A joint initiative of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute for Economic Research

CESifo Area Conference

on

Macro, Money and International Finance

14-15 March 2008

CESifo Conference Centre, Munich

Robust Model Averaging

Gernot Doppelhofer and Melvyn Weeks

CESifo GmbH
Poschingerstr. 5
81679 Munich
Germany

Phone: +49 (0) 89 9224-1410
Fax: +49 (0) 89 9224-1409
E-mail: office@cesifo.de
Web: www.cesifo.de

Robust Model Averaging*

Gernot Doppelhofer[†]
NHH, CESifo and CIMF

Melvyn Weeks[‡]
University of Cambridge

March 2008

Abstract

Bayesian Model Averaging (BMA) is used in many applications characterized by a large number of models and relatively few observations. Benchmark BMA uses linear regression models with independent normal sampling and homoscedastic errors. Although a number of studies have examined robustness of BMA in the context of prior model size and the specification of Zellner's g-prior, there has been a relative neglect of the impact of parameter heterogeneity and outliers. We evaluate the robustness of the normal benchmark BMA in the context of cross-country growth regressions and find that inference is significantly affected by considering deviations from the benchmark BMA.

Keywords: Model Uncertainty, Robustness, Heteroscedasticity, Outliers, Determinants of Economic Growth

JEL Classifications: C11, C21, C52, O20, O50

---

*We thank Jesus Crespo-Cuaresma, Hashem Pesaran, Xavier Sala-i-Martin, Arnold Zellner and seminar participants in Cambridge, Innsbruck, NHH, Norges Bank, Rome and IHS Vienna for helpful comments. Doppelhofer thanks the Institute for Advanced Studies (IHS) Vienna for their hospitality. All errors are our own.

[†]Department of Economics, Norwegian School of Economics and Business (NHH), Helleveien 30, N-5045 Bergen; CIMF and CESifo. Email: Gernot.Doppelhofer@nhh.no, Tel: +47 55959360, Fax: +47 55959543.

[‡]Faculty of Economics and Clare College, University of Cambridge, Cambridge CB3 9DD, UK. Email: mw217@econ.cam.ac.uk.

1 Introduction

The construction and evaluation of economic models is an extremely difficult task given that the analyst must confront and deal with many types of uncertainty. Fisher's likelihood theory, the workhorse of empirical analysis for over 50 years, explicitly incorporates both *parameter* and *stochastic* uncertainty. However the whole apparatus of both point and interval estimation rests upon the assumption that the model structure is known with the objective of finding the best estimates of an unknown parameter vector.

Recognition of the dangers of basing inference on a single model has spurned a substantial literature. Jeffreys (1961) laid the formulation for model averaging, with a significant contribution by Leamer (1978), resulting in the development of extreme bounds analysis (EBA). The principle idea behind EBA is that inference on one or more parameters of interest may be sensitive to the set of additional controls, $\mathbf{x}$. Whereas standard classical confidence intervals account for parameter uncertainty engendered by sample variation, the upper and lower bounds computed using EBA reflect additional uncertainty as a result of estimating a number of different regressions over different subsets of $\mathbf{x}$.

Following significant developments in computing power and the related developments in Monte Carlo Markov Chain (MCMC) sampling, a Bayesian approach to inference allows a relatively seamless integration of model uncertainty and parameter uncertainty. Model averaging addresses issues of model uncertainty explicitly, and there is a recent and growing literature on this topic. Recent economic applications of model averaging include policy evaluation (Brock, Durlauf and West 2003)), monetary policy (Levin and Williams (2003)), macroeconomic forecasting (Garratt, Lee, Pesaran and Shin (2003)) and economic growth (Sala-i-Martin, Doppelhofer and Miller (2004), henceforth SDM (2004)).[1] SDM (2004) conduct *Bayesian Averaging of Classical Estimates* or *BACE* to estimate the importance of explanatory variables. This BACE approach estimates the posterior distribution of parameters of interest and introduces a minimum of prior information by using Classical ordinary least squares (OLS) estimates.

The principle advantage of a Bayesian approach to inference when the model space $\mathcal{M}$ is large, is that by integrating out model uncertainty the marginal posterior distribution for model parameters will result in parameter estimates that are more robust to the effects of misspecification than procedures which place all support on a single model. Although there is a growing literature on the use of BMA to applications characterised by a large number of candidate variables and a relatively small number of observations,[2] there has been relatively little empirical analysis on how the posterior objects of interest are affected by other features of the data. Here

[1]Hoeting, Madigan, Raftery and Volinsky (1999) surveys the literature on Bayesian model averaging (BMA). Burnham and Anderson (2002) and Hjort and Claeskens (2003) discuss frequentist model averaging.

[2]See, for example, Fernandez, Ley and Steel (2001a, 2001b), Stone and Weeks (2001), and SDM (2004).

we note that in the case of the linear regression model with normal, homoscedastic errors, model uncertainty is equivalent to the problem of variable selection, and therefore other misspecification issues are ignored.

In a review of the application of econometric techniques to empirical growth models, Temple (2000) highlights three key problems that are likely to affect the validity of inference: model uncertainty, parameter heterogeneity and outliers. In this context one of the principle objections to BMA is that much of the advances in developing robust inference when faced with a large space of potential models, has been conditioned on an overly restrictive linear regression model with independent normal sampling and homoscedastic errors.

Our point of departure is what we refer to as the *benchmark BMA* which assumes that the residuals in the linear regression model are normally distributed and conditionally homoscedastic. By weighting parameter estimates over a class of models, BMA propagates model uncertainty into the posterior distribution of parameters. However, given the assumption of normal errors, there is limited robustness of model weights to outliers and parameter heterogeneity. At a general level outliers can be modeled by a mean shift, for example the introduction of dummy variables (see Hendry), or by a shift in the variance. The mean-shift model can be used to identify outliers for further study. In contrast the variance-inflation model approach is used when the object is to accommodate outliers, and thereby make parameters of interest robust to aberrant observations.

In this paper we incorporate robustness with respect to outliers and unequal variances into model averaging by allowing for independently distributed, but heteroscedastic errors $\varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{\Omega})$. We maintain the assumption that the covariance matrix is diagonal, $\mathbf{\Omega} \equiv diag(w_1, ..., w_T)$ with independent variances w_t, $t = 1, ..., T$. Geweke (1993) proposes a prior Chi-squared distribution with degrees of freedom v for $(w_1, ..., w_T)$. Intuitively, the degrees of freedom v determine the fatness of the tails of the Student-t distribution and the prior weight on outliers. Lower values of v imply a more skewed distribution with a higher probability of outliers and relatively larger variances. High values of v on the other hand imply errors drawn from a distribution close to the homoscedastic normal benchmark case.

As Geweke (1993) demonstrates the normal mixture model with independence prior is equivalent to a model with independent Student-t errors with v degrees of freedom. Fernandez and Steel (2000) also examine Bayesian inference based upon the linear regression model, focussing on the theoretical basis of independent sampling from a scale mixtures of normals. The authors also refer to a well developed literature in applied statistics containing examples of applications including Harvey, Ruiz and Shepherd (1994) and Bauwens and Lubrano (1998).

Despite the emergence of this literature it would appear that much of the advances in robust methods in Bayesian inference have been confined to single models. A notable exception is the work Hoeting, Raftery, and Madigan (1996), who develop an approach that simultaneously accounts for model uncertainty and outlier identification. A critical observation here and the principal motivation behind the study is that it is generally not possible to construct model averaging and outlier analysis separately. The authors develop a technique known as *simultaneous Bayesian variable selection and outlier identification* (SVO) which circumvents the problem that the order of methods determines the choice of outliers and variables. Recently, Gottardo and Raftery (2007) have proposed a unified approach to Bayesian robust variable and transformation selection that also uses t-distributions to robustify against outlying observations.

Another example of an approach that incorporates a number of dimensions of robustness within a model averaging framework, is a recent study by Le Sage and Parent (2006). The authors extends the BMA approach to a class of models, popular is spatial econometrics, where dependence across geographical observations calls for an explicit treatment of error dependence. The presence of such dependence can result in bias and inconsistent least-square estimates, and therefore precludes the use of vanilla BMA. Extending earlier contributions by Hepple (1995a, 1995b), the authors present a useful approach to model averaging over a large class of models defined with respect to both uncertainty over the set of regressors and how to best represent spatial dependence.

In this paper we evaluate the robustness of benchmark BMA in the context of cross-country growth regressions. We focus on this area for two reasons. First, there is a long lineage of applied work which has examined the robustness of inference on the determinants of economic growth to changes in the set of conditioning variables. Levine and Renelt (1992) carried out an EBA and concluded that many results were sensitive to the particular subset of regressors. Bayesian extension of EBA using BMA have also been constructed in the area of economic growth including Fernandez, Ley and Steel (2001a) and SDM (2004). Although a number of the aforementioned studies have examined robustness of BMA in the context of prior model size and the specification of Zellner's g-prior, there has been a relative neglect of the impact of parameter heterogeneity and outliers. These two issues are likely to be important when conducting inference in cross-country regressions.

The remainder of the paper is organized as follows: Section 2 describes the statistical method of model averaging, and Section 3 introduces robust model averaging. Section 4 applies robust model averaging to determinants of economic growth and section 5 concludes.

2 Benchmark Model Averaging

Consider the following general linear regression model

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \tag{1}$$

where $\mathbf{y}$ is a $(T \times 1)$ vector, $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_k]$ is a $(T \times k)$ matrix of explanatory variables (including an intercept α), β is a $(k \times 1)$ vector of unknown parameters and ε is a $(T \times 1)$ vector of residuals which are assumed to be normally distributed, $\varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, and *conditionally* homoscedastic.

