performed for three different model priors: (1) the DPC, (2) the BB, and (3) the DIL model prior. Standard BMA analysis is performed for each model prior to obtain a sample \( \Gamma \) from the posterior distribution of covariate inclusion vectors \( \gamma \). In the following detailed results are provided where the prior expected model size for all three model priors is set to 3 (under the assumption of orthogonal covariates for the DIL prior) and the hyper-\( g \)-prior for the regression coefficients is used. The sample \( \Gamma \) is drawn using three million iterations after two million burn-in iterations of the MC\(^3\) algorithm implemented in the R package BMS (Zeugner & Feldkircher, 2015).

The convergence of the MC\(^3\) algorithm is assessed using the correlation between the posterior model probabilities determined using the marginal likelihoods as well as using the empirical frequencies how often the models were visited. On average across all artificial data sets and the different \( \sigma \) specifications, these correlations are high indicating a strong congruence between those posterior model probability estimates and suggesting convergence of the chains. For the three different model priors the average correlations vary between 0.980 for the DPC model prior, 0.971 for the BB model prior, and 0.973 for the DIL model prior. As an additional check to investigate the mixing of the chains, two chains are initialized using either of the true models and combined. Results obtained with this initialization scheme are almost the same for the Dirichlet process clustering and the BB model prior. Slightly better clustering performance is obtained in this case for the DIL model prior. This indicates that mixing of the chains is worse when using the DIL model prior and more care is needed when using this model prior. This also suggests that the performance of the DIL model prior observed might not only be influenced by the characteristics of the model prior imposed, but also by the difficulties of the sampling scheme employed to obtain suitable draws from the posterior of the model specifications.

To increase the computational efficiency in the Dirichlet process clustering postprocessing step the sample \( \Gamma \) is reduced to only 10,000 sampled covariate inclusion vectors \( \gamma^m \) such that these vectors have the highest posterior weights determined based on the marginal likelihoods and that their inclusion frequencies are proportional to these posterior weights. This reduced sample is analyzed using DPC as implemented in package PReMiuM (Liverani et al., 2015). For this analysis a Gamma(2, 1) prior for \( a \) and an expected model size corresponding to the prior expected model size of the model priors are used, that is, \( a = 1 \) and \( b = (K - \bar{k})/\bar{k} \) with \( \bar{k} = 3 \) and \( K = 20 \).

Table 1 summarizes the model specifications used as input for the postprocessing procedure which are obtained with MC\(^3\) for the three model priors consisting of (1) the Dirichlet process clustering (DPC) model prior, (2) the BB model prior, and (3) the DIL model prior. The table contains for each of the \( \sigma \) specifications the average performance over the 100 artificial data sets with
TABLE 1 Correct estimation, that is, fraction of “true” models for the different model priors and $\sigma$ specifications including either the first three ($k = 1, 2, 3$) or the next three ($k = 4, 5, 6$) covariates (top part, columns 2–4); overestimation, that is, fraction of models which contain either the first three or the next three covariates plus additional covariates (top part, columns 5–7); missingness, that is, fraction of models which do include neither $x_1$ to $x_3$ nor $x_4$ to $x_6$ (bottom part, columns 2–4); mean model size (bottom part, columns 5–7)

| $\sigma$ | Correct estimation | Overestimation | Missingness | Mean model size |
|----------|-------------------|----------------|-------------|-----------------|
|          | DPC   | BB   | DIL | DPC   | BB   | DIL | DPC   | BB   | DIL |
| 0.01     | 0.94 (0.00) | 0.94 (0.00) | 0.95 (0.00) | 0.06 (0.00) | 0.06 (0.00) | 0.05 (0.00) | 3.06 (0.00) | 3.06 (0.00) | 3.05 (0.00) |
| 0.1      | 0.92 (0.02) | 0.91 (0.02) | 0.92 (0.01) | 0.08 (0.02) | 0.09 (0.02) | 0.08 (0.01) | 3.08 (0.02) | 3.09 (0.02) | 3.08 (0.02) |
| 0.5      | 0.58 (0.14) | 0.54 (0.15) | 0.61 (0.14) | 0.42 (0.14) | 0.46 (0.15) | 0.39 (0.14) | 3.54 (0.23) | 3.66 (0.33) | 3.48 (0.22) |
| 1        | 0.34 (0.15) | 0.28 (0.15) | 0.38 (0.16) | 0.56 (0.16) | 0.62 (0.17) | 0.52 (0.16) | 3.85 (0.37) | 4.18 (0.63) | 3.73 (0.35) |
| 2        | 0.04 (0.06) | 0.03 (0.05) | 0.04 (0.07) | 0.13 (0.19) | 0.14 (0.22) | 0.11 (0.17) | 2.79 (0.80) | 2.27 (1.43) | 2.62 (0.76) |

Abbreviations: BB, beta-binomial model prior; DIL, collinearity-adjusted dilution model prior; DPC, Dirichlet process clustering model prior.

The proposed DPC postprocessing approach is then applied to these sampled model specifications regardless of the model prior specified. First, the results for the automatic procedure to select a suitable number of clusters based on the average silhouette width in combination with
Table 2 contains the number of clusters selected based on this criterion for each of the 100 artificial data sets for the different values of \( \sigma \) as well as the three different model priors. If \( \sigma \) is at most 0.5, all three model priors correctly identify the two clusters for almost all data sets, with a slightly worse performance for the DIL prior. The two clusters are less frequently correctly identified if \( \sigma \) is increased beyond a value of 0.5. Higher values of \( \sigma \) lead to more clusters being identified for a large proportion of data sets regardless of the model prior employed. For small values of \( \sigma \) the number of clusters is underestimated in rare cases for any of the model priors. This might be due to poor mixing of the Markov chain for small \( \sigma \) values where switching between the two groups of covariates which predict the dependent variable equally well is less likely. In fact an improved clustering performance is observed if a different sampling scheme is employed which consists of combining the sampled model specifications using two different initializations where each of the initializations corresponds to one of the two true solutions.

