GRIPS Discussion Paper 15-16

Endogeneity and Panel Data in Growth Regressions: A Bayesian Model Averaging Approach

Roberto Leon-Gonzalez
Daniel Montolio

October 2015

GRIPS
NATIONAL GRADUATE INSTITUTE FOR POLICY STUDIES

National Graduate Institute for Policy Studies
7-22-1 Roppongi, Minato-ku,
Tokyo, Japan 106-8677

Endogeneity and Panel Data in Growth Regressions: A Bayesian Model Averaging Approach

Roberto León-González[a,b,*] and Daniel Montolio[c,d]

This version: July 2015
(First version: August 2012)

Abstract. Bayesian model averaging (BMA) has been successfully applied in the empirical growth literature as a way to overcome the sensitivity of results to different model specifications. In this paper, we develop a BMA technique to analyze panel data models with fixed effects that differ in the set of instruments, exogeneity restrictions, or the set of explanatory variables in the regression. The large model space that typically arises can be effectively analyzed using a Markov Chain Monte Carlo algorithm. We apply our technique to investigate the effect of foreign aid on per capita GDP growth. We show that BMA is an effective tool for the analysis of panel data growth regressions in cases where the number of models is large and results are sensitive to model assumptions.

[a] National Graduate Institute for Policy Studies (GRIPS)
[b] Rimini Centre for Economic Analysis, Italy.
[c] University of Barcelona
[d] Barcelona Institute of Economics (IEB)

Keywords: Bayesian Model Averaging, Instrumental variables, Panel Data, Empirics of Growth, Effectiveness of Aid.

JEL Classification: O1, O2, O4, C30, C11

*Corresponding author: Roberto León-González, rlg@grips.ac.jp.

1 Introduction

The technique of Bayesian model averaging (BMA) was popularized as a method to overcome model uncertainty in growth regressions by Fernández et al. (2001a, 2001b) and Sala-i-Martín et al. (2004). It was proposed as a method to overcome the sensitivity of results with respect to the set of explanatory variables that is included in a regression. Since then, BMA has been applied widely in the empirical growth literature (e.g., Durlauf et al., 2008; Prüfer and Tondl, 2008; Winford and Papageorgiou, 2008; Ciccone and Jarocinski, 2010; Crespo-Cuaresma et al., 2011) and in other areas of economics (e.g., Koop and Tole, 2003; Tobias and Li, 2004). Recent papers have contributed towards the development of summary measures of the output (Ley and Steel, 2007; Doppelhofer and Weeks, 2009); led to greater understanding of prior assumptions (e.g., Ley and Steel, 2009; 2012); and extended the technique in ways that are relevant to growth regressions such as threshold models (Crespo-Cuaresma and Doppelhofer, 2007), heteroscedasticity (Doppelhofer and Weeks, 2008), endogeneity (Cohen-Cole et al., 2009; Lenkoski et al., 2014; Koop et al., 2012; Karl and Lenkoski, 2012), and panel data models (León-González and Montolio, 2004; Chen et al. 2009; 2011; Moral-Benito, 2012; 2014).

As has been well documented, some of the regressors in growth regressions may be endogenous. This problem is particularly relevant in empirical studies looking at the impact of foreign aid on the economic growth of developing countries. The importance of the policy implications that can be derived from such studies makes the issue of properly addressing causality between aid and growth the cornerstone of this literature. Only robust and reliable results can be transferred into effective foreign aid policies that ultimately contribute towards the development of countries and their citizens.

However, while most previous research in growth regressions has dealt with uncertainty regarding the set of explanatory variables in the regression, little attention has been given to the uncertainty regarding the choice of instruments and exogeneity restrictions. In this sense it is also noteworthy that empirical results can be greatly affected by the choice and number of instruments that are used to tackle the endogeneity problem, as we further illustrate in Section 5. Moreover, although in a panel data context instruments can be easily constructed using lags, it has been argued that it is not good practice to use the whole set of available instruments (e.g., see Roodman, 2009a). As a consequence, there are no clear guidelines to choose among models with different sets of identifying restrictions.

In this paper we develop a new BMA strategy to deal with a model space that includes models that differ in the set of regressors, instruments, and exogeneity restrictions in a panel data context. To deal with the large number of models that arise in a typical application (in our application, we deal with approximately 2^{46} models) we use the reversible jump algorithm developed by Koop et al. (2012, KLS henceforth) for BMA in the instrumental variable regression model. We show how this framework can be adapted to deal with dynamic panel data models with endogenous (or predetermined) regressors and the large instrument set that is typically available in a panel data context (e.g., Arellano and Bond, 1991). Our work differs from other BMA analyses of dynamic panel data models with endogenous regressors (i.e. Chen et al. (2009, 2011) and Moral-Benito (2014)) in that we allow uncertainty regarding not only the set of controlling variables but also in other dimensions such as the set of instruments and exogeneity restrictions of regressors. Although Moral-Benito (2014) does not explicitly use instruments, the reduced form for endogenous regressors is written as an autoregressive process with the maximum number of lags. Because the likelihood could also be specified with a smaller number of lags, the problem of choosing the number of lags arises and this problem is analogous to choosing the set of instruments. Similarly, Chen et al. (2009, 2011) use GMM style moment restrictions and choosing the set of moment restrictions is also analogous to choosing the set of instruments.

We make use of the original dataset of Burnside and Dollar (2000, BD henceforth), as extended by Easterly et al. (2004, ELR henceforth), who used instrumental variable regression to analyze the impact of foreign aid (an endogenous regressor) on the per capita GDP growth of developing countries. The work of BD generated a lot of interest and was followed by a large number of papers that (using different estimation methods, set of control variables/instruments, definition of variables, slightly different datasets, etc.) found similar (e.g., Collier and Dollar, 2002) and sometimes different results (e.g., Hansen and Tarp, 2001). Furthermore, it still generates open debate in the aid effectiveness literature today.[1]

Our novel methodology provides a useful diagnostic tool to study whether foreign aid increases the growth rate of per capita GDP, thereby contributing substantially to the aid effectiveness debate. Furthermore, we show what we can learn from the approach adopted in BD if we appropriately consider the problem of model uncertainty in the set of regressors, in the exogeneity restrictions, and in the choice of instruments typically used in panel data growth regressions with fixed effects. We find that there is no strong evidence that foreign aid increases the growth rate of per capita GDP, not even when interacted with an index of good policies. Instead we find that good policies such as low inflation and openness have a clear role in improving the economic growth of recipient countries.

The paper is organized as follows. Section 2 describes the context of our

[1]For a broad review of this literature, see for example, the meta-analysis produced by Doucouliagos and Paldam (2009, 2010).

empirical contribution in reference to previous literature. Section 3 describes the model space in the context of panel data growth regressions with endogenous regressors. Section 4 briefly presents the main concepts regarding prior/posterior probabilities and computation. Section 5 presents the results of applying our new BMA strategy. Finally, Section 6 concludes.

2 Aid, Policies, and Economic Growth

The impact of foreign aid and macroeconomic policies on economic growth is still an interesting and open debate, in both developing countries (recipients) and developed countries (donors). The seminal work by BD became influential because of the policy implications of their results, which could be summarized as follows: donor countries should direct aid to developing countries with "good" macroeconomic (i.e. fiscal, monetary, and trade) policies. The "policy selectivity" result in BD has been questioned by various authors and for different reasons: i) data issues; ii) selection of regressors and of instrumental variables (when endogeneity of aid is accounted for); and iii) the econometric technique chosen. As Roodman (2007) states, "*The diversity of conclusions within this literature, arising from roughly similar specifications applied within the same data universe, alone suggests that many of the results in question are fragile. That should concern policymakers and researchers alike. Yet among research papers favoring one story or another, robustness testing is rare.*"

Because of the high relevance of this topic, and the controversy of some of the results found, the literature dealing with the effectiveness of foreign aid is vast and shows often contradictory results, which is a sign that the debate is still open and demands further research. The literature is so extensive that it is already the focus of meta-analysis techniques (e.g. Mekasha and Tarp (2013), Doucouliagos and Paldam (2009, 2010, 2015)). In a recent meta-analysis exercise Doucouliagos and Paldam (2015) report that the empirical aid effectiveness literature consists of more than 200 papers containing more than two thousand estimates of aid effectiveness with results that vary greatly. They found that this literature suffers from publication selection bias, distorting somehow the ultimate aim of such literature: to guide important policy decisions regarding where to invest the resources devoted to development aid. However, the controversy has always been present in this literature since the end of the 2WW and following one of the early influential studies on the subject by Mosley (1980) many improvements have been introduced to overcome some of the criticisms to the empirical estimates. Many improvements have been introduced, such as enhanced data sets (covering more countries and time periods); consideration of new variables as potential growth determinants (especially after the development of the so-called endogenous growth models), or new econometric techniques (as identification strategies) such as 2SLS or GMM estimation to tackle crucial problems of endogeneity that, if present, invalidate the policy implications of the results. In this line the influential contribution of BD reported evidence suggesting that the effect of aid is delimited by macroeconomic policies on inflation, budget deficit and the degree of openness of the recipient country. Furthermore, they found that aid alone does not seem to be a growth determinant for developing countries but it becomes significant when interacted with an index representing the macroeconomic policies of the recipient country. The finding had clear policy implications: aid stimulates economic growth in countries with "good policies" or with "good policy environments". This claim has been confirmed by some authors such as Collier and Dehn (2001) or Collier and Dollar (2002), but it has been also challenged by Dalgaard and Hansen (2001), Guillaumont and Chauvet (2001), Hansen and Tarp (2001), Lu and Ram (2001), Roodman (2007) and Easterly, Levine and Roodman (2004).

BD used a general model with aid and quadratic aid variables interacted with a policy index. This brought to the empirical estimation further regressors, which in turn gave rise to a concern about the robustness of the model specification. Indeed, using a similar framework, Dalgaard and Hansen (2001) (DH hereafter) provided a critical analysis of the growth regressions in BD. Firstly, they studied the theoretical relation between foreign aid and economic growth in a modified neoclassical growth model. Secondly, using the same database as BD, they emphasized that the crucial interactions between aid and good policies in the growth regression are fragile, being extremely data dependent, and concluding: "*It appears that 5 influential observations, which are excluded in Burnside and Dollar's preferred regressions, have a critical influence on the parameter of main interest. In a simple counter example it is shown that one may, on an equally valid statistical basis i.e., excluding 5 influential observations, claim that aid spurs growth unconditionally.*" Moreover, they highlight the possible endogeneity problem of aid and the importance of the choice of instruments in a 2SLS estimation.

Therefore, from our point of view, the literature on foreign aid and policy effectiveness warrants for improvements on the econometric methodology, such as the BMA approach developed in this paper, so as to be able to provide robust evidence that can reliably inform policy making.

