European Journal of Political Economy 27 (2011) 586–599

Contents lists available at ScienceDirect

European Journal of Political Economy

journal homepage: www.elsevier.com/locate/ejpe

Institutions and economic performance: A meta-regression analysis

Adnan Efendic [a,*], Geoff Pugh [b], Nick Adnett [b]

[a] University of Sarajevo, School of Economics and Business, 71000 Sarajevo, Trg Oslobodjenja A.I. 1, Bosnia and Herzegovina
[b] Staffordshire University Business School, Stoke-on-Trent, Leek Road, ST4 2DF, UK

ARTICLE INFO

Article history:
Received 17 December 2009
Received in revised form 13 December 2010
Accepted 15 December 2010
Available online 23 December 2010

JEL classifications:
E02
O43

Keywords:
Institutions
Economic performance
Meta-regression analysis

ABSTRACT

This paper applies meta-regression analysis to the empirical literature that investigates the effect of institutions on economic performance. Although studies with growth-theoretic foundations do not yield robust evidence for an authentic empirical effect, we find more robust evidence of positive and large institutional effects on output levels. The partial correlations between institutional and economic performance variables are also influenced by model specification choices and, in particular, (non)treatment of the potential endogeneity of institutions. A corollary of such pronounced heterogeneities is that we cannot report a representative estimated effect size, although the evidence overwhelmingly suggests a positive influence of institutional quality on economic outcomes.

1. Introduction

Jones and Romer (2009) argue that institutions are now at the centre of mainstream economic theory, while Economides and Egger (2009) regard institutions as a primary determinant of economic performance. In particular, the proposition that institutions affect economic performance (North, 1990) is no longer controversial. This consensus is supported by the good number of econometric studies that have investigated the link between institutions and economic performance. Yet many economists argue that institutional analysis is still in its development stage (Brousseau and Glachant, 2008; Chang, 2006; Furubotn and Richter, 2005; Kirman, 2007), is still "well short of a unified analytic framework" (Potts, 2007, p. 34) and that more research needs to be done before the institutional perspective can be fully operationalized (Pelikan, 2003; Rodrik, 2004a). The contribution of this paper is to report a meta-regression analysis (MRA) of the empirical literature on institutional quality and economic performance; and, hence, to suggest ways in which this evidence base may be criticised and developed to better inform institutional perspectives on economic performance.

Conventional narrative review establishes that this literature typically reports statistically significant and positive effects of institutional improvement on economic performance. MRA moves beyond such qualitative judgements. MRA is "the regression analysis of regression analysis" (Stanley and Jarrell, 1989). In this paper, we apply statistical tools developed by MRA practitioners to test for and measure the extent of selection or publication bias in the empirical institutional literature (Stanley, 2005; 2008); to identify and quantify the authentic effects of institutions net of publication bias; and to explain and quantify the heterogeneity of reported institutional effects. Accordingly, a particular contribution of this paper is to raise awareness about the consequences of specification choices, especially with respect to modelling the potential endogeneity of institutions.

The econometric investigation of institutions and economic performance is a relatively new area and MRA is a relatively new econometric tool. The paper most closely related to our work is Doucouliagos (2005), which investigates the existence of

* Corresponding author. Tel.: +387 61 162 249; fax: +387 33 275 900.
E-mail addresses: A.Efendic@staffs.ac.uk, Adnan.Efendic@efsa.unsa.ba (A. Efendic).

0176-2680/$ – see front matter
doi:10.1016/j.ejpoleco.2010.12.003

publication bias in the economic-freedom and economic-growth literature. However, because the primary studies in this literature do not necessarily articulate this relationship within the framework of institutional economics, whereas the present paper investigates the explicitly institutional literature on institutions and economic performance, there is only limited overlap in the studies considered by Doucouliagos (2005) and the present paper. Moreover, overlap exists at all only because "freedom indicators" have been used as institutional proxies and in papers with a theoretical basis explicitly in institutional economics. Another difference between the two studies is that the primary focus of Doucouliagos (2005) is on publication bias, whereas we focus mainly on heterogeneity within the institutional literature.