Consider the problem of making inference on the determinants of the dependent variable, given data $\mathbf{D} \equiv [\mathbf{y}, \mathbf{X}]$. Suppose that there are K potential regressors, then the model space $\mathcal{M}$ is the set of all 2^K linear models. Each model M_j is described by a $(k \times 1)$ binary vector $\gamma = (\gamma_1, ..., \gamma_K)'$, where a one (zero) indicates the inclusion (exclusion) of a variable $\mathbf{x}_i$ in regression (1). Let $\mathbf{X}_j$ be the set of regressors included in model M_j. For a given model M_j, the unknown parameter vector β represents the effects of the variables included in the regression model. We can estimate its density $p(\beta|M_j)$ conditional on model M_j.

Given the (potentially large) space of models $\mathcal{M}$ the *unconditional* effects of model parameters can be derived by integrating out all aspects of model uncertainty, including the space of models $\mathcal{M}$. A maintained assumption throughout is that the explanatory variables $\mathbf{X}$ are predetermined (weakly exogenous) and independent of parameters β and σ. The posterior density $p(\beta|\mathbf{y})$ can then be expressed as function of sample observations of $\mathbf{y}$.

The unconditional posterior distribution of the slope coefficient β is given by

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

where $p(\beta|M_j, \mathbf{y})$ is the conditional distribution of β given model M_j. The posterior model probability $p(M_j|\mathbf{y})$ propagates model uncertainty into the posterior distribution of model parameters. By Bayes' rule $p(M_j|\mathbf{y})$ can be written as

$$\begin{aligned} p(M_j|\mathbf{y}) &= \frac{l(\mathbf{y}|M_j) \cdot p(M_j)}{p(\mathbf{y})} \\ &\propto l(\mathbf{y}|M_j) \cdot p(M_j) \end{aligned} \tag{3}$$

such that the posterior model probability (weight) is proportional to the product of the model-specific marginal likelihood $l(\mathbf{y}|M_j)$ and the prior model probability $p(M_j)$. The model weights are converted into probabilities by normalizing relative to the set of all 2^K models:

$$p(M_j|\mathbf{y}) = \frac{l(\mathbf{y}|M_j) \cdot p(M_j)}{\sum\limits_{i=1}^{2^K} l(\mathbf{y}|M_i) \cdot p(M_i)} \tag{4}$$

We follow the Bayesian model averaging literature by assuming the following prior structure for parameters in each model. The prior slope coefficients β are normally distributed with mean zero and variance $\sigma^2 \mathbf{V}_{0j}$:

$$p(\beta|\sigma^2, M_j) \sim N(\mathbf{0}, \sigma^2 \mathbf{V}_{0j}) \tag{5}$$

The prior variance matrix $\mathbf{V}_{0j}$ is assumed to be proportional to the sample covariance

$$\mathbf{V}_0 = (g_0 \mathbf{X}_j' \mathbf{X}_j)^{-1} \tag{6}$$

with factor of proportionality g_0. This *g-prior* was first suggested by Zellner (1986), and is a convenient way to specify the prior variance matrix, in particular in the presence of considerable model uncertainty. Different values of the g-prior parameter g_0 have been proposed in the literature. Fernandez, Ley and Steel (2001b) conduct an extensive Monte Carlo study and propose to set $g_0 = \min(1/T, 1/K^2)$. This implies that the prior distribution of the slope parameter β becomes more diffuse if the number of candidate variables K is sufficiently large, so that $K^2 > T$. Alternatively, we assume in the benchmark case that the prior distribution is dominated by the sample information and the prior variance is sufficiently diffuse which implies that g_0 approaches zero in (6)[3].

For the two parameters that are common across all model, the prior error variance σ^2 and the intercept α, we assume non-informative (improper) priors that impose a minimum of prior information:

$$p(\sigma^2) \propto \frac{1}{\sigma^2} \tag{7}$$

$$p(\alpha) = 1 \tag{8}$$

Alternatively, one could assume a proper, inverse-Gamma prior distribution for the error variance σ^2 which is the natural conjugate prior for the normal regression model (see for example, Koop (2003).

The prior probability for model M_j is

$$p(M_j) = \prod_{i=1}^{K} \pi_i^{\gamma_i} (1 - \pi_i)^{1-\gamma_i} \tag{9}$$

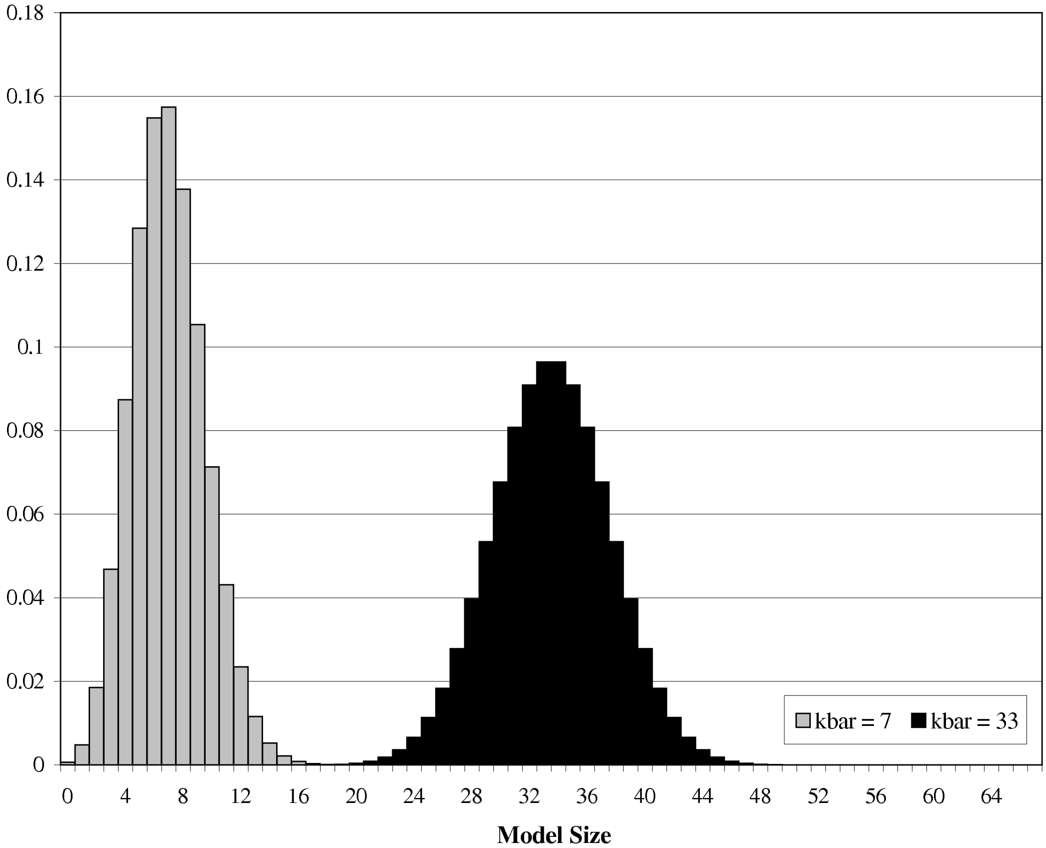
where π_i is the (independent) prior inclusion probability of variable $\mathbf{x}_i$ in M_j, with corresponding indicator γ_i. The assumption that each variable has an equal prior probability of inclusion corresponds to the adoption of uniform priors and setting the hyper-parameter $\pi_i = 1/2$ for all i (see George and McCulloch, 1993). However, with a relatively large number of regressors K, a uniform prior implies that the great majority of prior probability is allocated to models with a large number of variables. As an alternative, we follow SDM (2004) and introduce a prior expected model size $\bar{k}$ and corresponding prior inclusion probability $\pi_i^{BACE} = \bar{k}/K$. Figure 1 shows the prior distribution over model sizes for the benchmark case with $\bar{k} = 7$ (implied prior inclusion probability $\pi_i = 7/67 = 0.104$) and for the uniform prior case with $\pi_i = 1/2$ (and implied prior model size $\bar{k} = 33.5$).[4]

The assumed prior structure introduces a minimum of prior information into the estimation. In the limit, when the sample information dominates the prior

[3]For a large number of candidate variables K, the value hyperparameter $g_0 = 1/K^2$ recommended by Fernandez, Ley and Steel (2001b) approaches our benchmark case of sample-dominated prior information. We investigate this issue further in section 3.3.

[4]A disadvantage of this approach is that the notion of what constitutes a reasonable prior model size may vary across analysts. In response to this criticism Godsill, Stone and Weeks (2004) introduce another layer of prior information by combining independent Bernoulli sampling for each variable with a conjugate Beta prior for the binomial proportion parameter π_i.

Figure 1: Prior Probabilities by Model Size: Benchmark Case with Prior Model Size $\bar{k} = 7$ and Uniform Prior with $\bar{k} = 33$.



information, Leamer (1978) shows that the marginal likelihood of model M_j may be written as

$$l(\mathbf{y}|M_j) \propto T^{-k_j/2} \cdot SSE_j^{-T/2} \tag{10}$$

where k_j is the number of regressors and $SSE_j = (\mathbf{y} - \mathbf{X}_j\beta)'(\mathbf{y} - \mathbf{X}_j\beta)$ is the sum of squared errors in model M_j. The posterior model probability of model M_j is obtained by pre-multiplying (31) by the prior model probability $p(M_j)$ and dividing by the sum over all 2^K possible models:

$$p(M_j|\mathbf{y}) = \frac{p(M_j) \cdot T^{-k_j/2} \cdot SSE_j^{-T/2}}{\sum_{r=1}^{2^K} p(M_r) \cdot T^{-k_r/2} \cdot SSE_r^{-T/2}} \tag{11}$$

The posterior model weights (11) equal the prior model weights times the (exponentiated) Bayesian Information Criterion (BIC) developed by Schwarz (1978). The BIC weights depend on the likelihood, but penalizes relatively large models through the penalty term $T^{-k_j/2}$. The implied preference for smaller models addresses to a certain extent collinearity among regressors. Explanatory variables that are very similar explain relatively less of the variation of the dependent variable which implies less weight on such models.