3 Specification of Models

Let the per capita GDP growth rate of country i in period t, g_{it}, depend on strictly exogenous regressors (x_{it}), a set of possibly endogenous or predetermined regressors (y_{it}) and a fixed effect (f_i):

$$g_{it} = f_i + \gamma' y_{it} + \beta' x_{it} + u_{it} \qquad t = 1, ..., T_i \qquad i = 1, ..., N \tag{1}$$

where $g_i : 1 \times 1$, $y_i : m \times 1$ and $x_{it} : k_X^j \times 1$. The subindex j stands for the j^{th} model, and j varies from 1 to N^{mod}, where N^{mod} is the total number of models. For simplicity of notation, we do not attach j subindices to any of the parameter matrices (e.g. β) although their dimension varies over models. The only exception on this is the vector γ, whose dimension is always m, even though some of its elements might be restricted to be zero. In this way we can keep the dimension of y_{it} and the corresponding number of equations in the system of equations described below constant over models. The purpose of this notation is to make clear that each model conditions on the same set of observed data, which is a requirement for correct Bayesian Model Averaging. In our empirical application the vector y_{it} will contain the begining of period log of per capita GDP, aid, the squared of aid, and the interaction of aid with other regressors.

In order to eliminate the fixed effect, we use the forward orthogonal deviations operator (Arellano, 2003; p. 17), which applied to a variable u_{it} gives by definition

$$u_{it}^* = \left(\frac{T_i - t}{T_i - t + 1} \right)^{1/2} \left[u_{it} - \frac{1}{T_i - t} (u_{i,(t+1)} + ... + u_{iT_i}) \right].$$

Applying this operator to equation (1) yields

$$g_{it}^* = \gamma' y_{it}^* + \beta' x_{it}^* + u_{it}^* \qquad t = 1, ..., T_i - 1. \tag{2}$$

An advantage of this transformation over taking first differences is that if u_{it} is homoskedastic with no serial correlation, so is u_{it}^*. However, this transformation also introduces endogeneity in a dynamic model such as ours. To see this, note that if y_{it} contains the beginning of period log GDP, then y^*_{it} will necessarily be correlated with u^*_{it}, even if y_{it} is not correlated with u_{it}. For this reason we treat the (orthogonally transformed) beginning of period GDP as an endogenous variable. We show in the appendix that from a Bayesian perspective, this transformation arises from integrating out the individual effects from the posterior density in a system of equations using a flat prior for the individual effects. Note that we assume homoskedasticity for the vector $(u_{it}, v'_{it})'$, where v_{it} is the error term in the reduced form equations for y_{it}.

$$\Sigma = E\left(\begin{pmatrix} u_{it} \\ v_{it} \end{pmatrix}\begin{pmatrix} u_{it} \\ v_{it} \end{pmatrix}'\right) = E\left(\begin{pmatrix} u^*_{it} \\ v^*_{it} \end{pmatrix}\begin{pmatrix} u^*_{it} \\ v^*_{it} \end{pmatrix}'\right)$$

Hayashi and Sims (1983) used the forward orthogonal transformation in a time series model and proposed instrumental variable estimation with predetermined instruments. A predetermined instrument z^p_{it} is assumed to be uncorrelated with current and future values of u_{it} (and therefore, uncorrelated also with u^*_{it}), but allowed to be correlated with past values of u_{it} (and u^*_{it}). The correlation of z^p_{it} with past values of u_{it} affects neither the consistency nor the asymptotic variance of the instrumental variable estimator. Thus, for our purposes, we use the Bayesian analogue of the 2SLS/LIML estimators by adding auxiliary equations for y^*_{it} as follows:

$$\begin{aligned} g^*_{it} &= \gamma' y^*_{it} + \beta' x^*_{it} + u^*_{it}, \qquad (3) \\ y^*_{it} &= \Pi_{2x} x^*_{it} + \Pi_{2z^s} z^*_{it} + \Pi_{2z^p} z^p_{it} + v^*_{it}, \end{aligned}$$

where z^*_{it} are strictly exogenous instruments (in forward orthogonal deviations) and z_{it}^p are predetermined instruments. Strictly exogenous instruments are those that are uncorrelated with $(u_{it}\ ,v_{it})$ at all lags and leads. Therefore, the assumption needed for strictly exogenous instruments to be valid is stronger than for predetermined instruments, but note that our approach is still valid if only predetermined instruments are available. Note that z_{it}^p appears in levels in (3), so that the assumption $cov(z_{it}^p, u_{it}^*) = 0$ is satisfied because z_{it}^p is predetermined[2]. In contrast, the assumption $cov(z_{it}^*, u_{it}^*) = 0$ requires z_{it} to be strictly exogenous. We also assume that x_{it} is strictly exogenous, which implies that x_{it}^* is uncorrelated with $(u_{it}^*\ ,v_{it}^*)$ at all lags and leads.

Even though our instrument set includes predetermined instruments, we form the likelihood function of the model defined by equations in (3) (which we refer to as the pseudo-likelihood function, as in Gourieroux et al. 1984) as if (z_{it}^*, z_{it}^p) were uncorrelated with $(u_{it}^*,\ v_{it}^*)$ contemporaneously and at all lags and leads. The limited information maximum likelihood (LIML) estimator that maximizes the pseudo-likelihood of equation (3) has been proposed by Alonso-Borrego and Arellano (1999) to obtain estimates in the dynamic linear panel data model[3]. Thus, the pseudo-likelihood function that we use is a proper density function for the data, which is asymptotically maximized at the true value of the parameters. In the case in which there are no dynamics and no predetermined instruments in the model, the pseudo-likelihood function corresponds to a likelihood function in the strict sense. Previous papers offer alternative ways of specifying a likelihood function for the dynamic linear panel data model (e.g. Hsiao et al. (1999, 2002),

[2]In the appendix we show that z_{it}^p can be written as the forward orthogonal transformation of a vector $z_{it}^{(-p)}$, such that $z_{it}^p = z_{it}^{(-p)*}$. The value of $z_{it}^{(-p)}$ is a function of current and past values of z_{it}^p, that is, a function of $(z_{it}^p, z_{i(t-1)}^p, ..., z_1^p)$.

[3]Alonso-Borrego and Arellano (1999) argue that this estimator is more efficient than GMM alternatives and show that it can be made robust to heteroskedasticity (see also Arellano (2003, p.p. 169-174)). However, as mentioned before, in our Bayesian approach we will use the assumption of homoskedasticity for simplicity.

Lancaster (2002), Juárez and Steel (2010) and Moral-Benito (2012, 2014)). The advantage of the approach in our paper is that the pseudo-likelihood that we use has a simpler form, implying that the conditional posterior densities of the parameters belong to standard families of distributions (such as normal or inverted Wishart). This in turn implies that we can carry out BMA over a much larger model space than in previous papers (e.g. Moral-Benito (2012, 2014) and Chen et al. (2009, 2011)) and thus evaluate the validity of intruments and exogeneity restrictions.

The predetermined instruments z_{it}^{p} typically include lags of y_{it}. If y_{it} is, for instance, the beginning of period GDP then we have that $cov(y_{it}, u_{it}^{*}) = 0$ and one could choose $z_{it}^{p} = y_{it}$. If y_{it} is, however, foreign aid, we might have that $cov(y_{it}, u_{it}^{*}) \neq 0$ but still be able to assume that $cov(y_{i,(t-1)}, u_{it}^{*}) = 0$. In that case, we can fix $z_{it}^{p} = y_{i,(t-1)}$. This is the method suggested by Anderson and Hsiao (1982) to select instruments. However, later work suggested that more efficient estimates might be obtained by using a larger number of moment conditions. Various studies used further lags as instruments in a GMM framework (e.g., Holtz-Eakin et al. (1988) and Arellano and Bond (1991)). However, using further lags in our framework might imply losing time observations. To avoid this, we follow the strategy in Roodman (2009b), and define GMM-style predetermined instruments as follows:

$Zy_{it}^{1,0} = y_{i,t}$ if $t = 1$, and 0 otherwise;

$Zy_{it}^{2,0} = y_{i,t}$ if $t = 2$, and 0 otherwise;

$Zy_{it}^{2,1} = y_{i,t-1}$ if $t = 2$, and 0 otherwise;

$Zy_{it}^{3,0} = y_{i,t}$ if $t = 3$, and 0 otherwise;

$Zy_{it}^{3,1} = y_{i,t-1}$ if $t = 3$, and 0 otherwise;

$Zy_{it}^{3,2} = y_{i,t-2}$ if $t = 3$, and 0 otherwise;

and in general,

$Zy_{it}^{h,l} = y_{i,t-l}$ if $t = h$, and 0 otherwise, for $h = 1, ..., T_i - 1$ and $l = 0, ..., h-1$.

Thus, by using $Zy_{it}^{h,l}$ as instruments we can effectively use further lags as instruments without reducing the size of the sample. The use of these instruments mimics the common practice in GMM of creating a moment condition $E(u_{it}^{*}y_{i,(t-l)}) = 0$ separately for each period t and for each lag l. To see this, note that our likelihood embeds the assumption $E(u_{it}^{*}Zy_{it}^{h,l}) = 0$, whose sample analogue is

$$\sum_{i,t} u_{it}^{*}Zy_{it}^{h,l} = \sum_{i} u_{ih}^{*}y_{i,h-l} = 0,$$

which is also the sample analogue of the GMM moment condition ($E(u_{ih}^{*}y_{i,(h-l)}) = 0$). As for the model space, it is defined as follows.

- Set of instruments: the strictly exogenous instruments in z_{it} are a subset of a larger group of potential instruments denoted by Z^{*}. There is uncertainty as to which subset of Z^{*} should be entered in the model, and hence, uncertainty about the column dimension of matrix $\Pi_{2z^{s}}$. The predetermined instruments z_{it}^{p} are a subset of a larger group of predetermined instruments denoted by Z^{*p}. Thus there is also uncertainty regarding the column dimension of $\Pi_{2z^{p}}$.

- Restrictions on the coefficients of exogenous regressors: the strictly exogenous regressors x_{it} are a subset of X^{*}. Thus there is uncertainty regarding the dimension of β. Note that in this framework we are not allowing instruments to be entered in x_{it}.

- Restrictions on the coefficients of endogenous regressors: some coefficients in γ are restricted to be zero in some models. As explained before the parameter vector γ has the same dimension in all models, so that y_{it} has also always the same dimension.

- Exogeneity restrictions: some of the covariances between u^*_{it} and v^*_{it} are restricted to be zero in some models.[4]

Note that this framework allows us to evaluate whether the instruments are weak or strong. If an instrument was uncorrelated with the endogenous regressor (thus violating one of the conditions for being a valid instrument), it will not appear in (z_{it}, z^p_{it}), thus dropping out completely from the model. However, we are always assuming that all instruments are uncorrelated with u_{it}, so we are not testing the exogeneity of the instruments.

