

*What have we learned from a decade of empirical research on growth?*

Comment on “Growth Empirics and Reality,”  
by William A. Brock and Steven N. Durlauf

*Xavier Sala-i-Martin*

William Brock and Steven Durlauf’s article nicely summarizes some of the recent research on Bayesian model averaging. They make a number of important points. One is that the empirics of growth face three key problems: model uncertainty, parameter uncertainty, and endogeneity. They argue that theory uncertainty can be dealt with using Bayesian model averaging methods. Their key equations are 16, 17, and 18, for which the interpretation is as follows. Suppose you are interested in the distribution of the partial derivative of the growth rate with respect to variable  $z$ ,  $b_z$ . Let each set of every possible combination of explanatory variables be called a “model.” Conditional on each model there is a distribution of  $b_z$  for a given data set. Equation 17 says that the posterior distribution of  $b_z$  is a weighted average of all these individual distributions, where the weights are proportional to the likelihoods of the models. Equation 18 says that the mean of this distribution is the weighted average of the ordinary least squares (OLS) estimates of all these models, where the weights are proportional to the likelihoods. Equation 19 makes a similar claim about the variance.

The assumption that weights are proportional to the likelihoods is an important one. In fact, it may drive the authors’ first key empirical result—that Easterly and Levine’s (1997) regression of growth on ethnolinguistic fractionalization (ELF) is “robust” to Bayesian model averaging analysis. It is important to remember that models with more explanatory variables have larger likelihoods. It is also important to remember that Brock and Durlauf perform Bayesian model averaging analysis by combining the explanatory variables of the Easterly and Levine paper in all possible ways: sets of one right-hand-side variable, sets of two, sets of three, and, eventually, one set with all the right-hand-side variables. This last model is the one run by Easterly and Levine and the largest model run by Brock and Durlauf (and therefore the one that is likely to have the largest likelihood and that gets the largest weight).

Hence, it is not surprising that the weighted average of all the models is similar to that for Easterly and Levine’s model, because most of the weight of the average goes to Easterly and Levine’s specification, by construction. In other words, the finding that Easterly and Levine’s regression results (column 1 in table

Xavier Sala-i-Martin is at Columbia University and UPF.

© 2001 The International Bank for Reconstruction and Development / THE WORLD BANK

1) are “robust” to the Bayesian model averaging analysis because the weighted average of models (column 2) is virtually identical is likely to be an artifact of the weights used.

I should confess that these are also the weights I used in a 1997 paper (which has equations 16 and 17 in exactly the same form). However, in that work I averaged only regressions with a fixed set of explanatory variables, so I did not have the problem that I am pointing out here. Doppelhofer, Miller, and Sala-i-Martin (2000) derive an alternative weighting scheme. The posterior density of model  $M_m$  is proportional to the likelihood (sum of squares of residuals, or  $SSE_m^{-T/2}$ ), multiplied by  $T^{-k_m/2}$ , where  $T$  is the number of observations and  $k_m$  is the number of explanatory variables in model  $m$ :

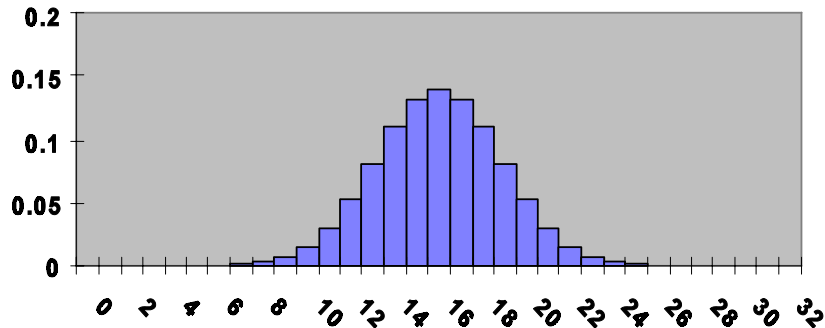
$$\mu(M_m | D) = \frac{\mu(M_m) T^{-k_m/2} . SSE_m^{-T/2}}{\sum_{i=1}^{2^K} \mu(M_i) T^{k_i/2} . SSE_i^{-T/2}}$$

Note that this weighting scheme penalizes larger models. It would be interesting to see whether column 1 still looks very much like column 2 when these alternative weights are used.

A second important assumption is the prior that allows Brock and Durlauf to eliminate the  $\mu(M_m)$  from equation 15 to derive equation 16. They use the prior that “all models are equally likely.” Imagine that we had 32 possible right-hand-side variables. If we believe that all models are equally likely, the prior distribution of model sizes is as shown in figure 1. The average model size is 16. If instead we had 10 explanatory variables, the implicit assumption would be that the average model size of the prior distribution of cross-country regressions is 5.

