Bayesian Model Averaging and Endogeneity Under Model Uncertainty: An Application to Development Determinants*

BY THEO S. EICHER
Department of Economics, University of Washington
ALEX LENKOSKI
Department of Statistics, University of Washington
AND ADRIAN E. RAFTERY
Department of Statistics, University of Washington

Abstract

Recent approaches to development accounting reflect substantial model uncertainty at both the instrument and the development determinant level. Bayesian Model Averaging (BMA) has been proven useful in resolving model uncertainty in economics, and we extend BMA to formally account for model uncertainty in the presence of endogeneity. The new methodology is shown to be highly efficient and to reduce many-instrument bias; in a simulation study we found that IVBMA estimates reduced mean squared error by 60% over standard IV estimates. We also introduce Bayesian over and under-identification tests that are based on model averaged predictive p-values. This approach is shown to mitigate the reduction in power these tests experience as dimension increases. In a simulation study where the exogeneity of the instrument is compromised we show that the classical Sargan test has a power of 0.2% while our Bayesian over-identification test has a power of 98% at detecting the violation of the exogeneity assumption. An application of our method to a prominent development accounting approach leads to new insights regarding the primacy of institutions. Using identical data and robustness specifications we find support not only for institutions, but also for geography and integration, once both model uncertainty and endogeneity have been jointly addressed.

*We thank David Albouy and Francesco Trebbi for kindly sharing their data as well as Chris Papageorgiou for helpful comments.

1 Introduction

The hallmark of the recent development literature is the distinction between proximate and fundamental development determinants. Proximate causes (technology, physical and human capital, etc) are the focus of cross-country growth regressions that are subject to substantial model uncertainty. The degree of model uncertainty is expressed by the 140 or so candidate regressors that have been suggested by competing theories.[1] Fundamental determinants (e.g., geography, institutions, and culture) are examined in the recent development literature, which is hampered by instrument uncertainty since a multitude of competing theories motivate numerous candidate instruments for identification.[2]

Raftery (1995) argued that uncertainty surrounding particular theories should be addressed explicitly by the statistical approach. Standard errors based on a single model can be underestimated when the uncertainty surrounding the validity of theories has been ignored. Bayesian model averaging (BMA) has been used extensively to account for model uncertainty in growth regressions.[3] To date, instrument uncertainty has been addressed only in standard robustness analyses that juxtapose one particular Instrumental Variable (IV) theory/specification against another. In one of the most prominent examples, 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 growth determinants and instruments requires a methodology that is rooted in statistical theory. Durlauf et al. (2007) introduced an IV model selection procedure to evaluate coefficient estimates according to t-statistics, while warning of the tenuous nature of the underlying 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 Two Stage Least Squares (2SLS), but noted that their heuristic approach lacked statistical justification.[4] Strictly speaking, Durlauf et al. (2008) also did not allow for instrument uncertainty, but provided a model averaging approach to instrument candidate regressors in the second stage only. We extend the Durlauf et al. (2008) approach and develop formal statistical foundations for an instrumental variable BMA (IVBMA) methodology that addresses model uncertainty in the presence of endogeneity.

We conduct a thorough exploration of the properties of IVBMA as a valid IV estimator and show that the procedure is a consistent methodology that reduces the well known many-instrument bias in standard IV regressions. A simulation study con-

[1]See Brock and Durlauf (2001) and Durlauf et al. (2005) for reviews of the growth empirics literature.

[2]See e.g., Rodrik (2003), Acemoglu (2008) for extensive surveys of both proximate and fundamental causes of growth.

[3]See e.g., Fernandez et al. (2001), Sala-i-Martin et al. (2004), Ciccone and Jarocinski (2007), and Eicher et al. (2007).

[4]A similar heuristic panel approach is introduced by Hineline (2007) to examine the growth/inflation relationship.

ducted below shows this reduction in bias and also a 60% reduction in mean squared error for estimating regression coefficients. Instrumental variable estimation of any kind requires a number of assumptions that relate to the identification of the implied structural model.

Frequently, as is the case in the Sargan (1958) and Cragg and Donald (1993) tests, a test statistic is compared to a reference distribution. The test statistic is often only asymptotically distributed according to the reference distribution, which frequently has a degrees of freedom related to the size of the model estimated. The nature of these statistics proves problematic when working with growth data, for which sample size is small and dimension is continually increasing.

The Bayesian approach provides a direct interpretation of the efficacy of an instrumentation strategy, by examining posterior inclusion probabilities. However, we also provide alternative measures to verify IV assumptions that are based on model averaged Bayesian predictive p-values. Rubin (1984) and Gelman et al. (1996) discussed the use of posterior predictive p-values for a single model.

Here we introduce the concept of model averaged p-values. We provide Bayesian tests for over-identification (based on the Sargan test), and a Bayesian test for under-identification (based on Cragg and Donald 1993) to examine instrument conditions within IVBMA. In a simulation study of moderate dimension in which a proposed instrument does not satisfy the exogeneity assumption, we found that the Bayesian over-identification test had a power of 98% at detecting this failure, while the traditional Sargan test had a power of only 0.2%.

IVBMA is then applied to a prominent approach to development accounting that features both instrument and determinant uncertainty. We first replicate the robustness analysis of RST, whose analysis led the authors to endorse the "primacy of institutions" over all other alternative theories. Using their own data and robustness specifications, but allowing for a principled approach to determinant and instrument uncertainty, we find that strong conclusions regarding any primacy of institutions must be modified. Not only institutions but also integration and geography are shown to have a clear effect on long term development once instrument and determinant uncertainty is addressed as part of the statistical approach.[5] At the instrument level, we find that, once we allow for instrument uncertainty evidence, Settler Mortality may, at times, not be robust to the inclusion of alternative instruments suggested by RST in their robustness specification. The exercises highlight that the results presented in RST may have relied on specific robustness specifications that do not generate the greatest model performance given the entire instrument and covariate space suggested by the authors. The resolution of model uncertainty at both the development determinant and the instrument level allows us to isolate additional models that receive stronger support from the same data.

