

# Two-Stage Bayesian Model Averaging in Endogenous Variable Models\*

BY ALEX LENKOSKI

*Institute for Applied Mathematics, Heidelberg University*

THEO S. EICHER

*Department of Economics, University of Washington*

AND ADRIAN E. RAFTERY

*Department of Statistics, University of Washington*

## Abstract

Economic modeling in the presence of endogeneity is subject to model uncertainty at both the instrument and covariate level. We propose a Two-Stage Bayesian Model Averaging (2SBMA) methodology that extends the Two-Stage Least Squares (2SLS) estimator. By constructing a Two-Stage Unit Information Prior in the endogenous variable model, we are able to efficiently combine established methods for addressing model uncertainty in regression models with the classic technique of 2SLS. To assess the validity of instruments in the 2SBMA context, we develop Bayesian tests of the identification restriction that are based on model averaged posterior predictive  $p$ -values. A simulation shows that 2SBMA has the ability to recover structure in both the instrument and covariate set, and substantially improves the sharpness of resulting coefficient estimates. The Bayesian Sargan test is shown to have a power of 50 percent in detecting a violation of the exogeneity assumption, while the method based on 2SLS has negligible power. We apply our approach to the problem of development accounting, and find support not only for institutions, but also for geography and integration as development determinants, once both model uncertainty and endogeneity have been jointly addressed.

*Keywords:* Bayesian Model Averaging, Endogeneity, Instrumental variables, Posterior Predictive P-Values

---

\*We thank David Albouy and Francesco Trebbi for kindly sharing their data as well as Chris Papageorgiou for helpful comments. Raftery's research was supported by NIH-NICHD Grant no. HD-54511, NSF grant no. IIS-0534094, and NSF grant no. ATM-0724721.

# 1 Introduction

In modeling economic interactions, applied economists are often faced with a surfeit of theories and a number of different variables that proxy a given theory. Raftery (1995) argued that uncertainty surrounding particular theories should be addressed explicitly by the statistical approach. Uncertainty about a parameter can be underestimated when it is based on a single model and there is uncertainty about which model or theory to use. Bayesian model averaging (BMA) has been used extensively to account for model uncertainty in regression models, including in cross-country growth regressions and development accounting.<sup>1</sup> However, these methods are constrained to modeling associations between covariates and the dependent variable, ignoring issues of endogeneity and uncertainty surrounding instrument specifications.

We propose a new methodology that combines the BMA approach to regression variable uncertainty and the Two-Stage Least Squares (2SLS) procedure. Our new method, Two-Stage BMA (2SBMA), enables the applied economist to address uncertainty in instrument and covariate specifications in a natural manner while simultaneously yielding an estimator that addresses endogeneity in the covariates.

To date, instrument uncertainty has been addressed only in standard robustness analyses that juxtapose instruments associated with one particular theory/specification against another. In a prominent example in the development accounting literature, Rodrik et al. (2004), henceforth RST, motivate their work by a “horse race” among alternative theories that propose candidate instruments and regressors.

Accounting for uncertainty about both covariates and instruments requires a methodology that is rooted in statistical theory. Durlauf et al. (2007) introduced an instrument model selection procedure based on evaluating coefficient estimates according to  $t$ -statistics, but warned of the tenuous nature of the underlying statistical theory. The most comprehensive approach to addressing endogeneity in growth regressions has previously been proposed by Durlauf et al. (2008), who built on Tsangarides (2004). The authors introduced a model averaged version of 2SLS, but noted that their heuristic approach lacked statistical justification.<sup>2</sup> Strictly speaking, Durlauf et al. (2008) did not allow for instrument uncertainty, but provided a model averaging approach to instrument candidate regressors in the second stage only.

The 2SBMA approach extends the method of Durlauf et al. (2008) and provides a sta-

---

<sup>1</sup>See e.g., Fernandez et al. (2001), Sala-i-Martin et al. (2004), Ciccone and Jarocinski (2007), and Eicher et al. (2010).

<sup>2</sup>A similar heuristic panel approach is introduced by Hine (2007) to examine the growth/inflation relationship. Morales-Benito (2009) provides statistical foundations for a panel BMA approach.

tistical foundation for the procedure. An alternative to theory-specified instruments is the atheoretic use of lagged dependent variables in a BMA-GMM approach (see Morales-Benito (2009)). We explore the properties of 2SBMA as a valid two-stage estimator and show that the procedure is a consistent methodology that reduces the well known many-instrument bias present in 2SLS. A simulation study conducted below shows a 45% reduction in many instrument bias, as well as a 46% reduction in mean squared error for estimating regression coefficients.

Instrumental variable estimation requires assumptions that relate to the identification of the implied structural model. Tests of the identification restrictions, such the Sargan (1958) and Cragg and Donald (1993) tests, compare a test statistic to a reference distribution. The test statistics are often asymptotically distributed according to the reference distribution, which frequently has a number of degrees of freedom that is related to the size of the model estimated. The nature of these statistics proves problematic in some economic applications, such as the study of growth and development determinants, where the sample size is small and the dimension is continually increasing.

The 2SBMA approach provides a direct interpretation of the efficacy of an instrumentation strategy, by examining posterior inclusion probabilities. However, we also provide measures to verify instrument conditions based on Bayesian model averaged posterior predictive  $p$ -values. We introduce these  $p$ -values as a simple extension of the work of Rubin (1984), Raftery (1988) and Gelman et al. (1996), who motivated the use of posterior predictive  $p$ -values for a single model. These  $p$ -values are then used to derive Bayesian versions of the Sargan and Cragg and Donald tests to test the identification restrictions.

A simulation study quantifies the efficiency of 2SBMA as compared to 2SLS. We find that the Bayesian over-identification test had a power of 50% at detecting this failure, while the traditional Sargan test had a power of only 0.8%.

Finally, we apply 2SBMA to a prominent approach to development accounting that is known to be subject to substantial instrument and determinant uncertainty.<sup>3</sup> We use RST’s own data and the variables motivated by their empirical approach to span the model space for our 2SBMA approach. RST found strong evidence for the “primacy of institutions” over all other alternative theories after conducting an elaborate “horse race” among all alternative theories and their associated candidate regressors. The 2SBMA results suggest a qualifica-

---

<sup>3</sup>We use the term development accounting in the broad sense, referring to studies that seek to examine the determinants of differences in levels of per capita income. Previous development accounting approaches differ in their emphases, such as physical capital (King and Levine (1994)), human capital (Klenow et al. (1997)), as well as TFP (Caselli (2005)). In our application we focus only on studies that sought to explain to differences in per capita incomes based on integration, institutions, and geography since the above approaches do not feature instrument uncertainty.

tion of RST’s strong conclusions. We find that not only institutions, but also integration and geography explain differences in per capita incomes across countries. At the instrument level, we find that the colonial origins (proxied by Settler Mortality) may not be robust to the inclusion of alternative instruments. This suggests that the 2SLS robustness results in the previous literature may reflect the specific robustness specifications used. We find, however, that a number of alternative instruments (and their associated theories) also receive support from the data, once instrument and determinant uncertainty have been addressed simultaneously. These results are consistent with the findings of Durlauf et al. (2008) that many growth theories/variables are not robust once model uncertainty is taken into account.

The article proceeds as follows. Section 2 outlines the statistical foundations of the 2SBMA approach. Section 3 introduces Bayesian tests of the identification restrictions and Section 4 describes the efficiency of 2SBMA and the power of Bayesian tests in simulation studies. Section 5 applies 2SBMA to a prominent approach to development accounting to highlight the importance of both determinant and instrument uncertainty. Section 6 concludes.

## 2 Statistical Foundations of 2SBMA

This section develops the theoretical foundations for two-stage Bayesian Model Averaging (2SBMA). First, we sketch the properties of the 2SLS estimator that we extend. Then we formulate the endogeneity problem to express first and second stage likelihoods in a model averaging context. After developing a two-stage prior that allows for an efficient approximation of the model probabilities, we then derive the complete 2SBMA methodology.

### 2.1 Marginal and Conditional Likelihoods for Two-Stage Applications

A standard approach to addressing endogeneity is to apply 2SLS and impose identification restrictions. We consider the model,

$$Y = (W \ X)\beta + \eta, \tag{1}$$

$$W = (Z \ X)\theta + \epsilon, \tag{2}$$

where  $Y$  is the  $n \times 1$  dependent variable,  $X$  is an  $n \times p_X$  set of covariates,  $W$  is the set of endogenous variables and  $Z$  is the  $n \times p_Z$  set of instruments. The notation  $(A \ B)$  denotes a matrix placing A and B together, provided they have the same number of rows. To simplify exposition we assume that  $W$  is  $n \times 1$ . Assuming that

$$\begin{pmatrix} \eta \\ \epsilon \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\eta^2 & \sigma_{\eta\epsilon} \\ \sigma_{\eta\epsilon} & \sigma_\epsilon^2 \end{pmatrix} = \Sigma \right),$$

the classical endogenous variable situation arises when  $\sigma_{\eta\epsilon} \neq 0$ , causing  $W$  to violate the standard regression assumption of independence of the error term,  $\eta$ .

Fundamental to the 2SBMA methodology is the specification of two-stage marginal and conditional likelihoods. Letting  $U = (W \ X)$  and  $V = (Z \ X)$ , we adopt the notation of Kleibergen and Zivot (2003), and rewrite Equations 1 and 2 as

$$Y = \tilde{U}(\theta)\beta + \nu, \tag{1'}$$

$$W = V\theta + \epsilon, \tag{2'}$$

where  $\tilde{U}(\theta) = (V\theta \ X)$  is the replacement of  $W$  in  $U$  with its fitted value, and the OLS estimate regressing  $W$  on  $V$  is given by  $\hat{U} = \tilde{U}(\hat{\theta})$  where  $\hat{\theta} = (V'V)^{-1}V'W$ . Since  $\nu = \beta\epsilon + \eta$ , we have

$$\text{Var} \begin{pmatrix} \nu \\ \epsilon \end{pmatrix} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix}' \Sigma \begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix}.$$

Kleibergen and Zivot (2003) reparameterize (1') and (2') using  $\phi = \omega_{22}^{-1}\omega_{21}$  to yield

$$Y = \tilde{U}(\theta)\beta + \xi + \epsilon\phi, \tag{1''}$$

$$W = V\theta + \epsilon, \tag{2''}$$

where  $\text{var}(\xi) = \omega_{11.2} = \omega_{11} - \omega_{12}^2\omega_{22}^{-1}$ .

In the presence of endogeneity, it is well known that the determination of  $W$  leads to inconsistent estimates of the entire coefficient vector,  $\beta$ , under standard Ordinary Least Squares (OLS). The 2SLS estimator,

$$\hat{\beta}^{2SLS} = \left( \hat{U}'\hat{U} \right)^{-1} \hat{U}'Y,$$

may resolve the inconsistency problem when instruments,  $Z$ , exist that are independent of  $Y$ , given  $W$  and the vector of covariates,  $X$ . The model is identified only when these conditional independence assumptions are valid, resulting in a consistent estimator. The 2SLS estimate of  $\beta$  is well known to be biased, and the extent of this bias increases with the number of terms that are added in the first stage with coefficients equal, or close to, zero. Under such conditions, 2SLS estimates may not reduce the bias of the OLS results

(Davidson and MacKinnon, 2004). It is therefore of practical interest to develop a 2SBMA methodology that quantifies the degree to which each proposed covariate and instrument enters into the equations above.