In the following the DPC postprocessing procedure is applied assuming that it is known that there are two clusters in the data. This means not the average silhouette width in combination with the variant of the \( C \) index as implemented in Charrad et al. (2015) are assessed. Table 2 contains the marginal posterior inclusion probabilities (PIPs) on the aggregate level for all three model priors as well as the cluster-specific PIPs determined based on the results from the DPC analysis applied to \( \Gamma \) after fixing the number of clusters to two. Only the tables for the results for \( \sigma \in \{0.01, 0.5, 2\} \) are reported. Mixing of the MC\(^3\) algorithm deteriorates for small values of \( \sigma \), for example, \( \sigma = 0.01 \). For \( \sigma = 0.5 \) or larger, the correlation between the posterior model probabilities induced by the marginal likelihoods as well as by the empirical frequencies is at least 0.98 regardless of the model prior employed, indicating good convergence behavior. Average results over the 100 artificial data sets are shown. In addition only the results for the first 10 covariates are included because the results for covariates 11–20 are similar to those obtained for covariates seven to 10 and these rather redundant results were thus omitted.

Table 3 contains the results for \( \sigma = 0.01 \) and clearly shows that results are very similar for the DPC and BB model priors. The aggregate results indicate that the first six covariates have a PIP of about 0.5 with a posterior mean (PM) of also about 0.5 or \(-0.5\). The two clusters of covariates which both equally well predict the dependent variable are correctly identified when applying the DPC postprocessing procedure to \( \Gamma \) for both model priors. This implies that if the information in the data is very strong, results are similar regardless of if a BB or DPC model prior is employed. However, the situation is different for the DIL model prior which fails to identify the
### Table 3
Simulation results averaged over 100 datasets for $\sigma = 0.01$

|        | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ | $\beta_6$ | $\beta_7$ | $\beta_8$ | $\beta_9$ | $\beta_{10}$ |
|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-------------|
| DPC    | PIP       | PM        | C1        | PIP       | PM        | C2        | PIP       | PM        | C1        | PIP         |
| Aggregate | 0.50     | 0.50     | 0.50      | 0.50      | 0.50      | 0.00      | 0.00      | 0.00      | 0.00      | 0.00         |
|        | 0.50      | 0.50     | 0.50      | -0.50     | -0.50     | -0.50     | 0.00      | 0.00      | 0.00      | 0.00         |
|        | 1.00      | 1.00     | 1.00      | 0.00      | 0.00      | 0.00      | 0.00      | 0.00      | 0.00      | 0.00         |
|        | 1.00      | 1.00     | 1.00      | 0.00      | 0.00      | 0.00      | 0.00      | 0.00      | 0.00      | 0.00         |
|        | 0.00      | 0.00     | 0.00      | 1.00      | 1.00      | 1.00      | 0.00      | 0.00      | 0.00      | 0.00         |
|        | 0.00      | 0.00     | 0.00      | -1.00     | -1.00     | -1.00     | 0.00      | 0.00      | 0.00      | 0.00         |
| BB     | PIP       | PM        | C1        | PIP       | PM        | C2        | PIP       | PM        | C1        | PIP         |
| Aggregate | 0.51     | 0.51     | 0.51      | 0.50      | 0.50      | 0.00      | 0.00      | 0.00      | 0.00      | 0.00         |
|        | 0.51      | 0.51     | 0.51      | -0.49     | -0.49     | -0.49     | 0.00      | 0.00      | 0.00      | 0.00         |
|        | 1.00      | 1.00     | 1.00      | 0.00      | 0.00      | 0.00      | 0.00      | 0.00      | 0.00      | 0.00         |
|        | 1.00      | 1.00     | 1.00      | 0.00      | 0.00      | 0.00      | 0.00      | 0.00      | 0.00      | 0.00         |
|        | 0.01      | 0.01     | 0.01      | 0.99      | 0.99      | 0.99      | 0.00      | 0.00      | 0.00      | 0.00         |
|        | 0.01      | 0.01     | 0.01      | -0.99     | -0.99     | -0.99     | 0.00      | 0.00      | 0.00      | 0.00         |
| DIL    | PIP       | PM        | C1        | PIP       | PM        | C2        | PIP       | PM        | C1        | PIP         |
| Aggregate | 0.63     | 0.63     | 0.63      | 0.38      | 0.38      | 0.38      | 0.00      | 0.00      | 0.00      | 0.00         |
|        | 0.62      | 0.62     | 0.62      | -0.38     | -0.38     | -0.38     | 0.00      | 0.00      | 0.00      | 0.00         |
|        | 0.83      | 0.83     | 0.82      | 0.24      | 0.24      | 0.24      | 0.01      | 0.00      | 0.00      | 0.00         |
|        | 0.76      | 0.76     | 0.76      | -0.24     | -0.24     | -0.24     | 0.00      | 0.00      | 0.00      | 0.00         |
|        | 0.39      | 0.39     | 0.39      | 0.76      | 0.69      | 0.67      | 0.01      | 0.01      | 0.01      | 0.01         |
|        | 0.39      | 0.39     | 0.39      | -0.61     | -0.61     | -0.61     | 0.00      | 0.00      | 0.00      | 0.00         |

*Note:* The number of clusters is set to two when extracting cluster memberships from the DPC output.

Abbreviations: BB, beta-binomial model prior; C1, cluster 1; C2, cluster 2; DIL, collinearity-adjusted dilution model prior; DPC, Dirichlet process clustering model prior; PIP, posterior inclusion probability; PM, posterior mean.

Group structure in the covariate inclusion vectors. In fact, for both identified clusters the inclusion probabilities are spread over the six data generating covariates $x_1, \ldots, x_6$. The DIL model prior fails because the covariates are not only correlated between groups of covariates, but also within one group of covariates in this simulation setup.

Table 4 indicates how the performance of the different model priors changes if $\sigma$ is increased from 0.01 to 0.5. In this case the aggregate PIP and PM deteriorate due to the worse signal-to-noise ratio of this setting with increased error variance. However, the cluster-specific estimates for all three model priors suggest that these two groups of covariates are well identified and thus primarily the group sizes vary in dependence of the model priors. We can infer from Table 5 that for $\sigma = 2$ the performance of estimating the aggregate PIPs deteriorates further compared with $\sigma = 0.5$ for all three model priors, with the DPC model prior having slightly better performance. This observation also applies to the cluster-specific estimates where the estimated PIPs for the DPC model prior relate better to the true groups of covariate inclusion than those estimated for the other two model priors.