The paper is structured as follows. Section 2 discusses issues in the analysis of institutional effects on economic performance together with associated heterogeneities that may affect empirical findings. Section 3 explains our approach to MRA. We then report and discuss our findings on publication bias and authentic empirical effect: first from graphical analysis and then from bivariate MRA (Section 4); and finally from multivariate MRA (Section 5). Section 6 investigates the robustness of our results, focussing on differences between studies of output growth and of output levels. Concluding remarks reported in Section 7 focus on the pronounced heterogeneities discovered in this literature and their implications for future research.

2. Institutional economics and empirical heterogeneities

In order to explain differences in economic performance, economists once focused on physical and/or human capital and technological change as sources of economic growth. However, experience from, for example, transition at the end of the last decade suggested that without efficient institutions the standard factors of production are not capable of delivering rapid growth (Eicher et al., 2006). The New Institutional Economics (NIE) contributes to explaining economic growth and income differences by considering not only the standard factors of production but also institutions. Hence, while mainstream neoclassical theory assumes the institutional framework within which markets operate, institutional theory and the recent experience of transition economies in particular suggest that market-promoting institutions cannot be taken for granted.

Institutional economists are recognized by their attempts to integrate institutions, which are generally defined as formal and informal "rules of the game" and their enforcement characteristics (North, 1990), into economic theory and applied work (Bates, 1995). The importance of institutions is linked to their direct influence on economic outcomes (e.g. through the effects on transaction costs; hence, on the total costs of production) or indirect economic channels (e.g. certainty or incentives to invest in physical or human capital). Accordingly, institutional economists have been trying to identify the consequences of "imperfect institutions" (Eggertsson, 2005, p. 1), suggesting that development should be redefined as economic growth plus appropriate institutional change to sustain growth (Toye, 1995). Institutionalists are trying "to complete the neoclassical program" (Khan, 1995, p. 71) by creating a theory of non-market institutions based on neoclassical foundations (Bates, 1995). It has been argued that the NIE does not fundamentally challenge the progress of neoclassical economics nor reject the main analytical tools it has developed (Brousseau and Glachant, 2008; Joskow, 2008). Yet nowadays some general changes producing a more "relaxed" neoclassical view are visible (Eggertsson, 1990; Stein, 1995), with at least the institutions of property rights being recognized as important preconditions for a successful economy (Chang, 2006; North, 2005).

While many researchers have been inspired by North's (1990) championing of the importance of institutions, empirical "institutional" research has given rise to a whole range of heterogeneities and corresponding questions: How to measure/proxy institutions? Which specification to use? Whether or not and, if so, how to address the potential endogeneity of institutions? What timing-effects to investigate? Which sample of countries to analyze? We now consider each of these questions in turn. In the next section, we explain how these sources of heterogeneity in the empirical literature are addressed by MRA.

Disparity among institutional proxies in empirical research, with consensus yet to emerge, suggests that there is no single variable "ideally" representing institutions (Aidis et al., 2009; Shirley, 2008). However, institutional performance has been proxied mainly by using aggregated "institutional" indices (e.g., the Heritage Foundation Index of Economic Freedom and the EBRD Structural and Institutional Change Indicators). In addition, researchers have investigated particular components of those indices (e.g., components measuring Property Rights and Rule of Law); other indicators (e.g. the level of corruption and shadow economy); or variables constructed from surveys. Although studies using all these different institutional proxy variables find that improving institutions positively affects economic performance, MRA will help to ascertain whether or not – and, if so, to what extent – choice of institutional proxy is associated with systematically different empirical findings.

