Bayesian Semiparametric Inference for Model Averaging: An Application to Growth Determinants

Gernot Doppelhofer* NHH

Melvyn Weeks† University of Cambridge

June 29, 2011

The empirical analysis of the determinants of economic growth has generated a large literature among economists and applied researchers. This literature is plagued by a number of problems that can have important consequences for statistical inference and economic implications.[1] In a critical survey, Brock and Durlauf (2001) argue that much empirical work in this area suffers from "incredible" assumptions that are difficult to defend. Temple (2000) highlights three key issues: (i) model uncertainty, (ii) parameter heterogeneity, and (iii) outliers. This research proposal investigates the robustness of inference about growth determinants to these specification problems, and specifically operationalises the Bayesian bootstrap which provides a semi-parametric analysis of the linear model. In doing so we accommodate data uncertainty within a model averaging framework that requires no distributional assumptions other than multinomial sampling.

A recent and quickly growing literature has utlised model averaging techniques[2] to address model uncertainty and the effect on inference and policy analysis on growth determinants. Early papers that address model uncertainty in growth regressions include Fernandez, Ley and Steel (2001), and Sala-i-Martin, Doppelhofer and Miller (2004). Recently, model averaging has been applied in the context of growth empirics to investigate the sensitivity to prior information (Ley and Steel, 2009), and predictive performance (Eicher, Papageorgiou and Raftery, 2009).

Studies which have considered uncertainty due to parameter heterogeneity in growth empirics include Durlauf and Johnson (1995), Brock and Durlauf (2001), an entire special issue of the *Journal of Macroeconomics* edited by Papageorgiou (2007), Masanjala and Papageorgiou (2008) and Tan (2009). A common feature of these studies is the use of prior knowledge which generally involves conditioning on a particular mechanism thought to generate heterogeneity in the distribution of coefficients of interest. Statis-

*Department of Economics, Norwegian School of Economics (NHH), Helleveien 30, N-5045 Bergen. Email: Gernot.Doppelhofer@nhh.no, Tel: +47 55959360, Fax: +47 55959543.

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

[1] Durlauf, Johnson and Temple (2008) give an excellent critical survey over the econometric methods and challenges in empirical research on economic growth.

[2] For recent surveys see Hoeting, Madigan, Raftery and Volinsky (1999) or Doppelhofer (2008).

tical inference and economic implications in these papers are therefore conditional on correctly specifying the source of parameter heterogeneity.

Outliers due to measurement errors or specification problems is a major problem encountered in the empirical growth literature. Schultz (1999, p. 71) notes that "Macroeconomic studies of growth often seek to explain differences in economic growth rates across countries in terms of [several variables]. However, these estimates are plagued by measurement error and specification problems." Deaton (2010) critically discusses measures of income across countries and over time. In a recent paper, Ciccone and Jarocinski (2010) investigate the sensitivity of inference on growth determinants when using different vintages of the Penn World Tables. Despite these data and specification problems, Zaman, Rousseeuw and Orhan (2001) note that a remarkably low number of papers address issues of robustness in the economic literature.

If specification problems are known *a priori*, it would be straightforward to adapt the growth model accordingly. In the more realistic case when the precise form of model misspecification are not known, inference should be conducted in a robust manner. One possible approach to modeling outlying observations is the introduction of *mean-shifts*. For example, Hendry and Santos (2005) propose to saturate the regression model by introducing a large number of dummy variables. Treating each outlying observation differently is problematic in the context of data limitations and model uncertainty in the empirical growth literature. Furthermore, economic theory offers little guidance about the appropriate form of parameter heterogeneity.

1 Robust Model Averaging

A flexible approach to robust estimation becomes all the more important in the presence of model uncertainty, when theory gives us little guidance about the correct model. Empirical researchers were initially restricted by limitations in statistical techniques and computing resources. Leamer (1982) and Leamer and Leonard (1983) propose so-called extreme bounds analysis (EBA) to test the sensitivity of parameters of interest to changes in the set of alternative models, represented by different combinations of additional control variables. Sturm and de Haan (2005) apply a version of EBA that uses re-weighted least squares first developed by Rousseeuw (1984). Zaman et al. (2001) highlight the importance of robust inference when applying least trimmed squares to a simple growth regression.

Much of the advances in robust methods in Bayesian inference have been confined to single models. A notable exception is the work of Hoeting, Raftery, and Madigan (1996), who develop an approach that simultaneously accounts for model uncertainty and outlier identification by introducing a prior for the proportion of outlying observations. Recently, Gottardo and Raftery (2007) adopt a unifying approach to Bayesian robust variable and transformation selection. Magnus, Wan and Zhang (2010) use a version of the weighted average least squares (or WALS) estimator with nonspherical disturbances in an analysis of the Hong Kong housing market.