The reparameterization of (1'') and (2'') proves useful to derive the two-stage marginal and conditional likelihoods.

**Theorem 1.** *The two-stage marginal and conditional likelihoods are given by*

$$L(\theta|W, V, \omega_{22}) = \omega_{22}^{-n/2} \exp\left(-\frac{1}{2} \frac{(W - V\theta)'(W - V\theta)}{\omega_{22}}\right) \quad (3)$$

$$L(\beta|\phi, \omega_{11.2}, Y, \hat{\theta}) = \omega_{11.2}^{-n/2} \exp\left(-\frac{1}{2} \frac{(Y - \hat{U}\beta + \hat{\epsilon}\phi)'(Y - \hat{U}\beta + \hat{\epsilon}\phi)}{\omega_{11.2}}\right), \quad (4)$$

where  $\hat{\epsilon} = W - V\hat{\theta}$ . The marginal likelihood is maximized at  $\hat{\theta} = (V'V)^{-1}V'W$ , while the conditional likelihood is maximized at  $\hat{\beta}^{2SLS}$ .

*Proof* See Appendix.

## 2.2 Constrained Two-Stage Likelihoods

When juxtaposing competing theories using 2SBMA, some coefficients in either  $\theta$  or  $\beta$  are constrained to be zero. As a consequence, these members of either  $X$  or  $Z$  need not be included in estimation of the associated parameter. This section derives the constrained marginal and conditional likelihoods for this case.

Let  $\mathcal{M}$  be a collection of individual first stage models,  $\{M_1, \dots, M_I\}$ . Associated with each first stage model  $M_i$  is a submatrix  $V^{(i)}$  of the matrix  $V$ . When the variables  $Z_i$  and  $X_m$  are in  $M_i$ , then these columns are retained in  $V^{(i)}$ , while those not in  $M_i$  are excluded. For  $M_i$ , the least squares estimate is  $\hat{\theta}^{(i)} = (V^{(i)'}V^{(i)})^{-1}V^{(i)'}W$ .

Similarly, let  $\mathcal{L}$  be a set of individual second stage models,  $\{L_1, \dots, L_J\}$ . Each model  $L_j$  again corresponds to a submatrix  $U^{(j)}$  of the matrix  $U$ . For  $W \in L_j$ ,  $\tilde{U}^{(j)}(\theta^{(i)})$  denotes the replacement of  $W$  in  $U^{(j)}$  with the fitted value  $V^{(i)}\theta^{(i)}$ , and when  $W \notin L_j$ , we write  $\tilde{U}^{(j)}(\theta^{(i)}) = U^{(j)}$ . To ease notation, we write  $\hat{U}^{(i,j)}$  for  $\tilde{U}^{(j)}(\hat{\theta}^{(i)})$ .

Analogous to the representation in Equations 1'' and 2'', the sampling model under  $M_i$  and  $L_j$  takes the form

$$Y = \tilde{U}^{(j)}(\theta^{(i)})\beta^{(j)} + \xi^{(i,j)} + \epsilon^{(i)}\phi^{(i,j)}, \quad (5)$$

$$W = V^{(i)}\theta^{(i)} + \epsilon^{(i)}, \quad (6)$$

where  $\text{var}(\epsilon^{(i)}) = \omega_{22}^{(i)}$ ,  $\text{var}(\xi^{(i,j)}) = \omega_{11.2}^{(i,j)}$ . The constrained 2SLS estimator based on models  $M_i$  and  $L_j$  is then calculated as

$$\hat{\beta}^{(i,j)} = \left( \hat{U}^{(i,j)'} \hat{U}^{(i,j)} \right)^{-1} \hat{U}^{(i,j)'} Y. \quad (7)$$

Theorem 2 highlights how the marginal and conditional likelihoods are affected when a variable contained in  $U^{(j)}$  is excluded from  $V^{(i)}$ .

**Theorem 2.** *Under models  $M_i$  and  $L_j$  the expressions in Equations (5) and (6) yield the following marginal and conditional likelihoods:*

$$L \left( \theta | W, V, \omega_{22}^{(i)} \right) = \left( \omega_{22}^{(i)} \right)^{-n/2} \exp \left( -\frac{1}{2} \frac{(W - V^{(i)} \theta^{(i)})' (W - V^{(i)} \theta^{(i)})}{\omega_{22}^{(i)}} \right), \quad (8)$$

$$L \left( \beta | \phi^{(i,j)}, \omega_{11.2}^{(i,j)}, Y, \hat{\theta}^{(i)} \right) = \left( \omega_{11.2}^{(i,j)} \right)^{-n/2} \exp \left( -\frac{1}{2} \frac{(Y - \hat{U}^{(j)} \beta^{(j)} + \hat{\epsilon}^{(i)} \phi^{(i,j)})' (Y - \hat{U}^{(j)} \beta^{(j)} + \hat{\epsilon}^{(i)} \phi^{(i,j)})}{\omega_{11.2}^{(i,j)}} \right), \quad (9)$$

where  $\hat{\epsilon}^{(i)} = W - V^{(i)} \hat{\theta}^{(i)}$ . Equation (8) is maximized at  $\hat{\theta}^{(i)}$ , while Equation (9) is maximized at

$$\hat{\beta}^{(i,j)} + \phi^{(i,j)} \hat{\Pi}^{(i,j)}, \quad (10)$$

where  $\hat{\Pi}^{(i,j)} = (\hat{U}^{(j)'} \hat{U}^{(j)})^{-1} \hat{U}^{(j)'} \hat{\epsilon}^{(i)}$ .

*Proof* See appendix.

Theorem 2 implies that the mode of the conditional likelihood for  $\beta$  equals  $\hat{\beta}^{(i,j)}$  plus the term  $\phi^{(i,j)} \hat{\Pi}^{(i,j)}$ . To interpret this difference we note that  $\hat{\Pi}^{(i,j)}$  represents the regression coefficients of  $\hat{U}^{(j)}$  on the first stage residuals. Hence for the case where all covariates in  $L_j$  are contained in  $M_i$ ,  $\hat{\Pi}^{(i,j)} = 0$  and Equation (10) becomes  $\hat{\beta}^{(i,j)}$ . For the case where variable  $X_l$  is in  $L_j$  but excluded from  $M_i$ , and where  $X_l$  has little explanatory power for  $W$  given the remaining elements of  $M_i$ ,  $\hat{\Pi}^{(i,j)}$  is, by definition, negligible. The 2SBMA assumptions outlined below that govern the model pair  $(M_i, L_j)$  constrain  $\hat{\Pi}^{(i,j)}$  to zero.

The 2SBMA methodology is based on calculating the quantities in Equation (7). As opposed to Limited Information Maximum Likelihood (LIML), the focus of our estimator is on maximizing the marginal, then the conditional likelihoods, yielding estimators of the form  $\hat{\beta}^{(i,j)}$ , while LIML maximizes the joint likelihood of the parameters. This causes our methodology to be more similar to 2SLS than LIML. Indeed, note that when  $M_i$  and  $L_j$  are the full models for the first and second stage, respectively,  $\hat{\beta}^{(i,j)} = \hat{\beta}^{2SLS}$ . Furthermore, if  $X_l$  is the only variable excluded from  $M_i$  and  $L_j$ ,  $\hat{\beta}^{(i,j)}$  is simply the 2SLS estimate with this

covariate excluded. Similarly, if  $Z_m$  is the only quantity excluded from  $M_i$ , and  $L_j$  is the full model,  $\hat{\beta}^{(i,j)}$  is the 2SLS estimate with instrument  $Z_m$  excluded.

The formulation above also allows for two occurrences not considered in standard 2SLS. First, we may have variable  $X_l \in M_i$ , but  $X_l \notin L_j$ . This is equivalent to stating that a variable originally considered to be a covariate is now considered conditionally independent of  $Y$ , given the remaining variables in  $L_j$ . However, the variable still has some explanatory power for  $W$ , which causes it to be included in  $M_i$ . Second, we may include variable  $X_l \in L_j$ , but exclude it from  $M_i$ . This case is not considered in standard 2SLS analysis, as it is well known to result in bias if the effect of  $X_l$  on  $W$  is substantial, as outlined by the term  $\hat{\Pi}^{(i,j)}$  in Theorem 2 above. In 2SBMA we do allow for these occurrences, since a large  $\hat{\Pi}^{(i,j)}$  would imply a failure in the model selection methodology during the first stage.

### 2.3 The Two-Stage Unit Information Prior

2SBMA requires a two-stage prior distribution over the parameters of the model. In the context of Bayesian Instrumental Variable estimation, Dreze (1976) suggested the improper prior

$$pr(\beta, \theta, \Sigma) \propto |\Sigma|^{-1/2},$$

for Equations (1) and (2). This two-stage prior is in part motivated by the standard improper prior  $pr(\beta, \sigma) \propto 1/\sigma$  placed on Bayesian regression problems.

Kleibergen and Zivot (2003) note that the Dreze prior may be poorly behaved as  $\theta \rightarrow 0$ . Instead they suggest

$$pr(\beta, \theta, \phi, \omega_{11.2}, \omega_{22}) \propto \omega_{11.2}^{-1} |\Omega|^{-1/2} |\theta' V' V \theta|^{1/2}, \quad (11)$$

a prior that includes dependence on the first stage coefficients,  $\theta$ . Kleibergen and Zivot (2003) note that the posterior from this prior bears a number of similarities to the 2SLS estimator. In particular, the conditional posterior of  $\beta$  when  $\theta = \hat{\theta}$  has mean and mode equal to  $\hat{\beta}^{2SLS}$ .

While the Kleibergen and Zivot prior has a number of desirable properties, it is still improper in several parameters. This may not have undue influence on standard Bayesian posterior parameter estimation. Kass and Raftery (1995) note, however, that prior choices can have a greater impact in the context of model comparison. Hence we seek to develop a proper two-stage prior and choose to base our analysis on the Unit Information Prior (UIP), which has been motivated in the context of BMA (Kass and Wasserman (1995), Raftery (1995)). The UIP is a normal prior with a mean centered at the maximum likelihood estimate and a variance equal to the inverse of the average information contained in one



observation.

We construct a prior of the form

$$pr(\beta, \theta, \omega_{11.2}, \omega_{22}, \phi) = pr(\beta, \omega_{11.2}, \phi | \theta, \omega_{22}) pr(\theta, \omega_{22}),$$

and specify a standard UIP on  $pr(\theta, \omega_{22})$ , based on the likelihood in Equation (8). In the second stage, we construct a prior that is conditional on the first stage,  $pr(\beta, \omega_{11.2}, \phi | \theta, \omega_{22})$ , and specify a UIP for given values of  $(\theta, \omega_{22})$ , based on the likelihood in Equation (9). By Theorem 1, the resulting two-stage UIP centers  $\beta$  at  $\hat{\beta}^{2SLS}$  when  $\theta$  is set to its mode,  $\hat{\theta}$ .