BIC model weights (11) have been extensively discussed in the literature. Alternative derivations include the so-called “unit information prior” discussed in Kass and Wassermann (1995), approximation to Bayes Factors by Kass and Raftery (1995) and Raftery (1995), benchmark priors by Fernandez, Ley and Steel (2001b), or the limiting case of a non-informative Jeffreys prior for the error variance with a particular choice of normalizing constant (see for example Wasserman, 2000). Klein and Brown (1984) show that by minimizing the so-called Shannon information in the prior distribution, the BIC model weights (11) can be used in small samples. We adopt the BIC posterior model weights since they provide a reasonable approximation to proper Bayesian model weights and are consistent in large samples.

2.1 Estimation of Posterior Objects

The unconditional mean and variance of slope parameters β can be calculated in a straightforward manner from conditional (model specific) parameter estimates (see Leamer (1978), p. 118). The mean of the posterior distribution of slope parameter β_i associated with variable $\mathbf{x}_i$, unconditional with respect to space of models $\mathcal{M}$, is given by

$$E(\beta_i|\mathbf{y}) = \sum_{j=1}^{2^K} p(M_j|\mathbf{y}) \cdot \widehat{\beta}_{ij} \tag{12}$$

where $\hat{\beta}_{ij} = E(\beta_i|\mathbf{y}, M_j)$ is the OLS estimate for slope parameter β_i given model M_j. The posterior variance of slope β_i is given by

$$V(\beta_i|\mathbf{y}) = \sum_{j=1}^{2^K} p(M_j|\mathbf{y}) \cdot V(\beta_i|\mathbf{y}, M_j) + \sum_{j=1}^{2^K} p(M_j|\mathbf{y}) \cdot \left[\widehat{\beta}_{ij} - E(\beta_i|\mathbf{y})\right]^2 \tag{13}$$

where the conditional variance is estimated by the maximum likelihood[5] estimator $V(\beta_i|M_j) = \hat{\sigma}_j^2(\mathbf{X}_j'\mathbf{X}_j)^{-1}_{ii}$, with error variance estimate $\hat{\sigma}_j^2 \equiv SSE_j/(T-k_j)$. Notice that the posterior variance (13) of coefficient β_i consists of two terms: the weighted sum of conditional (model-specific) variances and an additional term taking into account the difference between conditional and posterior estimates of mean coefficients.

We also contrast the Bayesian approach with a simple Frequentist approach that estimates a heteroscedasticity-consistent covariance matrix, developed by White (1980). For each model , White showed that the covariance matrix for the slope coefficient β can be consistently estimated by

$$V(\beta_i|\mathbf{y}, M_j) = (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{\Omega}_j\mathbf{X}_j(\mathbf{X}_j'\mathbf{X}_j)^{-1} \tag{14}$$

This estimator is unbiased asymptotically, but has been shown to be biased in small samples. We therefore use the simple finite-sample correction of scaling the covariance estimator $\hat{\Omega}_j = SSE_j \cdot T/(T-k_j)$.[6]

[5]Conditional on each model M_j, the maximum likelihood estimator is optimal, given that we consider a prior structure that is dominated by sample information. Alternatively, the variance estimator would result from diffuse priors starting with a Student prior for β (see Leamer (1978), p.79-80). We check the sensitivity of our estimates to using informative priors in section 3.2.

[6]See Davidson and MacKinnon (1993, pp. 552-6) for a textbook discussion.

A policymaker might be interested to know how important variables are in explaining the dependent variable. The *posterior inclusion probability* of variable $\mathbf{x}_i$

$$p(i|\mathbf{y}) = \sum_{j=1}^{2^K} \mathbf{1}(\gamma_i = 1|\mathbf{y}, M_j) \cdot p(M_j|\mathbf{y}) \tag{15}$$

represents the probability that, conditional on the data, but unconditional with respect to the model space $\mathcal{M}$, variable $\mathbf{x}_i$ is relevant in explaining the dependent variable (see Leamer (1978) and Mitchell and Beauchamp (1988)). This measure is therefore a model-weighted measure of relative importance of including a variable in the regression.

Alternatively, we might be interested in estimates of the mean and variance of the slope coefficients *conditional* on a variable's inclusion, but unconditional with respect to M. The conditional posterior mean for β_i is obtained by dividing the unconditional posterior mean (12) by the posterior inclusion probability (15):

$$E(\beta_i|\gamma_i = 1, \mathbf{y}) = \frac{E(\beta_i|\mathbf{y})}{p(i|\mathbf{y})} \tag{16}$$

Similarly, the variance conditional on including variable $\mathbf{x}_i$ is calculated from the unconditional posterior estimates of moments (12), (13), and the posterior inclusion probability (15):

$$V(\beta_i|\gamma_i = 1, \mathbf{y}) = \frac{V(\beta_i|\mathbf{y}) + [E(\beta_i|\mathbf{y})]^2}{p(i|\mathbf{y})} - [E(\beta_i|\gamma_i = 1, \mathbf{y})]^2 \tag{17}$$

To facilitate comparison across variables, we calculate the *posterior standardized coefficient* associated with variable $\mathbf{x}_i$ conditional on its inclusion,

$$E\left(\beta_i/s_i|\gamma_i = 1, \mathbf{y}\right) \equiv \frac{E(\beta_i|\gamma_i = 1, \mathbf{y})}{\sqrt{V(\beta_i|\gamma_i = 1, \mathbf{y})}} \tag{18}$$

calculated by dividing the posterior mean coefficient (12) by its posterior standard deviation s_i the square root of the posterior variance (13). We use the posterior standardized coefficient to assess the relative size and significance of the effect of variable $\mathbf{x}_i$ in explaining the dependent variable $\mathbf{y}$.[7]

All the posterior statistics presented in this section – inclusion probability (15) and standardized coefficient estimate (18) – are estimated unconditionally with respect to the model space and thereby taking model uncertainty into account. However, the benchmark model averaging framework does not take into account other forms of specification uncertainty, such as the sensitivity of model weights and inference to outliers and the assumed prior distribution. This will be addressed in the following section.

[7] Brock and Durlauf (2001) provide a decision-theoretic foundation to using such standardized coefficients for policy analysis.

3 Robust Model Averaging

This section introduces robustness into the model averaging framework. We consider a number of different approaches to robust model averaging.

Perhaps the most general distinction in terms of approaches is the choice of whether outliers are modelled by a mean shift, for example the introduction of dummy variables, or via a shift in the variance. This paper focusses on variance-inflation since we would like to accommodate outlying observations, but have little guidance regarding the most appropriate likelihood function. Furthermore, we wish to average across many different specifications and a flexible distribution can be obtained by mixing together a number of different distributions. In what follows we consider scale mixtures of normal distributions.

3.1 Scale Mixtures of Normals

To examine a number of alternate approaches consider first a canonical form given by

$$p(y_i|\beta, \sigma, M_j) = \int_0^\infty \frac{\omega_t^{1/2}}{(2\pi)^{1/2}\sigma} exp\{-\frac{\omega_t}{2\sigma^2}(y_t - \mathbf{x}_t'\beta)^2\} g(\omega_t|\tau) d\tau \tag{19}$$

which combined a normal kernel with the mixing distribution $g(\omega_t|\tau)$, where τ may be vector valued. Note that specification of the mixing distribution is equivalent to choosing a prior specification on the error variances ω_t. Conditional on ω_t, the distribution is normal (see Fernandez and Steel (2000)). In the following discussion we first consider the question of robustness to the choice of prior specification of the error distribution conditional on a given model M_j.[8]

The mixing distribution $g(\omega_t|\tau)$ may be chosen on a number of grounds including conjugacy or the dimension of the hyperparameter τ describing the mixing distribution. For example, the mixture prior model might use prior information in the form of two hyperparameters with $\tau = (\pi, \rho)$. π is used to identify a subset of observations as potential outliers, and conditional on this subset, the parameter ρ, controls the degree of variance-inflation parameter. Assuming the kernel is normal we have

$$p(y_i|\beta, \sigma^2, \pi, \rho, M_j) = (1 - \pi)N(y_i|\beta, \sigma^2) + (1 - \pi)N(y_i|\beta, \rho^2\sigma^2) \tag{20}$$

If the mixing distribution $g(\omega_t|\tau)$ is $Gamma(a, b)$ or $(G(a, b))$, the conjugacy of Normal-Gamma is such that the full conditional of each of the ω_t is also $Gamma$ distributed, making Gibbs sampling relatively simple. Particular values of the shape and scale parameters, a and b respectively generate a class of Gamma mixing distribution. Given that $G(a = v/2, b = 2)$ is equivalent $\chi^2(v)$, we have have a mixing distribution controlled by a single degrees of freedom parameter v. Moreover,

[8] This can be contrasted with Lange et al (1989) when estimating a model by maximum likelihood.

Geweke (1993) shows that the normal mixture model with independence (elaborate) prior for error variances

$$v/w_t \sim \chi^2(v), \quad t = 1, ..., T \tag{21}$$

is equivalent to a model with independent Student-t errors with v degrees of freedom. Lower values of v imply a more skewed distribution with a higher probability of outliers and relatively larger variances. High values of v on the other hand imply errors drawn from a distribution close to the homoscedastic normal benchmark case described in section 2. Intuitively, the degrees of freedom v determine the fatness of the tails of the Student-t distribution and the prior weight on outliers.