To assess the compactness of the identified clusters the average deviation from the true cluster-specific covariate inclusion vector is determined for each cluster separately. The true cluster-specific covariate inclusion vectors $c_1$ and $c_2$ correspond to the vectors $c_1 = (1, 1, 1, 0, \ldots, 0)$ and $c_2 = (0, 0, 0, 1, 1, 1, 0, \ldots, 0)$, respectively. The deviation for cluster $j = 1, 2$ is
TABLE 4 Simulation results averaged over 100 datasets for $\sigma = 0.5$

|       | Aggregate PIP | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 |
|-------|---------------|----|----|----|----|----|----|----|----|----|-----|
| DPC   | 0.54          | 0.54 | 0.54 | 0.48 | 0.48 | 0.48 | 0.03 | 0.06 | 0.03 | 0.03 |      |
|       | 0.52          | 0.53 | 0.52 | -0.46 | -0.47 | -0.47 | 0.00 | 0.00 | 0.00 | 0.00 |      |
| C1    | 1.00          | 1.00 | 1.00 | 0.03 | 0.03 | 0.03 | 0.02 | 0.06 | 0.03 | 0.03 |      |
|       | 0.99          | 1.00 | 0.98 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |      |
| C2    | 0.03          | 0.04 | 0.03 | 1.00 | 1.00 | 1.00 | 0.02 | 0.05 | 0.03 | 0.03 |      |
|       | 0.00          | 0.01 | 0.00 | -0.97 | -0.99 | -0.99 | 0.00 | 0.00 | 0.00 | 0.00 |      |
| BB    | 0.55          | 0.54 | 0.54 | 0.49 | 0.49 | 0.49 | 0.03 | 0.07 | 0.04 | 0.04 |      |
|       | 0.52          | 0.53 | 0.52 | -0.46 | -0.47 | -0.47 | 0.00 | 0.00 | 0.00 | 0.00 |      |
| C1    | 1.00          | 1.00 | 1.00 | 0.03 | 0.03 | 0.03 | 0.03 | 0.07 | 0.04 | 0.04 |      |
|       | 0.99          | 1.00 | 0.99 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |      |
| C2    | 0.04          | 0.05 | 0.04 | 1.00 | 1.00 | 1.00 | 0.03 | 0.06 | 0.04 | 0.04 |      |
|       | 0.00          | 0.01 | 0.00 | -0.97 | -0.99 | -0.99 | 0.00 | 0.00 | 0.00 | 0.00 |      |
| DIL   | 0.63          | 0.63 | 0.63 | 0.39 | 0.39 | 0.38 | 0.02 | 0.05 | 0.03 | 0.03 |      |
|       | 0.62          | 0.63 | 0.62 | -0.36 | -0.37 | -0.37 | 0.00 | 0.00 | 0.00 | 0.00 |      |
| C1    | 1.00          | 1.00 | 1.00 | 0.03 | 0.03 | 0.01 | 0.02 | 0.05 | 0.03 | 0.03 |      |
|       | 0.99          | 1.00 | 0.98 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |      |
| C2    | 0.02          | 0.03 | 0.02 | 1.00 | 0.99 | 1.00 | 0.02 | 0.05 | 0.03 | 0.03 |      |
|       | 0.00          | 0.01 | 0.00 | -0.97 | -0.99 | -0.99 | 0.00 | 0.00 | 0.00 | 0.00 |      |

Note: The number of clusters is set to two when extracting cluster memberships from the DPC output. Abbreviations: BB, beta-binomial model prior; C1, cluster 1; C2, cluster 2; DIL, collinearity-adjusted dilution model prior; DPC, Dirichlet process clustering model prior; PIP, posterior inclusion probability; PM, posterior mean.

given by $D_j = \sum_{k=1}^{K} |\text{PIP}_{jk} - c_{jk}|$, with $K$ denoting the total number of covariates, and $\text{PIP}_{jk}$ the PIP of variable $k$ for cluster $j$. The clusters are labeled to minimize the deviation. Results are shown in Table 6 for the three considered model priors and the different values of $\sigma$. The values are averaged over 100 artificial data sets and the clusters are identified assuming that two clusters are present. Again increasing $\sigma$ leads to less compact clusters regardless of if a DPC or BB model prior is used. However, for the same value of $\sigma$ the DPC model prior in general gives better results than the BB model prior. Although according to Table 1 the DIL model prior slightly outperforms the DPC prior in terms of correct estimation, the results of Table 6 show for the DIL model prior that this prior gives particularly bad results in case $\sigma$ is small. This may be caused by the fact that the DIL model prior down-weights higher correlated models. This deteriorates the mixing of the MCMC sampler and the sampler is more likely to be stuck in one “good” region of the model space.

Table 6 also contains the average absolute differences between the two cluster sizes for the simulation exercise. Cluster sizes correspond to the proportion of model specifications in $\Gamma$ being

2For MCMC sampling the correct prior model weights are used. Alternatively, one could consider to use different weights for sampling and then reweight the sampled models after the MCMC procedure. As a further alternative one could reduce the penalty term of the DIL prior, for example, by using $\sqrt{|C|}$ instead of $|C|$. This may also improve the mixing of the MCMC chain for the DIL prior in this case.
### TABLE 5 Simulation results averaged over 100 data sets for $\sigma = 2$