A prior over parameters of restricted models,  $(M_i, L_j)$ , is specified in essentially the same manner

$$pr\left(\beta^{(j)}, \theta^{(i)}, \omega_{11.2}^{(i,j)}, \omega_{22}^{(i)}, \phi^{(i,j)}\right) = pr\left(\beta^{(j)}, \omega_{11.2}^{(i,j)}, \phi^{(i,j)} | \theta^{(i)}, \omega_{22}^{(i)}\right) pr\left(\theta^{(i)}, \omega_{22}^{(i)}\right),$$

where a UIP is again specified on  $pr(\theta^{(i)}, \omega_{22}^{(i)})$ . Then, conditional on a value of  $\theta^{(i)}$ , we specify a UIP on  $\beta^{(j)}$ .

By imposing the conditional independence assumption on the pair of models  $(M_i, L_j)$ , we can set  $\hat{\Pi}^{(i,j)} = 0$ . Therefore, when  $\theta^{(i)} = \hat{\theta}^{(i)}$ , the prior distribution for  $\beta^{(j)}$  is centered about  $\hat{\beta}^{(i,j)}$ , and the two-stage UIP retains all properties of the Kleibergen and Zivot (2003) prior. That is, the prior features a built-in dependence on  $\theta$ , while its posterior mimics the 2SLS estimator.

## 2.4 Statistical Foundations of 2SBMA

2SBMA combines the 2SLS methodology discussed above with the standard BMA methodology reviewed below. 2SBMA processes the data much like a two-stage estimator while addressing model uncertainty in both stages. The first stage is a simple application of BMA to identify effective instruments. It is helpful to review the properties of BMA that are implied in stage 1.

Let  $\Delta$  be a quantity of interest. In BMA, the posterior distribution of  $\Delta$  given the data,  $D$ , is given by the weighted average of the predictive distribution under each model, weighted by the corresponding posterior probabilities,

$$pr(\Delta | D) = \sum_{i=1}^I pr(\Delta | M_i, D) pr(M_i | D),$$

where  $pr(\Delta | M_i, D)$  is the predictive distribution given model  $M_i$  and  $pr(M_i | D)$  is the posterior model probability of model  $M_i$ . The posterior model probability  $\pi_i$ , for each first stage

model  $M_i$  is given by

$$\begin{aligned}\pi_i &= pr(M_i|D) \\ &\propto pr(D|M_i)pr(M_i),\end{aligned}$$

where

$$pr(D|M_i) = \int pr(D|\theta^{(i)}, M_i)pr(\theta^{(i)}|M_i)d\theta^{(i)}$$

is the integrated likelihood of model  $M_i$  with parameters  $\theta^{(i)}$ . The prior densities for parameters and models are  $pr(\theta^{(i)}|M_i)$  and  $pr(M_i)$ , respectively.

Under BMA, the posterior mean of  $\theta$  is the sum of the posterior means of each model in the collection  $\mathcal{M}$ , weighted by their posterior probabilities

$$\hat{\theta}^{BMA} = \sum_{i=1}^I \pi_i \hat{\theta}^{(i)}.$$

Similarly, the posterior variance is

$$\sum_{i=1}^I \pi_i \hat{\sigma}_i^2 + \sum_{i=1}^I \pi_i \left( \hat{\theta}^{(i)} - \hat{\theta}^{BMA} \right)^2.$$

The posterior variance highlights how BMA methodology accounts for model uncertainty. The first term is the weighted variance for each model,  $\hat{\sigma}_i^2 = Var(\hat{\theta}^{(i)}|M_i, D)$ , averaged over all relevant models, and the second term indicates how stable the estimates are across models. The more the estimates differ between models, the greater is the posterior variance.

The posterior distribution for a parameter is a mixture of a regular posterior distribution and a point mass at zero, which represents the probability that the parameter equals zero. The sum of the posterior probabilities of the models that contain the variable is called the inclusion probability. For instance, for instrument  $Z_k$  we may write,

$$\mu^{BMA}(\theta_{Z_k}) = pr(\hat{\theta}_{Z_k} \neq 0|D) = \sum_{i \in \mathcal{M}_k} \pi_i,$$

where  $\mathcal{M}_k$  is a collection of indices for which  $i \in \mathcal{M}_k$  implies model  $M_i$  does not restrict the parameter  $\theta_{Z_k}$  to zero. Standard rules of thumb for interpreting  $\mu^{BMA}$  have been provided by Kass and Raftery (1995), modifying an earlier proposal of Jeffreys (1961). They suggested the following effect thresholds: <50%: evidence against the effect, 50-75%: weak evidence for the effect, 75-95%: positive evidence, 95-99%: strong evidence, and >99%: very strong evidence<sup>4</sup>.

---

<sup>4</sup>From a decision theory perspective, one can imagine cases where a policy-maker might be interested in a variable even when its posterior inclusion probability is below 50% ; see Brock and Durlauf (2001) and Brock et al. (2003)

In the case of 2SLS estimation in the presence of model uncertainty, the BMA framework must be extended to account for model uncertainty at both stages. For models  $M_i$  and  $L_j$ , we are now interested in

$$pr(M_i, L_j|D) = pr(L_j|M_i, D)pr(M_i|D). \quad (12)$$

The decomposition in Equation (12) indicates the dependence that the probability of model  $L_j$  has on the particular choice of model  $M_i$ , as each model  $M_i$  yields a slightly different distribution for the fitted value of  $W$ . Furthermore, for a given value of  $\theta^{(i)}$ , the quantity  $pr(L_j|\theta^{(i)}, M_i, D)$  is itself a regression problem with a particular UIP.

Denoting by  $\nu_{i,j}$  the probability of model  $L_j$  given model  $M_i$  yields the 2SBMA estimator

$$\hat{\beta}^{2SBMA} = \sum_{i=1}^I \sum_{j=1}^J \pi_i \nu_{i,j} \hat{\beta}^{(i,j)}. \quad (13)$$

Equation (13) shows that the 2SBMA estimate is formed as the average of each constrained 2SLS estimate that results from the combination of model  $M_i$  in the first stage, and model  $L_j$  in the second stage, weighted by both the first and second stage model probabilities. To calculate the posterior variance of  $\hat{\beta}^{2SBMA}$  we have the following result.

**Theorem 3.** *Let  $\hat{\beta}^{(i,\cdot)} = \sum_{j=1}^J \nu_{i,j} \hat{\beta}^{(i,j)}$  be the model averaged estimate of  $\beta$  for a given first stage model  $M_i$ . Then the variance of the estimate  $\hat{\beta}^{2SBMA}$  is*

$$\sigma_{2SBMA}^2(\beta) = \sum_{i=1}^I \pi_i Var(\beta|M_i) + \sum_{i=1}^I \pi_i \left( \hat{\beta}^{(i,\cdot)} - \hat{\beta}^{2SBMA} \right)^2, \quad (14)$$

where

$$Var(\beta|M_i) = \sum_{j=1}^J \nu_{i,j} Var(\beta|M_i, L_j) + \sum_{j=1}^J \nu_{i,j} \left( \hat{\beta}^{(i,j)} - \hat{\beta}^{(i,\cdot)} \right)^2,$$

is the BMA variance associated with second stage estimates for a fixed first stage model.

*Proof* See Appendix.

Theorem 3 shows that the variance of 2SBMA estimates can be decomposed into two parts that yield interpretations similar to standard BMA variances. The first term is the average of BMA variances associated with the models in the first stage, and the second term represents the variation of a given first stage model's BMA estimates relative to the overall 2SBMA estimate.

The posterior distribution of  $\hat{\beta}^{2SBMA}$  is again a mixture of a regular posterior distribution and a point mass at zero, which represents the probability that the parameter equals zero.

The sum of these posterior probabilities that contain the variable is then the inclusion probability at the second stage. For instance, for the variable  $X_l$  we have

$$\mu^{2SBMA}(\beta_{X_l}) = pr(\hat{\beta}_{X_l}^{2SBMA} \neq 0|D) = \sum_{i=1}^I \sum_{j \in \mathcal{L}_l} \pi_i \nu_{i,j},$$

where  $\mathcal{L}_l$  is the subset of  $\mathcal{L}$  for which the coefficient  $\beta_{X_l}$  is not constrained to zero. We continue to follow the standard rules of thumb for interpreting effect thresholds in the second stage, as suggested by Jeffreys (1961).

A desirable feature of the 2SBMA estimator is that asymptotically it will resemble the 2SLS estimator. This is due to the fact that any proposed instrument or covariate that has a zero coefficient in either the first or second stage will be dropped in the limit by the model selection procedure. This yields the following result.

**Theorem 4.** *The 2SBMA estimate is consistent under identification in the full model.  $\hat{\beta}^{2SBMA} \rightarrow_p \beta$  when  $\hat{\beta}^{2SLS} \rightarrow_p \beta$ .*

*Proof* See Appendix. This result mainly implies that one does not sacrifice consistency when employing 2SBMA rather than 2SLS.

We further note that coefficients derived from instrumental variables regressions become increasingly biased as the number of instruments increases (see, e.g., Hall (2005)). The bias is aggravated when the proposed instruments have little explanatory power for the endogenous variable (Bound et al. (1995)). While it is difficult to assess in general, the 2SBMA methodology has the potential to limit this bias, by dropping unnecessary terms. The simulation study in Section 4 shows an instance in which this occurs.

### 3 Bayesian Tests of Identification Restrictions

Various tests have been developed to examine the instruments' conditional independence (on  $Y$ ) and explanatory power (for  $W$ ). Below we develop Bayesian equivalents of the Sargan (1958) test of over-identification and the Cragg and Donald (1993) test of under-identification that Stock and Yogo (2005) used to propose a weak instruments test. We then show how model averaged versions of these tests can be used in the 2SBMA framework to verify model assumptions, and we discuss the properties of such techniques.

To develop Bayesian versions of the Sargan test and the Cragg and Donald test we rely on the notion of posterior predictive  $p$ -values. Rubin (1984) argued that posterior predictive

$p$ -values are useful tools in applied Bayesian statistics to verify model assumptions. Given a hypothesis,  $H$ , about the model, the posterior predictive  $p$ -value given data is calculated as  $pr(H|D)$ . It quantifies the degree to which the hypothesis is supported by the posterior distribution of the model parameters given the data  $D$ .

In the case of 2SBMA, when the hypothesis relates to the model identification, we consider the following decomposition for model averaged posterior  $p$ -values

$$pr(H|D) = \sum_{i=1}^I \sum_{j=1}^J \pi_i \nu_{i,j} pr(H|M_i, L_j, D). \quad (15)$$

The decomposition in Equation (15) outlines the motivation for model averaged posterior  $p$ -values. Since the hypotheses of primary interest are related to the identification of the 2SBMA estimate, it is natural to ask whether each sub-model which is used to form the 2SBMA estimate is appropriately identified. Naturally, different specifications yield varying degrees of confidence in each specification's identification.