3.2 Gamma Mixing

To robustify the estimation against outliers and unequal variances, we allow for independently distributed, but heteroscedastic errors, $\varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{\Omega})$. The covariance matrix is assumed diagonal, $\mathbf{\Omega} \equiv diag(w_1, ..., w_T)$ with independent variances w_t, $t = 1, ..., T$. The parameters of interest conditional on each model M_j are estimated by generalized least squares (GLS) which downweights observations by their variances.

The degrees of freedom v therefore play an important role in robustifying models against outliers. In the limiting normal case, the degrees of freedom are $v \to \infty$ and the regression errors become conditionally homoscedastic with covariance matrix equal the identity matrix, $\mathbf{\Omega} = \mathbf{I}$. For the degrees of freedom parameter v we consider two alternatives. Either we assume the degrees of freedom are drawn form a distribution with fixed hyper-parameters or alternatively we estimate the parameters characterizing the distribution of the degrees of freedom.

1. We can either draw the degrees of freedom v from a *Gamma* prior distribution:

$$v \sim G(a, b) \tag{22}$$

   where a and b are hyper-parameters. The Gamma distribution implies that the mean degrees of freedom equal $E(v) = ab$ and the variance $Var(v) = ab^2$. In the application shown in section 4, we consider the following values:

a	1000	500	100	20
b	1/10	1/10	1/4	1/2
$E(v)$	100	50	25	10
$Var(v)$	10	10	6	5

2. Alternatively, we can assume an analytically convenient prior distribution for the degrees of freedom. We can the fact that the exponential distribution is equivalent to a *Gamma* distribution with $b = 2$ and prior mean v_0:

$$p(v) \sim \exp(v|v_0) \tag{23}$$

The posterior distribution of the degrees of freedom v is then given by $p(v|w_t, M_j) \sim p(w_t|v, M_j) \cdot p(v)$.

3.2.1 Estimation of Posterior Objects

These prior assumptions imply the following convenient hierarchical structure:[9]

1. The posterior distribution of the slope parameters β in model M_j conditional on other parameters is given by

$$p(\beta|\sigma^2, \mathbf{\Omega}, M_j) \sim N(\tilde{\beta}_j, \sigma^2 \tilde{\mathbf{V}}_j) \tag{24}$$

The conditional mean of β is estimated using the generalized least squares (GLS) estimator

$$\tilde{\beta}_j = \tilde{\mathbf{V}}_j \left(\mathbf{X}_j' \mathbf{\Omega}^{-1} \mathbf{y} \right) \tag{25}$$

The posterior variance of β is given by $\sigma^2 \tilde{\mathbf{V}}_j$, where

$$\tilde{\mathbf{V}}_j = \left(\mathbf{X}_j' \mathbf{\Omega}^{-1} \mathbf{X}_j + \sigma^2 \mathbf{V}_{0j}^{-1} \right)^{-1} \tag{26}$$

2. The posterior distribution of the error variance parameter σ^2 conditional on the other parameters is given by

$$\left[\sum_{t=1}^{T} \left(e_{j,t}^2 / w_t \right) \Big/ \sigma^2 \right] |(\beta, \mathbf{\Omega}) \sim \chi^2(T) \tag{27}$$

where $e_{j,t} = y_t - \mathbf{x}_{j,t}' \beta_j$. Notice that the degrees of freedom equal T and not $T - k_j$, since we condition on β_j.

3. The posterior distribution of the elements of the error variance matrix $\mathbf{\Omega}$ conditional on the other parameters is proportional to

$$\left[\left(\sigma^{-2} e_{j,t}^2 + v \right) / \, w_t \right] |(\beta_j, \sigma^2) \sim \chi^2(v+1), \quad t = 1, ..., T \tag{28}$$

The $v + 1$ degrees of freedom follow from combining the prior distribution of w_t (21) with v degrees of freedom with terms $\sigma^{-2} e_{j,t}^2 / w_t$ from the likelihood function with a $\chi^2(1)$ kernel.

4a. Either we draw the degrees of freedom from the prior distribution (22) with fixed hyperparameters a, b.

4b. Alternatively, we draw degrees of freedom v from its posterior distribution:

$$p(v|w_t, M_j) \sim p(w_t|v, M_j) \cdot p(v) \tag{29}$$
$$\propto \left(\frac{v}{2} \right)^{Tv/2} \cdot \Gamma(v/2)^{-T} \cdot \exp(-\eta v) \tag{30}$$

[9]See Geweke (1993) for further details. Koop (2003) contains an excellent discussion of the case of t-distributed errors used here and of other error distributions resulting from mixed normal distributions.

where $\eta \equiv \sum_{t=1}^{T} [\ln w_t + 1/w_t]/2 + 1/v_0$. This distribution has no convenient analytic from, and we therefore introduce a Metropolis step to draw candidate values of v' from (30) and accept them with probability p_α.[10]

The hierarchical structure of steps 1-4 leads naturally to estimation by the Gibbs sampler. Starting from initial values, the Gibbs sampler estimates parameters iteratively by drawing from the conditional posterior distributions of parameters, given by (24)-(26) for mean and variance of the slope coefficient β, (27) and (28) for the error variance $\sigma^2\mathbf{\Omega}$, and (22) or (23) for the degrees of freedom v. Details of the Gibbs sampling procedure, including numerical convergence criteria, are discussed in the Computational Appendix A.

3.3 Sensitivity Analysis

We examine the robustness of the benchmark case with respect to the following deviations from the benchmark prior structure. First, we compare the results and inference under the benchmark case with sample-dominated priors and proper Bayesian priors with "nearly" normally distributed priors[11]. Conditional on $\mathbf{\Omega}$ and the proper normal-inverse Gamma prior, the marginal likelihood of model M_j is proportional to

$$l(\mathbf{y}|M_j) \propto \left(\frac{|\bar{\mathbf{V}}_j|}{|\mathbf{V}_{oj}|}\right)^{1/2} (T \cdot \sigma^2)^{-T/2} \tag{31}$$

where $\mathbf{V}_j$ is the posterior variance of the slope coefficient β_j and σ^2 is the common term of the posterior error variance. When using proper prior distributions, the marginal likelihood is the posterior model weights, corresponding to 11 in the benchmark case.

Second, we also contrast the estimates of the posterior inclusion probability (4) and posterior standardized coefficient (18) under the (normal) benchmark case with models that allow for heteroscedastic errors in the regression models. The errors are drawn from t-distributions with different degrees of freedom $v = 10, 25, 50$ and 100. Lower values of v imply fatter tails of the t-distribution allowing for more extreme outlying observations of the regression errors.

Third, we also contrast the standard errors estimated under the normal benchmark case with the heteroscedasticity-consistent estimates (14), proposed by White (1980).

4 Robustness of Growth Determinants

This section presents results of applying robust model averaging to the determinants of economic growth dataset of SDM (2004). In particular, we investigate if statistical

[10] For an introduction to the Metropolis-Hastings algorithm see Chib (2001) or Koop (2003). For details of our sampler see Computational Appendix.

[11] The values of hyper-parameters $a = 1000$ and $b = 1/10$ are chosen to imply $E(v) = 100$ and $Var(v) = 10$.

inference and economic importance of the effects associated with these explanatory variables is robust with respect to outliers and heteroscedasticity.

The explanatory variables are chosen from regressors found to be related to economic growth in earlier studies (see for example the list in Durlauf and Quah, 1999). SDM (2004) select variables that represent "state variables" in economic growth models and measure them as close as possible to the start of the sample period in 1960. Furthermore, the dataset is restricted to be balanced, i.e. without missing observations. Under these criteria the total number of explanatory variables is $K = 67$ and with observations for $T = 88$ countries.[12] The results are based on approximately 60 million randomly drawn regressions (see Computation Appendix A for details).

The dependent variable, *average growth rate of GDP per capita between 1960-96*, and the 67 explanatory variables are listed in the Data Appendix B. Also shown are short variable names, a brief variable description, and sample mean and standard deviations. In the Data Appendix B and tables of results, explanatory variables are ranked by posterior inclusion probability (15) in the benchmark normal BACE case, which is shown in the fourth column of the Data Appendix B table. The posterior can be compared to the prior probability of inclusion, which in the benchmark case with prior model size $\bar{k} = 7$ equals $\pi_i^{BACE} = \bar{k}/K = 7/67 = 0.104$. SDM (2004) call the 18 highest ranked explanatory variables with posterior inclusion probability greater than the prior, "significantly" related to economic growth. SDM (2004) find the next three variables ranked 19 to 21 to be "marginally" partially related to economic growth.

Table 1 contrasts the the posterior inclusion probabilities for the benchmark normal BACE case (shown in column 3) with estimates that allow for heteroscedastic errors and outlying observations *a priori*. Column 4 shows the posterior inclusion probability that uses the exponential prior distribution for the degrees of freedom (23) with *prior* mean degrees of freedom $v_0 = 25$. Note that the the posterior mean degrees of freedom $E(v|\mathbf{y})$ equal 19.5, implying important deviations from normality and evidence for outlying observations (see also the posterior variance estimates shown in Figure 2). As robustness check, we also show posterior inclusion probabilities for the case of fixed prior degrees of freedom, with increasing *prior* weight on outliers as we move from column 5 to column 8.

Table 2 shows the corresponding standardized coefficients defined in equation (18) that can be used to assess the economic importance of growth determinants. Table 2 shows standardized coefficients for the benchmark normal BACE case (column 3), and specifications that allowing for heteroscedasticity a priori, using either case 4b with the exponential distribution for the degrees of freedom (column 5), or fixed hyperparameters (columns 6 to 9, respectively). Column 4 also shows the standardized coefficients using White's robust standard errors (14) and BACE posterior

[12]The dataset, including a list of data sources and countries with complete observations, is available at `http://www.econ.cam.ac.uk/faculty/doppelhofer/research/bace.htm`. We are addressing data issues, such as missing observations, in ongoing research.

model weights.[13]

The main empirical findings can be summarized as follows.