In our empirical application we use 35 GMM-style predetermined instruments: 20 using current values and lags of *gdp* (i.e., beginning of period log GDP per capita) and 15 using lags of *eda* (i.e., foreign aid over GDP). The number of models can be calculated as $2^{2m}N^B$, where N^B is defined as

$$N^B = 2^{k^T_X} \sum_{j=m}^{k^T_Z + k^T_{ZP}} C_j^{k^T_Z + k^T_{ZP}},$$

where k^T_Z is the number of elements in Z^*, k^T_{ZP} is the number of elements in Z^{*p} and k^T_X is the number of elements in X^*. The notation C^n_j refers to the number of sets of j elements chosen without replacement from a set of n elements. In our empirical application we have 5 endogenous regressors (in y, and thus, $m = 5$), 42 potential instruments (all of them in Z^{*p}), and 9 exogenous regressors (X^*). Note that the instruments z^p_{it} are entered into the system of equations (3) without being transformed into orthogonal deviations. We have fewer regressors than in typical

[4]Note that the variance-covariance matrix of $(u_{it}, v'_{it})'$ is the same as that of $(u^*_{it}, v^{*\prime}_{it})'$. Thus, restrictions on the covariances betwen u_{it} and v_{it} are equivalent to restrictions on the covariances between u^*_{it} and v^*_{it}.

cross-section growth regressions because the time-invariant regressors drop out when we take orthogonal deviations. Even then, because of the large number of instruments, the number of models increases substantially. Taking into account that we force the beginning of period GDP to be endogenous and to be entered in all possible models, and that we also force time dummies to enter all models, the total number of possible models is of the order of 2^{46}.

4 Bayesian Model Averaging: Priors, Posterior Model Probabilities, and Computation

The BMA methodology has a number of optimal theoretical properties under the assumption of the correct specification of the prior for parameters (for a brief review see e.g. Raftery and Zheng (2003)). For example, BMA point estimates minimize the Mean Squared Error (MSE), BMA estimation intervals have correct coverage and the out of sample predictive performance of BMA is optimal in the log score sense. Furthermore, even without assuming that the prior distribution of parameters is correct, numerous Monte Carlo simulations (e.g. Fernandez et al. (2001b), Ley and Steel (2009, 2012)) and forecast evaluations with real data (e.g. Madigan et al. (1995), Hoeting et al. (2002), Wright (2008)) have confirmed that BMA performs satisfactorily and often better than competing methods.

In our case the prior for the set of parameters Θ within a model M_j, which we denote by $\pi(\Theta|M_j)$, involves normal and inverted Wishart densities and it is described in detail in the appendix. The marginal likelihood given model M_j is defined as

$$\pi(Y|M_j) = \int \pi(Y|\Theta, M_j)\pi(\Theta|M_j)d\Theta,$$

where Y represents all observed data and $\pi(Y|\Theta, M_j)$ is the likelihood. The weights for Bayesian model averaging are equal to the posterior model probabilities, which are defined as

$$\pi(M_j|Y) = \frac{\pi(M_j)\pi(Y|M_j)}{\sum_j \pi(M_j)\pi(Y|M_j)}, \tag{4}$$

where $\pi(M_j)$ is the prior probability of model M_j and the summation is over the whole model space.

It has been shown that the posterior probabilities could be sensitive to the value of the prior variance of parameters (e.g. Marin and Robert 2007, ch. 3). In our setup the prior variance depends on two scalars: $\underline{g}$ and $\underline{g}_e$. The parameter $\underline{g}$ controls the prior variance of slope parameters whereas $\underline{g}_e$ controls the prior variance of the covariances between u_{it}^* and v_{it}^*. To reduce the sensitivity of results to prior assumptions, we follow the recommendations in Ley and Steel (2012) to specify hyper-priors on $\underline{g}$ and $\underline{g}_e$ (see appendix for details). In our empirical application we also verify the robustness of the results with respect to different hyper-priors on $\underline{g}$ and $\underline{g}_e$.

The setup in this paper allows for any prior probability for models $\pi(M_j)$ that has been proposed in the previous literature. In our empirical application we follow the approach of Ley and Steel (2009) which first defines the prior probability (θ) of a restriction and then uses a hyper-prior on θ ($p(\theta)$) (see appendix for details). Compared to fixing θ to a particular value, this strategy relaxes the prior information on model size and tends to increase the predictive performance of the BMA. In our empirical application we carry out sensitivity analysis with respect to the hyper-prior $p(\theta)$.

The prior on θ ($p(\theta)$) has a convenient form (i.e. beta distributions) such that it is possible to obtain an analytical expression for $\pi(M_j)$ by performing the integral:

$$\pi(M_j) = \int \pi\left(M_j|\theta\right) p(\theta) d\theta \tag{5}$$

However, there are two challenges regarding computation. First, the number of models in our empirical application is very large and of the order of 2^{46}. Second, there is no analytical expression for the marginal likelihood $\pi(Y|M_j)$, which could only be calculated using computationally intensive numerical methods. That is, not only the number of terms to be calculated in the denominator of equation (4) is too large but also the calculation of each term is computationally expensive. To surmount these problems, we use the reversible jump algorithm proposed by KLS.[5] This algorithm is a Markov Chain algorithm that iteratively samples values for parameter Θ and model M_j. Given arbitrarily fixed initial values for (Θ, M_j), after a sufficient number of iterations, the generated values can be used as a sample from the posterior of (Θ, M_j). This sample is used to calculate quantities of interest such as posterior model probabilities (using the proportion of times that the chain visits a particular model) and credible intervals for parameters.

5 Results

To apply our new methodology to the aid and growth literature, we use the data from ELR, who updated the original dataset from BD from 1970-93 to 1970-97, as well as fill in missing data for the original period, 1970-93. Thus, we are using 7 four-year periods, denoted as $t = 2, ..., 8$. Table A.1 in the data appendix gives the variable/instrument definitions and the group to which they belong (i.e., Z^*, Z^{*p}, X^*, or y). In addition to all the time-variant regressors in BD, we also include two more regressors proposed by Dalgaard and Hansen (2001): the policy index

[5]We adapted this algorithm to allow $(\underline{g}, \underline{g}_e)$ to be random. The computer code was written in GAUSS language and can be downloaded freely at the authors' website.

(see below for its definition) squared and aid squared. As for instruments, we use the set of time-variant instruments in BD and we add predetermined instruments using lags of aid and log GDP per capita (see Table A.1 for details).

The vector of potentially endogenous regressors y includes aid (eda), squared aid and interactions of these two variables with the policy index. The squared term of aid allows for a non-linear impact of aid on growth whereas the interaction terms imply that the effectiveness of aid might depend on the value of the policy index. The variable aid might be endogenous because donors might react to unexpected recessions or expansions in the recipient country within the same period for humanitarian or strategic and commercial purposes (e.g. BD). The list of potential instruments are variables thought a priori to be correlated with aid allocation but not to have a direct impact on growth.

Recall that we include all the (time-variant) instruments in Z^{*p} (and hence none of the instruments are transformed into orthogonal deviations). We force the time dummies to be entered in all models.[6] We run each BMA separately for the whole sample and for the sample of low income countries. Following BD, for the latter sample we select those countries whose real GDP per capita in the year 1970 was below USD 1,900 (in constant 1985 dollars) and also Nicaragua.[7] Hence, in the full sample there are 63 countries (with 359 country-period observations), and in the low-income sample there are 44 countries (with 244 country-period observations).

Regarding the policy index (pol), we construct it following the methodology proposed by BD. BD create an index covering aspects of fiscal, monetary, and

[6]Note that system (3) does not include a constant and hence we proceed as such to avoid the model with no explanatory variables and no constant being visited by the algorithm.

[7]Although the GDP per capita of Nicaragua was over USD 1,900 in 1970, it then decreased over time and was below USD 1,900 in 1982. For this reason, BD included Nicaragua in the low-income sample.

trade policies. Fiscal policy is measured by the budget surplus over GDP (bb). The success or failure of monetary policy is measured by the level of inflation (infl), while trade policy is represented by a binary (0/1) openness indicator (sacw) constructed by Sachs and Warner (1995). To avoid collinearity problems, BD create an index using a weighted average of the three measures. The weights for the policy index are the estimated coefficients in a regression of GDP growth on the three measures and other exogenous regressors. Following this methodology, we construct two policy indices: one for the whole sample and another for the sample of low-income countries.[8] The index of policy is increasing with budget surplus and trade openness but decreasing with inflation.

Before we apply the BMA methodology, we show an example of the sensitivity of results to model specification. Tables 1 and 2 correspond to the GMM estimation of the dynamic panel model with fixed effects using alternative specifications for the set of regressors and instruments. The specifications differ on whether eda2 or polaid are included as regressors, differ also on the number of lags used as instruments (i.e. one, two or all possible lags) and on whether lags of eda2 are used as instruments in addition to the lags of eda. Table 1 shows system GMM estimates (Blundell and Bond (1998)) whereas Table 2 shows difference GMM estimates (Arellano and Bond (1991))[9]. The p-value of the marginal impact of aid at sample mean values changes from (0.007) to (0.58) in Table 1 and from (0.00) to (0.79) in Table 2, which illustrates that significance testing is highly sensitive to minor aspects of the model specification. Note that some models in Tables 1 and 2 have a p-value for the Hansen test equal or very close to one, which is a sign of overfitting bias caused by the use of too many instruments

[8]Given that our dataset, as previously noted, differs slightly from the original BD dataset, our policy index is also slightly different from the original one.

[9]Rajan and Subramanian (2008) use difference GMM and system GMM in an aid-growth regression and briefly discuss the relative strengths of both approaches.

(e.g. Roodman (2009a)). As we show below the BMA approach chooses few instruments, thus avoiding the overfitting bias.

<INSERT TABLE 1 AROUND HERE >

<INSERT TABLE 2 AROUND HERE >

We run the proposed reversible jump algorithm for 800,000 iterations after discarding the initial 30,000 iterations. As one of the checks for convergence, we estimate the total visited probability (George and McCulloch, 1997), which is an estimate of the proportion of the total probability mass that is visited by the algorithm.[10] This is over 98% in all cases, indicating good convergence. In addition to calculating the posterior model probabilities using the relative frequency of visits of the algorithm, we construct it by numerically calculating the marginal likelihood of each model visited by the algorithm. As a measure of convergence (Fernández et al., 2001b), we calculate the correlation between the two measures. It is over 0.86, again indicating good convergence. We also carry out several runs with randomly chosen initial values and obtain the same results. We perform some sensitivity analysis with respect to the prior specification and the results that we report do not change qualitatively (please see below for further details on the prior sensitivity analysis).

The best 10 models only accumulate about 5% probability, indicating that there is a great deal of model uncertainty. The number of models visited by the chain was 11748.

