TECHNICAL APPENDIX

for

DEMOCRACY AND ECONOMIC GROWTH: A META-ANALYSIS

Hristos Doucouliagos and Mehmet Ali Ulubaşoğlu

American Journal of Political Science
Vol. 52, No. 1., January 2008, Pp. 60-82.

AN INTRODUCTION TO META-ANALYSIS TECHNIQUES

Meta-analysis is a set of statistical techniques that, when applied to an empirical literature, are designed to: (a) summarize and integrate research findings; (b) evaluate and (c) explain the between-study differences in research findings.

1. Collection of studies

Meta-analysis requires the identification of primary studies and the coding of information from them.

Which studies should be included? Three approaches are adopted in the literature. First, analysts take the effort to identify as many studies as possible. This involves a comprehensive search of relevant databases, analysis of citations and careful study of references. The aim is to construct a dataset that represents the population of all publicly available estimates on a given topic. This is our preferred approach. Second, analysts make take a random sample of the available studies. This is easier to do if there is a database available from which samples can be drawn randomly. Such databases are currently unavailable in economics and political science and, hence, this approach is problematic (i.e., what makes the selection of studies random is not clear). Third, analysts may take the population of studies from a given year. For example, analysts may decide to include only studies published since 2000. While this has the benefit of making the data construction process easier, it does raise issues about the representativeness of the sample. Under certain circumstances, however, this approach has its merits. In all cases, care should be taken to ensure that the studies are comparable.

Which estimates should be included? Empirical studies usually report more than one set of results. Hunter and Schmidt (2004) recommend using the average of the different estimates for each study (this can be a weighted average). Doucouliagos and Ulubasoglu (2006) recommend using the best-set, although it can be difficult in some cases to identify which is the author's preferred set of results. A third approach is to use all the comparable estimates. Alternatively, the meta-analysis can be applied to the average-set, best-set and all-set and the results compared. The all-set is particularly useful if interest lies on identifying the source of heterogeneity in results (moderator analysis).

Should unpublished studies be included? For the democracy-growth meta-analysis, we chose to use only published studies. Published studies can be taken to be the final version of the empirical findings and under certain circumstances are higher quality estimates in that they have gone through the refereeing process. There may be situations, however, where the inclusion of unpublished studies is warranted. For example, in a newly emerging literature, a significant portion of the extant evidence will still be in the form of unpublished dissertations and working papers. It may make a lot of sense to include these in a research synthesis.

2. Descriptive statistics in meta-analysis

Descriptive statistics are used to summarize research findings. They are used also for the purpose of research synthesis. Two methods are in general use: Counts of coefficients with different signs and significance, and calculations of various averages.

2.1 Vote counting

All existing reviews of the democracy-growth literature have either explicitly or implicitly used vote counting. This involves counting the number of coefficients (estimated effects) that fall under certain categories (e.g. positive or negative, elastic or inelastic, statistically significant or non-significant). One advantage of vote counting is that it offers insights on the observed distribution of the reported findings. For example, Table 1 shows the distribution of our results in the form of vote counts.

Table 1. Vote counts of the published estimates of democracy-growth effects

		Negative		Positive	
		Significant	Insignificant	Insignificant	Significant
All-set	483 Estimates	72 (15%)	101 (21%)	180 (37%)	130 (27%)

However, counting the number of signs should not be given too much weight, as it does not provide a method for research *synthesis*. Moreover, it ignores information provided by the confidence intervals. For example, consider the coefficients and t-statistics associated with the following four studies:

Barro (2000) reports a coefficient of +0.05 (t-statistic of +1.83)
Leblang (1997) reports a coefficient of +0.12 (t-statistic of +2.18)
Dawson (1998) reports a coefficient of -0.003 (t-statistic of -0.05)
Gasiorowski (2000) reports a coefficient of -0.12 (t-statistic of -1.25).

Taken together there is one positive and statistically significant effect (Leblang), one positive and weakly statistically significant effect (Barro) and two negative but not statistically significant effects. However, once sampling error is considered in the form of confidence intervals, all four studies overlap significantly. The 95% confidence intervals for each of these studies are, respectively:

-0.004 to +0.11, +0.01 to +0.23, -0.11 to +0.10 and -0.32 to +0.07.