First, we find that several variables considered robustly partially correlated by with economic growth by SDM (2004) are sensitive to allowing for outliers and heteroscedastic errors. When comparing the posterior inclusion probabilities in Table 1under the normal benchmark case in column 3 with the cases that allowing for heteroscedasticity *a priori* in columns 4 to 8 we can see that some entries fall below the prior inclusion probability equal to 0.10. For convenience we have emphasized these entries in *italics*. In particular, we find that among the top ranked 18 variables half of them (nine) are considered not to be significantly related under *some* cases of departure from normally distributed errors. Two of these variables, the *Fraction of GDP in Mining* and the *Government Consumption Share*, are not considered significantly related under *any* deviation from the normal benchmark case. Similarly, the three marginal variables, *Population Density*, *Real Exchange Rate Distortions* and *Fraction Speaking a Foreign Language* are not robustly related when allowing for any degree of heteroscedasticity of the regression errors. However, we confirm the finding of SDM (2004) that none of the variables considered not significantly related by them is becoming significantly partially correlated once we allow for departure from the normal benchmark case.

Second, we find that *conditional on inclusion*, some of the variables considered robust by SDM (2004) do not appear to have statistically significant effects measured by a standardized coefficient. A value much below two in absolute value indicates that the coefficient is not considered to be statistically significant according to classical hypothesis tests. Also one might be interested in cases where the standardized coefficient indicates lower statistical significance than the normal benchmark case, i.e. when inference would be affected. We highlight such entries by setting them in italics. Table 2 shows the standardized coefficients defined in equation (18) for the normal benchmark case in column 3 and for cases that allow for heteroscedasticity *a priori* in columns 5 to 9. We find that nine of the 18 variables considered robust by SDM (2004) are not significant in some cases that deviate from the normal benchmark case. Conditional on inclusion, two variables of these variables, the *Fraction of GDP in Mining* and the *Government Consumption Share*, are considered not significantly related when allowing for *any* degree of heteroscedasticity of the regression errors. Also all marginally related variables are not significantly related with economic growth under any deviations from normality of errors. Conditional on being included in the regression model, we find no instances where allowing for heteroscedasticity changes the inference for the variables ranked 22nd and below that are considered not significantly related to economic growth by SDM (2004).

[13]Posterior model weights in the normal benchmark case are not affected by variance estmates, so the estimates in Table 1 using White's robust standard errors agree with the BACE estimates up to sampling variations.

4.1 A Closer Look at Some Growth Determinants

To further investigate the robustness of particular growth determinants we can contrast the inference based on the benchmark normal case with the estimates that allow for prior heteroscedasticity. The following discussion focusses on cases where statistical inference is affected and a decision-maker may wish to take the sensitivity of results with respect to changes in the prior error structure into account.

The *Price of Investment Goods* has a negative and significant partial effect on economic growth according to SDM (2004) with posterior inclusion probability equal to 0.77 in Table 1 (the variable is ranked third overall). For relatively strong prior heteroscedasticity with $E(v) = 10, V(v) = 5$ the inclusion probability is falling to 0.10 which is equal to the prior inclusion probability. On the other hand, when using the exponential prior for the degrees of freedom, we find that the posterior inclusion probability for investment goods prices is robust to allowing for outliers a priori. Conditional on inclusion in the regression model, the standardized coefficient falls from -3.37 in the normal benchmark case to -2.84 under the strongly heteroscedastic prior above, implying a statistically significant partial relation with economic growth. A policy-maker can therefore conclude that investment goods are a relatively robust determinant of economic growth, even though allowing for sufficiently large outliers *a priori*, can diminish the significance of its relationship with growth.

Inference on dummy variables for *Sub-Saharan African* and *Latin American* dummy, and population *Fraction Muslim* and *Fraction Buddhist*, are affected by allowing for different degrees of heteroscedasticity. This is perhaps not too surprising, because the dummy variables might to some extent capture outlying country observations. Interestingly, the Fraction of GDP Mining might also fall in this category, since one country in Sub-Saharan Africa, Botswana, had a much higher growth rate than its neighboring countries in Sub-Saharan Africa, but has a large mining share of GDP.[14] To investigate the extent of outlying observations more explicitly, Figure 2 shows the estimated posterior error variance matrix $E(\Omega|\mathbf{y}) = E(diag(w_t)|\mathbf{y})$. Figure 2 shows that Botswana has a posterior variance approximately three times larger than the average country in this sample. Other outlying observations are the Philippines with posterior variance about twice the average, and also the Central African Republic, Gabon, Zaire and Zambia. It is noticeable that five out of six countries with strong outliers are located in Africa, and in particular Sub-Saharan Africa. Notice that these outliers are present *despite* the inclusion of the Sub-Saharan Africa dummy as regressor in many models M_j.

In addition, the posterior inclusion probabilities for other variables labeled robust by SDM(2004) are affected by a more heteroscedastic prior. The inclusion probability for the *Number of Years an Economy has Been Open* falls from 0.11 for the benchmark case to 0.03 for the most heteroscedastic prior in Table 1. Conditional on inclusion, the standardized coefficients falls from 1.94 in the normal

[14] We thank Andrew Warner for pointing out this fact.

benchmark case to 1.70 when allowing for strong outliers a priori. A policymaker might again conclude that this openness measure is not very robust to outlying observations. A similar argument can be made for Ethnolinguistic Fractionalization which is ranked 17th by posterior inclusion probability in the normal benchmark case, but becomes statistically insignificant when allowing for prior heteroscedasticity.

A few variables, such the *Openness between 1965-74*, measured by total trade over GDP, *Higher Education Enrollment*, and the *Government Share of GDP*, have a larger standardized coefficient in Table 2 if we allow for strong outlying observations and fix the degrees of freedom *a priori*. This finding does not appear to be robust to estimating the posterior degrees of freedom as shown in column 5 of Table 2. We therefore conclude that no variable that is labeled unimportant in explaining growth under the benchmark normal BACE case is considered robust when allowing for heteroscedasticity *a priori*.

5 Conclusion and Outlook

This paper investigates the sensitivity of the benchmark Bayesian Model Averaging (BMA) procedure that assumes restrictive linear models with independent normal sampling and homoscedastic errors. We evaluate the robustness of the benchmark BMA in the context of cross-country growth regressions and find that inference on the robustness of growth determinants is significantly affected by considering deviations from the benchmark BMA, such as outliers and heteroscedastic errors.

We are working on extending the robust model averaging approach to a broader set of prior parameters. We plan to investigate the forecasting performance of robust model averaging and apply it in other cases that are likely to exhibit important deviations from the standard normal model averaging approach.

References

[1] Akaike H. 1973. Information Theory and an Extension of the Maximum Likelihood Principle. In *Second International Symposium on Information Theory*, Petrov B, Csake F. (eds). Akademiai Kiado: Budapest.

[2] Bauwens L, Lubrano M. 1998. Bayesian Inference on GARCH models using the Gibbs sampler. *Econometrics Journal* 1: C23-C46.

[3] Brock W, Durlauf S. 2001. Growth Empirics and Reality. *World Bank Economic Review* 15(2): 229-72.

[4] Brock W, Durlauf S, West K. 2003. Policy Evaluation in Uncertain Economic Environments (with Comments and Discussion.) *Brookings Papers of Economic Acitivity*: 235-322.

[5] Burnham KP, Anderson DR. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* 2nd edition. Springer: New York.

[6] Chib S. 1993. "Markov Chain Monte Carlo Methods: Computation and Inference." In *Handbook of Econometrics*, Vol. 5, Heckman J.J. and Leamer E. (eds). Elsevier: North Holland.

[7] Davidson, R. and J.G. MacKinnon. 1993. *Estimation and Inference in Econometrics.* Oxford University Press.

[8] Fernandez C, Ley E, Steel MFJ. 2001a. Model Uncertainty in Cross-Country Growth Regression. *Journal of Applied Econometrics* 16: 563-76.

[9] Fernandez C, Ley E, Steel MFJ. 2001b. Benchmark Priors for Bayesian Model Averaging. *Journal of Econometrics* 100(2): 381-427.

[10] Fernandez C, Steel MFJ. 2000. Bayesian Regression Analysis with Scale Mixtures of Normals. *Econometric Theory* 16: 80-101.

[11] Garratt A, Lee K, Pesaran MH, Shin Y. 2003. Forecasting Uncertainties in Macroeconometric Modelling: An Application to the UK Economy. *Journal of the American Statistical Association* 98: 829-38.

[12] George E, McCulloch R. 1993. Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association* 88(423): 881-89.

[13] Geweke J. 1993. "Bayesian Treatment of the Independent Student-t Linear Model." *Journal of Applied Econometrics* 8: S19-S40.

[14] Godsill S, Stone M, Weeks M. 2004. Assessing the Impact of Private sector Balance Sheets on Financial Crises: A Comparison of Bayesian and Information-Theoretic Measures of Model Uncertainty. *Working Paper*. University of Cambridge.

[15] Gottardo R, Raftery AE. 2007. Bayesian Robust Variable and Transformation Selection: A Unified Approach. *Technical Report* no. 508. Department of Statistics. University of Washington.

[16] Harvey AC, Ruiz E, Shepard, NG. 1994. Multivariate Stochastic Variance Models. *Review of Economic Studies* 61: 247-64.

[17] Hepple, L.W. 1995a. Bayesian techniques in spatial and network econometrics: 1 Model Comparison and posterior odds. *Environment and Planning* A, Volume 27, pp. 447-469.