The IVBMA approach thus highlights the extent to which results may be affected by focusing on particular specifications without accounting for the complexity of the model uncertainty that may be present in both stages of the IV approach. Our result

[5]Glaeser et al. (2004) rejected the primacy of institutions, contending that the proximate causes in previous approaches measured policy choices. Measurement error is not addressed in our approach.

is similar to the findings of Durlauf et al. (2008), who document that many theories/variables are not robust once model uncertainty is integrated into the statistical framework. Our approach differs, however, as we develop a formal methodology that is specifically designed to resolve endogeneity in the presence of model uncertainty.

Previous approaches to resolving endogeneity and identifying proximate and fundamental growth determinants are not limited to the papers and datasets we explore below. The purpose of our paper is to introduce IVBMA methodology and provide applications that highlight the importance of model uncertainty at both the determinant and instrument stages. Alternative approaches to 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 colonial origins 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).

The article proceeds as follows. Section 2 outlines the statistical approach that underlies IVBMA and discusses theoretical properties of the technique. Section 3 describes a simulation study, while Section 4 revisits key robustness results to highlight the importance of both determinant and instrument uncertainty in the recent development literature. Section 5 concludes.

2 Theoretical Properties of IVBMA

It is standard to address endogeneity by applying two-stage least squares (2SLS) and imposing over-identification and instrument restrictions as expressed by

$$Y = \beta' \begin{pmatrix} W \\ X \end{pmatrix} + \eta, \tag{1}$$

$$W = \theta_Z' Z + \theta_X' X + \epsilon, \tag{2}$$

where Y is the dependent variable, X is a set of covariates, W is the set of endogenous variables, and Z is the set of instruments. Both X and θ_X have dimension p_X, while Z and θ_Z have dimension p_Z. To simplify exposition we assume that W is univariate. Assuming that

$$\begin{pmatrix} \eta \\ \epsilon \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\epsilon^2 & \sigma_{\eta\epsilon} \\ \sigma_{\eta\epsilon} & \sigma_\epsilon^2 \end{pmatrix}\right), \tag{3}$$

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, η.

In the presence of endogeneity, it is well known that the determination of W leads to inconsistent estimates of the entire coefficient vector, β, under standard Ordinary

Least Squares (OLS). The 2SLS estimator solves the consistency problem, but relies on the existence of a set of instruments, Z, which are independent of Y, given W and the vector of covariates, X. The IV estimates derived in the second stage, $\hat{\beta}^{IV}$, obtained using the fitted values from the first stage, $\tilde{w}$, are consistent only if the conditional independence assumptions are valid.

Several problems can arise. A key concern in IV estimation is that the estimates of $\hat{\beta}^{IV}$ are biased and that 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. In that case, the instrumental variable results are not necessarily better than the biased OLS results (Davidson and MacKinnon, 2004). Another concern is that IV is a large-sample procedure, implying that even when all assumptions are met, there exists the distinct possibility of finite sample biases (see Bound et al. (1995)). The prospect of this type of bias looms large in development accounting applications, where samples rarely exceed 100 observations. A third concern for IV estimation is that economic data rarely present clear-cut instruments that have both strong explanatory power on the endogenous variables and unquestionable conditional independence properties in relation to the dependent variable. Over-identification tests such as the one proposed by Sargan (1958) help verify the validity of instrument assumptions, but can often become ineffective as the dimension of the problem increases in small sample situations.

2.1 Statistical Foundations

IVBMA combines the IV and BMA methodologies. It processes the data much like a two stage estimator while also addressing model uncertainty in both stages. The first stage is a simple application of BMA to identify effective instruments. As we introduce notation, it is helpful to review the properties of BMA that are implied in stage 1.

Let Δ be a quantity of interest and let the set of potential models in the first stage, $\mathcal{M}$, be comprised of individual models $\{M_1, \ldots, M_I\}$. The posterior distribution of Δ given the data, D, is given by the weighted average of the predictive distribution under each model, weighted by the corresponding posterior probabilities,

$$pr(\Delta|D) = \sum_{i=1}^{I} pr(\Delta|M_i, D) pr(M_i|D), \tag{4}$$

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 π_i, for each first stage model M_i is given by

$$\pi_i = pr(M_i|D) \propto pr(D|M_i) pr(M_i) \tag{5}$$

where

$$pr(D|M_i) = \int pr(D|\theta^{(i)}, M_i) pr(\theta^{(i)}|M_i) d\theta^{(i)} \tag{6}$$

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.

Previous approaches that outline Bayesian instrumental variable methods are provided by Geweke (1996) and Kleibergen and Zivot (2003). There is also a literature that attempts to "derive" IV, in some cases using automatic, information-theory-based methods to avoid explicit priors (see Kitamura and Stutzer (1997), Zellner et al. (1997), and Kim (2002)). Related work by Chao and Phillips (1998) pursues the use of Jeffreys priors, which are another automatically generated class of priors.

We simplify matters below by using the BIC approximation to the integrated likelihood. In general, Schwarz (1978) showed that

$$pr(D|M_i) = \log pr(D|\hat{\theta}^{(i)}, M_i) - (p_{Z,i} + p_{X,i}) \log n + O(1), \tag{7}$$

where $p_{Z,i}$ and $p_{X,i}$ are, respectively, the number of Z and X variables included in model M_i. Furthermore, when a unit information prior is used the $O(1)$ term may be replaced by $O(n^{-1/2})$; see Kass and Wasserman (1995) and Raftery (1995).