The problem is that Brock and Durlauf propose that when analyzing (or discussing) a paper like Easterly and Levine (1997), we take the key regression in that paper and perform Bayesian model averaging analysis with it. If we take

FIGURE 1. Prior Probabilities by Model Size: Equal Model Probabilities



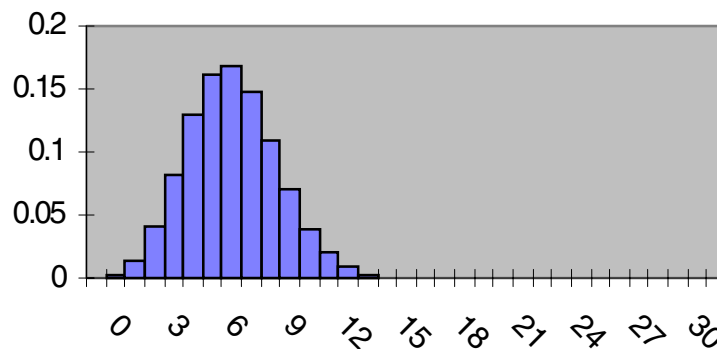
this proposal literally, we would implicitly assume that the average model size of “the growth regression” is 5 when the original paper had 10 variables, and 16 when the original paper had 32 variables. Besides being arbitrary, this assumption does not make sense: The prior model size should be invariant to the paper being discussed.

One solution to this problem, following Doppelhofer, Miller, and Sala-i-Martin (2000), would be to specify the model prior probabilities by choosing a prior mean model size,  $k$ , with each variable having a prior probability  $\kappa/K$  of being included, independent of the inclusion of any other variables, where  $K$  is the total number of potential regressors (figure 2). Equal probability for each possible model is the special case in which  $\kappa = K/2$ . The prior distribution of model sizes would be invariant to the paper analyzed. Moreover, the robustness of this prior could be checked by redoing the Bayesian model averaging exercise (or better yet, Bayesian averaging of classical estimates) for different values of  $\kappa$ .

My third comment relates to the treatment of parameter uncertainty. I agree with the authors that this problem is analogous to that of theory uncertainty. But if so, why do they propose a different solution? If we think that Africa needs a different slope for variable  $z$ , all we need to do is to construct a new variable ( $z$  times one for countries in Africa and  $z$  times zero otherwise) and put this new variable in the pool of potential variables to be included in the Bayesian model averaging analysis. Rather than columns 3–6, table 2 should include a row presenting the distribution of the  $\beta_{[j<2]}$  for this new variable, as a regular additional variable subject to theory uncertainty.

When we think of parameter uncertainty as another form of theory uncertainty, an additional problem comes to mind. Why do we think that Africa needs its own slope? Why don't we have a special slope for Christian countries? Or hot countries? Or small countries? Of course, we do not know whether or not special slopes are needed (we do not have a theory, or we can have many open-ended theories that would call for a special slope for each of these country groups). However, in the spirit of Durlauf and Johnson (1995), shouldn't we then per-

FIGURE 2. Prior Probabilities by Model Size ( $\kappa = 7$ )



form Bayesian model averaging or Bayesian averaging of classical estimates for each group of countries? How would we go about that?

A perhaps related question is that of nonlinearities, which Brock and Durlauf do not allow for in their article. It is clear that African countries have both lower average growth and greater ethnolinguistic fractionalization. The conditional data might therefore look like figure 3. If we think about the implications of figure 3, we arrive at the conclusion that if we could somehow reduce ELF for African countries, Africa will conditionally grow faster than the rest of the world forever (that is, we would move the African data points to the left along the steeper regression line). Because we do not have a theory of ELF, we do not know whether this is sensible or not.

Alternatively, we could think that the partial relationship between growth and ELF looks like figure 4. In fact, the data points in figures 3 and 4 are exactly the same. The only thing that differs is the functional form of the regression curve. Under this interpretation, if Africa manages to get the same ELF as the rest of the world, its growth rate will also be similar. Hence the economic implications of a separate slope for Africa are very different from those of a nonlinear relationship. It would have been interesting to incorporate nonlinearities in the analysis.

Finally, the claim that growth economists have not dealt with parameter uncertainty is not quite true. In fact, parameter uncertainty is a particular form of what economists usually label *interaction terms*. For example, suppose a claim is made that the partial derivative of growth with respect to  $z$  depends on variable  $y$ :

$$\frac{\partial g_i}{\partial z_{ji}} = \beta_z + \beta_{z,yj} y_j$$

FIGURE 3.