[18] Hepple, L.W. 1995b. Bayesian techniques in spatial and network econometrics: 2. Computational methods and algorithms, *Environment and Planning A,* Volume 27, pp. 615-644.

[19] Hjort N, Claeskens G. 2003. Frequentist Model Average Estimators. *Journal of the American Statistical Association* 98(464): 879-99.

[20] Hoeting J, Raftery A, Madigan D. 1996. A Method for Simultaneous Variable Selection and Outlier Identification in Linear Regression. *Computational Statistics and Data Analysis* 22: 251-70

[21] Hoeting J, Madigan D, Raftery A, Volinsky C. 1999. Bayesian Model Averaging: A Tutorial. *Statistical Science* 14: 382-417.

[22] Lange KL, Little RJA, Taylor JMG. 1989. Robust Statistical Modeling Using the t Distribution. *Journal of the American Statistical Association* 84(408): 881-96.

[23] Levin A, Williams J. 2003. Robust Monetary Policy with Competing Reference Models. *Journal of Monetary Economics* 50: 945-75.

[24] Levine RE, Renelt D. 1992. A Sensitivity Analysis of Cross-Country Growth Regressions. *American Economic Review* 82(4): 942-63.

[25] Ley E, Steel MFJ. 2006. Jointness in Bayesian Variable Selection With Applications to Growth Regressions. *World Bank Policy Research Working Paper* 4063, November.

[26] Jeffreys H. 1961. *Theory of Probability.* Third edition. Oxford University Press: London.

[27] Kass R, Raftery A. 1995. Bayes Factors. *Journal of the American Statistical Association* 90(430): 773-95.

[28] Kass R, Wasserman L. 1995. A Reference Bayesian Test for Nested Hypotheses and its Relationship to the Schwarz Criterion. *Journal of the American Statistical Association* 90: 928-34.

[29] Klein RW, Brown SJ. 1984. Model Selection When There is "Miminal" Prior Information. *Econometrica* 52(5): 1291-321.

[30] Koop G. 2003. *Bayesian Econometrics*. Wiley: Chichester.

[31] Leamer EE. 1978. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. Wiley: New York.

[32] LeSage, J.P. and Parent, O. 2006. Bayesian Model Averaging for Spatial Econometric Models. mimeo Department of Finance and Economics, Texas State University, USA.

[33] Mitchell T, Beauchamp J. 1988. Bayesian Variable Selection in Linear Models. *Journal of the American Statistical Association* 83(404): 1023-32.

[34] Raftery A. 1995. Bayesian Model Selection in Social Research. *Sociological Methodology* 25, 111-63.

[35] Sala-i-Martin X, Doppelhofer G, Miller RI. 2004. Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. *American Economic Review* 94(4): 813-35.

[36] Schwarz G. 1978. Estimating the Dimensions of a Model." *Annals of Statistics* 6(2): 461-4.

[37] Stone M, Weeks M. 2001. Systemic Financial Crises, Balance Sheets, and Model Uncertainity. *IMF Working Paper* 01/162.

[38] Temple J. 2000. Growth Regressions and what the Textbooks Don't Tell You. *Bulletin of Economic Research* 52(3): 181-205.

[39] Wasserman L. 2000. "Bayesian Model Selection and Model Averaging." *Journal of Mathematical Psychology* 44: 92-107.

[40] White H. 1980. A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica* 48(4): 817-38.

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

A Computational Appendix

This appendix discusses some of the details of the sampling and model averaging procedure. First, we discuss the random and stratified sampling procedures for the benchmark case presented in section 2. Second, we discuss details of the Gibbs sampler used in section 3.2.

A.1 Random and Stratified Sampling

For the benchmark normal case of sections 2 we can analytically calculate the marginal likelihood (31) and OLS estimates of mean and variance of β_j conditional on each model M_j. The *random sampler* therefore draws directly from the posterior distribution of $p(\beta|\mathbf{y})$. With $K = 67$ regressors the number of possible regression models equals $2^{67} \approx 1.48 \times 10^{20}$. Each regression takes approximately 0.0005 seconds using GAUSS on a recent PC. An exhaustive search over all models is therefore not feasible. Instead we draw from the posterior distribution and related parameters of interest until the parameters have converged numerically (see convergence criteria below).

The *stratified sampler* introduced by SDM (2004, Technical Appendix) allows for sampling inclusion probabilities to differ from prior inclusion probabilities $\pi = \bar{k}/K$. After every 100,000 regressions, the sampling inclusion probability for each variable π_i^S are set equal to a weighted average (with weight 0.5 in the benchmark case) of the initial inclusion probability and the posterior inclusion probability (4) estimated on those runs. To avoid sampling only a very small set of variables, the sampling inclusion probabilities are restricted to lie in the interval $[0.1, 0.85]$. The stratified sampler over-samples models that include variables with high inclusion probability which greatly speeds up numerical convergence. To correct for differences in sampling probabilities of different regressors, we scale the posterior weights by the ratio of prior to sampling probabilities π/π_i^S for $i = 1, ..., K$.

To check the *numerical convergence* in the benchmark case, we monitor changes in the posterior mean of the slope coefficient β, normalized by the ratio of the standard deviation of $\mathbf{y}$ to the standard deviation of $\mathbf{X}$. Numerical convergence is achieved if the maximum change of the normalized coefficients falls below 10^{-6} for ten consecutive sets of 10,000 runs.

A.2 Gibbs Sampling

For the mixture-normal case of section 3.2, analytic expressions of the marginal likelihood are not available. However, we can make a sufficiently large number of draws from the *conditional* posterior distributions (24)-(26) for mean and variance of the slope coefficient β, (27) and (28) for the error variance $\sigma^2\mathbf{\Omega}$, and fixed hyperparameter (22) or alternatively prior distribution (23) for the degrees of freedom v. Under certain regularity conditions the chain of draws from the Gibbs sampler converges to the posterior distribution as the number of draws becomes large (see for example Chib, 2001).

Let $s = 1, .., S$ be number of draws from the posterior distributions. We discard S_0 burn-in draws and estimate the parameters using the remaining $S_1 = S - S_0$ draws.

- The posterior mean $\beta_j \equiv E(\beta_j | M_j, \mathbf{y})$ is estimated by $\frac{1}{S_1}\sum_{s=S_0+1}^{S} \tilde{\beta}_{j,s}$ using draws from (24).

- The posterior variance matrix $\mathbf{V}_j \equiv Var(\beta_j | M_j, \mathbf{y})$ can be calculated either numerically (n) or analytically (a) using

$$\mathbf{V}_j^{\mathrm{n}} = \frac{1}{S_1}\sum_{s=S_0+1}^{S} \tilde{\beta}_{j,s}^2 - [E(\beta_j | M_j, \mathbf{y})]^2 \tag{32}$$

$$\mathbf{V}_j^{\mathrm{a}} = \left(\mathbf{X}_j' \boldsymbol{\Omega}^{-1} \mathbf{X}_j + \sigma^2 \mathbf{V}_{0j}^{-1}\right)^{-1} \tag{33}$$

  $\sigma^2 \equiv E(\sigma^2 | M_j, \mathbf{y})$ is estimated by $\frac{1}{S_1}\sum_{s=S_0+1}^{S} \sigma_s^2$ using draws from (27).

- The posterior error variance $\boldsymbol{\Omega} \equiv E(\boldsymbol{\Omega} | M_j, \mathbf{y})$ is estimated using draws from (28)

$$\frac{1}{S_1}\sum_{s=S_0+1}^{S} \boldsymbol{\Omega}_s = \frac{1}{S_1}\sum_{s=S_0+1}^{S} diag(w_{1,s}, ..., w_{T,s}) \tag{34}$$

We combine the estimates from the Gibbs sampler for the conditional mean, β_j and conditional variance, $\sigma^2 \mathbf{V}_j$ with posterior model weights (31) to estimate the posterior mean and variance unconditionally over the model space $\mathcal{M}$.

To check for *numerical convergence* of model averaging with Gibbs sampling, we observe that for a large enough number of draws S from the Gibbs sampler, the numerical (32) and analytic (33) estimates of the posterior variance should be similar. The default values are 1050 draws from the Gibbs sampler (excluding 50 draws for burnin) and 1 million draws from the model averaging loops. We use the difference between posterior standardized coefficients using the numerical and analytic posterior variance as numerical convergence criterion in this case. In particular, we choose the number of Gibbs draws $\bar{S}$, so that

$$\left| \beta_i / \sqrt{\mathbf{V}_i^{\mathrm{n}}} - \beta_i / \sqrt{\mathbf{V}_i^{\mathrm{a}}} \right|_{S=\bar{S}} < 0.1, \quad i = 1, ..., K \tag{35}$$