Another challenge confronting institutional applied work is to decide which empirical specification to use to investigate the effect of institutions on economic performance. Again, there is no consistency in the empirical strategies used by researchers. From conventional literature review, we can identify specifications in which institutions are the only explanatory variable of economic performance, hence estimated in a simple bi-variate specification (e.g., Dreher et al., 2007; Sachs, 1996). Other authors argue that such specifications are just a reduced form of a fully specified model that should include standard growth factors together with institutions (Gwartney et al., 2004; Jones and Romer, 2009). Accordingly, if one applies an "extended regression" that combines institutional proxies with traditional economic growth factors, by "borrowing" from neoclassical economics (Kirman, 2007, p. 34), we come to some form of the "extended production function" specification (Baliamoune-Lutz and Ndikumana, 2007; Glaeser et al., 2004; Gwartney et al., 2006; Ulubasoglu and Doucouliagos, 2004; Vijayarghavan and Ward, 2009). Finally, there is a whole range of "other specifications" that combine institutions with factors like geography, trade, and macroeconomic policy. Seminal research by Rodrik et al. (2004) may be representative for such specifications, in which authors analyze the so-called "deep determinants of income" by combining institutional, integration, and geographical explanatory variables (see also Alcala and Ciccone, 2004; Easterly and Levine, 2003; Jacob and Osang, 2007; Sachs, 2003). The choice between all these different types of specifications is not straightforward and some authors even choose to specify their models on the basis of the general-to-specific approach (e.g., Klomp and de Haan, 2009). Accordingly, our MRA will investigate the implications for the reported findings of this considerable heterogeneity among empirical specifications. In turn, these implications should help the authors of future studies by raising awareness of the possible consequences of specification choices in modelling institutional effects on economic performance.

Many economists agree that institutional quality is endogenous to economic performance (e.g., Acemoglu and Johnson, 2005; Chang, 2006; Rodrik, 2004a; Shirley, 2008). The problem of the endogeneity or exogeneity of institutions is one of the most serious difficulties in empirical institutional work (Nye, 2008) and remains an important econometric issue in this literature (Ahlerup et al., 2009). However, we have identified a number of approaches to addressing the potential endogeneity of institutions, mainly by employing different instrumental variable techniques, together with papers in which the potential endogeneity of institutions has not been considered at all. Nevertheless, if potential endogeneity of institutions exists, and if it is not "properly" investigated and controlled for, then this might bias estimates. Accordingly, our MRA will investigate how different approaches to the problem of the endogeneity of institutions may affect findings on the link between institutions and economic performance.

Historical evidence and contemporary research suggest that institutional change is gradual in most cases. However the experience of transition economies provides an example in which institutions changed practically overnight and have continued to evolve rapidly (Eicher and Schreiber, 2010; Opper, 2008; Raiser et al., 2001; Redek and Susjan, 2005). Empirical studies, especially for non-transition countries where data availability is better (and generally institutional change less dynamic), typically emphasize the long-range influence of institutions on economic performance (Gwartney et al., 2006; Richter and Timmons, 2008), though others examine the short-run effects (Eicher and Schreiber, 2010; Sachs, 1996). Accordingly, our MRA will examine potential differences of institutional effects between transition and non-transition samples as well as the timing of institutional effects.

Finally, a notable aspect of heterogeneous specifications in the empirical literature is the choice of dependent variable; primarily, either output growth or output level. Accordingly, it might be the case that institutions display different correlations with, respectively, output-level and output-growth. Accordingly, a particular feature of the MRA reported below is an investigation of the implications of choice of dependent variable for reported institutional effects.

3. Meta-analysis

Following the methodology applied by Stanley (2001), the search process for relevant empirical studies included: the EconLit database; working paper series; Google Scholar; and references obtained from the literature up to the end of 2008. Key words used in the search were: "institutions + economic growth"; "institutions + economic development"; "institutions + economic performance". These approaches identified more than 200 papers, including 40 econometric studies together reporting 112 regressions of interest, which form the observations in the MRA database.[1]

Most authors report several regressions arising from testing down procedures and/or from comparative analysis of different specifications (e.g. institutions in different combinations with other variables). However, our selection criterion was in each case to take authors' preferred or base specification(s) for the MRA database, although in some cases the reader is left to identify these. Even base specifications usually permitted more than one regression per study, since the author(s) often applied their preferred specification(s) to different samples, different institutional indices or sub-indices, different time periods, and different endogeneity assumptions. These were recorded as independent regressions, because we wanted to investigate the implications of these heterogeneities for the reported effects. This is common practice among MRA researchers (Doucouliagos and Stanley, 2009). Consequently, the number of reported regressions per study ranges from one (as in: Beck and Laeven, 2006; Chousa et al., 2005; De la Croix and Delavallade, 2009; Hall and Jones, 1999; and Redek and Susjan, 2005) up to six (Aixala and Fabro, 2008; Moers, 1999), with an "average" of 3.1. The full list of studies included in the MRA is available on-line.[2] Since institutional applied research in economics started quite recently, the oldest research reported in this sample was published in 1995 (i.e. Knack and Keefer, 1995; Mauro, 1995) while the latest studies are from 2008 (i.e. Aixala and Fabro, 2008; Richter and Timmons, 2008).