Rather than an inconclusive result, the four studies taken together actually share a common interval range of +0.01 to +0.07. With meta-analysis we can combine all studies and avoid the potential problems of sign counting.

2.2 Average effects

The effect between two variables (holding other effects constant) established by a literature can be derived as a weighted average of the associated estimates:

(1) $$\varepsilon = \sum [N_i \varepsilon_i] / \sum N_i$$

where ε is the *standardized* effect (such as an elasticity, a t-statistic or a partial correlation) from the i[th] study and N is the associated weight.

It is important that a standardized measure of an effect is used. In general, the use of regression coefficients is not appropriate. Elasticities are preferable but in most empirical literatures, studies do not provide enough information from which elasticities can be calculated (the calculation of elasticities depends on the functional form. Formulae can be found in most econometrics textbooks). In such cases, the t-statistic or the associated partial correlations can be used.

The partial correlation coefficients are calculated by using the *t*-statistics reported in the primary studies. Where t-statistics are not reported, they can be approximated from the reported levels of statistical significance, or from the reported regression coefficients and standard errors. The formula used to calculate partial correlations is:

$\sqrt{t^2/(t^2+df)}$, where t is the t-statistic and df is degrees of freedom

Note that this will always produce a positive number, so it is necessary to convert it to a negative number if the regression coefficient is negative (see Greene 2000, chapter 6).

For the purposes of weighting, N is normally some measure of precision or research quality. Measures of precision include the sample size or the estimate's standard error. Measures of research quality include the number of citations received and the journal's impact factor. It is a good idea to use different weights and explore the sensitivity of the results. In practice, however, the different weights appear to have little effect.

In addition to weighted averages, the findings can be summarized also by the median and the unweighted mean.

2.3 Confidence intervals

Confidence intervals in meta-analysis can be calculated in several ways. Hunter and Schmidt (2004) derive the formula for the standard error in the mean correlation for a homogenous group of studies, as well as the standard error in the mean correlation for a heterogenous group of studies. Hedges and Oklin (1985) use a slightly different procedure. We prefer to follow Adams *et al.* (1997) and use resampling techniques to construct bootstrap confidence intervals. Bootstrap confidence intervals are more conservative. The 95% confidence intervals of partial correlations were constructed using the bootstrap, with 1000 iterations (with replacement) to generate the distribution of democracy-growth effects (see Efron and Tibshirani, 1993). The lower and upper 2.5 percent of the values of the generated distribution are used to construct the 95 percent confidence intervals.

2.4 Credibility intervals

A credibility interval is the Bayesian equivalent of a confidence interval and is based on the idea that the underlying population correlations (in our case between democracy and economic growth) may vary across studies (see Hunter and Schmidt 2004). That is, there may be a distribution of parameter values, rather than a single value. Confidence intervals are constructed around a single population value, while credibility intervals highlight the *distribution* of population values. In this sense, credibility intervals are more important and informative than confidence intervals. In order to construct a credibility interval, it is necessary to compare the observed variance of the correlations to the variance expected from sampling error. Hunter and Schmidt (2004, p. 83) show that the observed variance across studies results from two sources: (a) variation in the population correlations and (b) variation in sample correlations produced by sampling error. That is:

(2) $$\sigma_r^2 = \sigma_\rho^2 + \sigma_e^2$$

where σ denotes variance, and the subscripts r, ρ and e denote observed, population and sampling error, respectively. Thus the observed variance in correlations should be corrected for the impact of sampling error. Following this logic, we calculated the observed variance of partial correlations drawn from each study, with each partial correlation weighted by the associated sample size. See Hunter and Schmidt (2004) for the formula for the observed variance and the variance expected from sampling error.

2.5 Graphical representation

Three excellent ways to visualize the findings from an empirical literature are histograms, the funnel plot and time series plots.

The funnel plot is a scatter of the standardized effects and a measure of precision (sample size or standard errors). The funnel plot offers three important pieces of information. (a) *Ceteris paribus*, the more symmetrical is the plot the more representative is the observed distribution of findings and confidence with descriptive statistics is increased. (b) The funnel plot shows the degree to which empirical results converge towards one underlying population effect and the extent to which the literature has reported heterogeneous findings. (c) The center of a symmetrical funnel plot is an unbiased estimate of the underlying population effect.