|       | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ | $\beta_6$ | $\beta_7$ | $\beta_8$ | $\beta_9$ | $\beta_{10}$ |
|-------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| DPC   |            |           |           |           |           |           |           |           |           |           |
| Aggregate | PIP       | 0.28      | 0.30      | 0.29      | 0.19      | 0.23      | 0.50      | 0.08      | 0.11      | 0.09      | 0.09      |
|        | PM        | 0.26      | 0.32      | 0.27      | -0.15     | -0.21     | -0.34     | 0.00      | 0.01      | 0.00      | 0.01      |
| C1     | PIP       | 0.54      | 0.52      | 0.54      | 0.06      | 0.06      | 0.14      | 0.07      | 0.10      | 0.08      | 0.08      |
|        | PM        | 0.54      | 0.57      | 0.53      | -0.01     | -0.02     | -0.07     | -0.01     | 0.01      | 0.00      | 0.01      |
| C2     | PIP       | 0.10      | 0.12      | 0.08      | 0.29      | 0.34      | 0.76      | 0.05      | 0.08      | 0.07      | 0.07      |
|        | PM        | 0.07      | 0.11      | 0.06      | -0.28     | -0.36     | -0.55     | 0.00      | 0.01      | 0.00      | 0.00      |
| BB    |            |           |           |           |           |           |           |           |           |           |           |
| Aggregate | PIP       | 0.25      | 0.27      | 0.26      | 0.18      | 0.21      | 0.44      | 0.08      | 0.11      | 0.09      | 0.10      |
|        | PM        | 0.17      | 0.23      | 0.19      | -0.19     | -0.25     | -0.36     | 0.00      | 0.02      | 0.00      | 0.00      |
| C1     | PIP       | 0.39      | 0.41      | 0.40      | 0.06      | 0.07      | 0.24      | 0.05      | 0.08      | 0.06      | 0.08      |
|        | PM        | 0.31      | 0.38      | 0.33      | -0.11     | -0.13     | -0.22     | -0.01     | 0.01      | -0.01     | 0.01      |
| C2     | PIP       | 0.10      | 0.12      | 0.09      | 0.24      | 0.28      | 0.61      | 0.05      | 0.07      | 0.06      | 0.06      |
|        | PM        | 0.08      | 0.11      | 0.07      | -0.24     | -0.31     | -0.46     | 0.00      | 0.01      | 0.00      | -0.01     |
| DIL   |            |           |           |           |           |           |           |           |           |           |           |
| Aggregate | PIP       | 0.28      | 0.30      | 0.29      | 0.15      | 0.19      | 0.46      | 0.07      | 0.10      | 0.08      | 0.08      |
|        | PM        | 0.27      | 0.34      | 0.28      | -0.12     | -0.17     | -0.31     | 0.00      | 0.01      | 0.00      | 0.01      |
| C1     | PIP       | 0.51      | 0.50      | 0.50      | 0.06      | 0.07      | 0.14      | 0.06      | 0.09      | 0.07      | 0.07      |
|        | PM        | 0.53      | 0.56      | 0.52      | -0.01     | -0.02     | -0.07     | -0.01     | 0.01      | 0.00      | 0.01      |
| C2     | PIP       | 0.08      | 0.10      | 0.08      | 0.23      | 0.29      | 0.74      | 0.05      | 0.08      | 0.06      | 0.06      |
|        | PM        | 0.06      | 0.10      | 0.06      | -0.24     | -0.32     | -0.52     | 0.00      | 0.02      | 0.00      | 0.00      |

*Note:* The number of clusters is set to two when extracting cluster memberships from the Dirichlet process clustering output.

Abbreviations: BB, beta-binomial model prior; C1, cluster 1; C2, cluster 2; DIL, collinearity-adjusted dilution model prior; DPC, Dirichlet process clustering model prior; PIP, posterior inclusion probability; PM, posterior mean.

... assigned to either cluster 1 or 2. For the DPC prior with $\sigma = 0.01$ we can infer that the mean absolute difference between the cluster sizes is (approximately) 0.00. This implies that roughly 50% of the sampled models are assigned to cluster 1 and 50% to cluster 2, which is in line with our data generating process. On the other hand, we do not observe this pattern for the DIL prior, where the average difference in size is 0.72 for $\sigma = 0.01$. This result is also in line with the worse mixing of the MC$^3$ algorithm for the DIL model prior. This effect vanishes, if $\sigma$ is increased, leading to all model priors having a similar performance with respect to the cluster sizes with an absolute difference of about 0.36–0.39 in cluster sizes.

The simulation study was repeated using a different prior expected model size. The results reported so far correspond to the case where the prior expected model size corresponds to the correct model size. Investigating the results for a prior expected model size of 8 indicates the performance if a wrong prior expected model size is imposed. In this case the BB prior performs best regarding the sampled models and in fact comparable to the case where the prior expected model size corresponds to the “correct” number of 3. The correct estimation decreases for the DPC prior while the overestimation increases. The performance of the DIL prior with respect to the sampled models is in-between the DPC and BB model priors, while clustering results with respect to the number of clusters selected are rather stable regardless of prior expected model size...
Table 6. Within-cluster deviation from the true cluster-specific covariate inclusion vectors and the absolute difference in cluster size (Δ) averaged over 100 artificial data sets

| σ   | Cluster 1 | Cluster 2 | Δ   | Cluster 1 | Cluster 2 | Δ   | Cluster 1 | Cluster 2 | Δ   |
|-----|-----------|-----------|-----|-----------|-----------|-----|-----------|-----------|-----|
| 0.01| 0.061     | 0.061     | 0.00| 0.072     | 0.122     | 0.01| 1.324     | 2.139     | 0.72|
| 0.1 | 0.384     | 0.098     | 0.25| 0.210     | 0.113     | 0.23| 0.990     | 2.030     | 0.67|
| 0.5 | 0.558     | 0.579     | 0.37| 0.686     | 0.703     | 0.37| 0.502     | 0.510     | 0.44|
| 1   | 1.056     | 1.222     | 0.35| 1.478     | 1.578     | 0.34| 0.962     | 1.129     | 0.41|
| 2   | 2.768     | 2.879     | 0.39| 3.126     | 3.057     | 0.36| 2.764     | 2.901     | 0.36|

Abbreviations: Δ, absolute difference in cluster size; BB, beta-binomial model prior; DIL, collinearity-adjusted dilution model prior; DPC, Dirichlet process clustering model prior.

specified. If two clusters are specified, the cluster profiles determined for the smallest σ value and the DPC model prior reflect the true grouping structure well even though also irrelevant covariates have higher inclusion probabilities due to the larger a posteriori expected model sizes. However, these inclusion probabilities are evenly distributed over all irrelevant covariates thus not negatively impacting the main interpretations drawn. But these profiles obviously lead to worse within-cluster deviations.