B Data Appendix

Rank	Short Name	Variable Description	$p(i\|\mathbf{y})$	Mean	S.D.
Depend. Variable	**GROWTH**	**Average Growth Rate of PPP-adjusted GDP per Capita between 1960–1996**	–	0.0182	0.019
1	**EAST**	**East Asian Dummy**	0.823	0.1136	0.3192
2	**P60**	**Primary Schooling Enrollment**	0.796	0.7261	0.2932
3	**IPRICE1**	**Investment Price**	0.774	92.47	53.68
4	**GDPCH60L**	**Initial ncome (Log GDP in 1960)**	0.685	7.3549	0.9011
5	**TROPICAR**	**Fraction of Tropical Area**	0.563	0.5702	0.4716
6	**DENS65C**	**Population Coastal Density**	0.428	146.87	509.83
7	**MALFAL66**	**Malaria Prevalence**	0.252	0.3394	0.4309
8	**LIFE060**	**Life Expectancy**	0.209	53.72	12.06
9	**CONFUC**	**Fraction Confucian**	0.206	0.0156	0.0793
10	**SAFRICA**	**Sub-Saharan Africa Dummy**	0.154	0.3068	0.4638
11	**LAAM**	**Latin American Dummy**	0.149	0.2273	0.4215
12	**MINING**	**Fraction GDP in Mining**	0.124	0.0507	0.0769
13	**SPAIN**	**Spanish Colony Dummy**	0.123	0.1705	0.3782
14	**YRSOPEN**	**Years Open 1950-94**	0.119	0.3555	0.3444
15	**MUSLIM00**	**Fraction Muslim**	0.114	0.1494	0.2962
16	**BUDDHA**	**Fraction Buddhist**	0.108	0.0466	0.1676
17	**AVELF**	**Ethnolinguistic Fractionalization**	0.105	0.3476	0.3016
18	**GVR61**	**Government Consumption Share**	0.104	0.1161	0.0745
19	**DENS60**	**Population Density**	0.086	108.07	201.44
20	**RERD**	**Real Exchange Rate Distortions**	0.082	125.03	41.71
21	**OTHFRAC**	**Fraction Speaking Foreign Language**	0.080	0.3209	0.4136
22	**OPENDEC1**	**Openness 1965-74**	0.076	0.5231	0.3359
23	**PRIGHTS**	**Political Rights**	0.066	3.8225	1.9966
24	**GOVSH61**	**Government Share of GDP**	0.063	0.1664	0.0712
25	**H60**	**Higher Education Enrollment**	0.061	0.0376	0.0501
26	**TROPPOP**	**Fraction Population In Tropics**	0.058	0.3000	0.3731
27	**PRIEXP70**	**Primary Exports**	0.053	0.7199	0.2827
28	**GGCFD3**	**Public Investment Share**	0.048	0.0522	0.0388
29	**PROT00**	**Fraction Protestant**	0.046	0.1354	0.2851
30	**HINDU00**	**Fraction Hindu**	0.045	0.0279	0.1246
31	**POP1560**	**Fraction Population Less than 15**	0.041	0.3925	0.0749
32	**AIRDIST**	**Air Distance to Big Cities**	0.039	4324	2614
33	**GOVNOM1**	**Nominal Govertnment Share**	0.036	0.1490	0.0584
34	**ABSLATIT**	**Absolute Latitude**	0.033	23.21	16.84
35	**CATH00**	**Fraction Catholic**	0.033	0.3283	0.4146
36	**FERTLDC1**	**Fertility**	0.031	1.5620	0.4193
37	**EUROPE**	**European Dummy**	0.030	0.2159	0.4138
38	**SCOUT**	**Outward Orientation**	0.030	0.3977	0.4922
39	**COLONY**	**Colony Dummy**	0.029	0.7500	0.4355
40	**CIV72**	**Civil Liberties**	0.029	0.5095	0.3259
41	**REVCOUP**	**Revolutions and Coups**	0.029	0.1849	0.2322
42	**BRIT**	**British Colony Dummy**	0.027	0.3182	0.4684
43	**LHCPC**	**Hydrocarbon Deposits**	0.025	0.4212	4.3512
44	**POP6560**	**Fraction Population Over 65**	0.022	0.0488	0.0290
45	**GDE1**	**Defense Spending Share**	0.021	0.0259	0.0246
46	**POP60**	**Population in 1960**	0.021	20308	52538
47	**TOT1DEC1**	**Terms of Trade Growth in 1960s**	0.021	-0.0021	0.0345
48	**GEEREC1**	**Public Education Spending Share**	0.021	0.0244	0.0096
49	**LANDLOCK**	**Landlocked Country Dummy**	0.021	0.1705	0.3782
50	**HERF00**	**Religion Measure**	0.020	0.7803	0.1932
51	**SIZE60**	**Size of Economy**	0.020	16.15	1.82
52	**SOCIALIST**	**Socialist Dummy**	0.020	0.0682	0.2535
53	**ENGFRAC**	**English Speahing Population**	0.020	0.0840	0.2522
54	**PI6090**	**Average Inflation 1960-90**	0.020	13.13	14.99
55	**OIL**	**Oil Producing Country Dummy**	0.019	0.0568	0.2328
56	**DPOP6090**	**Population Growth Rate 1960-90**	0.019	0.0215	0.0095
57	**NEWSTATE**	**Timing of Independence**	0.019	1.0114	0.9767
58	**LT100CR**	**Land Area Near Navigable Water**	0.019	0.4722	0.3802
59	**SQPI6090**	**Square of Inflation 1960-90**	0.018	394.54	1119.70
60	**WARTIME**	**Fraction Spent in War 1960-90**	0.016	0.0695	0.1524
61	**LANDAREA**	**Land Area**	0.016	867189	1814688
62	**ZTROPICS**	**Tropical Climate Zone**	0.016	0.1900	0.2687
63	**TOTIND**	**Terms of Trade Ranking**	0.016	0.2813	0.1904
64	**ECORG**	**Capitalism**	0.015	3.4659	1.3809
65	**ORTH00**	**Fraction Othodox**	0.015	0.0187	0.0983
66	**WARTORN**	**War Particpation 1960-90**	0.015	0.3977	0.4922
67	**DENS65I**	**Interior Density**	0.015	43.37	88.06

Variables ranked by Posterior Inclusion Probability $p(i|\mathbf{y})$. Prior inclusion probability for benchmark case equals $\bar{k}/67 = 7/67 = 0.10$.

Table 1: Posterior Inclusion Probabilities

Rank	Variable	BACE	MH $(v_0) = 25$	$E(v) = 100$ $V(v) = 10$	$E(v) = 50$ $V(v) = 10$	$E(v) = 26$ $V(v) = 6$	$E(v) = 10$ $V(v) = 5$
1	**EAST**	0.82	0.88	0.94	0.98	0.94	0.97
2	**P60**	0.80	0.83	0.68	0.64	0.85	0.75
3	**IPRICE1**	0.77	0.77	0.48	0.46	0.32	0.10
4	**GDPCH60L**	0.68	0.67	0.44	0.42	0.51	0.79
5	**TROPICAR**	0.56	0.63	0.50	0.49	0.41	0.29
6	**DENS65C**	0.43	0.46	0.28	0.19	0.27	0.21
7	**MALFAL66**	0.25	0.25	0.45	0.47	0.52	0.61
8	**LIFE060**	0.21	0.16	0.17	0.24	0.11	0.29
9	**CONFUC**	0.21	0.17	0.10	*0.07*	0.10	*0.06*
10	**SAFRICA**	0.15	*0.09*	*0.06*	*0.07*	0.12	0.27
11	**LAAM**	0.15	*0.10*	*0.06*	*0.05*	0.15	0.27
12	**MINING**	0.12	*0.08*	*0.04*	*0.02*	*0.02*	*0.02*
13	**SPAIN**	0.12	0.14	0.17	0.17	0.30	0.38
14	**YRSOPEN**	0.12	*0.10*	*0.07*	*0.06*	*0.02*	*0.03*
15	**MUSLIM00**	0.11	0.11	*0.07*	*0.06*	*0.08*	*0.05*
16	**BUDDHA**	0.11	*0.08*	*0.06*	*0.03*	*0.05*	*0.01*
17	**AVELF**	0.10	0.11	*0.05*	*0.11*	*0.02*	*0.01*
18	**GVR61**	0.10	*0.08*	*0.07*	*0.11*	*0.04*	*0.08*
19	**DENS60**	0.09	0.07	0.02	0.04	0.01	0.01
20	**RERD**	0.08	0.09	0.05	0.04	0.04	0.05
21	**OTHFRAC**	0.08	0.08	0.04	0.03	0.02	0.07
22	**OPENDEC1**	0.08	0.07	0.05	0.05	0.02	0.02
23	**PRIGHTS**	0.07	0.06	0.03	0.02	0.03	0.01
24	**GOVSH61**	0.06	0.04	0.03	0.03	0.07	0.05
25	**H60**	0.06	0.07	0.05	0.05	0.06	0.02
26	**TROPPOP**	0.06	0.05	0.03	0.04	0.02	0.01
27	**PRIEXP70**	0.05	0.04	0.05	0.05	0.04	0.04
28	**GGCFD3**	0.05	0.04	0.01	0.02	0.01	0.02
29	**PROT00**	0.05	0.08	0.03	0.02	0.06	0.05
30	**HINDU00**	0.04	0.04	0.02	0.02	0.02	0.02
31	**POP1560**	0.04	0.04	0.03	0.03	0.02	0.02
32	**AIRDIST**	0.04	0.03	0.02	0.01	0.02	0.05
33	**GOVNOM1**	0.04	0.03	0.02	0.02	0.03	0.08
34	**ABSLATIT**	0.03	0.03	0.02	0.04	0.02	0.01
35	**CATH00**	0.03	0.02	0.02	0.01	0.01	0.01
36	**FERTLDC1**	0.03	0.04	0.02	0.02	0.04	0.01
37	**EUROPE**	0.03	0.02	0.02	0.01	0.02	0.01
38	**SCOUT**	0.03	0.03	0.02	0.02	0.03	0.08
39	**COLONY**	0.03	0.03	0.02	0.04	0.02	0.02
40	**CIV72**	0.03	0.02	0.02	0.05	0.03	0.07
41	**REVCOUP**	0.03	0.02	0.01	0.01	0.01	0.01
42	**BRIT**	0.03	0.03	0.02	0.02	0.01	0.01
43	**LHCPC**	0.02	0.02	0.01	0.01	0.01	0.01
44	**POP6560**	0.02	0.02	0.02	0.01	0.01	0.01
45	**GDE1**	0.02	0.03	0.01	0.01	0.03	0.01
46	**POP60**	0.02	0.02	0.01	0.01	0.01	0.01
47	**TOT1DEC1**	0.02	0.02	0.01	0.02	0.02	0.01
48	**GEEREC1**	0.02	0.02	0.01	0.01	0.01	0.01
49	**LANDLOCK**	0.02	0.02	0.01	0.01	0.01	0.01
50	**HERF00**	0.02	0.02	0.01	0.01	0.00	0.01
51	**SIZE60**	0.02	0.02	0.01	0.01	0.01	0.01
52	**SOCIALIST**	0.02	0.02	0.01	0.01	0.01	0.01
53	**ENGFRAC**	0.02	0.02	0.01	0.01	0.01	0.01
54	**PI6090**	0.02	0.02	0.01	0.01	0.01	0.01
55	**OIL**	0.02	0.02	0.01	0.01	0.01	0.01
56	**DPOP6090**	0.02	0.02	0.01	0.02	0.01	0.01
57	**NEWSTATE**	0.02	0.02	0.01	0.01	0.01	0.01
58	**LT100CR**	0.02	0.03	0.01	0.01	0.01	0.01
59	**SQPI6090**	0.02	0.02	0.01	0.01	0.01	0.01
60	**WARTIME**	0.02	0.01	0.01	0.01	0.03	0.01
61	**LANDAREA**	0.02	0.01	0.01	0.01	0.01	0.01
62	**ZTROPICS**	0.02	0.02	0.01	0.01	0.01	0.01
63	**TOTIND**	0.02	0.01	0.01	0.01	0.01	0.01
64	**ECORG**	0.02	0.01	0.01	0.01	0.00	0.01
65	**ORTH00**	0.02	0.01	0.01	0.01	0.01	0.01
66	**WARTORN**	0.02	0.01	0.01	0.01	0.01	0.01
67	**DENS65I**	0.02	0.01	0.01	0.01	0.01	0.01