Doppelhofer and Weeks (2011) contribute to the existing literature by dealing with model uncertainty and allowing for heterogeneity of unknown form, generated either by outliers or neglected parameter heterogeneity. They adopt a *variance-inflation* approach that accommodates outliers and robustifies inference against *unknown* aberrant observations. The variance-inflation model has the advantage of being parsimonious and flexible, which makes it attractive given the numerous specification and data problems that plague the empirical growth literature. As an example, consider a combination of two distributions, with low and high variance, and within these distributions observations are identically and independently distributed. Combining these two gives a mixture distribution with different variances. Heterogeneous parameters can be handled through mixture distributions over one or more parameters (random coefficients model), but this would quickly get cumbersome with a large number of parameters. Fernandez and Steel (2000) examine Bayesian inference within the confines of the linear regression model, focussing on the theoretical basis of independent sampling from a scale mixture of normal distributions of the regression errors.

In both Classical and Bayesian settings parameter estimates will be sensitive to the particular set of assumptions which underlie the approach. The robust model averaging approach introduced in Doppelhofer and Weeks (2011) combines model averaging with a flexible and parsimonious mixture model that allows for fat-tailed errors compared to the normal benchmark case. Inference and economic analysis are made robust with respect to outliers and unequal variances by allowing *a priori* for thicker tails of the distribution of regression errors compared to normal benchmark model averaging. Suppose that the regressions errors are independently normally distributed:

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{\Omega})$$

with diagonal covariance matrix $\boldsymbol{\Omega} \equiv diag(\omega_1, ..., \omega_N)$. The errors for each observation (or country $i = 1, ..., N$) are scaled by a variance inflation term ω_i with an independent mixing distribution. Geweke (1993) demonstrates the equivalence of such a normal mixture model with an independent Chi-square prior to a model where errors are drawn from an independent Student-t distribution, where the degrees of freedom determine the fatness of the tails and the prior weight on outliers.

This approach to robustness, and that of alternative approaches based on mixture models, depends on parametric (prior) assumptions about the distribution of errors. In the context of model uncertainty about potential growth determinants, it is our goal to accommodate observations with different degrees of reliance within a robust framework that also addresses other aspects of model uncertainty. A standard approach to robustify inference is the use of heteroscedasticity-consistent standard errors (HCSE). Although biased in finite samples, HCSE represents an improvement upon OLS estimates using a minimal set of assumptions (see White, 1980; MacKinnon and White, 1985). Lancaster (2009) demonstrates that HCSE are reasonable approximations to the posterior standard deviations around the OLS estimator (see also Poirier, 2008).

Conditional on having estimated the covariance matrix $\mathbf{\Omega}$, all other quantities of interest for a given model are estimated by using Generalized Least Squares (GLS) instead of Ordinary Least Squares (OLS). The parameters of the model are drawn from their respective conditional distributions using the Gibbs sampler.

Below we introduce the Bayesian Bootstrap that we will use to address both the question of data uncertainty for a given model, with minimal prior information.

2 The Bayesian Bootstrap

An alternative approach to robust inference is to use the Bayesian bootstrap developed by Rubin (1981). One variant of this approach, so-called *Bagging* (short for Bootstrap Aggregating) is used to generate robust predictions accounting for data problems (see Breiman, 1996; and Clyde and Lee, 2001). The logic of the Bayesian bootstrap is to consider parameters as functionals of the data (i.e. moment conditions) and to sample directly from the posterior distribution of the data.

In the context of a *single* model, Poirier (2008) shows that posterior weights can then be used to weight individual observations such that the resulting parameter estimates have a weighted least squares representation. Consider first a particular model and associated set of regressors $\mathbf{X}$. Let $\mathbf{Z} = \{y_i, \mathbf{x}_i\}$ denote rows of the data matrix, then we may consider rows of $\mathbf{Z}$ as realizations of multinomial variates on, for example, $G+1$ points of support with probabilities $\{\omega_g\}$. The Bayesian bootstrap assigns an appropriate prior distribution to the data, namely to the vector of probabilities $\boldsymbol{\omega} = \{\omega_g\}$. An obvious candidate is the Dirichlet distribution which is conjugate to the multinomial distribution in the same way that the beta is conjugate to the binomial distribution. We can then construct the posterior distribution of the data, which in turn can be used to obtain the posterior distribution of the parameters $\boldsymbol{\beta} = \boldsymbol{\beta}(\boldsymbol{\omega})$, given that $\boldsymbol{\beta}$ has been defined as a functional on the distribution of $\mathbf{Z}$. As Lancaster (2009) emphasises, this estimation framework does not impose restrictions on the conditional distribution of y given $\mathbf{X}$, accommodating both non-linearity and heteroscedasticity.