We first report the BMA estimates of the first derivative of the growth rate with respect to aid (g_A), the policy index (g_P), and the logarithm of the beginning of period GDP per capita ($g_{initial}$) in Table 3. Because some potential regressors

[10]See the technical appendix for details of how this total visited probability is actually constructed.

are defined as interactions of other regressors, these partial derivatives might depend on several parameters in the model and on the value of some regressors. In that case we evaluate the partial derivatives at the sample average of the regressor. Because g_A depends on the coefficients of several regressors, the probability that g_A is different from 0 cannot be derived directly using only the probabilities of inclusion of individual regressors. Instead, as an approximation to this probability, we calculate the percentage of times that g_A becomes 0 when using the draws from the BMA algorithm. As shown in Table 3, the posterior probability of g_A being zero is 68% when using the sample of low income countries but only 9% when using the whole sample. However, in this latter case the sign of g_A cannot be established with certainty, because the probabilities of g_A being positive or negative are both substantial (67% and 24%, respectively). Furthermore, in our sensitivity analysis we find that when using an inverse gamma prior, instead of an inverse beta prior, for the parameters $\underline{g}$ and $\underline{g}_e$ (as in Zellner and Siow (1980)), the probability of g_A being greater than 0 drops to 18%, while the probability of g_A being equal to zero increases to 82%. Therefore, we find no strong evidence to conclude that g_A is positive with either the low-income sample or the whole sample of countries. In contrast, the posterior probability of g_P being positive is close to 100% in all cases, which confirms that good policies have a positive impact on economic growth.

<INSERT TABLE 3 AROUND HERE >

The probabilities of $g_{initial}$ being positive or negative are both substantial (never smaller than 40% each), which indicates that the sign cannot be established with certainty. Note that we are forcing the beginning of period GDP to be entered in all models, and that these results apply to the sample period in question, as in the long run we would expect $g_{initial}$ to be negative, so as to ensure the convergence of the GDP processes among countries.

We also report g_{AP}, which is the cross derivative of growth with respect to aid and policies. This derivative measures the extent to which a higher policy index increases the effectiveness of aid. The posterior probability of g_{AP} being equal to 0 is higher than 97% in all cases. The results are similar to those found by Eris (2008) who applied BMA to the dataset of BD assuming all regressors to be exogenous in a pooled regression.

The marginal impacts presented in Table 3 are consistent for the two sub-samples used: all countries and low-income countries. Therefore, our analysis fails to find strong evidence of aid being effective, even when interacted with the so-called "good policies."In contrast, we find strong evidence that good policies themselves matter for growth; good policies in the spirit of BD. Given that one of the main results of our empirical exercise is that macroeconomic policy making really matters for economic growth, we perform a final robust estimation. Table 4 presents the BMA estimates for a reduced form equation for growth in which the policy index components are entered as separate regressors. Note that the signs are as expected and that the two policy variables with higher posterior probabilities of inclusion are inflation and trade openness. Thus, good policies in our context should be mostly understood as policies that relate not so much to budget surplus but to inflation and trade openness.

<INSERT TABLE 4 AROUND HERE >

Tables 5 and 6 show more detailed output from the BMA estimation (which was used to compute the marginal effects in Table 3). BMA has a preference for parsimony. The regressors with posterior inclusion probability close to one are policy (pol), policy squared (pol2), and m21 (lag M2 over GDP). However, the sign of the coefficient of m21 cannot be established with certainty, as the 95% credible interval contains both positive and negative values. The significance of the policy index squared indicates that the impact of inflation (and budget deficit) is non-linear, possibly capturing threshold effects.

<INSERT TABLE 5 AROUND HERE >

<INSERT TABLE 6 AROUND HERE >

In Table 6, we observe that out of the large number of potential instruments, only a few are chosen, and among these very few were constructed with lags. The GMM-style instruments that are chosen are the nearest available lags (e.g. $Zgdp^{3,0}$, $Zgdp^{3,1}$, $Zgdp^{6,0}$, $Zgdp^{6,1}$, $Zeda^{4,1}$, $Zgdp^{4,1}$ (low-income), $Zgdp^{4,2}$ (low-income) and $Zeda^{5,2}$ (all countries)), which are normally more strongly correlated with the endogenous regressors than further lags. Using the full sample the average number of instruments in the models visited by the algorithm was 13.5, and the maximum number of instruments chosen for a model was 20. In the low income countries sample the average number of instruments was slightly smaller (12.8) and the maximum number of instruments chosen was 17. This fits well with recent literature that concludes that models that use fewer but strong instruments are better for inference (e.g., see Roodman, 2009a for a review of this literature).

We carried out several robustness checks, finding that the main results did not change qualitatively. Regarding the specification of the prior we tried both inverse gamma and inverted beta distributions (with different prior means and variances) for the parameters on $\underline{g}$ and $\underline{g}_e$. We also carried out a BMA analysis that fixed θ and ($\underline{g}$, $\underline{g}_e$) to particular values (as opposed to being estimated).[11] This changed the evidence on endogeneity, but the main conclusions shown in

[11] We fixed θ equal to 1/2, so that all models become equally likely a priori, and $\underline{g} = \underline{g}_e = N\bar{T} = \sum_{i=1}^{N}(T_i - 1)$. We also tried other values such as $\underline{g} = \underline{g}_e = (N\bar{T})^2$ and $\underline{g} = \underline{g}_e = (N\bar{T})^3$.

Table 3 were unchanged. We also carried out the estimation with a different policy index, which was constructed using the weights obtained from the BMA estimation of the reduced form growth regression.[12] Finally we also carried out the BMA analysis without GMM-style instruments, using instead just one lag of gdp and eda as instruments (in the style of Anderson and Hsiao, 1982). In all of these robustness checks g_{AP} was very likely to be zero, the probability of g_A being positive was never higher than 70% whereas g_P was always positive with a probability near to 1.

6 Conclusions

BMA has been widely used in the empirical growth literature but the focus has been mostly on uncertainty regarding the set of explanatory variables. However, typical growth regressions use panel data with endogenous regressors, where the available instrument set tends to be very large. Although results could be sensitive to the instrument set chosen, there are currently no clear guidelines on how to choose the instruments. The purpose of the present paper is to develop a new BMA methodology that allows panel regression with fixed effects and endogenous regressors, while simultaneously allowing uncertainty regarding the set of instruments, regressors, and exogeneity restrictions. In our empirical application, we show that the large model space that typically arises can be effectively analyzed with the reversible jump algorithm proposed by Koop et al. (2012) and that the BMA methodology selects models with fewer but stronger instruments.

This methodology is then applied to perform an independent replication in a widely debated area of the empirics of economic growth—the impact of foreign aid on the economic growth of developing countries. By using well-known datasets,

[12] The BMA estimation of the reduced form growth regression is that shown in Table 4.

we obtain that once all the model uncertainty in growth regressions has been accounted for, we find no strong evidence that foreign aid has an impact on the growth rate of recipient countries. Moreover, the interaction term of aid with the index of good policies proposed by BD has no impact on growth. From our BMA results, it emerges that it is macroeconomic policy making that has a higher posterior probability of inclusion in a growth regression, and hence, a greater potential for explaining the GDP growth rates of developing countries.

Acknowledgments

Roberto León-González thanks the Japan Society for the Promotion of Science for financial support (Start-up Grant #20830025 and Young-Scientist (B) Grant #23730214). Daniel Montolio gratefully acknowledges support from the Generalitat of Catalonia (2014SGR420). We also thank an anonymous referee and the editor for useful comments that helped us to improve the paper.

References

Alonso-Borrego, C. and M. Arellano (1999) "Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data," *Journal of Business & Economic Statistics*, 17, 36-49.

Anderson, T.W. and C. Hsiao (1982) "Formulation and estimation of dynamic models using panel data," *Journal of Econometrics* 18, 47-82.

Arellano, M. (2003) *Panel Data Econometrics,* Oxford University Press.

Arellano, M. and S. Bond (1991) "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277-297.

Bauwens, L., M. Lubrano, and J.F. Richard (1999) *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press, Oxford.

Blundell, R. and S. Bond (1998) "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics* 87, 115-143.

Burnside, C. and D. Dollar (2000) "Aid, policies and growth," *American Economic Review*, 90, 847-868.

Chen, H., A. Mirestean, and C.G. Tsangarides (2009), "Limited Information Bayesian Model Averaging for Dynamic Panels with Short Time Periods," *IMF Working Paper*, WP/09/74. Washington: International Monetary Fund.

Chen, H., A. Mirestean, and C.G. Tsangarides (2011) "Limited Information Bayesian Model Averaging for Dynamic Panels with an Application to a Trade Gravity Model," *IMF working paper*, WP/11/230.

Ciccone, A. and M. Jarocinski (2010) "Determinants of Economic Growth: Will Data Tell?," *American Economic Journal: Macroeconomics*, 2, 222-246.

Cohen-Cole, E., S. Durlauf, J. Fagan and D. Nagin (2009) "Model uncertainty and the deterrent effect of capital punishment," *American Law and Economics Review*, 11, 335-369

Collier, P. and J. Dehn, (2001) "Aid, shocks and growth," *Policy Research Working Paper* 2608, Development Research Group, World Bank, Washington, DC.

Collier, P. and D. Dollar (2002) "Aid Allocation and Poverty Reduction," *European Economic Review*, 8, 1475-1500

Crespo-Cuaresma, J. and G. Doppelhofer (2007) "Nonlinearities in cross-country growth regressions: a Bayesian averaging of thresholds (BAT) approach," *Journal of Macroeconomics*, 29, 541-554

Crespo-Cuaresma, J., N. Fosterd and R. Stehrer (2011) "Determinants of Regional Economic Growth by Quantile," *Regional Studies*, 45, 809-826.

Dalgaard, C-J and H. Hansen (2001) "On Aid, Growth and Good Policies," *Journal of Development Studies*, 37, 17-41.

Doppelhofer G, M. Weeks (2008) "Robust Growth Determinants," *Cambridge Working Papers in Economics*, CWPE1117.

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

Doucouliagos, H. and M. Paldam (2009) "The Aid Effectiveness Literature: The Sad Results Of 40 Years Of Research," *Journal of Economic Surveys*, 23, 433-461.

Doucouliagos, H. and M. Paldam (2010) "Conditional Aid Effectiveness: A Meta-Study," *Journal of International Development*, 22, 391-410.

Doucouliagos, H. and M. Paldam (2015) "Finally a breakthrough? The recent rise in the size of the estimates of aid effectiveness," in Mak Arvin (Ed.), *Handbook on the Economics of Foreign Aid*, Edward Elgar, Cheltenham, Ch. 20.

Dreze, J. (1976) "Bayesian limited information analysis of the simultaneous equations model," *Econometrica*, 44, 1045-1075.

Durlauf, S.N., A. Kourtellos and C. Tan (2008) "Are Any Growth Theories Robust?," *Economic Journal*, 118, 329-346.