In a further extension of the simulation study, we varied the regression parameter priors using priors where g is fixed, in particular the unit information prior (UIP) with g = n and the risk inflation criterion prior (RIC) with g = K² in addition to the hyper-g-prior. Results indicate that changing the regression parameter priors from a hyper-g-prior to a UIP prior or a RIC prior has a similar effect as increasing the prior expected model size. For the RIC prior and even more pronounced for the UIP prior the posterior mean model size is increased compared with the hyper-g-prior and higher than 3 even if the prior expected model size is set to 3. While this leads to a decrease in correct models being sampled, the correct number of clusters is selected similarly well and the interpretation of the identified cluster profiles is unchanged as the increased PIPs are equally spread across the irrelevant covariates, thus only slightly increasing the individual PIPs of the irrelevant covariates. However, the within-cluster deviation criterion deteriorates. Changing the regression parameter priors changes the performance of the model priors, but their relative performance remains comparable with the DPC model prior performing best. These results indicate that clear advantages of the hyper-g-prior for the regression parameters compared with the UIP or RIC prior are discernible and the additional flexibility of the hyper-g-prior allows to better capture the correct model size a posteriori.

The results of the simulation study indicate that the DPC model prior performs particularly well in case not only different groups of covariates are relevant, but also if reliable guidance on the size of these groups is available.

5 | Empirical Application

The economic growth data set presented in Sala-i-Martin et al. (2004) is used to illustrate the difference in results obtained using either of the three model priors when performing BMA analysis combined with DPC postprocessing. This data set has been previously used by several authors (Crespo Cuaresma et al., 2016; Doppelhofer & Weeks, 2009; Ley & Steel, 2007) when discussing
jointness analysis for postprocessing BMA analysis results to identify complements and supplements in the covariates and thus provide a richer picture on robust determinants for long-term economic growth.

The data set contains income per capita growth rates for 88 countries over the period 1960–1996 as well as 67 covariates which have been proposed as potential determinants of income growth in the literature. In line with Sala-i-Martin et al. (2004) the BMA analysis is performed with a prior expected model size of 7 (under the assumption of orthogonal covariates for the collinearity-adjusted dilution prior) and the hyper-$g$-prior for the regression coefficients. We base our inference concerning the PIPs of covariates on three million MCMC iterations, after discarding the first two million draws as burn-in. The MCMCsampler uses the MC$^3$algorithm with birth and death as well as swap moves.

Table 7 compares the BMA results obtained for the DPC, the BB, and the DIL model priors. The estimated PIPs as well as the ranks assigned according to the PIPs are provided for the 20

| Covariate  | PIP DPC | PIP BB | PIP DIL | Rank DPC | Rank BB | Rank DIL |
|------------|---------|--------|---------|----------|---------|----------|
| EAST       | 0.87    | 0.91   | 0.89    | 1        | 1       | 1        |
| P60        | 0.85    | 0.70   | 0.71    | 2        | 2       | 2        |
| IPRICE1    | 0.81    | 0.60   | 0.60    | 3        | 3       | 3        |
| TROPICAR   | 0.65    | 0.53   | 0.51    | 5        | 4       | 4        |
| GDPCH60L   | 0.71    | 0.51   | 0.30    | 4        | 5       | 7        |
| MALFAL66   | 0.24    | 0.38   | 0.37    | 7        | 6       | 5        |
| DENS65C    | 0.52    | 0.37   | 0.31    | 6        | 7       | 6        |
| LIFE60     | 0.14    | 0.13   | 0.05    | 8        | 8       | 18       |
| SPAIN      | 0.12    | 0.12   | 0.16    | 11       | 9       | 8        |
| CONFUC     | 0.13    | 0.11   | 0.14    | 9        | 10      | 9        |
| LAAM       | 0.11    | 0.09   | 0.08    | 12       | 11      | 12       |
| SAFRICA    | 0.12    | 0.08   | 0.05    | 10       | 12      | 15       |
| YRSOPEN    | 0.07    | 0.07   | 0.05    | 16       | 13      | 17       |
| GVR61      | 0.06    | 0.06   | 0.08    | 18       | 14      | 13       |
| AVELF      | 0.08    | 0.06   | 0.05    | 13       | 15      | 16       |
| BUDDHA     | 0.07    | 0.05   | 0.09    | 14       | 16      | 10       |
| MUSLIM00   | 0.07    | 0.05   | 0.08    | 15       | 17      | 11       |
| MINING     | 0.06    | 0.05   | 0.03    | 17       | 18      | 20       |
| OPENDEC1   | 0.04    | 0.04   | 0.06    | 20       | 19      | 14       |
| H60        | 0.05    | 0.04   | 0.04    | 19       | 20      | 19       |

Note: Results are sorted by the beta-binomial model prior. For a detailed description of the covariates see Sala-i-Martin et al. (2004).

Abbreviations: BB, beta-binomial model prior; BMA, Bayesian model averaging; DIL, collinearity-adjusted dilution model prior; DPC, Dirichlet process clustering model prior; PIP, posterior inclusion probability.
## Table 8
Average silhouette width values for different numbers of clusters and the three different model priors

| Number of clusters | 2     | 3     | 4     | 5     |
|--------------------|-------|-------|-------|-------|
| DPC                | 0.445 | 0.499 | 0.536 | 0.408 |
| BB                 | 0.428 | 0.455 | 0.482 | 0.450 |
| DIL                | 0.286 | 0.361 | 0.402 | 0.434 |

Abbreviations: BB, beta-binomial model prior; DIL, collinearity-adjusted dilution model prior; DPC, Dirichlet process clustering model prior.

covariates with the highest PIP values. These 20 covariates are the same regardless of which model prior is specified for BMA analysis.