The Linear Model and the Bayesian Bootstrap

An application of Bayesian bootstrap to the linear model is provided below.

Let $z_i = (y_i, \mathbf{x}_i)$ where $\mathbf{x}_i$ is the i'th row of $\mathbf{X}$. Define the functional $\boldsymbol{\beta}$ by the condition that

$$E(\mathbf{X}'(y - \mathbf{X}'\boldsymbol{\beta}) = 0.$$

Thus,

$$\beta = [E(\mathbf{X}'\mathbf{X})]^{-1}E(\mathbf{X}'y)$$

where a typical element of $E(\mathbf{X}'\mathbf{X})$ is $\sum_{i=1}^{n} \mathbf{x}_{il}\mathbf{x}_{ij}\omega_i$ and a typical element of $E(\mathbf{X}'y)$ is $\sum_{i=1}^{n} \mathbf{x}_{il}y_i\omega_i$. (Defined in this way $\boldsymbol{\beta}$ is the coefficient vector in the linear projection of

y on $\mathbf{X}$). Thus we can write $\boldsymbol{\beta}$ as

$$\beta = (X'P\mathbf{X})^{-1}\mathbf{X}'Py$$

where $P = diag\{\omega_i\}$. As P varies from realisation to realisation so does $\boldsymbol{\beta}$ and this variation is the Bayesian bootstrap (posterior) distribution of $\boldsymbol{\beta}$.[3]

Suppose we now add uncertainty over the model $m = 1, ..., M$, represented by regressors $\mathbf{X}_m$. Combining the Bayesian bootstrap with model averaging, weights are applied both at the level of the model, and at the level of the observation the linear predictor for an outcome Y can then be written as

$$\widehat{Y} = \frac{1}{R}\sum_{r=1}^{R}\widehat{Y}^r, \tag{1}$$

where $\widehat{Y}^r = \sum_m^M \pi(m|\mathbf{X}_m, Y, \boldsymbol{\omega}^r)\widehat{Y}_m^r$ denotes the predictor for the r^{th} bootstrap sample adjusted for model uncertainty. $\widehat{Y}_m^r = \mathbf{X}_m'\widehat{\boldsymbol{\beta}}_m^r$ is the predicted value for the m^{th} model and the r^{th} bootstrap sample, with $\pi(m|\mathbf{X}_m, Y, \boldsymbol{\omega}^r)$ the attendant posterior weight. Note that parameter estimates $\widehat{\boldsymbol{\beta}}_m^r$ are calculated conditional on model m and using weights $\boldsymbol{\omega}^r$. By averaging over the R bootstrap samples, the overall predictor $\widehat{Y}$ accounts for both uncertainty at the level of the model, and uncertainty on the reliability of individual observations. Inference is then unconditional with respect to a space of models, and also made robust to outliers and parameter heterogeneity.

This research project is designed to extend the robust model averaging introduced by Doppelhofer and Weeks (2011), and in doing so integrate the Bayesian Bootstrap within a modelling framework that admits uncertainty over the true model. Our approach can be summarised by the following layers of model and data uncertainty:

1. uncertainty over correct models $m = 1, ..., M$, representing, for example, uncertain growth determinants $\mathbf{X}_m$

2. uncertainty over distribution of errors and possibility of outliers: here we will use Bayesian bootstrap to free us from strong parametric priors; alternatively, use some parametric prior (as in Doppelhofer and Weeks, 2011)

3. uncertainty over missing observations: refer to weighted hot-deck imputations that samples from posterior distribution of the data $p(Y, \mathbf{X}_m)$

We plan to relate the implied posterior distributions back to the three layers of uncertainty which are important in the context of robustness of growth determinants:

- Robust inference on economic growth: Sala-i-Martin, Doppelhofer and Miller (2004); Doppelhofer and Weeks (2011)

[3] The exact details of how the Bayesian bootstrap is implemented, and how the posterior distribution of $\boldsymbol{\omega}$ is generated, will be developed in the proposal.

- Outliers: Temple (2000); Zaman, Rousseeuw and Orhan (2001); Ciccone and Jarocisnski (2010)
- Missing data: Deaton (2010)

References

[1] Breiman, L. 1996. Bagging Predictors. *Machine Learning* 24: 123-140.