Under BMA, the posterior mean of θ is

$$\hat{\theta}^{BMA} = \sum_{i=1}^{I} \pi_i \hat{\theta}^{(i)}, \tag{8}$$

which is the sum of the posterior means of each model in the collection $\mathcal{M}$, weighted by their posterior probabilities. Similarly, the posterior variance of the BMA estimate is calculated as

$$\hat{\sigma}^{BMA}(\theta) = \sum_{i=1}^{I} \pi_i \hat{\sigma}_i^2 + \sum_{i=1}^{I} \pi_i \left(\hat{\theta}^{(i)} - \hat{\theta}^{BMA}\right)^2. \tag{9}$$

This variance has a clear interpretation that highlights how model uncertainty is accounted for in the standard errors by the BMA methodology. 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 and can then be taken as a measure of the importance of a variable. For instance, for instrument Z_k we may write,

$$\mu^{BMA}(\theta_{Z_k}) = pr(\hat{\theta}_{Z_k} \neq 0|D) = \sum_{i \in \mathcal{M}_k} \pi_i, \tag{10}$$

where $\mathcal{M}_k$ is collection of indices for which $i \in \mathcal{M}_k$ implies model M_i does not restrict the parameter θ_{Z_k} to zero. Standard rules of thumb for interpreting μ^{BMA} have been provided by Kass and Raftery (1995). They establish 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.

In the case of IV estimation in the presence of model uncertainty, the BMA framework must be extended to account for the two stages in which estimation is performed. IVBMA is a nested approach that first determines the posterior model probabilities in the first stage according to the BMA methodology, determining both π_i as well as $\tilde{w}_i$, the first-stage fitted value according to model M_i, for all models in $\mathcal{M}$. Denoting by $\mathcal{L} = \{L_1, \ldots, L_J\}$ the set of second stage models, IVBMA then uses the fitted value, $\tilde{w}_i$ to derive second stage posterior model probabilities, $\nu_j(\tilde{w}_i)$, and estimates, $\hat{\beta}^{(j)}(\tilde{w}_i)$ for each model $L_j \in \mathcal{L}$. The IVBMA estimate of β is calculated as

$$\hat{\beta}^{IVBMA} = \sum_{i=1}^{I} \sum_{j=1}^{J} \pi_i \nu_j(\tilde{w}_i) \hat{\beta}^{(j)}(\tilde{w}_i). \tag{11}$$

Equation 11 shows that the IVBMA estimate is formed as the average of each IV estimate that results from using 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 probabilities.

Furthermore, for the estimated variance we have the following result.

Theorem 1. *Let $\hat{\beta}_{i*} = \sum_{j=1}^{J} \nu_j(\tilde{w}_i) \beta^{(j)}(\tilde{w}_i)$ be the model averaged estimate of β for a fixed first stage model M_i. Then the variance of the estimate $\hat{\beta}^{IVBMA}$ is*

$$\sigma^2_{IVBMA}(\beta) = \sum_{i=1}^{I} \pi_i Var(\beta|M_i) + \sum_{i=1}^{I} \pi_i (\hat{\beta}_{i*}(\tilde{w}_i) - \hat{\beta}^{IVBMA})^2, \tag{12}$$

where

$$Var(\beta|M_i) = \sum_{j=1}^{J} \nu_j(\tilde{w}_i) \hat{\beta}^{(j)}(\tilde{w}_i) + \sum_{j=1}^{J} \nu_j(\tilde{w}_i)(\hat{\beta}^{(j)}(\tilde{w}_i) - \hat{\beta}_{i*})^2 \tag{13}$$

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

Proof See Appendix.

Theorem 1 shows that the variance of IVBMA estimates has a similar separation property to standard BMA variances, containing a part which is the average of BMA variances associated with a single first stage model and another part that quantifies the variation in the BMA estimates obtained by fixing a particular first stage model relative to the overall IVBMA estimate.

The posterior distribution of $\hat{\beta}^{IVBMA}$ 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 in the second stage, which indicates the importance of a variable. For instance, for the variable X_l we may write,

$$\mu^{IVBMA}(\beta_{X_l}) = pr(\hat{\beta}_{X_l} \neq 0|D) = \sum_{i=1}^{I} \sum_{j \in \mathcal{L}_l} \pi_i \nu_j(\tilde{w}_i). \tag{14}$$

Where $\mathcal{L}_l$ indicates the subset of $\mathcal{L}$ for which the coefficient β_{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 Kass and Raftery (1995).

2.2 Properties of IVBMA

2.2.1 Consistency of IVBMA

The driving motivation underlying IV estimation is the fact that, in the presence of endogeneity, IV estimates are consistent, in contrast to OLS estimates. This consistency is retained by the IVBMA estimates.

Theorem 2. *The IVBMA estimate is consistent, in that* $\hat{\beta}^{IVBMA} \to_p \beta$.

Proof See Appendix.

2.2.2 Many Instrument Efficiency

In what is sometimes called the "many instruments" problem (Hall, 2005), IV estimates become increasingly biased as the number of proposed instruments increases, especially when these proposed instruments have little explanatory power on the endogenous variable. Sawa (1969) derives this bias explicitly as[6]

$$\beta_W - E\left[\hat{\beta}_W^{2SLS}\right] = (\beta_W - \sigma_{\eta\epsilon}/\sigma_\eta^2)\left[1 - \frac{\tau^2}{K} F_{1;1}\left(1, \frac{K+2}{2}; -\frac{\tau^2}{2}\right)\right], \tag{15}$$

where K is the number of proposed instruments plus those proposed covariates that actually have coefficients equal to zero, $F_{1;1}(\cdot,\cdot;\cdot)$ is a confluent hypergeometric function and $\tau = \sum_{i=1}^{p_Z}\sum_{j=1}^{p_Z}\sigma_{Z_iZ_j}\theta_{Z_i}\theta_{Z_j} + \sum_{i=1}^{p_Z}\sum_{j=1}^{p_X}\sigma_{Z_iX_j}\theta_{Z_i}\theta_{X_j} + \sum_{i=1}^{p_X}\sum_{j=1}^{p_X}\sigma_{X_iX_j}\theta_{X_i}\theta_{X_j}$.

Bound et al. (1995) show that the many instrument bias in standard IV regressions increases as τ^2/K decreases. As τ^2 is a function of the first stage regression coefficients, we see that adding instruments with no explanatory power on W leads to a decline in τ^2/K, thereby creating larger bias.