For all three model priors the same seven covariates have ranks 1–7 even though the ordering is not exactly the same. These covariates have a PIP of at least 0.24 which is considerably higher than the prior inclusion probability of $\frac{7}{67} \approx 0.1$. The covariate with the highest PIP, EAST, is the same for all three model priors with a PIP of around 90%. Considering the other covariates with high PIPs, their inclusion is in general higher for the DPC model prior than for the BB or the DIL model prior. The only exception is MALFAL66 where the PIP for the BB and the DIL model prior is substantially higher than the PIP for the DPC model prior. Covariate LIFE060 has rank 8 for the DPC and the BB model prior with a similar PIP of 0.14 and 0.13, while its PIP is much lower for the DIL model prior. For the DIL model prior LIFE060 has a PIP of 0.05 which corresponds to rank 18. On the other hand, covariates measuring religious fractions (BUDDHA, MUSLIM00) increase their importance using the DIL model prior and they also obtain better ranks in comparison.

Depending on the model prior specifications, the BMA analysis gives different samples from the posteriors of the model specifications. The DPC postprocessing procedures is applied to the samples $\Gamma$ from the posterior of the model specifications to identify groups of determinants. For computational efficiency only at most 10,000 covariate inclusion vectors are retained from the sampled vectors. The retained covariate inclusion vectors correspond to the models with the highest posterior weights (based on the marginal likelihoods) and the frequencies of these covariate inclusion vectors are proportional to the posterior weights. The priors used in the DPC postprocessing procedure are a Gamma$(2, 1)$ prior for $\alpha$ and $a = 1$ and $b = (K - \bar{k})/\bar{k}$ with $\bar{k} = 7$ and $K = 67$ for the beta prior.

In a first step in the postprocessing procedure the number of clusters needs to be determined. Table 8 contains the average silhouette width values obtained for the clustering results into the different number of clusters for the different model priors varying the number of clusters from 2 to 5. The average silhouette width values are highest for the DPC model prior, followed by the BB prior and with the smallest values for the DIL prior. This indicates that for the DPC prior a better cluster solution is obtained. The number of clusters are selected using the maximum value of the average silhouette widths. This criterion clearly suggests to select four clusters for the DPC and the BB model prior. For the DIL model prior the average silhouette width value indicates five clusters.

The cluster-specific PIP estimates obtained using the selected number of clusters, that is, four for the DPC and the BB and five for the DIL model prior, are displayed in Figure 1. The results are visualized by cluster and model prior with clusters arranged by column and the model priors by row. For the 20 most important covariates (according to the aggregate PIP values), the bar plots
Figure 1 shows that overall the cluster profiles indicated by the cluster-specific PIP values are very similar for the DPC and the BB model prior. The correspondence between the cluster profiles even seems to be stronger than the similarity between the aggregate PIP values for the two model priors. Essentially the same clusters, that is, groups of covariates which occur together, are detected and only the size of the clusters differ between the model priors (see Table 9).
Cluster sizes, cluster-specific mean model sizes, and within-cluster mean absolute deviations obtained when applying Dirichlet process clustering postprocessing to $\Gamma$ obtained to approximate the posterior distribution of model specifications using as model priors the Dirichlet process clustering, the beta-binomial, and the collinearity-adjusted dilution priors.

| Cluster   | DPC | BB | DIL | DPC | BB | DIL | DPC | BB | DIL |
|-----------|-----|----|-----|-----|----|-----|-----|----|-----|
| Cluster 1 | 0.632 | 0.496 | 0.286 | 6.724 | 6.343 | 6.414 | 0.054 | 0.054 | 0.036 |
| Cluster 2 | 0.225 | 0.392 | 0.353 | 5.180 | 4.061 | 4.225 | 0.064 | 0.054 | 0.052 |
| Cluster 3 | 0.083 | 0.061 | 0.246 | 7.258 | 7.304 | 4.661 | 0.050 | 0.052 | 0.059 |
| Cluster 4 | 0.059 | 0.051 | 0.075 | 6.150 | 5.780 | 5.608 | 0.084 | 0.083 | 0.055 |
| Cluster 5 | 0.040 | 0.040 | 0.040 | 6.610 | 6.610 | 6.610 | 0.047 | 0.047 | 0.047 |

Table 9. Cluster sizes, cluster-specific mean model sizes, and within-cluster mean absolute deviations obtained when applying Dirichlet process clustering postprocessing to $\Gamma$ obtained to approximate the posterior distribution of model specifications using as model priors the Dirichlet process clustering, the beta-binomial, and the collinearity-adjusted dilution priors.

The interpretation of the clusters is essentially the same for the DPC and BB model priors. For example, covariate EAST, which is a dummy covariate, has a high PIP on the aggregate level (black line), as well as for the first two clusters, but loses importance in the third and forth cluster. While EAST loses importance in the third cluster, covariates like LAAM, SAFRICA, or CONFUC are gaining importance. This implies that the groups of covariates where either EAST or LAAM, SAFRICA, and CONFUC are contained constitute robust determinants of economic growth. Furthermore, inspecting Figure 1, we find a complementary relationship between primary school enrollment P60 and price of investment goods IPRI. For each cluster, both PIPs are either above or below the aggregate level. While Ley and Steel (2007) only found very weak evidence of jointness in the data at hand our results are perfectly in line with Doppelhofer and Weeks (2009).

The cluster profiles of the DIL model prior are harder to compare to the clustering solutions of the other two model priors because the solution contains five instead of four clusters. However, still some congruence between the solutions can be discerned. It seems that cluster 1 of the other two model priors is split into two clusters, cluster 1 and cluster 3, for the DIL model prior. This is also in line with the cluster sizes indicated in Table 9 where the sum of these two cluster sizes would be between the size of cluster 1 for the DPC model prior and the BB model prior. Cluster 2 has the same profile for all three model priors and also a similar size. Cluster 3 of the DPC and the BB model priors seems to re-emerge as cluster 5 in the DIL model prior results and the cluster size is again comparable. This leaves cluster 4 in all three model prior results to correspond to a similar group. The results of the DIL model prior also indicate that covariate EAST has high PIP for some clusters, in this case clusters 1–3, but has negligible PIP for other clusters, in this case clusters 4 and 5, whereas again covariates like LAAM, SAFRICA, or CONFUC gain importance for the latter two clusters.