It is then appropriate to weight posterior predictive  $p$ -values for each model by the extent to which each model contributes to the 2SBMA estimate. Suppose a given model receives the majority of the posterior model probability but tests indicate the specification is under-identified. Even if all other models used to form the 2SBMA estimate are identified, the fact that posterior means are largely formed from the underidentified model must be accounted for in the model averaging context. This logic motivates the decomposition in Equation (15), which gives rise to the model-averaged posterior predictive  $p$ -value.

### 3.1 A Bayesian Test of the Over-Identification Restrictions

By selecting first-stage models based on the integrated likelihood, we are explicitly choosing combinations of instruments based on model fit. Hall et al. (1996) note that any procedure that chooses the first stage based on goodness of fit measures may risk including endogenous variables. Therefore, it is important to develop a diagnostic that helps assess whether the model selection procedure has produced such an over-identification failure.

We base our test of over-identification on the Sargan (1958) test. Let  $\hat{\eta}^{(i,j)}$  be the residuals from the combination of models  $M_i$  and  $L_j$ , and let  $p_{i,j}$  be the total number of  $X$  and  $Z$  included in this combination. The Sargan  $p$ -value  $S^*$  is calculated as  $S^* = pr(nR_{**}^2 > \chi_{p_X+p_Z-1}^2)$  where  $R_{**}^2$  is the  $R^2$  associated with the regression of  $\hat{\eta}^{2SLS}$  on all  $X$  and  $Z$  variables.

Just as in the classical Sargan test, we consider the regression of  $\hat{\eta}^{(i,j)}$  on the subset of the variables  $X$  and  $Z$  that belong to either  $M_i$  or  $L_j$ , and determine  $R_{ij}^2$ , which is the  $R^2$

associated with this regression. Letting  $S^{(i,j)} = p(\chi_{p_{i,j}-1}^2 > nR_{ij}^2)$ , we define the Bayesian Sargan  $p$ -value to be

$$S^{2SBMA} = \sum_{i=1}^I \sum_{j=1}^J \pi_i \nu_{i,j} S^{(i,j)}.$$

$S^{2SBMA}$  is therefore the average of the Sargan  $p$ -values derived from the specific models  $M_i$  and  $L_j$ , weighted by their respective posterior probabilities.

The benefit of the Bayesian 2SBMA Sargan test is that the parsimony of 2SBMA effectively mitigates the reduction in power that the traditional Sargan test experiences as the dimensions of the  $X$  or  $Z$  variables grow. This increase in power can be big, as shown in the simulation study below.

### 3.2 Bayesian Tests of Under-Identification and Weak Instruments

While it is crucial to verify that proposed instruments do not violate the conditional independence assumption, it is also important to test that they have explanatory power for the endogenous  $W$ . When  $W$  is univariate, this may be done by considering an  $F$  test based on the first stage. However, when  $p_W > 1$ , Cragg and Donald (1993) derived an equivalent test for this. Here we derive a Bayesian analog of this test.

Consider fixed first and second stage models,  $M_i$  and  $L_j$  respectively. Let  $Z_{ij}$  be the instruments used in this combination, namely all those variables in  $Z$  used in  $M_i$  and those variables  $X$  used in  $M_i$  but excluded from  $L_j$ . Let  $X_j$  be those  $X$  contained in  $L_j$ , and let  $V_{ij}$  be the matrix of all  $X$  and  $Z$  variables included in either  $M_i$  or  $L_j$ . Define  $P_{V_{ij}} \equiv V_{ij}(V_{ij}'V_{ij})^{-1}V_{ij}'$  and  $M_{V_{ij}} \equiv I_n - P_{V_{ij}}$  where  $I_n$  is the  $n \times n$  identity matrix, and similarly define  $P_{X_j} \equiv X_j(X_j'X_j)^{-1}X_j'$  and  $M_{X_j} = I_n - P_{X_j}$ , and finally define  $G_{ij} \equiv \hat{\Sigma}_{ij}^{-1/2} \Theta_{ij} \hat{\Sigma}_{ij}^{-1/2}$  where  $\hat{\Sigma}_{ij} = W'M_{V_{ij}}W$  and  $\Theta_{ij} = (M_{X_j}W)'M_{X_j}Z_{ij}((M_{X_j}Z_{ij})'M_{X_j}Z_{ij})^{-1}(M_{X_j}Z_{ij})'M_{X_j}W$ . The Cragg and Donald statistic under model  $M_i$  and  $L_j$  can then be derived as the minimum eigenvalue of  $G_{ij}$ ,  $g_{ij} = \min \text{eigen} G_{ij}$ .

In practice, the statistic  $g_{ij}$  is used in two ways. Asymptotically, under the null hypothesis of under-identification,  $ng_{ij} \sim \chi_{p_{Z_{ij}}-1}^2$ , and this reference distribution is used to derive a posterior predictive  $p$ -value. Here we propose a Bayesian model-averaged version of this posterior  $p$ -value by considering

$$CD = \sum_{i=1}^I \sum_{j=1}^J \pi_i \nu_{i,j} Pr(\chi_{p_{Z_{ij}}-1}^2 > ng_{ij}).$$

A second use of  $g_{ij}$  was suggested by Stock and Yogo (2005), but their test statistic provides only critical values, not the necessary  $p$ -values that can be averaged when models

have different numbers of instruments. The Bayesian approach can also be used to assess an apparent weakness of an instrument. The approach is simple and direct: it simply requires an examination of the instruments' inclusion probabilities.

## 4 Simulation Study

We conduct a simulation study to quantify the estimation properties of 2SBMA and the behavior of the Bayesian tests of the identification restrictions.

In the following we consider a framework in which there are ten variables in  $Z$ , 15 in  $X$  and  $W$  is univariate. Our construction of the variables in  $X$  and  $Z$  is similar to the simulation study in Fernandez et al. (2001), which was based in turn on Raftery et al. (1997).

For constructing  $X$  we let  $(X_1 \dots X_{10})$  be an  $n \times 10$  matrix of independent draws from a  $N(0, 1)$  distribution. We set

$$(X_{11} \dots X_{15}) = (X_1 \dots X_5)(.3 \ .5 \ .7 \ .9 \ 1.1)'(1 \dots 1) + E, \quad (16)$$

where  $E$  is an  $n \times 5$  matrix of draws from the  $N(0, 1)$  distribution. Note that Equation (16) induces a correlation between the first five regressors and the last five regressors. It takes the form of small to moderate correlations between the first five variables in  $X$  and the last five where the theoretical correlation coefficients increase from 0.153 to 0.561.

The matrix  $Z$  is sampled in a similar manner.  $(Z_1 \dots Z_5)$  is an  $n \times 5$  matrix of standard normal variates and we then set

$$(Z_6 \dots Z_{10}) = (Z_1 \dots Z_5)(.3 \ .5 \ .7 \ .9 \ 1.1)'(1 \dots 1) + F,$$

where  $F$  is an  $n \times 5$  matrix of independent draws from the  $N(0, 1)$  distribution. Finally, we construct the model

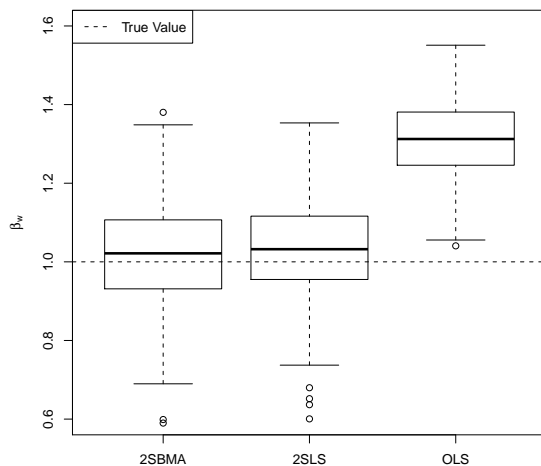
$$Y = W + 1.8X_1 + 1.5X_2 + X_{11} - 1.5X_{12} + 2\eta, \quad (17)$$

$$W = 1.1X_1 - .5X_3 + .75X_{12} + .75Z_1 - 2Z_8 + 3\epsilon. \quad (18)$$

Thus, we consider a situation in which four covariates along with  $W$  have explanatory power for  $Y$ . Furthermore, two variables in  $Z$  serve as instruments and two of the variables in  $X$  have explanatory power on both  $Y$  and  $W$ . Finally, one variable in  $X$  would be more properly classified as an instrument, as it has explanatory power on  $W$  but not on  $Y$ .

We introduce endogeneity by drawing  $\epsilon$  from a  $N(0, 1)$  distribution and setting  $\eta = \epsilon + \xi$ , with  $\xi$  drawn from a  $N(0, 1)$  distribution as well. We then consider two scenarios. The first

Figure 1: Finite Sample Bias under 2SBMA, 2SLS and OLS. Distribution of the estimate of the coefficient  $\beta_W$  across replications using 2SBMA, 2SLS and OLS, when  $\beta_W = 1$ . The average bias of  $\hat{\beta}_W$  across 500 replications was .019, .035 and .310 for 2SBMA, 2SLS and OLS respectively. The average mean squared error for estimating the entire vector  $\beta$  was .213, .395 and .840 for 2SBMA, 2SLS and OLS respectively.



scenario is one in which the IV model is correctly specified, i.e. the  $Z$  covariates have no effect on  $Y$ . In the second scenario we consider a misspecified model in which  $\eta = Z_1 + \epsilon + \xi$ , so that the instrument condition fails. In each scenario we simulate datasets of 100 observations and consider 500 replicates. The simulation study is structured to roughly resemble the data set we will be examining below.

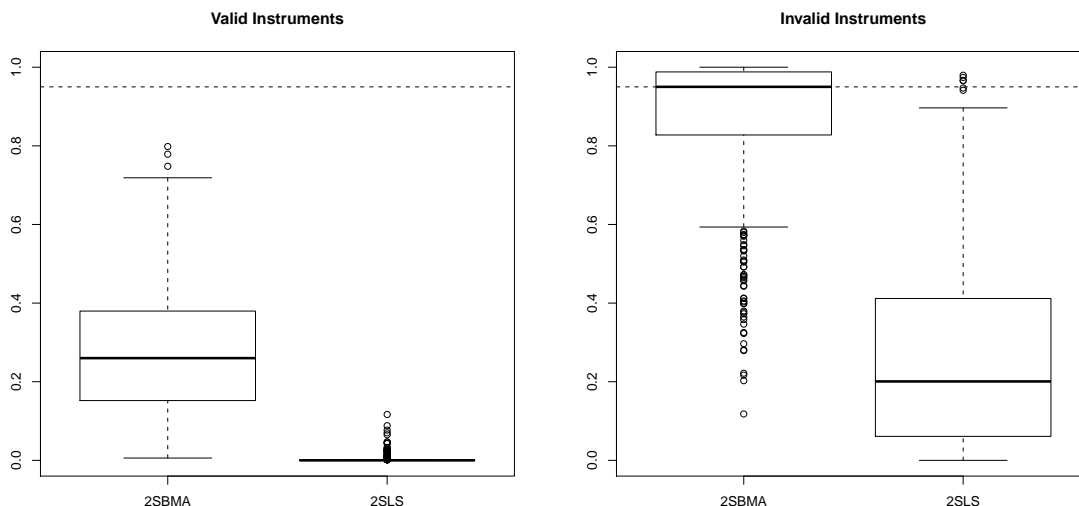
Figure 1 shows the distribution of the estimate of  $\beta_W$  across replications using 2SBMA, 2SLS and OLS. We see that the OLS estimates are centered about a value of 1.3. Indeed, in this case the OLS estimate will asymptotically approach this value. Both 2SBMA and 2SLS rectify this bias and are more closely centered about the true value of 1. However, the average bias and average mean squared error of 2SBMA are about 45% lower than those of 2SLS, so 2SBMA performs much better than OLS or 2SLS.