We now show that IVBMA mitigates the many instrument bias. Let B_{ij} denote the bias of the estimate $\hat{\beta}^{(j)}(\tilde{w}_i)$ from model M_i in the first stage and L_j in the second stage. Also define B_{**} the bias from $\hat{\beta}^{2SLS}$, equivalent to using the full models in both the first and second stages. Provided model M_i excludes some X or Z covariates, but contains at least one valid instrument from Z, it can be shown that $B_{ij} \leq B_{**}$, which implies that IVBMA will mitigate any many instruments problem that may have existed in the standard IV procedure.

Theorem 3. *The bias of* $\hat{\beta}^{IVBMA}$ *is less than or equal to* $\hat{\beta}^{2SLS}$ *provided that each first-stage model* M_i *with* $\pi_i > 0$ *contains as least one valid instrument.*

[6] The derivation of Sawa (1969) considers the case in which there are no exogenous covariates X and the instruments Z are considered to be independently distributed, but includes an outline of the straightforward updates necessary to incorporate additional covariates and dependence between X and Z. Our treatment takes these additional factors into account, as they are central to the purpose of modeling multiple growth theories.

Proof. See Appendix.

2.3 Bayesian Tests of Assumption Validity

As mentioned above, the IV framework requires the proposed instrument set to be conditionally independent of the variable Y and presumes that the instruments have some explanatory power on the endogenous variables W. Various tests have been developed to verify these assumptions, most notably the over-identification test of Sargan (1958) and the under-identification test of Cragg and Donald (1993) which Stock and Yogo (2002) use to propose a weak instruments test. In this section we show how model averaged versions of these tests can be used in the IVBMA framework to verify model assumptions and discuss the properties of such techniques.

2.3.1 A Bayesian Test of Over-identification

The most important assumption in IV regressions is that the instrument condition is satisfied, namely that the instrument is exogenous, $E(\eta|Z) = 0$, and that the instrument is relevant $Cov(Z, W) \neq 0$. To allow for an examination of whether these conditions are satisfied in the IVBMA context, we present a Bayesian over-identification test of the exogeneity assumption that provides similar information to that provided by the Sargan (1958) test for the standard IV procedure. Our test proceeds in a manner similar to the Sargan test, but is conducted at the model level and then averaged using model probabilities.

Let $\hat{\eta}_{ij}$ be the residuals from the combination of models M_i and L_j and let p_{ij} be the total number of X and Z included in this combination. Note that the Sargan p-value S^* is calculated as $S^* = pr(nR^2_{**} > \chi^2_{p_X + p_Z - 1})$ 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 Sargan test, we can then consider the regression of $\hat{\eta}_{ij}$ on the subset of the variables X and Z that belong to either M_i or L_j and determine R^2_{ij}, the R^2 associated with this regression. Letting $S_{ij} = p(\chi^2_{p_{ij}-1} > nR^2_{ij})$, we define the Bayesian Sargan p-value to be

$$S^{IVBMA} = \sum_{i=1}^{I} \sum_{j=1}^{J} \pi_i \nu_j(\tilde{w}_i) S_{ij}. \tag{16}$$

S^{IVBMA} 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 Sargan test is that it effectively mitigates the reduction in power that the traditional Sargan test experiences as the dimension of the X or Z variables grows. This increase in power can be marked, as shown in the simulation study below.

2.3.2 Bayesian Tests of Under-Identification and Weak Instruments

While it is crucial to verify that none of the proposed instruments violates the conditional independence assumption, it is also important to test that they have an appropriate level of explanatory power on 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) derive an equivalent test and test statistic to help verify this claim. Here we derive a Bayesian analog of this test.

Consider fixed first and second stage models, M_i and L_j respectively, and let Z_{ij} be the instruments used in this combination (thus all those variables in Z used in M_i and those variables X used in M_i but excluded from L_j) and let X_j be those X contained in L_j. 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, $ng_{ij} \sim \chi^2_{p_{Z_{ij}}-1}$, and this reference distribution is used to derive a p-value. Here we propose a Bayesian model-averaged version of this p-value by considering

$$CD = \sum_{i=1}^{I}\sum_{j=1}^{J} \pi_i \nu_j(\tilde{w}_i) pr(\chi^2_{p_{Z_{ij}}-1} > ng_{ij}). \tag{17}$$

A second use of g_{ij} was suggested by Stock and Yogo (2002), but their test statistic provides only critical values, not p-values that one can average over when models have different numbers of instruments. The apparent weakness of an instrument can, however, be directly assessed in a Bayesian way, using the inclusion probabilities in the first stage.

3 Simulation Study

We conduct a simulation study to show the estimation properties of IVBMA, as well as the behavior of the Bayesian over-identification test. In the following we consider a framework in which there are ten variables in Z, fifteen in X and W is univariate. We set $\beta_{X_1} = \beta_{X_2} = \beta_W = 1$ and the remaining elements of β to zero. In the first stage, we set $\theta_{Z_1} = \theta_{Z_2} = \theta_{X_1} = \theta_{X_3} = 1$ and the remaining elements of θ_Z and θ_X to zero.

Thus, we consider a situation in which two covariates along with W have explanatory power on Y. Furthermore, two variables in Z serve as instruments, one of the variables of X has 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. All variables in X and Z are determined by independent draws from a $N(0,1)$ distribution.

We introduce endogeneity by drawing ϵ from a $N(0,1)$ distribution and setting $\eta = \epsilon + \xi$, with ξ drawn from a $N(0,1)$ distribution as well. We then consider two scenarios. The first 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. This framework leads to an R^2 value of .89, which is typical of data sets currently considered in the growth literature. In each scenario we simulate datasets of 100 observations and consider 500 replicates. The simulation study is structured to roughly resemble the growth data set we will be examining below.

Figure 3 shows the distribution of the estimate of β_W across replications using IVBMA, 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 IVBMA and 2SLS rectify this bias and are more closely centered about the true value of 1. However, there is a distinct improvement in the quality of the estimator using IVBMA, evidence of the finite-sample bias reduction property of IVBMA as discussed above. The average bias of $\hat{\beta}_W$ was 0.330, 0.047 and 0.021 for OLS, IV and IVBMA respectively. The average mean squared error for estimating the entire vector β was 0.0399, 0.0244 and 0.0094 for OLS, IV and IVBMA respectively. IVBMA performs substantially better than OLS or IV.