Easterly, W., R. Levine, and D. Roodman (2004) "Aid, Policies, and Growth: Comment," *American Economic Review*, 94, 774-780.

Eris, M. (2008) "Foreign Aid and Growth," *Economics Bulletin*, 15, 1-14.

Fernández, C., E. Ley, and M.F.J. Steel (2001a) "Model uncertainty in cross-country growth regressions," *Journal of Applied Econometrics*, 16, 563-576.

Fernández, C., E. Ley, and M.F.J. Steel (2001b) "Benchmark priors for Bayesian model averaging," *Journal of Econometrics* 100, 381-427.

George, E. and R. McCulloch (1997) "Approaches for Bayesian variable selection," *Statistica Sinica* 7, 339-373

Gourieroux, C., A. Monfort, and A. Trognon (1984) "Pseudo-maximum Likelihood Methods: Theory," *Econometrica*, 42, 681-700.

Guillaumont, P. and L. Chauvet (2001) "Aid and Performance: A Reassessment," *Journal of Development Studies*, 37, 66-92.

Hansen, L. (1982) "Large sample properties of generalized method of moments estimators," *Econometrica,* , 50, 1029-1054.

Hansen, H. and F. Tarp (2001) "Aid and Growth Regressions," *Journal of Development Economics*, 64, 545-68.

Hayashi, F. and C. Sims (1983) "Nearly efficient estimation of time series models with predetermined, but not exogenous, instruments," *Econometrica*, 51, 783-798.

Holtz-Eakin, D., W. Newey, and H. Rosen (1988) "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371-1395.

Hoeting, J.A., A.E. Raftery and D. Madigana (2002) "Bayesian Variable and Transformation Selection in Linear Regression," *Journal of Computational and Graphical Statistics*, 11, 485-507.

Hsiao, C., M.H. Pesaran, and A.K. Tahmiscioglu (1999). Bayes estimation of short run coeffcients in dynamic panel data models. In: Hsiao et al. (Eds.), *Analysis of Panels and Limited Dependent Variable Models: In Honor of G.S. Maddala.* Cambridge University Press, Cambridge, p.p. 268–296.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002) "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods," *Journal of Econometrics*, 109, 107-150.

Juárez, M.A. and M.F.J. Steel (2010) "Non-Gaussian Dynamic Bayesian Modelling for Panel Data," *Journal of Applied Econometrics*, 25, 1128-1154.

Knack, S. and P. Keefer (1995) "Institutions and economic performance: Cross-country tests using alternative institutional measures," *Economics and Politics,* 7, 207–227.

Karl, A. and A. Lenkoski (2012) "Instrumental Variable Bayesian Model Averaging via Conditional Bayes Factors," working paper, *arXiv:1202.5846v3 [stat.ME]*

Koop G., R. León-González and R. Strachan (2012) "Bayesian Model Averaging in the Instrumental Variable Regression Model," *Journal of Econometrics*, 171, 237-250.

Koop, G. and L. Tole (2003) "Measuring the Health Effects of Air Pollution: To What Extent Can We Really Say that People are Dying from Bad Air?" *Journal of Environmental Economics and Management*, 47, 30-54.

Lancaster, T. (2002) "Orthogonal Parameters and Panel Data," *Review of Economic Studies*, 69, 647-666.

Lenkoski, A., T. S. Eicher and A. E. Raftery (2014) "Two-Stage Bayesian Model Averaging in the Endogenous Variable Model," *Econometric Reviews*, 33, 122-151.

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

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

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

León-González, R. and D. Montolio (2004) "Growth, Convergence and Public Investment. A Bayesian Model Averaging Approach," *Applied Economics*, 36, 1925-1936

Lu, S. and R. Ram (2001) "Foreign Aid, Government Policies and Economic Growth: Further Evidence from Cross-country Panel Data for 1970 to 1993," *International Economics*, 54, 15-29.

Madigan, D., J. Gavrin and A.E. Raftery (1995) "Eliciting Prior Information to Enhance the Predictive Performance of Bayesian Graphical Models," *Communications in Statistics - Theory and Methods*, 24, 2271-2292.

Marin, J. and C. Robert (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. Springer-Verlag, New York.

Mekasha, T.J. and F. Tarp (2013) "Aid and Growth: What Meta-Analysis Reveals" *Journal of International Development* 49, 564-583

Moral-Benito, E. (2012) "Determinants of Economic Growth: A Bayesian Panel Data Approach," *Review of Economics and Statistics,* 94, 566-579

Moral-Benito, E. (2014) "Growth Empirics in Panel Data under Model Uncertainty and Weak Exogeneity," *Journal of Applied Econometrics*, doi: 10.1002/jae.2429

Mosley, P. (1980) "Aid, Savings and Growth Revisited" *Oxford Bulletin of Economics and Statistics*, 42, 79–95.

Prüfer, P. and G. Tondl (2008) "The FDI-growth nexus in Latin America: the role of source countries and local conditions," *CentER Discussion Paper* No. 2008-61

Raftery, A.E. and Y. Zheng (2003) "Discussion: performance of Bayesian model averaging," *Journal of the American Statistical Association* 98, 931–938.

Rajan, R.G. and Sabramanian, A. (2008) "Aid and Growth: What Does the Cross-Contry Evidence Really Show?," *Review of Economic Studies*, 90, 643–665.

Roodman, D. (2007) "The Anarchy of Numbers: Aid, Development and Cross-country Empirics," *World Bank Economic Review*, 21, 255-277.

Roodman, D. (2009a) "A Note on the Theme of Too Many Instruments," *Oxford Bulletin of Economics and Statistics*, 71, 135–158.

Roodman, D. (2009b) "How to do xtabond2: An introduction to difference and system GMM in Stata," *The Stata Journal*, 9, 86-136.

Sachs, J. D. and A. M. Warner (1995) "Economic reform and the process of global integration," *Brooking Papers on Economic Activity* 1, 1–118.

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

Tobias, L. and M. Li (2004) "Returns to Schooling and Bayesian Model Averaging: A Union of Two Literatures," *Journal of Economic Surveys*, 18, 153-180

Winford H. M. and C. Papageorgiou (2008) "Rough and lonely road to prosperity: a reexamination of the sources of growth in Africa using Bayesian model averaging," *Journal of Applied Econometrics*, 23, 671-682.

Wright, J.H. (2008) "Bayesian Model Averaging and Exchange Rate Forecasts," *Journal of Econometrics,* 146, 329-341.

Zellner, A. (1971) *An Introduction to Bayesian Inference in Econometrics.* Wiley, New York.

Zellner, A. and A. Siow (1980). Posterior odds ratios for selected regression hypotheses (with discussion). In: Bernardo, J.M., DeGroot, M.H., Lindley, D.V., Smith, A.F.M. (Eds.), *Bayesian Statistics.* University Press, Valencia, pp. 585–603.

Data Appendix

The dataset comes from Easterly, Levine and Roodman (2004), who revised and extended the dataset of Burnside and Dollar (2000). In addition to variables related to foreign aid (eda and functions of eda) and GDP (gdp, gdpg, and interactions), the dataset includes variables to control for political instability of the recipient country: ethnic fractionalization (ethnf) and assassinations (assas), and their interaction (eth_a). Moreover, the variable icrge accounts for institutional quality and it is an index based on the evaluation of five different institutional indicators. It is constructed by the private international investment risk service "International Country Risk Guide." The five indicators are as follows: Quality of Bureaucracy, Corruption in Government, Rule of Law, Expropriation Risk, and Repudiation of Contracts by Government (for more details, see Knack and Keefer, 1995). The proxy for the development of financial markets is broad money relative to GDP (m2) while the lagged value of the share of imported arms on all imports (arms) accounts for the possible existence of conflicts in recipient countries. The dataset also includes the country's population (lpop and interactions) and dummy variables for location: Sub-Saharan Africa (ssa), East Asia (easia), Central America (centam), Egypt and Franc zone (frz). However, we only use the time-variant variables that are shown in Table A.1.

Table A.1: Variables used		
Name	Brief Description	Type
gdpg	Real GDP per capita growth (%)	g
eda	Aid (% of GDP)	y
eda2	Square of eda	y
polaid	eda*policy index	y
aid2pol	eda2*policy index	y
gdp	log of real GDP per capita, beginning of period	y
ethnf	Ethnic fractionalization	X^*
assas	Assassinations	X^*
m2	Lagged M2 (% of GDP)	X^*
eth_a	ethnf*assas	X^*
pol	Policy index	X^*
pol2	Squared policy index	X^*
dum3 to dum8	Time dummies	X^*
lpop	log of population	Z^{*p}
arms	Lagged Armed Imports (% of all imports)	Z^{*p}
polarms	Policy index*arms	Z^{*p}
polpop	Policy index*lpop	Z^{*p}
polpop2	Policy index*(lpop)^2	Z^{*p}
polgdp	Policy index*gdp	Z^{*p}
polgdp2	Policy index*(gdp)^2	Z^{*p}
$Zgdp_t^{h,l}$	gdp_{t-l} if $t = h$ and 0 otherwise. For $l = 0, ..., h-2$	Z^{*p}
$Zeda_t^{h,l}$	eda_{t-l} if $t = h$ and 0 otherwise. For $l = 1, ..., h-2$	Z^{*p}

Technical Appendix

Prior specification and convergence diagnostic

Several priors for the incomplete simultaneous equation model have been proposed in the Bayesian econometrics literature. Although the KLS algorithm can be used for many of those priors, here we have used a prior using the parameterization in Drèze (1976). We define the following.

$$\Pi_x = \begin{pmatrix} \gamma'\Pi_{2x} + \beta' \\ \Pi_{2x} \end{pmatrix} = \begin{pmatrix} \pi_{1x} \\ \Pi_{2x} \end{pmatrix}, \tag{6}$$

$$\Pi_z = \begin{pmatrix} \pi_{1z} \\ \Pi_{2z} \end{pmatrix} = \begin{pmatrix} \gamma' \\ I_m \end{pmatrix} \Pi_{2z}, \qquad \Sigma = E\left(\begin{pmatrix} u_{it} \\ v_{it} \end{pmatrix} \begin{pmatrix} u_{it} \\ v_{it} \end{pmatrix}' \right),$$

$$\Omega = \begin{pmatrix} 1 & \gamma' \\ 0 & I_m \end{pmatrix} \Sigma \begin{pmatrix} 1 & 0 \\ \gamma & I_m \end{pmatrix} = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \Omega_{22} \end{bmatrix},$$

$$\omega_{11\cdot 2} = var(v_{1it}|v_{2it}) = \omega_{11} - \omega_{12}\Omega_{22}^{-1}\omega_{21},$$

$$\widetilde{\omega}_{21} = \Omega_{22}^{-1}\omega_{21}.$$

where $\Pi_{2z} = (\Pi_{2z^s}, \Pi_{2z^p})$. Let $\gamma_{\widetilde{E}}$ be a $d_{\widetilde{E}} \times 1$ vector containing the non-zero elements of γ. We specify a normal prior on $(\gamma'_{\widetilde{E}}, vec(\Pi_x)', vec(\Pi_{2z})')'$ such that $vec(\Pi_x)|\Omega \sim N(0, \underline{g}\underline{V}_{\Pi_x} \otimes \Omega)$, $\gamma_{\widetilde{E}}|\Omega \sim N(0, \underline{g}\omega_{11\cdot 2}\underline{A})$, and $vec(\Pi_{2z})|\Omega \sim N(0, \underline{g}\underline{D} \otimes \Omega_{22})$, where $(\underline{g}, \underline{V}_{\Pi_x}, \underline{A}, \underline{D})$ are prior hyper-parameters. We set $\underline{A} = I_{d_{\widetilde{E}}}$. Further, we set $\underline{V}_{\Pi_x}$ as the inverse of the cross-products of exogenous regressors in the model, and $\underline{D}$ as the inverse of the cross-products of the instruments.