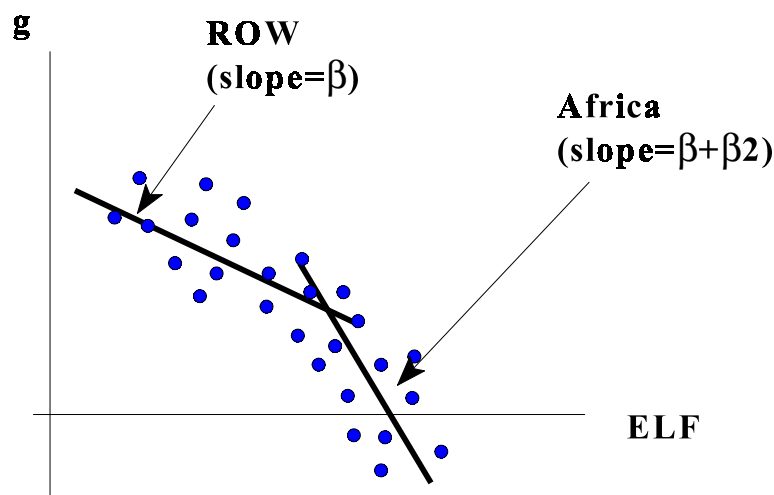
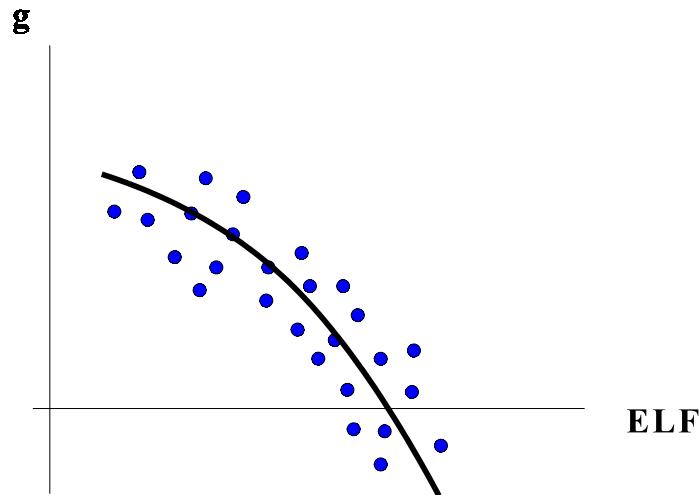


FIGURE 4.



The way to test this claim would be to run a regression of growth with  $z$  as an explanatory variable and with an additional variable that is a country-by-country product of  $z$  times  $y$ . That is, we should introduce interaction terms. It should be clear that parameter uncertainty is nothing but an interaction term when variable  $y$  is simply a dummy variable for a region (in this case, Sub-Saharan Africa). To the extent that growth economists have introduced interaction terms, therefore, they have allowed for parameter heterogeneity.

I conclude with two sources of disappointment about this otherwise excellent article. First, the article is not really about the empirics of economic growth. All empirical analyses are subject to the problems it discusses, especially those forced to use small data sets. In this sense the title, though cute, is highly misleading and, to the extent that it leads future researchers away from economic growth analysis, potentially damaging. A more appropriate title would be “Small-Sample Econometrics,” because the problems discussed are common to all empirical analyses with small samples (which include all cross-country analyses in any field). After all, if we had a huge data set with zillions of observations, we could simply throw in all potential variables, with particular slopes for each potential set of countries, with all potential nonlinearities, and so on—and the data would tell us which coefficients are zero and which are not. The fact that we have more potential variables than we have countries prevents us from following this strategy, and this is where the problem starts. But this is a problem of small samples, not growth econometrics.

Second, although the authors introduce endogeneity as an important problem early in their article, I was disappointed to find that they went no further. Given the authors’ reputation, I was excited when I started reading the article about the prospect of a potential solution, perhaps along the lines of Bayesian model averaging. But no solution was offered.

REFERENCES

- Doppelhofer, G., R. Miller, and X. Sala-i-Martin. 2000. "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach." NBER Working Paper no. 7750, National Bureau of Economic Research, Washington, D.C.
- Durlauf, S., and P. Johnson. 1995. "Multiple Regimes and Cross-Country Growth Behavior." *Journal of Applied Econometrics* 10:365–84.
- Easterly, W., and R. Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics* 112(4):1203–50.
- Sala-i-Martin, X. 1997. "I Just Ran Two Million Regressions." *American Economic Review, Papers and Proceedings* 87:178–83.