The first panel in Figure 3 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 lower. However, these scores are still sufficiently high that the exogeneity assumption is unlikely to be incorrectly rejected.

The second panel in Figure 3 shows the resulting Bayesian Sargan and classical Sargan p-values for the case of a misspecified exogeneity assumption. In the case of valid instruments, the size of both tests was 0. However, in the case of invalid instruments the power of the Bayesian Sargan test as 98%, whereas it was 0.2% using the traditional Sargan test, based on an $\alpha = .05$. We clearly see that the Bayesian Sargan test performs much more precisely in discerning the failure of the exogeneity assumption and it is far more likely to reject the hypothesis that the IV assumptions are valid than the classical Sargan test.

The previous figures show that IVBMA returns appropriate coefficient estimates and features 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 mean inclusion probability for each variable across the 500 replications in both stages. We see that in the first stage the two variables in Z as well as the two variables in X are given inclusion probabilities of essentially 1, while the remaining variables are given low inclusion probabilities. This remains true in the second stage as well, where W and the two covariates in X that have explanatory power are given inclusion probabilities of close to 1 and all others are given low inclusion probabilities.

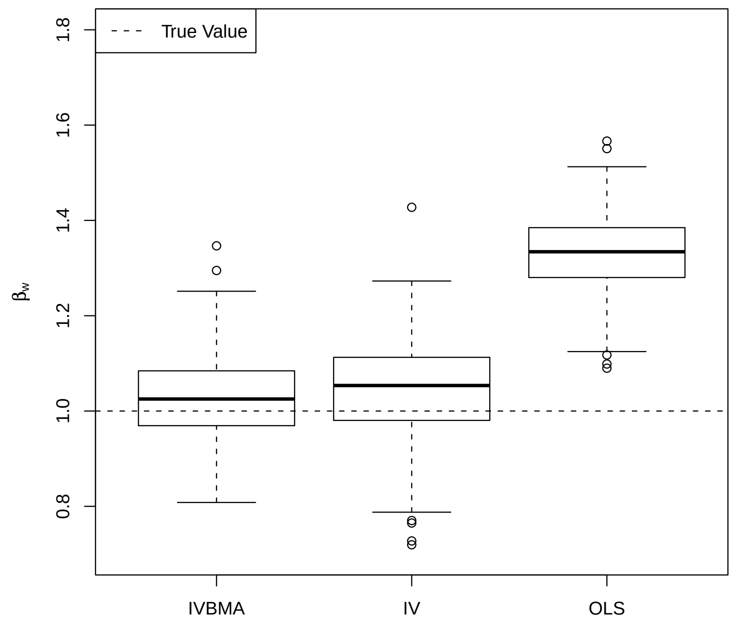

Figure 1: Finite Sample Bias under IVBMA, 2SLS and OLS. Distribution of the estimate for the coefficient β_W across replications using IVBMA, IV and OLS, when $\beta_W = 1$. The average bias of $\hat{\beta}_W$ across 500 replications was .021, .047 and .33 for IVBMA, BMA and OLS respectively. Furthermore, the average mean squared error for estimating the entire vector β was .0094, .0244 and .0399 for IVBMA, IV and OLS respectively.

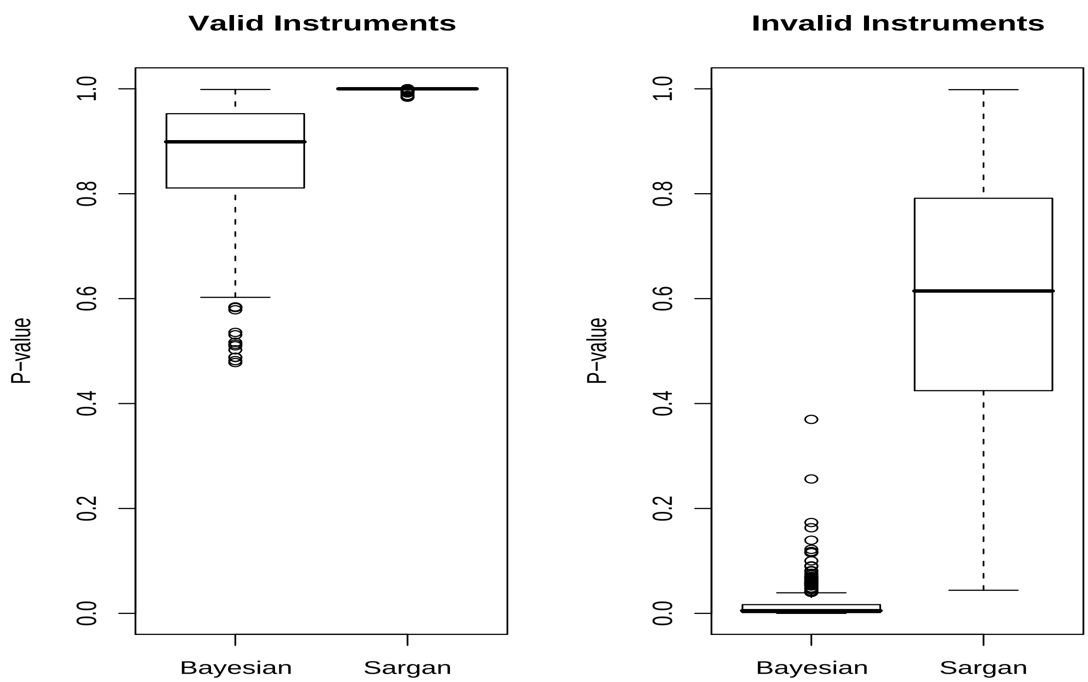

Figure 2: Distribution of scores returned by the Bayesian Sargan test and the Sargan test across replications when the IV assumptions hold (Valid) and when they do not (Invalid). In the case of valid instruments, the size of both tests was 0. However, in the case of invalid instruments the power of the Bayesian Sargan test as 98%, whereas it was 0.2% using the traditional Sargan test.