Arranging the estimates in a chronological order and plotting these in a time series graph informs on whether the findings are stable over time and whether structural breaks have occurred.

Descriptive statistics can be calculated for all studies or for different groups of studies.

3. Regression based tests (Meta-Regression Analysis)

The meta-regression model (known as MRA) has been developed to analyze the multi-dimensional nature of the research process (Stanley and Jarrell 1998). The impact of specification, data and methodological differences can be investigated by estimating an MRA of the following (linear) form:

(3) $$\varepsilon_i = \alpha + \gamma_1 X_{i1} + \ldots + \gamma_k X_{ik} + \delta_1 K_{i1} + \ldots + \delta_n K_{in} + u_i$$

where ε_i is the standardized effect derived from the ith study (in our study we use the partial correlation), α is the constant term, X_j are dummy variables representing characteristics associated with the ith study, K_j are continuous variables associated with the ith study, γ and δ are the unknown regression coefficients, and u_i is the disturbance term, with usual Gaussian error properties.

Equation 3 is a fixed effects MRA and assumes that variation in ε_i can be explained by sampling error and *systematic* differences between studies (the *X* and *K* study characteristics variables).

The random effects version of the MRA is given by:

$$(4) \qquad \varepsilon_i = \alpha + \gamma_1 X_{i1} + \ldots + \gamma_k X_{ik} + \delta_1 K_{i1} + \ldots + \delta_n K_{in} + u_i + e_i$$

Equation 4 assumes that in addition to sampling error, the source of some of the variation in ε_i is due to *random* differences among studies that cannot be identified. The regression coefficients in 3 and 4 quantify the impact of specification, data and methodological differences on reported study effects (ε_i).

It is recommended that both fixed effects and random effects models be estimated (see Hunter and Schmidt 2004).

Where the assumption of statistical independence of the estimates is uncertain, the MRA models can be estimated using the bootstrap to derive standard errors.

The dependent variable in most MRA models is a continuous variable. However, there may be cases where a binary variable is used. For example, rather than using a partial correlation, the analyst may code the results as 1 if a positive finding was reported and 0 otherwise. Such models can be estimated as meta-probit models.

Robustness

It is a good idea to check the sensitivity and robustness of the MRA. Examples of such testing include: (a) comparison of fixed effects and random effects models; (b) removing the largest and smallest estimates; (c) using only those studies that the analysts regards as superior according to some criterion (e.g. published in leading journals or used a particular estimation procedure).

4. Software

A number of software vendors have in-built routines for meta-analysis. All of the techniques listed in this appendix can be performed with *Stata*. Other options include *Metawin* and *Comprehensive Meta-Analysis*.

References

Adams, D.C., Gurevitch J., Rosenberg, M.S., 1997. Resampling Tests for Meta-analysis of Ecological Data, *Ecology* 78, 1277-83

Barro, Robert J. 2000. Inequality and growth in a panel of countries. *Journal of Economic Growth* 5: 5-32.

Dawson, John W. 1998. Institutions, investment and growth: New cross-country and panel data evidence. *Economic Inquiry* 36: 603-19.

Doucouliagos, Chris (Hristos), and Mehmet A. Ulubasoglu. 2006. Economic freedom and economic growth: Does specification make a difference? *European Journal of Political Economy* 22: 60-81.

Efron, B, Tibshirani, R., 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York

Gasiorowski, Mark. 2000. Democracy and macroeconomic performance in underdeveloped countries - An empirical analysis. *Comparative Political Studies* 33: 319-349.

Greene, William H. 2000. *Econometric analysis*. London:Prentice Hall, 4th edn

Hedges, L. V., Olkin, I, 1985, *Statistical Methods for Meta-Analysis*. Academic Press, Orlando, FL

Hunter, John, and Frank Schmidt. 2004. *Methods of meta-analysis: Correcting error and bias in research findings*. London: Sage.

Leblang, David A. 1997. Political democracy and economic growth: Pooled cross-sectional and time-series evidence. *British Journal of Political Science* 27: 453-72.

Stanley, T.D. and Jarell, S.B., 1998. Gender Wage Discrimination Bias? a meta-regression analysis. *The Journal of Human Resources* 33, 947-73