Regarding the variance-covariance matrix, we fix the following prior specification on $(\widetilde{\omega}_{21}, \Omega_{22}, \omega_{11\cdot 2})$:

$$\begin{aligned} \widetilde{\omega}_{21} &\sim N(0, \underline{g}^e \omega_{11\cdot 2} I_m), \\ \Omega_{22} &\sim IW_m(\underline{S}_{22}, \underline{v}_{22}), \\ p(\omega_{11\cdot 2}) &\propto |\omega_{11\cdot 2}|^{-1}, \end{aligned} \tag{7}$$

where $IW_m(\underline{S}_{22}, \underline{v}_{22})$ represents the inverted Wishart distribution with degrees of freedom equal to $\underline{v}_{22}$ and parameter matrix $\underline{S}_{22}$ (Bauwens et al., 1999, p. 305). We set $\underline{v}_{22} = m + 1$ and $\underline{S}_{22} = \underline{g}^{-1} I_m$.

The hyper-priors for $(\underline{g}, \underline{g}^e)$ are specified by first defining the shrinkage factors $\underline{\delta} = \underline{g}/(1 + \underline{g})$ and $\underline{\delta}^e = \underline{g}^e/(1 + \underline{g}^e)$ and then assuming that $\underline{\delta} \sim B(\underline{b}, \underline{c})$ and $\underline{\delta}^e \sim B(\underline{b}^e, \underline{c}^e)$, where $B(.)$ denotes the beta distribution. In the context of a linear regression with exogenous regressors, Ley and Steel (2012) recommend choosing $\underline{c} = 0.01$ and $\underline{b} = \underline{c}\max(N, k^2)$, where k is the number of regressors whose coefficients are subject to restrictions. Accordingly, in our empirical application we chose $\underline{c} = \underline{c}^e = 0.01$ and $\underline{b} = \underline{b}^e = 0.01(N\bar{T})$, where $\bar{T} = \sum_{i=1}^{N}(T_i - 1)/N$ such that $(N\bar{T})$ becomes the number of country-period observations after taking forward orthogonal deviations. As a robustness analysis we also considered the case $\underline{c} = \underline{c}^e = 1$ and $\underline{b} = \underline{b}^e = (N\bar{T})$.

Using the Jacobian of the transformation, the beta priors on $(\underline{\delta}, \underline{\delta}^e)$ imply that the prior densities for $\underline{g}$ and $\underline{g}^e$ are inverted beta distributions (Zellner, 1971, p. 375):

$$\begin{aligned}
\pi(\underline{g}) &= \frac{\Gamma(\underline{b}+\underline{c})}{\Gamma(\underline{b})\Gamma(\underline{c})}\underline{g}^{\underline{b}-1}(1+\underline{g})^{-(\underline{b}+\underline{c})} \\
\pi(\underline{g}^e) &= \frac{\Gamma(\underline{b}^e+\underline{c}^e)}{\Gamma(\underline{b}^e)\Gamma(\underline{c}^e)}\left(\underline{g}^e\right)^{\underline{b}^e-1}(1+\underline{g}^e)^{-(\underline{b}^e+\underline{c}^e)}
\end{aligned}$$

For sensitivity analysis we also specified inverse gamma priors on $(\underline{g}, \underline{g}^e)$: $\underline{g} \sim IG_2(\underline{s}, \underline{v})$ and $\underline{g}^e \sim IG_2(\underline{s}^e, \underline{v}^e)$, where $IG_2(.)$ denotes the inverted gamma-2 distribution (Bauwens et al., 1999, p. 292). In the context of a linear regression with exogenous regressors, Zellner and Siow (1980) proposed choosing $\underline{s} = N$ and

$\underline{v} = 1$. However, in our context this prior results in a posterior with very fat tails for both $\underline{g}$ and $\underline{g}^e$. For this reason we preferred to choose $\underline{v} = 3$ (instead of $\underline{v} = 1$) and $\underline{s} = (N\bar{T})$.

To define precisely the prior model probabilities in our context, let us define first the following prior probabilities of restrictions:

- θ_X is the prior probability that a variable from X^* should be in x_{it}. That is, it is the prior probability of inclusion of exogenous regressors. Recall that k_X^T is the total number of variables in X^* and let k_X^j be the number of variables from X^* that are included in x_{it} in model M_j.

- θ_Z is the prior probability that a variable from $Z^* \cup Z^{*p}$ should be in the model (as an instrument). Let the total number of variables in $Z^* \cup Z^{*p}$ be denoted by $k_Z^T + k_{ZP}^T$ and let k_Z^j be the number of variables from $Z^* \cup Z^{*p}$ that are included in model M_j (either in z_{it} or in z_{it}^p).

- θ_Y is the prior probability that a coefficient in γ is different from zero. That is, it is the prior inclusion probability of potentially endogenous regressors. Recall that the number of potentially endogenous regressors is m and let k_y^j be the number of variables in y whose corresponding coefficient in γ is different from 0 in model M_j.

- θ_V is the prior probability that a covariance between u_{it} and v_{it} is different from zero. That is, it is the prior probability of endogeneity. Let k_v^j be the number of variables in v_{it} whose covariance with u_{it} is different from 0 in model M_j.

Using this framework, the prior probability of a model M_j conditional on the parameters ($\theta = (\theta_X, \theta_Z, \theta_Y, \theta_V)$) (i.e. $\pi\left(M_j|\theta\right)$) can be written as:

$$\begin{aligned}&(\theta_X)^{k_X^j}\left(1-\theta_X\right)^{k_X^T-k_X^j}\left(\theta_Z\right)^{k_Z^j}\left(1-\theta_Z\right)^{k_Z^T+k_{ZP}^T-k_Z^j}\times\\&(\theta_Y)^{k_y^j}\left(1-\theta_Y\right)^{m-k_y^j}\left(\theta_V\right)^{k_v^j}\left(1-\theta_V\right)^{m-k_v^j}\end{aligned}$$

We follow Ley and Steel (2009) to specify the following hyper-priors on $(\theta_X, \theta_Z, \theta_Y, \theta_V)$:

- $\theta_X \sim B(\underline{\alpha}_1, \underline{\beta}_1)$, $\underline{\alpha}_1 = 1$, $\underline{\beta}_1 = (k_X^T - \overline{m}_X)/\overline{m}_X)$, with $\overline{m}_X = (k_X^T)/2$.
- $\theta_Z \sim B(\underline{\alpha}_2, \underline{\beta}_2)$, $\underline{\alpha}_2 = 1$, $\underline{\beta}_2 = (k_Z^T - \overline{m}_Z)/\overline{m}_Z))$, with $\overline{m}_Z = (k_Z^T + k_{ZP}^T)/2$.
- $\theta_Y \sim B(\underline{\alpha}_3, \underline{\beta}_3)$, $\underline{\alpha}_3 = 1$, $\underline{\beta}_3 = (m - \overline{m}_Y)/\overline{m}_Y))$, with $\overline{m}_Y = m/2$.
- $\theta_V \sim B(\underline{\alpha}_4, \underline{\beta}_4)$, $\underline{\alpha}_4 = 1$, $\underline{\beta}_4 = (m - \overline{m}_V)/\overline{m}_V))$, with $\overline{m}_V = m/2$.

Properties of the Beta distribution imply that the expected value of $\theta = (\theta_X, \theta_Z, \theta_Y, \theta_V)$ is equal to $(\overline{m}_X/k_X^T, \overline{m}_Z/(k_Z^T + k_{ZP}^T), \overline{m}_Y/m, \overline{m}_V/m)$. Therefore, the parameters $(\overline{m}_X, \overline{m}_Z, \overline{m}_Y, \overline{m}_V)$ control the number of restrictions that we expect to hold on average a priori, or in some sense, the model size. Using properties of the Beta distribution, the vector of restriction probabilities θ can be integrated out (using (5)) so that the prior probability of a model is:

$$
\begin{aligned}
\pi(M_j) \quad = \quad & \frac{\Gamma(\underline{\alpha}_1 + \underline{\beta}_1)}{\Gamma(\underline{\alpha}_1)\Gamma(\underline{\beta}_1)} \frac{\Gamma(\underline{\alpha}_1 + k_X^j)\Gamma(\underline{\beta}_1 + k_X^T - k_X^j)}{\Gamma(\underline{\alpha}_1 + \underline{\beta}_1 + k_X^T)} \times \\
& \frac{\Gamma(\underline{\alpha}_2 + \underline{\beta}_2)}{\Gamma(\underline{\alpha}_2)\Gamma(\underline{\beta}_2)} \frac{\Gamma(\underline{\alpha}_2 + k_Z^j)\Gamma(\underline{\beta}_2 + k_Z^T + k_{ZP}^T - k_Z^j)}{\Gamma(\underline{\alpha}_2 + \underline{\beta}_2 + k_Z^T + k_{ZP}^T)} \times \\
& \frac{\Gamma(\underline{\alpha}_3 + \underline{\beta}_3)}{\Gamma(\underline{\alpha}_3)\Gamma(\underline{\beta}_3)} \frac{\Gamma(\underline{\alpha}_3 + k_y^j)\Gamma(\underline{\beta}_3 + m - k_y^j)}{\Gamma(\underline{\alpha}_3 + \underline{\beta}_3 + m)} \times \\
& \frac{\Gamma(\underline{\alpha}_4 + \underline{\beta}_4)}{\Gamma(\underline{\alpha}_4)\Gamma(\underline{\beta}_4)} \frac{\Gamma(\underline{\alpha}_4 + k_v^j)\Gamma(\underline{\beta}_4 + m - k_v^j)}{\Gamma(\underline{\alpha}_4 + \underline{\beta}_4 + m)}
\end{aligned}
$$

In order to calculate the total visited probability (George and McCulloch, 1997), we first define a large set of models A that encompasses the set B which is composed of the models visited by the algorithm. We then calculate the marginal likelihood for each model in A, so that we could obtain the estimated total visited probability as the joint posterior probability of B over that of A.