Table 1: Mean variable inclusion probabilities and the standard deviation of these inclusion probabilities across iterations. 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.

	First Stage		Second Stage	
Variable	$p \neq 0$	SD	$p \neq 0$	SD
W	–	–	**1**	**(0)**
$\mathbf{X_1}$	**1**	**(0)**	**0.974**	**(0.004)**
$\mathbf{X_2}$	0.087	(0.007)	**0.981**	**(0.003)**
$\mathbf{X_3}$	**1**	**(0)**	0.046	(0.003)
X_4	0.11	(0.009)	0.082	(0.005)
X_5	0.081	(0.007)	0.075	(0.004)
X_6	0.085	(0.008)	0.079	(0.005)
X_7	0.069	(0.006)	0.073	(0.004)
X_8	0.087	(0.008)	0.079	(0.005)
X_9	0.084	(0.008)	0.069	(0.004)
X_{10}	0.085	(0.008)	0.072	(0.004)
X_{11}	0.078	(0.007)	0.07	(0.004)
X_{12}	0.088	(0.008)	0.081	(0.005)
X_{13}	0.099	(0.009)	0.087	(0.005)
X_{14}	0.087	(0.008)	0.079	(0.004)
X_{15}	0.098	(0.008)	0.077	(0.004)
$\mathbf{Z_1}$	**1**	**(0)**	–	–
$\mathbf{Z_2}$	**1**	**(0)**	–	–
Z_3	0.091	(0.008)	–	–
Z_4	0.086	(0.008)	–	–
Z_5	0.079	(0.007)	–	–
Z_6	0.084	(0.008)	–	–
Z_7	0.082	(0.008)	–	–
Z_8	0.085	(0.008)	–	–
Z_9	0.092	(0.008)	–	–
Z_{10}	0.086	(0.007)	–	–

4 Instrument and Determinant Uncertainty in Development Accounting

We now apply IVBMA to a prominent dataset in the development accounting literature, where Rodrik et al., (RST) provide an explicit "horse race" of theories that pertain not only to development determinants (geography, integration and institutions), but also to a range of theories that suggest alternative instruments to resolve the endogeneity of the determinants. With less than 100 observations, the sample is a standard size of datasets in development accounting. Model uncertainty among development determinants is a defining feature of the literature and endogeneity is uniformly acknowledged to be rampant.

RST explored over 25 different robustness specifications with alternative candidate regressors suggested by a range of theories. Based on this, they claimed to resolve model uncertainty in a clear way. The claims of the paper are unambiguous and well captured by the title "Institutions rule: the primacy of institutions over geography and integration in economic development." While the previous literature had provided evidence of Trade and Geography effects on development (e.g., Hall and Jones (1999), Sachs (2003)), RST found that geography has at best weak direct effects on incomes, and Integration is found to be "always insignificant, and often enters the income equation with the 'wrong' sign."

Using their data, we reexamine RST's suggested robustness specifications to account for the model and instrument uncertainty that RST highlight so forcefully. The IVBMA first and second stages are reported in Tables 2 and 3. Geography is taken to be exogenous, so the upper panel in Table 2 represents the first stage for the institutions proxy (Rule of Law) and the lower panel is the first stage for Integration. Although it would be sufficient to present only the IVBMA results that explore the entire model space spanned by RST's determinants and instruments, we also provide two intermediate stages. Column 1 represents RST's "core specification" (their Table 2) and Column 2 is the first set of robustness exercises that RST introduce; it highlights the sensitivity of the core specification to even a slight increase in model uncertainty.

Column 1 in Table 3 provides the second stage of RST's preferred core specification (RST's Table 2). Both RST and IVBMA 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 IVBMA 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 IVBMA model is 0.53 versus 0.55 in RST's 2SLS approach.

Column 1 in Table 2 reports the IVBMA first stages for the core specification. They broadly confirm the 2SLS results although the IVBMA suggests slightly more parsimonious models. IVBMA 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 IVBMA (0.49). For Integration, the IVBMA first stage suggests only two strong instruments (Implied Trade Shares and Settler Mortality) while 2SLS produces statistically significant coefficient for an additional instrument (Fraction Speaking English). Nevertheless the R^2 of the IVBMA best model and of the 2SLS first stage are identical (0.58).

RST find that any core specifications with more than one instrument fails 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. Others might argue that RST's specifications do not contain the appropriate set of instruments. We examine this issue further below but note that already at this stage, the under-identification (as measured by the Bayes/Crag-Donald p-value) is easily rejected by IVBMA (not reported in RST).

The 2SLS and IVBMA 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. Columns 2 and 3 in Tables 2 and 3 report the first and second stages for additional robustness exercises suggested by RST. Column 2 adds regressors suggested by theories pertaining 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, IVBMA results start to diverge from the individual 2SLS regressions that juxtapose a particular theory against another. In other words, the disparities across results become more pronounced and extend beyond parsimony as model uncertainty increases.

IVBMA results that use the most comprehensive set of instruments and development determinants (Column 3 in Tables 2 and 3), cast doubt on the strong primacy of institutions result. Instead IVBMA finds that the "horse race" ends in a statistical three-way tie when model uncertainty is considered. Geography (as measured by Tropics), Institutions and Integration are shown to be highly effective development determinants. This result is particularly surprising since Geography is only occasionally weakly significant in RST, while Integration is never significant and often of the wrong sign. In IVBMA all three effects are strong and estimated with the correct sign. The results support the strong contentions of Sachs (2003) and Alcalá and Ciccone (2004) who report strong effects of Geography and Integration.

The divergence of 2SLS and IVBMA results originates in the first stages. [7] 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 is given nearly a 90 percent inclusion 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 Tem-

[7] RST report neither first stages nor tests of instrument restrictions beyond the core specification.

perature variables in both first stages. The increase in the model space of development determinants and instruments dramatically increases the fit of the IVBMA first stage. The best models in both IVBMA first stages (column 3) report R^2 that is at least 40 percent greater than those found in the core specification.