The first panel in Figure 2 shows the distribution of the  $p$ -values returned from the Bayesian Sargan test as well as the traditional Sargan test. We see that the  $p$ -values from the Bayesian Sargan test are much higher. However, these scores are still sufficiently low that the exogeneity assumption is unlikely to be incorrectly rejected.

The second panel in Figure 2 shows the resulting Bayesian Sargan and classical Sargan  $p$ -values for the case of a misspecified exogeneity assumption. In the case of valid instru-



Figure 2: Distribution of  $p$ -values returned by the Bayesian Sargan test and the Sargan test across replications when the IV assumptions hold (left) and when they do not (right). In all cases, a nominal level of 5% was used, implying that the assumption of valid instruments would be rejected when the  $p$ -value was above .95, shown by the dotted line in each plot. In the case of valid instruments, the size of both tests was 0 (no false rejections). However, in the case of invalid instruments the power of the Bayesian Sargan test was 50% (half the cases were correctly rejected), whereas it was 0.8% using the traditional Sargan test (4 cases out of 500 were correctly rejected).



ments, the size of both tests was 0. However, in the case of invalid instruments the power of the Bayesian Sargan test was 50%, whereas it was 0.8% using the traditional Sargan test, based on  $\alpha = .05$ . We see that the Bayesian Sargan test is more precise in discerning the failure of the exogeneity assumption and is far more likely to reject the hypothesis that the IV assumptions are valid than the classical Sargan test. For the case of valid instruments, both the classical and Bayesian Cragg-Donald tests correctly rejected the null hypothesis of no identification for all repetitions.

The previous figures show that 2SBMA returns appropriate coefficient estimates and yields dramatically improved power at detecting assumption violations over traditional methods. Table 1 shows that the technique also uncovers the pattern of interaction in both stages of the estimation. When the model is correctly specified, Table 1 shows the median inclusion probability for each variable across the 500 replications in both stages as well as the interquartile range of the inclusion probabilities. We see that in the first stage the two variables in  $Z$  as well as the three variables in  $X$  are given high inclusion probabilities, while the remaining variables are generally excluded. This remains true in the second stage as well, where  $W$  and the four covariates in  $X$  that have explanatory power are given large inclusion

probabilities and all others are given negligible inclusion probabilities.

## 5 Instrument and Determinant Uncertainty in Development Accounting

We now apply 2SBMA to one of the most prominent approaches to institutions in the development accounting literature.<sup>5</sup>

Rodrik et al. (2004) (RST) conducted an explicit “horse race” of theories that pertain not only to development determinants (geography, integration and institutions), but also to a wide range of theories that suggest alternative instruments. With less than 100 observations, the RST sample is a standard size of datasets in development accounting. The model uncertainty RST highlight among development determinants and instruments is a defining feature of the literature.

RST explored over 25 different robustness specifications with alternative candidate regressors that are suggested by a comprehensive set of theories that span the literature. Their results are so uniform and decisive across all specifications that their claim to have resolved model uncertainty is well captured by their title: “Institutions rule: the primacy of institutions over geography and integration in economic development.” At best RST find geography may have weak direct effects while Integration is found to be “always insignificant, and often enters the income equation with the ‘wrong’ sign.” Their results are, however, in stark contrast to previous evidence of Trade/Integration, and Geography effects on development (e.g., Hall and Jones (1999), Sachs (2003)); but these papers were discounted because they did not provide RST’s level of robustness analysis.

Using RST’s own data set, we reexamine their robustness specifications using 2SBMA to

---

<sup>5</sup>Our computations use the `bicreg` function from the `BMA` R package (Raftery et al. (2005)), since the Bayesian Information Criteria (BIC) closely approximates the posterior model probability under the UIP (Kass and Wasserman (1995), Raftery (1995)). Since the two-stage UIP is proper in all parameters, and being a nested set of UIPs, it allows for BIC approximations of the integrated likelihood in both the first and second stages. For a given value of  $\theta^{(i)}$ , the quantity  $pr(L_j|\theta^{(i)}, M_i, D)$  is given by a regression model with a particular UIP and therefore well approximated by the BIC. In addition we can approximate  $pr(L_j|M_i, D)$  with  $pr(L_j|\hat{\theta}^{(i)}, D)$  and calculate BIC relative to this  $\hat{\theta}^{(i)}$ . The conditional independence assumptions of  $(M_i, L_j)$  enable us to set  $\hat{\Pi}^{(i,j)} = 0$ . Thus in our BIC approximation the likelihood  $pr(\beta|L_j, \hat{\theta}^{(i)}, D)$  is maximized at  $\hat{\beta}^{(i,j)}$  defined in Equation (7). Note that BIC can be calculated from the associated  $R^2$  of the regression. By calculating the model probability  $\nu_{i,j}$  conditional on the fitted value  $V^{(i)}\hat{\theta}^{(i)}$  of  $W$ , we note that the resulting  $R^2$  is equivalent to using the “Generalized  $R^2$ ” suggested by Pesaran and Smith (1994) for scoring the second stage model  $L_j$ .

address the model and instrument uncertainty that is highlighted so forcefully in their paper. The 2SBMA first and second stages are reported in Tables 2 and 3, respectively. In terms of development determinants, RST assume that Geography is exogenous, so the upper panel in Table 2 represents the first stage for the endogenous institutions proxy (Rule of Law) and the lower panel in the same table is the first stage for the endogenous Integration variable.

Given the 2SBMA methodology, it would be enough to present only the 2SBMA results that explore the entire model space spanned by RST’s development determinants and the associated instruments. This specification is provided in Column 3. We first provide, however, two intermediate stages, where Column 1 represents RST’s “core specification” (RST’s Table 2) and Column 2 is the first set of RST’s robustness exercises. The three-step approach highlights the sensitivity of the core specification to the introduction of additional covariates that are associated with different theories (in Column 2), as well as sensitivity to different variables associated with the different theories (in Column 3).

Column 1 in Table 3 provides the second stage of RST’s preferred core specification (RST’s Table 2). Both RST and 2SBMA find that only Rule of Law shows an effect and the conditional posterior mean is nearly identical to RST’s 2SLS estimate. In this specification, the 2SBMA result confirms RST’s central finding that “the preferred specification accounts for about half of the variance in incomes across the sample, with institutional quality (instrumented by settler mortality) doing most of the work.” The generalized  $R^2$  for the best 2SBMA model is 0.53 versus 0.55 in RST’s 2SLS approach.

Column 1 in Table 2 reports the two 2SBMA first stages for RST’s core specification. They broadly confirm the 2SLS results in RST although 2SBMA suggests slightly more parsimonious models. 2SBMA suggests three strong instruments for Rule of Law (Settler Mortality, Latitude, and the Fraction Speaking English), while RST found significant coefficients for all five instruments across their various 2SLS exercises. This generates a slightly higher  $R^2$  for RST’s preferred 2SLS specification (0.55) as compared to the best model in 2SBMA (0.49). For Integration, the 2SBMA first stage suggests only two strong instruments (Implied Trade Shares and Settler Mortality) while 2SLS produces a statistically significant coefficient for an additional instrument (Fraction Speaking English). Nevertheless the  $R^2$  of the 2SBMA best model and of the 2SLS first stage are identical (0.58).

RST find that core specifications with more than one instrument fail to pass the Sargan test. This finding is confirmed by the Bayesian Sargan test in Column 1 of Table 3, which presents a similar  $p$ -value to that found by RST. One interpretation is that the Sargan test undermines alternative determinant and instrument strategies as suggested by RST. One could also argue, however, that RST’s specification does not contain the appropriate set of instruments. We examine this issue further below but note that already at this stage,

under-identification (as measured by the Bayesian Cragg-Donald  $p$ -value) is easily rejected by 2SBMA (not reported in RST). Weak instruments are not of concern in this data set.

The 2SLS results in RST’s core specification and the 2SBMA results in Column 1 are nearly identical because the core specification includes minimal model uncertainty at the determinant level and only a fraction of the standard instruments suggested by the development literature. Column 2 adds regressors suggested by alternative theories that pertain to Legal Origins and Religion, as well as regional dummies, while Column 3 represents the most comprehensive set of regressors that adds standard covariates related to alternative Geography theories (most notably Temperature, Malaria) as well as alternative Integration measures (such as Sea Access). As we allow for additional theories and the associated regressors, the 2SBMA results start to diverge from the results in RST’s individual 2SLS robustness regressions. In other words, the disparities across results become more pronounced as model uncertainty increases.

The 2SBMA results that use the most comprehensive set of instruments and development determinants (Column 3 in Tables 2 and 3), cast doubt on the primacy of institutions result. Instead 2SBMA finds that the “horse race” ends in a statistical three-way tie. Geography (as measured by Tropics), Institutions and Integration are shown to be highly effective development determinants. This result is particularly surprising in light of the fact that Geography is only occasionally weakly significant in RST, and Integration is never significant and often of the wrong sign. In 2SBMA all three effects are strong and estimated with the correct sign. Once model uncertainty is comprehensively addressed at both the development determinant, as well as the instrument level, the results thus support the contentions of Sachs (2003) and Alcalá and Ciccone (2004), who report strong effects of Geography and Integration.

The divergence of 2SLS and 2SBMA results can already be observed in the first stages.<sup>6</sup> Most importantly, the Implied Trade Share no longer receives support as a strong instrument for Integration. It is most strongly instrumented by EuroFrac in combination with the covariates PopGrowth, Oil, SeaAccess, Malaria94, EuroFrac, Tropics, Latitude, FrostArea, and PolicyOpenness. In contrast to the findings of RST, religion variables also play an important part in the first stage regression. In particular, Catholic has nearly a 90 percent inclusion probability in the first stage for Rule of Law and above 50 percent in the first stage for Integration. Similarly, the power of Settler Mortality as an instrument for Institutions is dominated by regressors such as EuroFrac and Temperature variables in both first stages.

Note that the increase in the model space of development determinants and instruments going from the specification in Column 1 to Column 3 dramatically improves the fit of the 2SBMA first stage. For Column 3, the best models in both 2SBMA first stages report an

---

<sup>6</sup>RST report neither first stages nor tests of instrument restrictions beyond the core specification.

$R^2$  that is at least 40 percent greater than those found in the core specification.