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

[3] Ciccone A, Jarocinski M. 2010. Determinants of Economic Growth: Will Data Tell? *American Economic Journal: Macroeconomics*, Forthcoming.

[4] Clyde MA, Lee H. 2001. Bagging and the Bayesian bootstrap. In *Artificial Intelligence and Statistics*, 169–174, Richardson T, Jaakkola T. (eds.), Proceedings of the Eighth International Workshop, Morgan Kaufmann Publishers.

[5] Conley TG, Hansen CB, McCulloch RE, Rossi PE. 2008. A Semi-Parametric Bayesian Approach to the Instrumental Variable Problem. *Journal of Econometrics* 144: 276-305.

[6] Deaton A. 2010. Price Indexes, Inequality, and the Measurement of World Poverty. *American Economic Review* 100(1): 5-34.

[7] Doppelhofer G. 2008. Model Averaging. in SN Durlauf and LE Blume (Eds.), *The New Palgrave Dictionary of Economics*, Second Edition. Palgrave Macmillan.

[8] Doppelhofer G, Weeks MJ. 2011. *Robust Growth Determinants*. NHH Dept. of Economics Discussion Paper No. 3/2011

[9] Durlauf SN, Johnson P. 1995. Multiple Regimes and Cross Country Growth Behaviour. *Journal of Applied Econometrics* 10(4): 365-84.

[10] Durlauf SN, Johnson PA, Temple, JRW. Growth Econometrics. in P Aghion and SN Durlauf (Eds.) *Handbook of Economic Growth* (Elsevier: North Holland, 2008)

[11] Eicher TS, Papageorgiou C, Raftery AE. 2009. Default Priors and Predictive Perfromence in Bayesian Model Averaging, with Applications to Growth Determinants. *Journal of Applied Econometrics*, Forthcoming.

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

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

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

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

[16] Hendry DF, Santos C. 2005. Regression Models with Data-Based Indicator Variables. *Oxford Bulletin of Economics and Statistics* 67(5): 571-95.

[17] Hjort NL, Holmes C, Müller P, Walker SG. 2010. *Bayesian Nonparametrics*. CUP.

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

[19] Lancaster, T. (2004). *An Introduction to Modern Bayesian Econometrics.* Blackwell Publishing.

[20] Lancaster, T. (1997). Bayes WESML Posterior inference from choice-based samples. *Journal of Econometrics* Vol. 79, pp. 291-303.

[21] Lancaster, T. (2009) "Explaining Bootstraps and Robustness", mimeo, Brown University

[22] Leamer E.E. 1982. Sets of Posterior Means with Bounded Variance Priors. *Econometrica* 50(3): 725-36.

[23] Leamer EE, Leonard H. 1983. Reporting the Fragility of Regression Estimates. *The Review of Economics and Statistics* 65: 306-17.

[24] Ley E, Steel MFJ. 2009. On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regressions. *Journal of Applied Econometrics* 24: 651-74.

[25] MacKinnon J.G, White H. 1985. Some Heteroscedasticity- Consistent Covariance Matrix Estimators with Improved Finite Sample Properties. *Journal of Econometrics* 29: 305-25.

[26] Magnus JR, Wan TK and Zhang X. 2010. WALS estimation with Nonspherical Disturbances and an Application to the Hong Kong Housing Market. mimeo.

[27] Masanjala W, Papageorgiou C. 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: 671-82.

[28] Papageorgiou C. (ed.) 2007. Special Issue on the Empirics of Growth Nonlinearities, *Journal of Macroeconomics* 29(3).

[29] Poirer, D. J. (2008) "Bayesian Interpretations of Heteroskedastic Consistent Covariance Estimators Using the Informed Bayesian Bootstrap", mimeo, University of California, Irvine

[30] Rousseeuw PJ. 1984. Least Median of Squares Regression. *Journal of the American Statistical Association* 79: 871-80.

[31] Rubin DB. 1981. The Bayesian Bootstrap. *Annals of Statistics* 9: 130-34.

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

[33] Schultz TP. 1999. Health and Schooling Investments in Africa. *Journal of Economic Perspectives* 13(3): 67-88.

[34] Sturm JE, de Haan J. 2005. Determinants of Long-Term Growth: New results Applying Robust Estimation and Extreme Bounds. *Empirical Economics* 30: 597-617.

[35] Tan CM. 2009. No One True Path: Uncovering the Interplay between Geography, Institutions, and Fractionalization in Economic Development. *Journal of Applied Econometrics*. Forthcoming.

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

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

[38] Zaman A, Rousseeuw P.J. and Orhan M. 2001. Econometric Applications of High-Breakdown Robust Regression Techniques. *Economics Letters* 71: 1-8.