This improvement in model fit is likewise observed in the second stage when considering the generalized R^2 (Pesaran and Smith, 1994) of the best model returned by IVBMA. 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). IVBMA has therefore uncovered combinations of instruments and growth determinants that fit the data substantially better and therefore produce different results in both the first and second stages than those presented in RST. These models confirm the strong effect institutions have on growth, but also suggest that the effects of integration and geography cannot be ignored when instrument and development determinant uncertainty is directly incorporated into the estimation strategy. The Bayesian Sargan and Bayesian Cragg and Donald tests clearly show, respectively, that over-identifcation is easily rejected with the improved set of instruments and that under-identification remains of no concern.

Table 2: First Stage Results for RST Example

	I			II			III		
	RST Table 2			RST Table 2, 4			RST Table 2, 4, 5, 6		
	Core Specification			I + LegalOrig, Relig, Region			II + Alt. Integr./Geo Measures		
	Stage 1, Depedent Variable: Rule of Law								
	$p \neq 0$	Mean	Sd	$p \neq 0$	Mean	Sd	$p \neq 0$	Mean	Sd
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		
	Stage 1, Depedent Variable: Integration								
	$p \neq 0$	Mean	Sd	$p \neq 0$	Mean	Sd	$p \neq 0$	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			II			III		
	RST Table 2			RST Table 2, 4			RST Table 2, 4, 5, 6		
	Core Specification			I + LegalOrig, Relig, Region			II + Alt. Integr./Geo Measures		
	Stage 2								
	$p \neq 0$	Mean	Sd	$p \neq 0$	Mean	Sd	$p \neq 0$	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 p value	0.0308			0.7591			0.8538		
Bayes/Cragg-Donald p value	0.0000			0.0000			0.0097		

5 Conclusion

The recent development literature focuses not only on competing theories that suggest alternative development determinants, but also on a different set of theories that motivate instruments that may resolve the endogeneity between development determinants and development outcomes. We develop a methodology to address model uncertainty in the presence of endogeneity and explore its properties as a valid IV estimator. The method is based on Bayesian Model Averaging (BMA), which has already been extensively used in economic growth applications. IVBMA is a two step BMA procedure that is shown to be a consistent methodology that also reduces many-instrument bias.

Instrumental variable estimation of any kind requires a number of assumptions that relate to the identification of the implied structural model. To enable assumptions to be verified in this setting we have proposed a new concept, that of model averaging Bayesian predictive p-values within the IVBMA framework. As shown in the simulation study, by using model averaged p-values we are able to reduce the effect increasing dimension has on the power of the proposed tests, while not affecting the size in any substantive manner.

We conclude our study by applying IVBMA to the dataset of Rodrik et al. (2004), who motivate their paper by the diversity of alternative theories of development and clearly outline the associated model uncertainty. Instead of resolving the model uncertainty in a horse race of alternative regressions, we use the formal IVBMA 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.

Among the number of potential implementations of the Bayesian model selection paradigm, we have chosen to focus on augmenting the BMA methodology outlined in Raftery (1995) and Raftery (1996) to accommodate the IV estimation problem. We made this particular choice because this methodology has become familiar to a broad range of econometricians and statisticians and has proven robust to the needs of applied researchers. However, recent research into the use of BMA in the development determinant literature has suggested several modifications of this paradigm that may help rectify the particulars of growth data with the goal of testing the strength of various growth theories. In particular, 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 Steele (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 IVBMA framework would help extend the application of IVBMA to the particular problem of testing growth theory robustness.

The IVBMA 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 heteroskadasticity. Future research into these areas will improve the applicability of the BMA framework to economic analysis, in growth economics and beyond.

References

Acemoglu, D. (2008). *An Introduction to Modeling Economic Growth*. Princeton University Press.

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: Geograph and Institutions in the Making of the Modern World Income Distribution." *The Quaterly Journal of Economics*, 117, 1231–1294.

Alcalá, F. and Ciccone, A. (2004). "Trade and Productivity." *The Quarterly Journal of Economics*, 119, 612–645.

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 Endogeneous 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.

Chao, J. C. and Phillips, P. C. (1998). "Posterior Distributions in Limited Information Analysis of the Simultaneous Equations Model Using the Jeffreys Prior." *Journal of Econometrics*, 87, 49–86.

Ciccone, A. and Jarocinski, S. (2007). "Determinants of Economic Growth: Will Data Tell?" *CERP working paper 6544*.

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).

Doppelhofer, G. and Weeks, M. (2009). "Jointness of Growth Determinants." *Journal of Applied Econometrics*, 24, 209–244.

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.

Durlauf, S. N., Johnson, P., and Temple, J. (2005). "Growth Econometrics." In *Handbook of Economic Growth*. P. Aghion and N. Durlauf, eds., North Holland, Amsterdam.

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. (2007). "Default Priors and Predictive Performance in Bayesian Model Averaging, with Application to Growth Determinants." *Journal of Applied Econometrics, forthcoming*.

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

Gelman, G., Meng, X., and Stern, H. (1996). "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies." *Statistica Sinica*, 6, 733–807.

Geweke, J. (1996). "Bayesian Reduced Rank Regression in Econometrics." *Journal of Econometrics*, 75, 121–146.

Glaser, E., La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (2004). "Do Institutions Cause Growth?" *Journal of Economic Growth*, 9, 271–303.

Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press.

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.

Hineline, D. R. (2007). "Examining the Robustness of the Inflation and Growth Relationship." *Southern Economic Journal*, 73, 1020–1037.

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.

Kim, J. Y. (2002). "Limited information likelihood and Bayesian analysis." *Journal of Econometrics*, 107, 175–193.

Kitamura, Y. and Stutzer, M. (1997). "An Information-Theoretic Alternative to Generalized Method of Moments Estimation." *Econometrica*, 65, 861–874.

Kleibergen, F. and Zivot, E. (2003). "Bayesian and Classical Approaches to Instrumental Variable Regression." *Journal of Econometrics*, 114, 29–72.

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.