A similar improvement in model fit can be observed in the second stage when considering the generalized  $R^2$  (Pesaran and Smith, 1994) of the best model returned by 2SBMA. In fact, none of the top 100 models' generalized  $R^2$  falls below .82, which greatly exceeds any model presented by RST (whose highest generalized  $R^2$  is .73). 2SBMA has therefore uncovered combinations of instruments and development determinants that fit the data substantially better. This is then the source of the difference in the 2SLS and 2SBMA results in both the first and second stages. The Bayesian Sargan and Bayesian Cragg and Donald tests clearly show, respectively, that over-identification is easily rejected with the improved set of instruments and that under-identification remains of little concern.

## 6 Conclusion

We have developed a methodology to address model uncertainty in the presence of endogeneity and explored its properties as a valid IV estimator. The method is based on Bayesian Model Averaging (BMA), which has already been extensively used in economic applications, particularly in modeling economic growth and development. 2SBMA is shown to be a consistent methodology that merges the BMA and 2SLS procedures in a natural and straightforward manner.

Resolving endogeneity via IV estimation may take a number of forms, such as LIML estimators or GMM estimation. We have chosen to focus on extending the 2SLS estimator for two reasons. Primarily, we feel that 2SLS is the IV technique most familiar to the applied economist, and the 2SBMA method is a manner of automatically performing the assessment of model uncertainty that frequently occurs “off-site” in economic modeling exercises. Secondly, the 2SBMA estimator builds naturally on the BMA methodology for regression variable uncertainty that is already well-established in the statistical and econometric literatures.

Alternative approaches, for instance addressing model uncertainty in LIML models, would have required reconsideration of many questions raised in model uncertainty and averaging exercises. For instance, prior elicitation, model search, the potential for approximations of the integrated likelihood, and posterior parameter estimation would have to be considered. While such exercises are certainly worth pursuing, by basing our framework on 2SLS we were able to refer to the resolution of these issues in regression variable uncertainty problems, an association that would be impossible in a framework such as LIML.

In order to extend the 2SLS paradigm to a Bayesian context, we introduced a Two-Stage UIP, which built on the development of the Bayesian two-stage prior of Kleibergen and Zivot (2003). Lancaster (2004) outlines methods for Bayesian IV estimation that are not designed to mimic 2SLS, but use proper priors over model parameters. Posterior parameter distributions are then determined through Markov Chain Monte Carlo (MCMC) methods. A potential alternative to the 2SBMA methodology would be to augment the model described by Lancaster (2004) to incorporate model uncertainty. This could involve use of the Reversible Jump MCMC algorithm (Green, 1995). We have avoided this approach, as trans-dimensional Markov Chains are often difficult to implement in high dimensional spaces and may have convergence issues. By using 2SBMA methodology, in contrast, we were able to leverage existing widely-used software for regression variable uncertainty.

A key assumption in our development was that the errors were normally distributed. Such an assumption is not necessary in order for standard 2SLS to function properly as a

consistent estimator, but it was important in specifying our two-stage UIP and is similar to most Bayesian developments of the endogenous variable model (Kleibergen and Zivot, 2003; Lancaster, 2004; Rossi et al., 2006). Relaxing the normality assumption therefore constitutes an important further step in the development of this methodology. One potential avenue for performing this relaxation is to consider extending the Dirichlet process approach employed by Conley et al. (2008).

In the context of GMM estimation, Andrews (1999) proposes a methodology for simultaneous model and moment selection. This methodology resembles that of the BIC calculation in that it has two parts: one pertaining to the estimator from this model/moment combination, and a penalty term. However, in place of the maximized log-likelihood that is included in BIC, Andrews (1999) uses the J-statistic first described by Hansen (1982), which is a generalization of the Sargan (1958) statistic for GMM models. Andrews (1999) then shows that this model selection criterion is a consistent technique when the BIC penalty is used, but is inconsistent using an AIC penalty. Andrews and Lu (2001) continue this development, applying GMM estimation to dynamic panel data models using these criteria. Donald and Newey (2001) propose an alternative methodology for selecting instruments in IV models which they base on minimizing the mean-squared error implied by the IV estimates. They show that this leads to diminished estimation bias similar to that displayed in the simulation study above. The authors note that their technique for selecting instruments differs from that proposed by Andrews (1999) in that their objective is to choose instruments from a small set known to be valid, while the procedure of Andrews (1999) is best suited for finding the largest set of valid instruments. Kapetanios (2006) extends this selection procedure by developing a search method based on genetic algorithms that finds the instrument set which minimizes the mean-squared error quickly, even in the presence of a large instrument set.

When there are many weak instruments asymptotic properties of IV and GMM estimators may not behave properly (Stock and Wright, 2000; Chao and Swanson, 2005). This has led to a number of techniques to handle many weak instruments (Chamberlain and Imbens, 2004; Flores-Lagunes, 2007; Okui, 2009). The 2SBMA methodology shares characteristics of this methodology, but differs in its focus and use. The frameworks considered in the weak instrument literature are typically motivated by datasets such as that considered by Angrist and Krueger (1991), who study the effect of compulsory schooling on academic performance. This dataset is quite large, with several hundred thousand observations. Furthermore, there are over 500 proposed instruments in this model, none of which is felt to have strong explanatory power. The methods above perform well in analyzing the Angrist and Krueger (1991) data, in that they mitigate the effect the extreme number of instruments has on estimation.

However, 2SBMA is developed with an alternative focus. We have created an estimation

procedure that performs well in datasets of small sample sizes and with considerably fewer proposed instruments than encountered in Angrist and Krueger (1991). This is due to our focus on situations such as development accounting and the modeling of growth determinants. Furthermore, as highlighted in the inclusion probabilities shown in Table 2, we are considering a situation in which we believe there exist instruments with strong explanatory power for the endogenous variable. Thus our primary goal is to characterize instrument and covariate relevance through posterior inclusion probabilities.

The 2SBMA methodology also offers a means of model averaging over a large class of models, while the methodology derived to handle many weak instruments typically focuses on finding a single coefficient estimate. Recent work has developed similar model averaging methodology in alternative classes of IV and GMM estimation problems (Chen et al., 2009; Morales-Benito, 2009), typically in the context of panel data models. These developments share many of the aspects of the 2SBMA estimator. However, because our focus is not on panel data estimation, we are able to derive model probabilities in a more straightforward manner than others. A particularly nice feature of 2SBMA is that it separates the problem into two stages, which can then be addressed directly using standard techniques for regression variable uncertainty.

Instrumental variable estimation of any kind requires a number of assumptions that relate to the identification of the implied structural model. To examine the validity of these assumptions in the context of 2SBMA, we propose Bayesian Sargan and Bayesian Cragg and Donald tests of the over- and under-identification restrictions. These tests are based on averaged posterior  $p$ -values within the 2SBMA framework. We show that the power of these tests is substantially greater than those of standard 2SLS tests. In addition, within the context of 2SBMA these tests are less affected by an increase in the number of potential instruments.

We apply 2SBMA to a prominent development accounting approach (Rodrik et al. (2004)), which was itself motivated by the vast model uncertainty associated with alternative theories of development and alternative instruments to control for potential endogeneity. Instead of resolving the model uncertainty in a horse race of alternative regressions, we use the formal 2SBMA approach. We find not only support for institutions, but also substantial support for geographic and trade factors, once model uncertainty in the presence of endogeneity is addressed. The latter two effects had been relegated to second order effects by RST.

The purpose of our paper is to introduce the statistical foundations of 2SBMA methodology and provide applications that highlight the importance of instrument and covariate uncertainty in economics. In the context of the RST example, alternative approaches to de-



development accounting include Mauro (1995), who first suggested ethnolinguistic fragmentation as a fundamental determinant of corruption, and Hall and Jones (1999), who introduced Latitude and Language indicators as instruments to measure Western influence. Acemoglu et al. (2001b; 2001a) suggested population density in 1500 and the type of colonial origin (indicated by settler mortality) as effective instruments, respectively. La Porta et al. (2004) presented yet another "horse race" of theories, in their case juxtaposing judicial independence and constitutional review. In RST the "horse race" is between three possible determinants: Institutions, Integration, and Geography. Geography-based theories of fundamental development determinants have previously been proposed by Bloom and Sachs (1998), Easterly and Levine (2003), and Sachs (2003).

Several modifications of the BMA paradigm have been proposed in the development determinant literature. Brock et al. (2003) and Durlauf et al. (2008) discuss priors on the model space that account for the fact that many variables may be collected to proxy one particular theory, while fewer may be available to proxy an alternative theory. Ley and Steel (2007) and Doppelhofer and Weeks (2009) develop metrics to quantify the degree to which development determinants act "jointly" to affect growth. Determining how these extensions of the BMA paradigm may be taken into account in the 2SBMA framework would help extend the application of 2SBMA to the particular problem of testing growth theory robustness.

The 2SBMA method allows researchers to incorporate concepts of model uncertainty and model averaging into the assessment of a diverse range of economic behavior where observations are subject to endogeneity. However, the current framework does not directly handle such concepts as panel data, mixed effects, random coefficient models, and heteroskedasticity. Future research into these areas will improve the applicability of the BMA framework to economic analysis, in growth economics and beyond.

Table 1: Median and interquartile range (IQR) of variable inclusion probabilities across 500 repetitions. Variables shown in bold are those that are included in either the first or second stage. This table shows that inclusion probabilities closely match the true structure of the system.

Variable	First Stage		Second Stage	
	Median	IQR	Median	IQR
$W$	–	–	<b>1</b>	<b>(1,1)</b>
$X_1$	<b>1</b>	<b>(1,1)</b>	<b>0.761</b>	<b>(0.573,0.999)</b>
$X_2$	0.017	(0,0.064)	<b>0.588</b>	<b>(0.222,0.962)</b>
$X_3$	<b>0.561</b>	<b>(0.167,0.943)</b>	0.065	(0.018,0.064)
$X_4$	0.031	(0,0.081)	0.092	(0.023,0.096)
$X_5$	0.031	(0,0.102)	0.09	(0.024,0.086)
$X_6$	0.019	(0,0.082)	0.091	(0.015,0.094)
$X_7$	0.014	(0,0.067)	0.086	(0.014,0.083)
$X_8$	0.026	(0,0.084)	0.077	(0.018,0.07)
$X_9$	0.018	(0,0.077)	0.077	(0.016,0.088)
$X_{10}$	0.014	(0,0.074)	0.072	(0.014,0.065)
$X_{11}$	0.021	(0,0.088)	<b>0.473</b>	<b>(0.1,0.874)</b>
$X_{12}$	<b>1</b>	<b>(1,1)</b>	<b>0.72</b>	<b>(0.492,0.987)</b>
$X_{13}$	0.016	(0,0.076)	0.094	(0.023,0.08)
$X_{14}$	0.02	(0,0.096)	0.107	(0.027,0.109)
$X_{15}$	0.022	(0,0.096)	0.112	(0.027,0.117)
$Z_1$	<b>0.908</b>	<b>(0.963,1)</b>	–	–
$Z_2$	0.074	(0,0.066)	–	–
$Z_3$	0.079	(0,0.063)	–	–
$Z_4$	0.077	(0,0.068)	–	–
$Z_5$	0.082	(0,0.073)	–	–
$Z_6$	0.079	(0,0.066)	–	–
$Z_7$	0.083	(0,0.066)	–	–
$Z_8$	<b>1</b>	<b>(1,1)</b>	–	–
$Z_9$	0.075	(0,0.063)	–	–
$Z_{10}$	0.1	(0,0.089)	–	–