The clustering solutions obtained are further characterized by the cluster sizes, the cluster-specific mean model sizes, and the within-cluster mean absolute deviations which are included in Table 9. The cluster sizes correspond to the fraction of model specifications in $\Gamma$ being assigned to each of the clusters by the DPC postprocessing procedure. For the DPC model prior cluster 1 is larger than for the BB prior. The reverse is true for cluster 2. The mean model sizes of the clusters indicate that the model size of cluster 1 is closer to the prior expected model size specified than cluster 2. These results are thus in line with the fact that the BB model prior distributes...
the prior mass over a larger range of model sizes. The compactness of the clusters are assessed using the within-cluster deviation from the cluster centroids where each cluster centroid is simply the mean over the covariate inclusion vectors assigned to the cluster. Similar compactness of clusters are observed for all considered model priors.

6 | DISCUSSION

BMA identifies robust determinants in regression applications where model uncertainty is faced due to the availability of a large set of potential covariates. Due to the lack of theory suggesting the relevant covariates data-driven methods are used to assess the relevance of the different covariates. In general interest is not only in the importance of single covariates independent of other covariates being included or not, but also in the joint inclusion or exclusion of covariates. These joint patterns indicate importance of theories (Durlauf et al., 2008) and determine if covariates act as complements or substitutes based on bivariate jointness measures (Doppelhofer & Weeks, 2009; Ley & Steel, 2007; Strachan, 2009) or on Dirichlet process clustering postprocessing (Crespo Cuaresma et al., 2016).

The DPC postprocessing approach proposed by Crespo Cuaresma et al. (2016) assumes that groups of covariates exist which interchangeably predict the dependent variable. These groups correspond to different theories, but neither the number of theories nor the covariates associated with each theory are a priori known. The approach proposed by Crespo Cuaresma et al. (2016) builds on the results of a standard BMA analysis and decomposes the dependency structure in the posterior distribution of covariate inclusion by splitting the inclusion patterns into clusters such that there is no dependency of covariate inclusion within clusters. Covariate importance measured by PIPs is then assessed on the cluster level.

Using a standard BMA analysis their approach does not impose a model prior which would correspond to the structure imposed a posteriori when clustering the covariate inclusion patterns. To obtain this congruence between prior and posterior structure imposed we propose to use a model prior in the BMA analysis which is based on Dirichlet process clustering. We show that the marginal distribution of this model prior is identical to that of the binomial model prior. This implies that collapsed sampling can be used to approximate the posterior distribution of covariate inclusion. This sampling scheme is identical to previously used sampling procedures. In a post-processing step the visited models are clustered detect latent class structures within the covariate inclusion patterns and characterize these latent groups through group-specific point estimates. This step is equivalent to the postprocessing analysis already proposed in Crespo Cuaresma et al. (2016).

This article provides the framework for a BMA analysis where the a priori belief of competing groups of covariates is explicitly included and taken into account. A DPC model prior is proposed where the clusters correspond to these competing groups of covariates. Interestingly this approach can be implemented re-using already available tools for sampling from the posterior of covariate inclusion and clustering the visited models in a post hoc analysis. The posterior model weights in BMA are obtained based on the marginal likelihoods which are available in closed form and thus the MCMC sampling only needs to employ a model prior which allows to suitably explore the relevant model space. The simulation study implies that for the regression parameter prior the hyper-g-prior outperforms the priors using fixed values for g, while for the model prior the hierarchical specification using a BB model prior is not dominating the DPC prior if groups of covariates are to be identified in a postprocessing step to take the jointness of covariates into
account when assessing the robustness of determinants. The empirical illustration indicates that similar clustering results are obtained regardless of the model prior imposed, even though slightly worse results are obtained for the DIL model prior. The simulation study suggests that this might be due to the data being informative of the clustering structure reducing the sensitivity on the model prior and leading to similar results for the DPC and the BB model priors.

In this article we only considered the MC³ algorithm to perform the BMA analysis. This algorithm, however, might be prone to poor mixing. In case of different groups of covariates being relevant, it is necessary for the sampler to visit the different modes corresponding to each of these groups in order for the proposed approach to be able to correctly identify the different groups in the postprocessing step. Alternative sampling approaches for BMA with a better mixing behavior, such as those proposed in Clyde et al. (2011) and Hubin and Storvik (2018), might thus be preferable to be used in case of groups of relevant covariates being suspected. The impact of their use should be investigated in future work.

ACKNOWLEDGMENT
The authors would like to thank the editor and the two anonymous reviewers for their constructive feedback and comments which helped improving the paper.

ORCID
Bettina Grün https://orcid.org/0000-0001-7265-4773
Paul Hofmarcher https://orcid.org/0000-0002-9206-6386

REFERENCES
Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. The Annals of Statistics, 2, 1152–1174.
Baran, S. (2014). Probabilistic wind speed forecasting using Bayesian model averaging with truncated normal components. Computational Statistics & Data Analysis, 75, 227–238.
Bayarri, M. J., Berger, J. O., Forte, A., & García-Donato, G. (2012). Criteria for Bayesian model choice with application to variable selection. The Annals of Statistics, 40, 1550–1577.
Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for sparse signals. Biometrika, 97, 465–480.
Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. Journal of Statistical Software, 61, 1–36.
Charrad, M., Ghazzali, N., Boiteau, V. and Niknafs, A. (2015) NbClust: Determining the best number of clusters in a data set. R package version 3.0. Retrieved from https://CRAN.R-project.org/package=NbClust
Chipman, H. (1996). Bayesian variable selection with related predictors. Canadian Journal of Statistics, 24, 17–36.
Clyde, M. A., Ghosh, J., & Littman, M. L. (2011). Bayesian adaptive sampling for variable selection and model averaging. Journal of Computational and Graphical Statistics, 20, 80–101.
Crespo Cuaresma, J., Grün, B., Hofmarcher, P., Humer, S., & Moser, M. (2016). Unveiling covariate inclusion structures in economic growth regressions using latent class analysis. European Economic Review, 81, 189–202.
Doppelhofer, G., & Weeks, M. (2009). Jointness of growth determinants. Journal of Applied Econometrics, 24, 209–244.
Durlauf, S., Kourteltas, A., & Tan, C. (2008). Are any growth theories robust? The Economic Journal, 118, 329–346.
Eicher, T. S., Papageorgiou, C., & Raftery, A. E. (2011). Default priors and predictive performance in Bayesian model averaging, with application to growth determinants. Journal of Applied Econometrics, 26, 30–55.
Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. The Annals of Statistics, 1, 209–230.
Fernández, C., Ley, E., & Steel, M. F. J. (2001a). Benchmark priors for Bayesian model averaging. Journal of Econometrics, 100, 381–427.
Fernández, C., Ley, E., & Steel, M. F. J. (2001b). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics, 16*, 563–576.