Variables ranked by Posterior Inclusion Probability $p(i|\mathbf{y})$.

Prior inclusion probability for benchmark case equals $\bar{k}/67 = 7/67 = 0.10$.

Table 2: Standardized Coefficients

Rank	**Variable**	**BACE**	**White**	**MH** $(v_0) = 25$	$E(v) = 100$ $V(v) = 10$	$E(v) = 50$ $V(v) = 10$	$E(v) = 26$ $V(v) = 6$	$E(v) = 10$ $V(v) = 5$
1	**EAST**	3.56	4.92	3.67	4.10	4.20	4.06	3.95
2	**P60**	3.37	4.96	3.37	3.02	2.77	3.21	3.62
3	**IPRICE1**	-3.37	-8.01	-3.04	-3.20	-3.07	-3.23	-2.84
4	**GDPCH60L**	-2.96	-4.81	-2.84	-2.75	-2.73	-2.59	-2.84
5	**TROPICAR**	-3.49	-5.31	-3.49	-3.34	-3.24	-3.12	-2.85
6	**DENS65C**	2.98	5.99	2.89	2.88	2.72	2.79	2.60
7	**MALFAL66**	-2.54	-3.30	-2.64	-3.20	-3.25	-3.51	-3.67
8	**LIFE060**	2.28	2.90	2.21	2.37	2.55	2.21	2.98
9	**CONFUC**	2.43	3.53	2.07	2.17	*1.90*	2.24	2.01
10	**SAFRICA**	-2.14	-2.62	-2.10	*-1.71*	*-1.86*	-2.08	-2.28
11	**LAAM**	-2.19	-2.70	*-1.94*	*-1.67*	*-1.58*	-2.24	*-1.86*
12	**MINING**	2.02	3.43	*1.37*	*1.86*	*1.77*	*1.64*	*0.53*
13	**SPAIN**	-2.13	-3.09	-2.37	-2.53	-2.61	-2.93	-2.59
14	**YRSOPEN**	*1.94*	3.10	*1.74*	*1.99*	*1.98*	*1.95*	*1.70*
15	**MUSLIM00**	2.02	2.81	2.02	*1.97*	2.13	*1.99*	*1.58*
16	**BUDDHA**	2.02	2.94	*-1.69*	2.07	*1.53*	*1.59*	*0.82*
17	**AVELF**	*-1.93*	-3.76	*-1.79*	*-1.85*	*-1.56*	*-1.64*	*-1.24*
18	**GVR61**	*-1.74*	-3.06	*-1.52*	*-1.86*	*-1.55*	*-1.76*	*-1.86*
19	**DENS60**	*1.86*	3.71	*1.44*	*1.38*	*1.45*	*1.27*	*0.12*
20	**RERD**	*-1.82*	-2.93	*-1.84*	*-1.84*	*-1.84*	*-1.85*	*-1.84*
21	**OTHFRAC**	*1.77*	2.80	*1.72*	*1.65*	*1.77*	*1.74*	*1.92*
22	**OPENDEC1**	1.70	2.59	1.56	1.92	1.85	1.70	1.35
23	**PRIGHTS**	-1.54	-2.41	-1.38	-1.40	-1.28	-1.13	-0.31
24	**GOVSH61**	-1.19	-1.91	-0.80	-0.97	-1.21	-1.98	-1.79
25	**H60**	-1.67	-2.76	-1.48	-1.85	-1.82	-2.06	-1.40
26	**TROPPOP**	-1.59	-2.15	-1.43	-1.46	-1.61	-1.23	-1.01
27	**PRIEXP70**	-1.51	-2.26	-1.35	-1.77	-1.62	-1.60	-1.87
28	**GGCFD3**	-1.43	-2.03	-1.34	-0.79	-0.91	-0.84	-0.85
29	**PROT00**	-1.28	-1.49	-1.76	-1.46	-1.54	-1.76	-1.89
30	**HINDU00**	1.40	2.20	1.20	1.26	1.34	1.25	1.30
31	**POP1560**	1.09	1.43	1.15	1.15	1.17	0.96	0.88
32	**AIRDIST**	-1.23	-1.91	-1.54	-1.00	-0.90	-1.17	-1.80
33	**GOVNOM1**	-1.23	-1.97	-1.26	-1.20	-1.10	-1.14	-1.93
34	**ABSLATIT**	0.58	0.65	0.41	0.88	0.92	0.67	0.97
35	**CATH00**	-0.99	-0.98	-0.67	-0.79	-0.44	-0.29	0.00
36	**FERTLDC1**	-0.74	-0.92	-1.12	-0.48	0.03	-1.01	-0.52
37	**EUROPE**	-0.22	-0.12	-0.15	0.07	0.20	0.44	1.00
38	**SCOUT**	-1.21	-2.85	-1.20	-1.18	-1.28	-1.45	-1.78
39	**COLONY**	-1.06	-1.46	-1.03	-1.26	-0.69	-1.20	-1.30
40	**CIV72**	-1.01	-1.40	-1.30	-1.18	-1.14	-1.40	-1.62
41	**REVCOUP**	-1.16	-2.33	-0.93	-1.09	-0.99	-0.92	-0.51
42	**BRIT**	1.01	1.46	0.95	1.07	1.20	0.96	0.20
43	**LHCPC**	0.73	1.03	0.63	0.21	0.07	0.20	0.47
44	**POP6560**	0.16	0.23	0.18	0.16	0.33	0.27	0.65
45	**GDE1**	0.59	0.99	0.64	0.86	0.79	0.59	0.36
46	**POP60**	0.86	1.42	0.83	0.77	0.70	0.83	0.72
47	**TOT1DEC1**	0.70	1.15	0.70	0.89	0.73	1.29	0.96
48	**GEEREC1**	0.75	1.22	0.39	0.59	0.45	0.12	0.27
49	**LANDLOCK**	-0.49	-0.90	-0.58	-0.23	-0.32	-0.47	-0.78
50	**HERF00**	-0.65	-1.29	-0.58	-0.43	-0.15	-0.29	0.16
51	**SIZE60**	-0.36	-0.43	-0.12	-0.68	-0.14	-0.62	0.19
52	**SOCIALIST**	0.80	2.22	0.82	0.82	1.02	0.83	1.11
53	**ENGFRAC**	-0.51	-0.75	-0.47	-0.88	-0.88	-1.14	-1.03
54	**PI6090**	-0.75	-1.56	-0.81	-0.87	-0.92	-0.82	-0.49
55	**OIL**	0.68	0.99	0.52	0.40	0.35	0.37	0.09
56	**DPOP6090**	0.07	0.10	0.05	0.32	0.21	0.01	0.01
57	**NEWSTATE**	0.56	0.85	0.67	0.61	0.64	0.56	0.41
58	**LT100CR**	-0.44	-0.64	-0.22	-0.37	-0.52	-0.44	-0.62
59	**SQPI6090**	-0.58	-1.07	-0.56	-0.79	-0.65	-0.87	-0.52
60	**WARTIME**	-0.15	-0.29	-0.01	-0.48	-0.29	0.10	-0.28
61	**LANDAREA**	0.20	0.30	0.27	-0.06	-0.01	-0.17	-0.06
62	**ZTROPICS**	-0.31	-0.53	-0.16	-0.41	-0.53	-0.32	-0.07
63	**TOTIND**	-0.39	-0.68	-0.12	-0.34	-0.35	-0.13	0.09
64	**ECORG**	-0.21	-0.40	-0.53	-0.05	-0.05	-0.12	-0.18
65	**ORTH00**	0.42	0.90	0.43	0.61	0.67	0.73	0.26
66	**WARTORN**	-0.25	-0.48	-0.13	-0.23	-0.27	-0.25	-0.36
67	**DENS65I**	-0.08	-0.08	-0.02	0.07	0.13	0.23	-0.14

Figure 2: Posterior Error Variances $\Omega = diag(w_t)$.