In each regression reported in the literature under investigation, the dependent variable is related to some measure of institutional quality as the variable of interest. The dependent variables chosen by researchers may be grouped into two categories: variables related to the level of output (GDP; GDP per capita; GNP per capita; and GDP per worker), which are represented by the dummy variable (*LEVELGDP*); and variables related to growth (GDP growth; GDP per capita growth; and GDP per worker growth), which are represented by another dummy (*GRGDP*). Accordingly, all MRA tests will be conducted on the whole sample (40 studies and 112 observations) and, separately, for the output-growth (20 studies and 52 observations) and output-level (21 studies and 60 observations) sub-samples.[3]

The studies in this MRA sample used a range of institutional indices that include different units, scales and components. Accordingly, we need a standardized measure of the estimated effect of the institutional variable on economic performance; i.e., a measure without dimensionality (Stanley and Jarrell, 1989). For each reported regression we calculate the partial correlation coefficient (henceforth *PCC*) between the dependent variable and the institutional variable of interest. Partial correlation is a

[1] In our MRA sample, we *include* all econometric studies of the effects of institutions on levels or growth of economic output, but *exclude* all studies with other dependent variables such as human capital, physical capital, trust, entrepreneurship, trade, etc.

[2] http://www.efsa.unsa.ba/~adnan.efendic/EJPE_-_List_of_studies_used_in_MRA_-_Efendic_et_al_2010.pdf.

[3] One study (Gwartney et al., 2004) reports two results: one for output growth; and the other for output levels. For the purpose of sub-sample comparison only, these were coded as separate studies. Hence, the combined number of subsample studies (41) is one greater than the whole sample.

standardized measure of the degree of association, controlling for the influences of other factors (Greene, 2008, p. 31; for an application in MRA, see Doucouliagos and Laroche, 2009). Hence, this standardization will enable direct comparison between different studies in the sample, as well as between different MRA subsamples (i.e. the output growth and output level studies).

The simplest MRA model regresses the standardized effect size (i.e. *PCC*) on an intercept ($\hat{\alpha}$) and the conventional measure of precision (*SEpcc*, the standard error of the *PCC*):

$$PCC_{s,i} = \hat{\alpha} + \hat{\beta}_0 \cdot SEpcc_{s,i} + \hat{\varepsilon}_{s,i} \tag{1}$$

where $s = 1,\ldots, 40$ indexes the 40 studies in our dataset, $i = 1,\ldots, 112$ indexes the 112 individual regressions reported by the studies, $\hat{\alpha}$ and $\hat{\beta}_0$ are estimated, and $\hat{\varepsilon}_{s,i}$ is the regression error term.

Eq. (1) is rarely estimated due to a common problem with heteroskedasticity (Stanley et al., 2008). Instead, Stanley (2008) recommends that this bivariate regression be weighted with the standard error of the *PCC* to reduce heteroskedasticity and yield more efficient estimates. Hence, Eq. (1) is divided by the standard error of the *PCC* to obtain Eq. (2), which is the weighted least squares (WLS) version of Eq. (1).

$$TSTAT_{s,i} = \hat{\alpha} \cdot \left(1 / SEpcc_{s,i}\right) + \hat{\beta}_0 + \hat{\varepsilon}_{s,i} \cdot \left(1 / SEpcc_{s,i}\right) \tag{2}$$

After the transformation of Eq. (1), the dependent variable in Eq. (2) ($TSTAT_{s,i}$) is the t-statistic measuring the significance of the coefficient on the variable of interest in the respective primary regression.[4] Yet, because Eq. (2) is a WLS regression, the interpretation should be in terms of Eq. (1); i.e., the estimated effects are to be interpreted in terms of partial correlations and *not* in terms of the t-statistics.