Integrating out the individual effect to obtain forward orthogonal deviations

Let us now show that equations in (3) can be obtained by first specifying a dynamic panel data model in levels and then integrating out the fixed effects from the posterior density. To see this, first complete equation (1) for g_{it} with auxiliary equations for y_{it}:

$$
\begin{aligned}
g_{it} &= f_i + \gamma' y_{it} + \beta' x_{it} + u_{it} \qquad t = 1, ..., T, \qquad (8) \\
y_{it} &= f_i^y + \Pi_{2x} x_{it} + \Pi_{2z^s} z_{it} + \Pi_{2z^p} z_{it}^{(-p)} + v_{it},
\end{aligned}
$$

where $z_{it}^{(-p)}$ is a vector such that the forward orthogonal deviation of $z_{it}^{(-p)}$ is equal to z_{it}^{p}, that is $z_{it}^{p} = z_{it}^{(-p)*}$. In the final part of this proof we explain how $z_{it}^{(-p)}$ can be constructed so that $z_{it}^{p} = z_{it}^{(-p)*}$. For simplicity in notation in this part we set $T_i = T$. The reduced form of equations in (8) is:

$$h_{it} = \begin{pmatrix} g_{it} \\ y_{it} \end{pmatrix} = \Pi_x x_{it} + \Pi_z \widetilde{z}_{it} + f_i^r + \varepsilon_{it}, \tag{9}$$

where Π_x and Π_z are defined as in (6) and $(f_i^r, \varepsilon_{it}, \widetilde{z}_{it})$ are defined as

$$\varepsilon_{it} = \begin{pmatrix} 1 & \gamma' \\ 0 & I_m \end{pmatrix} \begin{pmatrix} u_{it} \\ v_{it} \end{pmatrix}, \qquad f_i^r = \begin{pmatrix} 1 & \gamma' \\ 0 & I_m \end{pmatrix} \begin{pmatrix} f_i \\ f_i^y \end{pmatrix} \qquad \widetilde{z}_{it} = \begin{pmatrix} z_{it} \\ z_{it}^{(-p)} \end{pmatrix}$$

Since the variance-covariance matrix of ε_{it} is Ω, the likelihood function can be written as

$$|\Omega|^{-NT/2} |2\pi|^{-NT(m+1)/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{N} \sum_{t=1}^{T} tr[\Omega^{-1}(\widetilde{h}_{it} - f_i^r)(\widetilde{h}_{it} - f_i^r)'] \right),$$

where $\widetilde{h}_{it} = h_{it} - \Pi_x x_{it} - \Pi_z \widetilde{z}_{it}$. Using a flat prior on f_i^r, we can integrate this expression with respect to f_i^r and obtain

$$|\Omega|^{-N(T-1)/2} |2\pi|^{-N(T-1)(m+1)/2} T^{-N/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{N} tr[\Omega^{-1} \widetilde{h}_i' Q \widetilde{h}_i] \right), \tag{10}$$

where $\widetilde{h}_i = (\widetilde{h}_{i1}, ..., \widetilde{h}_{iT})'$, Q is the within-group operator (Arellano 2003, p. 15) $Q = I - (1/T)ii'$, i is a $T \times 1$ vector of ones, and I is the identity matrix. To see that this is the likelihood of the model defined by the equations in (3), first note that Q can be written as $Q = A'A$, where A is a $(T-1) \times T$ matrix known as the forward orthogonal operator (Arellano 2003, p. 17), such that $\widetilde{h}_i' Q \widetilde{h}_i = (A\widetilde{h}_i)' A \widetilde{h}_i$.

Hence, expression (10) can be written as

$$|\Omega|^{-N/2}\,|2\pi|^{-N/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{N}tr[\Omega^{-1}\widetilde{h}_i^{*\prime}\widetilde{h}_i^{*}]\right), \tag{11}$$

where

$$\widetilde{h}_i^{*}=A\widetilde{h}_i=A\begin{pmatrix}\widetilde{h}_{i1}^{\prime}\\ \vdots\\ \widetilde{h}_{iT}^{\prime}\end{pmatrix}=A\left[\begin{pmatrix}h_{i1}^{\prime}\\ \vdots\\ h_{iT}^{\prime}\end{pmatrix}-\begin{pmatrix}x_{i1}^{\prime}\\ \vdots\\ x_{iT}^{\prime}\end{pmatrix}\Pi_x^{\prime}-\begin{pmatrix}\widetilde{z}_{i1}^{\prime}\\ \vdots\\ \widetilde{z}_{iT}^{\prime}\end{pmatrix}\Pi_z^{\prime}\right].$$

Note also that A is a $(T-1)\times T$ matrix and hence $\widetilde{h}_i^{*}=A\widetilde{h}_i$ is a $(T-1)\times 1$ vector with the forward orthogonal deviations of $\widetilde{h}_i$:

$$\widetilde{h}_i^{*}=\begin{pmatrix}\widetilde{h}_{i1}^{*\prime}\\ \vdots\\ \widetilde{h}_{iT-1}^{*\prime}\end{pmatrix}=\begin{pmatrix}g_{i1}^{*} & y_{i1}^{*\prime}\\ \vdots & \vdots\\ g_{iT-1}^{*} & y_{iT-1}^{*\prime}\end{pmatrix}-\begin{pmatrix}x_{i1}^{*\prime}\\ \vdots\\ x_{iT-1}^{*\prime}\end{pmatrix}\Pi_x^{\prime}-\begin{pmatrix}\widetilde{z}_{i1}^{*\prime}\\ \vdots\\ \widetilde{z}_{iT-1}^{*\prime}\end{pmatrix}\Pi_z^{\prime}.$$

Using the properties for the trace operator, it is possible to write expression (11) as

$$\begin{aligned}&|\Omega|^{-N/2}\,|2\pi|^{-N/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{N}tr[\widetilde{h}_i^{*}\Omega^{-1}\widetilde{h}_i^{*\prime}]\right)\\ =\;&|\Omega|^{-N/2}\,|2\pi|^{-N/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T-1}[\widetilde{h}_{it}^{*\prime}\Omega^{-1}\widetilde{h}_{it}^{*}]\right),\end{aligned} \tag{12}$$

which is clearly the likelihood of the model defined by the equations in (3) if we take into account that $\widetilde{z}_{it}^{*\prime}=(z_{it}^{*\prime},z_{it}^{(-p)*\prime})=(z_{it}^{*\prime},z_{it}^{p\prime})$. To show that $z_{it}^{p}=z_{it}^{(-p)*}$ let us construct z_{it}^{-p} as follows:

$$
\begin{aligned}
z_{i1}^{(-p)} &= z_{i1}^{p}\sqrt{\frac{T}{T-1}} \\
z_{i2}^{(-p)} &= z_{i2}^{p}\sqrt{\frac{T-2}{T-1}} \quad \text{if } T \geq 3, \qquad z_{iT}^{(-p)} = 0 \quad \text{if } T = 2 \\
z_{it}^{(-p)} &= z_{it}^{p}\sqrt{\frac{T-t}{T-t+1}} - \sum_{h=2}^{t-1}\left(z_{ih}^{p}\sqrt{\frac{T-h+1}{T-h}}\frac{1}{T-h+1}\right) \qquad \text{for } t = 3, ..., (T-1) \text{ if } T \geq 4 \\
z_{iT}^{(-p)} &= -\sum_{h=2}^{T-1}\left(z_{ih}^{p}\sqrt{\frac{T-h+1}{T-h}}\frac{1}{T-h+1}\right) \quad \text{if } \ T \geq 3
\end{aligned}
$$

Note that $z_{it}^{(-p)}$ is a function of current and past values of z_{it}^{p}, that is, a function of $(z_{it}^{p}, z_{i(t-1)}^{p}, ..., z_{i1}^{p})$. Using this definition, it can be shown that $z_{it}^{p} = z_{it}^{(-p)*}$ by simply calculating the forward orthogonal deviations of $z_{it}^{(-p)}$ directly. Note also that this construction is not unique, because adding or subtracting an arbitrary constant to z_{it}^{-p} (the same constant for each value of t) also gives a vector whose forward orthogonal deviations are equal to z_{it}^{p}.

	Model 1		Model 2		Model 3		Model 4		Model 5	
	Coef	p-value	Coef	p-value	Coef	p-value	Coef	p-value	Coef	p-value
gdp	1.80	(0.04)	0.79	(0.38)	1.77	(0.03)	0.74	(0.35)	0.52	(0.62)
eda	0.19	(0.51)	-1.06	(0.05)	0.19	(0.46)	-1.55	(0.00)	-1.18	(0.00)
eda2	Exc		0.14	(0.00)	Exc		0.19	(0.00)	0.16	(0.00)
polaid	Exc		Exc		0.13	(0.36)	Exc		0.07	(0.64)
pol	1.11	(0.00)	1.15	(0.00)	0.48	(0.30)	1.04	(0.00)	1.03	(0.02)
assas	-0.40	(0.41)	-0.16	(0.71)	-0.24	(0.48)	-0.30	(0.31)	0.03	(0.94)
m21	-0.01	(0.81)	0.00	(0.86)	0.00	(0.89)	0.00	(0.84)	0.00	(0.86)
eth_a	0.05	(0.96)	-0.03	(0.97)	-0.04	(0.96)	0.11	(0.86)	-0.59	(0.41)
dum3	-0.20	(0.74)	-0.24	(0.78)	-0.38	(0.54)	0.39	(0.59)	0.16	(0.83)
dum4	-1.92	(0.00)	-1.53	(0.07)	-1.85	(0.01)	-0.87	(0.34)	-1.15	(0.15)
dum5	-3.40	(0.00)	-2.83	(0.00)	-3.59	(0.00)	-2.02	(0.03)	-2.25	(0.01)
dum6	-2.23	(0.00)	-1.57	(0.03)	-2.30	(0.00)	-1.01	(0.17)	-1.03	(0.16)
dum7	-2.33	(0.00)	-1.84	(0.06)	-2.30	(0.00)	-1.18	(0.18)	-1.30	(0.17)
dum8	-2.06	(0.00)	-1.73	(0.07)	-1.90	(0.01)	-0.87	(0.28)	-1.13	(0.23)
cons	-9.78	(0.12)	-1.89	(0.77)	-10.08	(0.08)	-1.78	(0.77)	-0.23	(0.98)
Marginal Impact of Aid at Sample Mean Values										
g_A	0.19	(0.51)	-0.70	(0.11)	0.14	(0.58)	-1.06	(0.007)	-0.80	(0.02)
Diagnostic Tests										
Hansen test	(1.00)		(1.00)		(1.00)		(0.88)		(1.00)	
AR(2) test	(0.18)		(0.11)		(0.18)		(0.08)		(0.11)	
Endogenous variables whose lags are used as instruments	gdp, eda pol, assas m21, eth_a		gdp, eda pol, assas m21, eth_a		gdp, eda pol, assas m21, eth_a		gdp, eda pol, assas m21, eth_a		gdp, eda pol, assas m21, eth_a	
No. of lags used as instruments	All		All		All		1		All	

Table 1: System GMM Estimates and Diagnostic Tests Under Five Model Specifications. p-values in brackets. g_A is the marginal impact of aid on growth derived from the estimates in the model. '*AR(2)*' is the p-value for a test of second order autocorrelation in the residuals (Arellano and Bond, 1991) and '*Hansen p*' is the p-value of a test of over-identifying restrictions (Hansen, 1982). The time dummies were used as exogenous instruments.