Frühwirth-Schnatter, S., & Malsiner-Walli, G. (2019). From here to infinity – Sparse finite versus Dirichlet process mixtures in model-based clustering. *Advances in Data Analysis and Classification, 13*, 33–64.

George, E. (1999). *Discussion of Bayesian model averaging and model search strategies by M. A. Clyde*. In *Bayesian statistics (Vol. 6, pp. 157–177)*. Oxford University Press.

George, E. I. (2010). Dilution priors: Compensating for model space redundancy. *IMS Collections Borrowing Strength: Theory Powering Applications – A Festschrift for Lawrence D. Brown, 6*, 158–165.

Green, P. J., & Richardson, S. (2001). Modelling heterogeneity with and without the Dirichlet process. *Scandinavian Journal of Statistics, 28*, 355–375.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science, 14*, 382–417.

Hofmarcher, P., Crespo Cuaresma, J., Grün, B., & Hornik, K. (2015). Last night a shrinkage saved my life: Economic growth, model uncertainty and correlated regressors. *Journal of Forecasting, 34*, 133–144.

Hofmarcher, P., Crespo Cuaresma, J., Grün, B., Humer, S., & Moser, M. (2018). Bivariate jointness measures in Bayesian model averaging: Solving the conundrum. *Journal of Macroeconomics, 57*, 150–165.

Hofmarcher, P., & Grün, B. (2020). Bayesian model averaging. In P. Fuleky (Ed.), *Macroeconomic forecasting in the era of big data. vol. 52 of Advanced Studies in Theoretical and Applied Econometrics* (pp. 359–388). Springer.

Hubin, A., & Storvik, G. (2018). Mode jumping MCMC for Bayesian variable selection in GLMM. *Computational Statistics & Data Analysis, 127*, 281–297.

Kalli, M., Griffin, J. E., & Walker, S. G. (2011). Slice sampling mixture models. *Statistics and Computing, 21*, 93–105.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data*. John Wiley & Sons.

Lazarsfeld, P. F. (1950). *The logical and mathematical foundations of latent structure analysis*. In S. A. Stouffer (Ed.), *Measurement and prediction* (pp. 362–412). John Wiley & Sons.

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

Ley, E., & Steel, M. F. J. (2009). On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics, 24*, 651–674.

Ley, E. and Steel, M. F. J. (2012) Mixtures of g-priors for Bayesian model averaging with economic applications. *Journal of Econometrics, 171*, 251–266.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association, 103*, 410–423.

Liverani, S., Hastie, D., Azizi, L., Papathomas, M., & Richardson, S. (2015). PReMiUM: An R package for profile regression mixture models using Dirichlet processes. *Journal of Statistical Software, 64*, 1–30.

Madigan, D., & Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam’s window. *Journal of the American Statistical Association, 89*, 1535–1546.

Madigan, D., York, J., & Allard, D. (1995). Bayesian graphical models for discrete data. *International Statistical Review, 63*, 215–232.

Malsiner-Walli, G., Frühwirth-Schnatter, S., & Grün, B. (2016). Model-based clustering based on sparse finite Gaussian mixtures. *Statistics and Computing, 26*, 303–324.

Malsiner-Walli, G., Hofmarcher, P., & Grün, B. (2019). Semi-parametric regression under model uncertainty: Economic applications. *Oxford Bulletin of Economics and Statistics, 81*, 1117–1143.

Man, G. (2018). Critical appraisal of jointness concepts in Bayesian model averaging: Evidence from life sciences, sociology, and other scientific fields. *Journal of Applied Statistics, 45*, 845–867.

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

McAuliffe, J. D., Blei, D. M., & Jordan, M. I. (2006). Nonparametric empirical Bayes for the Dirichlet process mixture model. *Statistics and Computing, 16*, 5–14.

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

Moral-Benito, E. (2015). Model averaging in economics: An overview. *Journal of Economic Surveys, 29*, 46–75.
Moser, M., & Hofmarcher, P. (2014). Model priors revisited: Interaction terms in BMA growth applications. *Journal of Applied Econometrics*, 29, 344–347.

Ouysse, R., & Kohn, R. (2010). Bayesian variable selection and model averaging in the arbitrage pricing theory model. *Computational Statistics & Data Analysis*, 54, 3249–3268.

Piironen, J., & Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, 11, 5018–5051.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92, 179–191.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.

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

Sethurman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4, 639–650.

Sloughter, J. M., Gneiting, T., & Raftery, A. E. (2010). Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *Journal of the American Statistical Association*, 105, 25–35.

Strachan, R. W. (2009). Comment on ‘Jointness of growth determinants’ by Gernot Doppelhofer and Melvyn Weeks. *Journal of Applied Econometrics*, 24, 245–247.

Zellner, A. (1986). *On assessing prior distributions and Bayesian regression analysis with g-prior distributions*. In P. Goel & A. Zellner (Eds.), *Bayesian inference and decision techniques: Essays in honor of Bruno De Finetti*. North-Holland.

Zeugner, S., & Feldkircher, M. (2015). Bayesian model averaging employing fixed and flexible priors: The BMS package for R. *Journal of Statistical Software*, 68, 1–37.

Zhou, X. K., Liu, F., & Dannenberg, A. J. (2012). A Bayesian model averaging approach for observational gene expression studies. *The Annals of Applied Statistics*, 6, 497–520.

**How to cite this article:** Grün, B., & Hofmarcher, P. (2021). Identifying groups of determinants in Bayesian model averaging using Dirichlet process clustering. *Scandinavian Journal of Statistics*, 48(3), 1018–1045. https://doi.org/10.1111/sjos.12541