Paolo, M. (1995). "Corruption and Growth." *Quarterly Journal of Economics*, 110, 681–712.

Pesaran, M. H. and Smith, R. J. (1994). "A Generalized R2 Criterion For Regression Models Estimated by the Instrumental Variables Method." *Econometrica*, 62, 705–710.

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

— (1996). "Approximate Bayes Factors and Accounting for Model Uncertainty in Generalised Linear Models." *Biometrika*, 83, 251–266.

Rodrik, D. (2003). *In Search of Prosperity: Analytic Narratives on Economic Growth*. Princeton University Press.

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.

Rubin, D. B. (1984). "Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician." *Annals of Statistics*, 12, 1151–1172.

Sachs, J. (2003). "Institutions Dont 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.

Sawa, T. (1969). "The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators." *Journal of the American Statistical Society*, 64, 923–937.

Schwarz, G. (1978). "Estimating the dimension of a model." *Annals of Statistics*, 6, 461–464.

Stock, J. and Yogo, M. (2002). "Testing for Weak Instruments in Linear IV Regression." *NBER Working Paper Number 284*.

Tsangarides, C. G. (2004). "A Bayesian Approach to Model Uncertainty." *IMF Working Paper, no. 0468*.

Zellner, A., Tobias, J., and Ryu, H. K. (1997). "Bayesian Method of Moments (BMOM) Analysis of Parametric and Semiparametric Regression Models." *Proceedings of the Section on Bayesian Statistical Science of the American Statistical Association*, 211–216.

Appendix

Proof of Theorem 1: Note that using the standard BMA results, the variance of $\hat{\beta}^{IVBMA}$ can be written as

$$\sigma^2_{IVBMA}(\beta) = \sum_{i=1}^{I}\sum_{j=1}^{J}\pi_i\nu_j(\tilde{w}_i)Var(\hat{\beta}^{(j)}(\tilde{w}_i)) + \sum_{i=1}^{I}\sum_{j=1}^{J}\pi_i\nu_j(\tilde{w}_i)(\hat{\beta}^{(j)}(\tilde{w}_i) - \hat{\beta}^{IVBMA})^2. \tag{A-1}$$

Rewriting this we have,

$$\sigma^2_{IVBMA}(\beta) = \sum_{i=1}^{I}\pi_i\left\{\sum_{j=1}^{J}\nu_j(\tilde{w}_i)\left[Var(\hat{\beta}^{(j)}(\tilde{w}_i)) + (\hat{\beta}^{(j)}(\tilde{w}_i) - \hat{\beta}^{IVBMA})^2\right]\right\}, \tag{A-2}$$

$$= \sum_{i=1}^{I}\pi_i\left\{\sum_{j=1}^{J}\nu_j(\tilde{w}_i)\left[Var(\hat{\beta}^{(j)}(\tilde{w}_i)) + (\hat{\beta}^{(j)}(\tilde{w}_i) - \hat{\beta}_{i*} + \hat{\beta}_{i*} - \hat{\beta}^{IVBMA})^2\right]\right\}, \tag{A-3}$$

$$= \sum_{i=1}^{I}\pi_i\left\{\sum_{j=1}^{J}\nu_j(\tilde{w}_i)\left[Var(\hat{\beta}^{(j)}(\tilde{w}_i)) + (\hat{\beta}^{(j)}(\tilde{w}_i) - \hat{\beta}_{i*})^2 + (\hat{\beta}_{i*} - \hat{\beta}^{IVBMA})^2\right]\right\}, \tag{A-4}$$

which results since

$$\sum_{j=1}^{J}\nu_j(\tilde{w}_i)(\hat{\beta}^{(j)}(\tilde{w}_i) - \hat{\beta}_{i*})(\hat{\beta}_{i*} - \hat{\beta}^{IVBMA}) = 0. \tag{A-5}$$

Reordering the terms we then receive,

$$\sigma^2_{IVBMA}(\beta) = \sum_{i=1}^{I}\pi_i Var(\beta|M_i) + \sum_{i=1}^{I}\pi_i(\hat{\beta}_{i*} - \hat{\beta}^{IVBMA})^2, \tag{A-6}$$

as desired. □

Proof of Theorem 2: For convenience, suppose that $M_1 \in \mathcal{M}$ is the true model for the first stage. Then,

$$\pi_1 \to_p 1 \text{ and } \pi_j \to_p 0, j \neq 1 \text{ as } n \to \infty. \tag{A-7}$$

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

$$\nu_1(\tilde{w}_1) \to_p 1 \text{ and } \nu_j(\tilde{w}_1) \to_p 0, j \neq 1 \text{ as } n \to \infty. \tag{A-8}$$

Therefore,

$$\hat{\beta}^{IVBMA} = \sum_{i=1}^{I}\sum_{j=1}^{J}\pi_i\nu_j(\tilde{w}_i)\hat{\beta}^{(j)}(\tilde{w}_i) \to_p \hat{\beta}^{(1)}(\tilde{w}_1). \tag{A-9}$$

Finally consider $\hat{\beta}^{2SLS}$. We know that $\hat{\beta}^{2SLS} \to_p \beta$ by the consistency of the technique. Furthermore, since the first and second stage estimates of 2SLS are individually consistent we have $\hat{\beta}^{2SLS} \to_p \hat{\beta}^{(1)}(\tilde{w}_1)$ provided M_1 and L_1 are the true first and second stage models. Thus, $\hat{\beta}^{(1)}(\tilde{w}_1) \to_p \beta$, showing the technique is consistent. $\square$

Proof of Theorem 3: Let B^{IVBMA} be the bias of $\hat{\beta}^{IVBMA}$. Note that $B^{IVBMA} = \sum_{i=1}^{I} \sum_{j=1}^{J} \pi_i \nu_j(\tilde{w}_i) B_{ij}$ and since $B_{ij} \leq B_{**}$ for all i and j, we immediately have that $B^{IVBMA} \leq B_{**}$, with equality only when $\pi_* = 1$, where M_* denotes the model including all X and Z variables. $\square$