	Model 1		Model 2		Model 3		Model 4		Model 5	
	Coef	*p*-value	Coef	*p*-value	Coef	*p*-value	Coef	*p*-value	Coef	*p*-value
gdp	-3.62	(0.49)	-3.74	(0.46)	-3.65	(0.47)	-3.67	(0.12)	-1.93	(0.25)
eda	0.92	(0.31)	0.12	(0.93)	0.95	(0.31)	-1.35	(0.00)	-1.19	(0.05)
eda2	Exc		0.08	(0.38)	Exc		0.14	(0.00)	0.14	(0.00)
polaid	Exc		Exc		0.08	(0.79)	Exc		Exc	
pol	0.73	(0.26)	0.81	(0.22)	0.58	(0.45)	0.91	(0.00)	1.25	(0.00)
assas	0.03	(0.98)	-0.06	(0.95)	0.02	(0.99)	-0.29	(0.64)	-0.35	(0.55)
m21	0.15	(0.38)	0.17	(0.31)	0.16	(0.36)	0.00	(0.98)	0.00	(0.97)
eth_a	-0.71	(0.78)	-0.53	(0.84)	-0.72	(0.78)	0.20	(0.88)	0.45	(0.68)
dum3	-0.45	(0.63)	-0.30	(0.76)	-0.50	(0.62)	0.68	(0.36)	0.21	(0.75)
dum4	-1.52	(0.38)	-1.24	(0.49)	-1.55	(0.37)	-0.27	(0.80)	-1.04	(0.14)
dum5	-3.33	(0.11)	-3.02	(0.15)	-3.42	(0.10)	-1.52	(0.06)	-2.23	(0.01)
dum6	-3.08	(0.21)	-2.70	(0.29)	-3.15	(0.21)	-0.15	(0.87)	-0.57	(0.57)
dum7	-3.38	(0.21)	-3.03	(0.26)	-3.38	(0.20)	-0.04	(0.97)	-0.66	(0.55)
dum8	-2.86	(0.35)	-2.65	(0.38)	-2.89	(0.33)	0.51	(0.65)	-0.62	(0.54)
Marginal Impact of Aid at Sample Mean Values										
g_A	0.92	(0.31)	0.32	(0.79)	0.91	(0.30)	-0.99	(0.00)	-0.82	(0.10)
Diagnostic Tests										
Hansen test	(0.06)		(0.05)		(0.04)		(0.92)		(1.00)	
AR(2) test	(0.29)		(0.25)		(0.30)		(0.11)		(0.14)	
Endogenous variables whose lags are used as instruments	gdp, eda pol, assas m21, eth_a		gdp, eda pol, assas m21, eth_a		gdp, eda pol, assas m21, eth_a		gdp, eda eda2 pol, assas m21, eth_a		gdp, eda eda2 pol, assas m21, eth_a	
No. of lags used as instruments	1		1		1		2		All	

Table 2: Difference GMM Estimates and Diagnostic Tests Under Five Model Specifications. *p*-values in brackets. g_A is the marginal impact of aid on growth derived from the estimates in the model. '*AR(2)*' is the p-value for a test of second order autocorrelation in the residuals (Arellano and Bond, 1991) and '*Hansen p*' is the *p*-value of a test of over-identifying restrictions (Hansen, 1982). The time dummies were used as exogenous instruments.

	All Countries				Low Income Countries			
	Median	$Pr < 0$	$Pr = 0$	$Pr > 0$	Median	$Pr < 0$	$Pr = 0$	$Pr > 0$
g_A	0.253	0.238	0.093	0.669	0.000	0.005	0.679	0.316
g_{AP}	0.000	0.013	0.977	0.009	0.000	0.006	0.990	0.003
g_P	0.927	0.005	0.000	0.995	0.743	0.035	0.000	0.965
$g_{Initial}$	0.627	0.396	0.000	0.604	-0.291	0.548	0.000	0.452

Table 3: Marginal Impacts: BMA Estimation with Fixed Effects. The column '*Median*' gives the posterior median and ($Pr < 0$, $Pr = 0$, $Pr > 0$) give the posterior probability of being smaller, equal and greater than 0. g_A is the first derivative of the growth rate with respect to aid. g_{AP} is the second derivative with respect to aid and the policy index. g_P is the first derivative with respect to policy index. $g_{Initial}$ is the first derivative with respect to the beginning of period log GDP per capita. Marginal derivatives are evaluated at sample means.

	All Countries		Low Income	
	mean	$Pr \neq 0$	mean	$Pr \neq 0$
constant	0.1	0.13	0.26	0.18
icrge	0.3	0.95	0.37	0.87
m21	0.0	0.06	0	0.1
ssa	-1.3	0.96	-1.51	0.96
easia	1.6	0.9	1.93	0.91
bb	0.3	0.08	0.1	0.07
infl	-2.2	1.00	-2.75	1.00
sacw	1.2	0.87	0.82	0.62
ethnf	0.0	0.07	-0.01	0.07
gdp	0.0	0.12	0.02	0.16
assas	-0.1	0.36	-0.43	0.65
eth_a	-0.1	0.17	-0.04	0.11
dum2	1.3	0.81	0.7	0.47
dum3	1.7	0.96	1.2	0.73
dum4	0.0	0.07	0.03	0.08
dum5	-1.2	0.83	-0.41	0.35
dum6	0.0	0.07	-0.01	0.07
dum7	0.0	0.07	-0.31	0.29
dum8	0.0	0.08	0.11	0.15

Table 4: BMA when policy variables (i.e. bb, infl and sacw) enter as separate regressors in a reduced-form regression on growth. $Pr \neq 0$ is the posterior probability that the coefficient is different from 0. *mean* is the posterior mean of the coefficient. All regressors are assumed to be exogenous. The sample sizes were 396 (all countries) and 262 (low-income). ssa and easia are dummies for Sub-Saharan Africa and East Asia, respectively. icrge is an index of institutional quality

	All Countries					Low Income				
	Pr≠ 0	Endg	2.5	50	97.5	Pr≠ 0	Endg	2.5	50	97.5
eda	0.52	0.97	-2.72	0.00	1.55	0.29	0.87	0.00	0.00	1.26
polaid	0.02	0.88	0.00	0.00	0.00	0.01	0.74	0.00	0.00	0.00
aid2pol	0.01	0.20	0.00	0.00	0.00	0.00	0.27	0.00	0.00	0.00
eda2	0.67	0.93	0.00	0.10	0.32	0.03	0.96	0.00	0.00	0.05
gdp	1.00	1.00	-4.46	0.63	5.45	1.00	1.00	-5.00	-0.28	4.73
assas	0.12		-0.56	0.00	0.00		0.33	-1.22	0.00	0.00
m21	1.00		-0.05	0.02	0.10		1.00	-0.10	-0.02	0.07
eth_a	0.04		0.00	0.00	0.00		0.07	0.00	0.00	167.64
pol	1.00		0.19	0.96	1.73		1.00	-0.09	0.76	1.61
pol2	1.00		-0.09	0.04	0.17		1.00	-0.09	0.02	0.13
dum3	1.00		-1.44	0.06	1.59		1.00	-1.80	-0.13	1.45
dum4	1.00		-3.70	-1.76	0.32		1.00	-3.48	-1.31	0.52
dum5	1.00		-5.00	-2.99	-0.71		1.00	-4.76	-2.48	-0.63
dum6	1.00		-4.20	-2.17	0.11		1.00	-4.01	-1.59	0.23
dum7	1.00		-4.39	-2.27	0.13		1.00	-4.99	-2.27	-0.42
dum8	1.00		-4.18	-1.91	0.55		1.00	-3.96	-1.15	0.83

Table 5: BMA estimation with fixed effects. Pr≠ 0 is the posterior probability of entering in the model as a regressor (in y_2 or x). '*Endg*' is the posterior probability of being endogenous. The (2.5%, 50%, 97.5%) percentiles of the posterior distribution are under the corresponding headings.

	All Countries	Low Income
lpop	0.99	0.89
arms1	0.56	0.01
polarms	1.00	1.00
polpop	0.84	0.65
polpop2	0.17	0.46
polgdp	1.00	0.92
polgdp2	1.00	0.97
$Zgdp^{3,0}$	0.88	0.50
$Zgdp^{3,1}$	0.99	0.91
$Zgdp^{4,0}$	0.07	0.05
$Zgdp^{4,1}$	0.15	1.00
$Zgdp^{4,2}$	0.06	1.00
$Zgdp^{5,0}$	0.01	0.02
$Zgdp^{5,1}$	0.03	0.02
$Zgdp^{5,2}$	0.02	0.02
$Zgdp^{5,3}$	0.04	0.05
$Zgdp^{6,0}$	0.90	1.00
$Zgdp^{6,1}$	0.79	0.95
$Zgdp^{6,2}$	0.06	0.01
$Zgdp^{6,3}$	0.05	0.01
$Zgdp^{6,4}$	0.49	0.07
$Zgdp^{7,0}$	0.10	0.10
$Zgdp^{7,1}$	0.02	0.02
$Zgdp^{7,2}$	0.25	0.16
$Zgdp^{7,3}$	0.02	0.01
$Zgdp^{7,4}$	0.01	0.00
$Zgdp^{7,5}$	0.01	0.00
$Zeda^{3,1}$	0.12	0.04
$Zeda^{4,1}$	1.00	1.00
$Zeda^{4,2}$	0.09	0.01
$Zeda^{5,1}$	0.16	0.46
$Zeda^{5,2}$	0.82	0.27
$Zeda^{5,3}$	0.38	0.02
$Zeda^{6,1}$	0.11	0.01
$Zeda^{6,2}$	0.04	0.02
$Zeda^{6,3}$	0.05	0.03
$Zeda^{6,4}$	0.07	0.27
$Zeda^{7,1}$	0.13	0.00
$Zeda^{7,2}$	0.03	0.00
$Zeda^{7,3}$	0.00	0.00
$Zeda^{7,4}$	0.00	0.00
$Zeda^{7,5}$	0.01	0.00

Table 6: BMA with Fixed Effects: Posterior Probability of Being an Instrument.