Table 2: First Stage Results for RST Example

	I RST Table 2 Core Specification			II RST Table 2, 4 I + LegalOrig, Relig, Region			III RST Table 2, 4, 5, 6 II + Alt. Integr./Geo Measures		
	<i>p</i> ≠ 0	Mean	Sd	<i>p</i> ≠ 0	Mean	Sd	<i>p</i> ≠ 0	Mean	Sd
	<i>Stage 1, Dependent Variable: Rule of Law</i>								
SettlerMortality	92.7	-0.21189	0.07113	25.5	-0.10696	0.07218	17.1	-0.02528	0.06317
EuroFrac	16.5	0.23	0.23093	99.9	1.66071	0.31244	100	1.03	0.2994
Catholic				14.9	-0.00478	0.00381	89.9	-0.01405	0.00576
MeanTemp							86.8	-0.05535	0.02794
PopGrowth							72.5	-0.1011	0.08351
SubSaharaAfrica				10.7	-0.34126	0.30796	55.5	-0.2287	0.2442
Muslim				13	-0.00434	0.00393	40.2	-0.00204	0.003
Latitude	89.7	0.02254	0.00792	99.1	0.02919	0.00732	20.5	0.00411	0.00938
LatinAmerica				99.9	-1.00161	0.27639	14.9	-0.1277	0.3428
Area							12.8	.02582	.07443
Oil							8.8	-0.03573	0.1377
FR_Trade Shares	38.5	0.17551	0.09726	99	0.28821	0.08545	8	-0.01918	0.08153
Tropics							7.9	-0.02392	0.1062
EngFrac	98.6	1.07778	0.28172	12.2	0.37306	0.35301	7.5	0.04617	0.201
FrostArea							6.3	0.04695	0.2127
Protestant				10.7	-0.00702	0.00619	3.8	0.00026	0.00159
FrostDays							1.9	0.00043	0.00725
LegalOrigFr				46.7	-0.30591	0.15008	1.8	-0.00387	0.03538
SeaAccess							1.4	0.00211	0.02705
PolicyOpenness							1.1	0.00268	0.03683
EastAsia				91.4	0.72988	0.25051	0	0	0
Malaria94							0	0	0
LegalOrigSocialist				63.9	-0.77728	0.36386	na	na	na
BIC best model	-41.53			-53.81			-51.42		
R2 best model	0.49			0.66			0.75		
	<i>Stage 1, Dependent Variable: Integration</i>								
	<i>p</i> ≠ 0	Mean	Sd	<i>p</i> ≠ 0	Mean	Sd	<i>p</i> ≠ 0	Mean	Sd
FR_Trade Shares	100	0.5985	0.0612	100	0.5769	0.0502	0.7	-0.086	0.1074
LegalOrigSocialist				20.1	-0.2561	0.2001	na	na	na
PopGrowth							100	-0.2735	0.0285
SeaAccess							94.8	-0.3023	0.106
Oil							94.4	0.3445	0.1284
Malaria94							91.8	-0.4383	0.1399
EuroFrac	14.4	-0.1053	0.1389	6.1	0.0563	0.1329	81.2	-0.5145	0.1826
Tropics							73.1	0.4392	0.1921
Latitude	23.5	-0.0065	0.005	4.9	0.0003	0.0039	72.7	-0.0164	0.007
FrostArea							65.3	0.497	0.2019
PolicyOpenness							59.1	0.3468	0.1391
Catholic				7.3	0.001	0.0014	52.5	-0.0036	0.0018
SettlerMortality	84.9	-0.1111	0.0408	9	-0.0349	0.0371	50.2	-0.1077	0.0579
EastAsia				100	0.8236	0.139	28.2	0.2917	0.1663
EngFrac	23.2	0.246	0.1865	83.6	0.382	0.1431	24.7	-0.6486	0.3051
FrostDays							17.1	0.0217	0.0125
LatinAmerica				6.3	-0.0448	0.1147	16.6	-0.4149	0.1899
MeanTemp							13.5	-0.0225	0.0118
SubSaharaAfrica				5.3	-0.033	0.0925	4	-0.1922	0.1627
LegalOrigFr				6	0.0464	0.1039	3.8	-0.095	0.0902
Protestant				11.9	0.0041	0.0035	0.7	0.0024	0.0029
Muslim				5.3	-0.0005	0.0012	0.2	-0.0007	0.0015
Area							0.2	0	0
BIC best model	-61.22			-84.37			-54.57		
R2 best model	0.58			0.71			0.81		

Table 3: Second Stage Results for RST Example

	I RST Table 2 Core Specification			II RST Table 2, 4 I + LegalOrig, Relig, Region			III RST Table 2, 4, 5, 6 II + Alt. Integr./Geo Measures		
	<i>p</i> ≠ 0	Mean	Sd	<i>p</i> ≠ 0	Mean	Sd	<i>p</i> ≠ 0	Mean	Sd
Rule of Law	100	1.2775	0.1772	100	0.9485	0.1323	96.4	0.7979	0.3155
Integration	20	0.1119	0.2578	7.4	0.0697	0.1451	84.7	0.9275	0.3803
Tropics							69	-0.7828	0.37
Area							57.1	.164	.171
SubSaharaAfrica				97	-0.7487	0.1998	50.7	-0.5319	0.3077
Catholic				36.2	0.0043	0.0028	50.6	0.01	0.0072
PolicyOpenness							49.4	0.6857	0.368
PopGrowth							46.7	0.2099	0.1473
Muslim				50.3	-0.0044	0.0025	43.8	-0.0043	0.0035
LatinAmerica				10.1	0.0984	0.2858	36.1	0.6529	0.3652
LegalOrigFr				29.5	0.2083	0.2065	34.6	0.29	0.1682
FrostArea							33.3	1.2204	0.8814
FrostDay							31.3	-0.0621	0.0383
MeanTemp							22.2	0.0323	0.0433
EastAsia				22.8	0.3345	0.3127	19.5	0.532	0.3898
Latitude	18.3	-0.0019	0.0143	10.8	-0.0058	0.0099	18.6	-0.0168	0.0162
Oil							18	0.323	0.2919
Malaria94							7.3	-0.243	0.4787
SeaAccess							5.6	-0.0698	0.3142
Protestant				8	-0.0027	0.006	1.9	-0.0016	0.0069
LegalOrigSocialist				41	-0.6144	0.4917	na	na	na
BIC best model	-57.34			-92.34			-77.12		
Generalized R2 best model	0.53			75.10			85.70		
Bayes/Sargan <i>p</i> -value	0.0308			0.7581			0.8438		
Bayes/Cragg-Donald <i>p</i> -value	0.0000			0.0000			0.0097		

## References

- Acemoglu, D., Johnson, S., and Robinson, J. A. (2001a). “The Colonial Origins of Comparative Development: An Empirical Investigation.” *The American Economic Review*, 91, 1369–1401.
- (2001b). “Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution.” *The Quarterly Journal of Economics*, 117, 1231–1294.
- Alcalá, F. and Ciccone, A. (2004). “Trade and Productivity.” *The Quarterly Journal of Economics*, 119, 612–645.
- Andrews, D. W. K. (1999). “Consistent Moment Selection Procedures for Generalized Method of Moments Estimators.” *Econometrica*, 67, 543–564.
- Andrews, D. W. K. and Lu, B. (2001). “Consistent Model and Moment Selection Procedures for GMM Estimation with Application to Dynamic Panel Data Models.” *Journal of Econometrics*, 101, 123–164.
- Angrist, J. D. and Krueger, A. B. (1991). “Does Compulsory School Attendance Affect Schooling and Earnings?” *The Quarterly Journal of Economics*, 106, 9979–1014.
- Bloom, D. and Sachs, J. (1998). “Geography, Demography and Economic Growth in Africa.” *Brookings Papers on Economic Activity*, 2, 207–289.
- Bound, J., Jaeger, D. A., and Baker, R. M. (1995). “Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Variable is Weak.” *Journal of the American Statistical Association*, 90, 443–450.
- Brock, W. and Durlauf, S. N. (2001). “Growth Empirics and Reality.” *World Bank Economic Review*, 15, 229–272.
- Brock, W., Durlauf, S. N., and West, K. (2003). “Policy Evaluation in Uncertain Economic Environments.” *Brookings Papers on Economic Activity*, 1, 235–322.
- Caselli, F. (2005). “Accounting for Cross-Country Income Differences.” In *Handbook of Economic Growth*, eds. P. Aghion and S. Durlauf, vol. 1, 1st ed., chap. 9, 679–741. Elsevier.
- Chamberlain, G. and Imbens, G. (2004). “Random Effects Estimators with Many Instrumental Variables.” *Econometrica*, 72, 295–306.

- Chao, J. C. and Swanson, N. R. (2005). “Consistent Estimation with a Large Number of Weak Instruments.” *Econometrica*, 73, 1673–1692.
- Chen, H., Mirestean, A., and Tsangarides, C. (2009). “Limited Information Bayesian Model Averaging for Dynamic Panels with Short Time Periods.” *IMF Working Paper WP/09/74*.
- Cicchone, A. and Jarocinski, S. (2007). “Determinants of Economic Growth: Will Data Tell?” *CERP working paper 6544*.
- Conley, T. G., Hansen, C. B., McCulloch, R. E., and Rossi, P. E. (2008). “A Semi-parametric Bayesian Approach to the Instrumental Variable Problem.” *Journal of Econometrics*, 144, 276–305.
- Cragg, J. G. and Donald, S. G. (1993). “Testing Identifiability and Specification in Instrumental Variable Models.” *Econometric Theory*, 9, 222–240.
- Davidson, R. and MacKinnon, J. G. (2004). *Econometric Theory and Methods*. Oxford University Press (New York).
- Donald, S. G. and Newey, W. K. (2001). “Choosing the Number of Instruments.” *Econometrica*, 69, 1161–1191.
- Doppelhofer, G. and Weeks, M. (2009). “Jointness of Growth Determinants.” *Journal of Applied Econometrics*, 24, 209–244.
- Dreze, J. H. (1976). “Bayesian Limited Information Analysis of the Simultaneous Equations Model.” *Econometrica*, 44, 1045–1075.
- Durlauf, S., Cohen-Cole, E., Fagan, J., and Nagin, D. (2007). “Model Uncertainty and the Deterrent Effect of Capital Punishment.” *American Law and Economics Review*, forthcoming.
- Durlauf, S., Kourtellos, A., and Tan, C. M. (2008). “Are Any Growth Theories Robust?” *Economic Journal*, 118, 329–346.
- Easterly, W. and Levine, R. (2003). “Tropics, Germs, and Crops: the Role of Endowments in Economic Development.” *Journal of Monetary Economics*, 50, 3–39.
- Eicher, T. S., Papageorgiou, C., and Raftery, A. E. (2010). “Default Priors and Predictive Performance in Bayesian Model Averaging, with Application to Growth Determinants.” *Journal of Applied Econometrics*, forthcoming.

- Fernández, C., Ley, E., and Steel, M. (2001). “Benchmark Priors for Bayesian Model Averaging.” *Journal of Econometrics*, 100, 381–427.
- Flores-Lagunes, A. (2007). “Finite Sample Evidence of IV Estimators Under Weak Instruments.” *Journal of Applied Econometrics*, 22, 677–694.
- Gelman, A., Meng, X. L., and Stern, H. (1996). “Posterior Predictive Assessment of Model Fitness via Realized Discrepancies.” *Statistica Sinica*, 6, 733–807.
- Green, P. J. (1995). “Reversible jump Markov chain Monte Carlo Computation and Bayesian Model Determination.” *Biometrika*, 82, 711–732.
- Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press.
- Hall, A. R., Rudebusch, G. D., and Wilcox, D. W. (1996). “Judging Instrument Relevance in Instrumental Variables Estimation.” *International Economic Review*, 37, 283–298.
- Hall, R. and Jones, C. I. (1999). “Why Do Some Countries Produce So Much More Output Per Worker Than Others?” *Quarterly Journal of Economics*, 114, 83–116.
- Hansen, L. P. (1982). “Large Sample Properties of Generalized Method of Moments Estimators.” *Econometrica*, 50, 1029–1054.
- Hineline, D. R. (2007). “Examining the Robustness of the Inflation and Growth Relationship.” *Southern Economic Journal*, 73, 1020–1037.
- Jeffreys, H. (1961). *Theory of Probability*. 3rd ed. Oxford, U.K.: Clarendon Press.
- Kapetanios, G. (2006). “Choosing the Optimal Set of Instruments from Large Instrument Sets.” *Computational Statistics & Data Analysis*, 51, 612–620.
- Kass, R. E. and Raftery, A. E. (1995). “Bayes Factors.” *Journal of the American Statistical Association*, 90, 773–795.
- Kass, R. E. and Wasserman, L. (1995). “A Reference Test for Nested Hypotheses with Large Samples.” *Journal of the American Statistical Association*, 90, 928–934.
- King, R. G. and Levine, R. (1994). “Capital Fundamentalism, Economic Development, and Economic Growth.” *Carnegie-Rochester Conference Series on Public Policy*, 40, 259–92.
- Kleibergen, F. and Zivot, E. (2003). “Bayesian and Classical Approaches to Instrumental Variable Regression.” *Journal of Econometrics*, 114, 29–72.

- Klenow, P. J. and Rodrigues-Clare, A. (1997). “The Neoclassical Revival in Growth Economics: Has It Gone Too Far?” *NBER Macroeconomics Annual*, 12.
- Lancaster, A. (2004). *An Introduction to Modern Bayesian Econometrics*. Blackwell.
- LaPorta, R., de Silanes, R. F. L., Pop-Eleches, C., and Shleifer, A. (2004). “Judicial Checks and Balances.” *Journal of Political Economy*, 112, 445–470.
- Ley, E. and Steel, M. F. (2007). “Jointness in Bayesian variable selection with applications to growth regression.” *Journal of Macroeconomics*, 29, 476–493.
- Morales-Benito, E. (2009). “Predetermined Variables: Likelihood-Based Estimation and Bayesian Averaging.” *CEMFI working paper*.
- Okui, R. (2009). “Instrumental Variable Estimation in the Presence of Many Moment Conditions.” *Journal of Applied Econometrics*, forthcoming.
- Paolo, M. (1995). “Corruption and Growth.” *Quarterly Journal of Economics*, 110, 681–712.
- Pesaran, M. H. and Smith, R. J. (1994). “A Generalized R<sup>2</sup> Criterion For Regression Models Estimated by the Instrumental Variables Method.” *Econometrica*, 62, 705–710.
- Raftery, A. E. (1988). “Analysis of a Simple Debugging Model.” *Journal of the Royal Statistical Society, series C - Applied Statistics*, 37, 12–22.
- (1995). “Bayesian Model Selection in Social Research (with Discussion).” *Sociological Methodology*, 25, 111–196.
- Raftery, A. E., Madigan, D., and Hoeting, J. (1997). “Bayesian Model Averaging for Linear Regression Models.” *Journal of the American Statistical Association*, 92, 179–191.
- Raftery, A. E., Painter, I., and Volinsky, C. T. (2005). “BMA: An R package for Bayesian Model Averaging.” *R News*, 5, 2, 2–8.
- Rodrik, D., Subramanian, A., and Trebbi, F. (2004). “Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development.” *Journal of Economic Growth*, 9, 131–165.
- Rossi, P. E., Allenby, G. M., and R., M. (2006). *Bayesian Statistics and Marketing*. New York: Wiley.
- Rubin, D. B. (1984). “Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician.” *Annals of Statistics*, 12, 1151–1172.



- Sachs, J. (2003). “Institutions Don’t Rule: Direct Effects of Geography on Per Capita Income.” *National Bureau of Economic Research Working Paper, No. W9490*.
- Sala-I-Martin, X., Doppelhofer, G., and Miller, R. (2004). “Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach.” *American Economic Review*, 94, 813–835.
- Sargan, J. (1958). “The Estimation of Economic Relationships using Instrumental Variables.” *Econometrica*, 26, 393–415.
- Stock, J. H. and Wright, J. H. (2000). “GMM With Weak Identification.” *Econometrica*, 68, 1055–1096.
- Stock, J. H. and Yogo, M. (2005). “Testing for Weak Instruments in Linear IV Regression.” In *Identification and Inference for Econometric Models*, eds. D. W. K. Andrews and J. H. Stock, 80–108. Cambridge University Press.
- Tsangarides, C. G. (2004). “A Bayesian Approach to Model Uncertainty.” *IMF Working Paper, no. 0468*.

## Appendix

*Proof of Theorem 1:* In what follows, for a matrix  $A$ , let  $\mathbf{P}_A = A(A'A)^{-1}A'$  and  $\mathbf{M}_A = I - \mathbf{P}_A$ . The likelihood equations follow directly from the marginal distribution of  $\epsilon$  and the conditional distribution of  $\xi$  given  $\epsilon$ . Furthermore, the fact that  $\hat{\theta}$  maximizes Equation 8 follows directly from standard results for the MLE of a regression equation.

Now, consider  $\beta$ . Note that if  $\hat{U}$  is in the column space of  $V$ , which occurs in this case, then  $\hat{U} = \mathbf{P}_V U$ . By moving to log likelihoods, and differentiating with respect to the vector  $\beta$  we see that

$$\beta = (\hat{U}'\hat{U})^{-1}\hat{U}'Y + \phi(\hat{U}'\hat{U})^{-1}\hat{U}'\hat{\epsilon} \quad (\text{A-1})$$

but note that

$$\hat{U}'\hat{\epsilon} = U'\mathbf{P}_V\mathbf{M}_V V = 0$$

Since  $\mathbf{P}_V\mathbf{M}_V = 0$ , thus Equation A-1 becomes  $\beta = \hat{\beta}^{2SLS}$ .  $\square$

*Proof of Theorem 2:* This follows directly from the proof of Theorem 1. However, we note that it is possible that  $\hat{U}^{(j)}$  may not be written as  $\mathbf{P}_{V^{(i)}}U^{(j)}$  for arbitrary  $M_i$  and  $L_j$ , which causes the additional  $\phi^{(i,j)}\Pi^{(i,j)}$  to be present.  $\square$

*Proof of Theorem 3:* Note that using the standard BMA results, the variance of  $\hat{\beta}^{2SBMA}$  can be written as

$$\sigma_{2SBMA}^2(\beta) = \sum_{i=1}^I \sum_{j=1}^J \pi_i \nu_{i,j} \text{Var}(\hat{\beta}^{(i,j)}) + \sum_{i=1}^I \sum_{j=1}^J \pi_i \nu_{i,j} (\hat{\beta}^{(i,j)} - \hat{\beta}^{2SBMA})^2. \quad (\text{A-2})$$

Rewriting this we have,

$$\sigma_{2SBMA}^2(\beta) = \sum_{i=1}^I \pi_i \left\{ \sum_{j=1}^J \nu_{i,j} \left[ \text{Var}(\hat{\beta}^{(i,j)}) + (\hat{\beta}^{(i,j)} - \hat{\beta}^{2SBMA})^2 \right] \right\}, \quad (\text{A-3})$$

$$= \sum_{i=1}^I \pi_i \left\{ \sum_{j=1}^J \nu_{i,j} \left[ \text{Var}(\hat{\beta}^{(i,j)}) + (\hat{\beta}^{(i,j)} - \hat{\beta}^{(i,\cdot)} + \hat{\beta}^{(i,\cdot)} - \hat{\beta}^{2SBMA})^2 \right] \right\}, \quad (\text{A-4})$$

$$= \sum_{i=1}^I \pi_i \left\{ \sum_{j=1}^J \nu_{i,j} \left[ \text{Var}(\hat{\beta}^{(i,j)}) + (\hat{\beta}^{(i,j)} - \hat{\beta}^{(i,\cdot)})^2 + (\hat{\beta}^{(i,\cdot)} - \hat{\beta}^{2SBMA})^2 \right] \right\}, \quad (\text{A-5})$$

which results since,

$$\sum_{j=1}^J \nu_{i,j} (\hat{\beta}^{(i,j)} - \hat{\beta}^{(i,\cdot)}) (\hat{\beta}^{(i,\cdot)} - \hat{\beta}^{2SBMA}) = 0.$$

Reordering the terms we then obtain

$$\sigma_{2BMA}^2(\beta) = \sum_{i=1}^I \pi_i \text{Var}(\beta|M_i) + \sum_{i=1}^I \pi_i (\hat{\beta}^{(i,\cdot)} - \hat{\beta}^{2SBMA})^2,$$

as desired.  $\square$

*Proof of Theorem 4:* For convenience, suppose that  $M_1 \in \mathcal{M}$  is the true model for the first stage. By true model, we do not mean the underlying model that gave rise to  $W$ , but the correct subset of  $V$  for which the associated elements in  $\theta$  are not zero. Then,

$$\pi_1 \rightarrow_p 1 \text{ and } \pi_j \rightarrow_p 0, j \neq 1 \text{ as } n \rightarrow \infty.$$

by the consistency of the model selection procedure. Furthermore, suppose that  $L_1 \in \mathcal{L}$  is the true second stage model. Then,

$$\nu_{1,1} \rightarrow_p 1 \text{ and } \nu_{1,j} \rightarrow_p 0, j \neq 1 \text{ as } n \rightarrow \infty.$$

Therefore,

$$\hat{\beta}^{2SBMA} \rightarrow_p \hat{\beta}^{(1,1)}.$$

Finally consider  $\hat{\beta}^{2SLS}$ . We know that  $\hat{\beta}^{2SLS} \rightarrow_p \beta$  by the consistency of the technique. Furthermore, since the first and second stage estimates of 2SLS are individually consistent we have  $\hat{\beta}^{2SLS} \rightarrow_p \hat{\beta}^{(1,1)}$  provided  $M_1$  and  $L_1$  are the true first and second stage models. Thus,  $\hat{\beta}^{(1,1)} \rightarrow_p \beta$ , showing the technique is consistent.  $\square$