In Eq. (2), the slope and intercept terms are now reversed compared to Eq. (1) and the inverse standard error becomes the independent variable in the MRA (Stanley et al., 2008). Accordingly, in Eq. (2): the dependent variable is the t-value ($TSTAT_{s,i}$) of the estimated coefficient on the institutional variable in each regression; $\hat{\alpha}$ is the same coefficient as in Eq. (1) and thus still provides an estimate of the “true” or underlying effect size in terms of the partial correlation coefficient; and $\hat{\beta}_0$, which is also from Eq. (1), measures publication bias. In Eq. (1), as the standard errors rise from zero, estimates of the effect size become ever more imprecise but, in the absence of systematic bias, should remain randomly distributed around the true effect estimated by $\hat{\alpha}$; in this case, $\hat{\beta}_0$ will not be significantly different from zero. However, in the presence of publication (selection) bias, estimates of the effect size are not randomly distributed around the true effect $\hat{\alpha}$ but are biased away from it, because more weakly estimated results are subject to greater selection efforts in accord with some publication selection criteria (e.g., that estimated coefficients should be positive or negative); in this case $\hat{\beta}_0 \neq 0$ (For explanation and visual display, see Stanley, 2005, pp. 314–316). Publication bias is discussed further in Section 3 below. Hence, in Eq. (2) $\hat{\alpha}$, the coefficient on the inverse of the standard error of the partial correlation coefficient, *measures the underlying effect of institutional quality on economic performance identified in the literature and is corrected for publication selection*. Correction for publication selection accounts for a potential source of bias on the reported estimates that would otherwise remain unacknowledged.

The estimated bivariate MRA regression can be biased if important explanatory variables are omitted (Doucouliagos and Stanley, 2008). Accordingly, a more developed version of Eq. (2) is standard. Multivariate MRA (Doucouliagos and Laroche, 2009) enables investigation of the sources of heterogeneity in the literature (discussed in Section 2 above) by adding independent (or moderator) variables to Eq. (2). These additional meta-independent variables are also divided by the standard error of *PCC*. Hence, the general multivariate MRA is specified as follows:

$$TSTAT_{s,i} = \hat{\alpha} \cdot \left(1 / SEpcc_{s,i}\right) + \hat{\beta}_0 + \sum_{1}^{K} \hat{\lambda}_k \cdot \left(1 / SEpcc_{s,i}\right) Z_{k,s,i} + \hat{u}_{s,i} \cdot \left(1 / SEpcc_{s,i}\right). \tag{3}$$

In Eq. (3), $Z_{k,s,i}$ are $k = 1,\ldots K$ moderator variables each weighted by $(1/SEpcc_{s,i})$ (henceforth, *SEINVpcc*); $\hat{\lambda}_k$ are k coefficients to be estimated, where each one measures the impact of the corresponding moderator variable on the underlying effect of institutions on economic performance (because the measure of precision, $SEpcc_{s,i}$, is interacted with each moderator variable, the effect of each $\hat{\lambda}_k$ is to be interpreted by additive combination with $\hat{\alpha}$); and $\hat{u}_{s,i}$ is the meta-regression disturbance term, which has the standard characteristics.

The “meta-independent” or “moderator” variables used in this MRA are listed and explained in Table 1. These comprise the initial set of variables that were identified above (Section 2) as potential sources of heterogeneity in the econometric literature on institutions and economic performance. The final specification of a MRA is determined by the data (Stanley and Jarrell, 1989) and, hence, by model diagnostics.

Variation from study to study in the number of reported regressions (from 1 to 6, as we report above) implies that some studies may be over-weighted in their influence on the MRA estimates. Accordingly, the regressions reported below are weighted so as to

[4] See Fisher (1954, p. 194) for the relationship in each underlying regression between the t-statistic on the estimated coefficient on the variable of interest, the *PCC* between the dependent variable and the variable of interest, and the standard error of the *PCC*. This relationship informs the transformation of Eq. (1) into Eq. (2); namely: $SEpcc_{s,i} = (PCC_{s,i}/TSTAT_{s,i})$, and $TSTAT_{s,i} = (PCC_{s,i}/SEpcc_{s,i})$. The t-statistics come from the reported regressions in the MRA sample while, in each case, the *PCC* and its standard error are calculated by the authors of the MRA.

Table 1
Potential explanatory (moderator) variables.

Variables	Short explanation of moderator variables	Original		Transformed	
		Mean	St. dev.	Mean	St. dev.
YEARSE	The year an article is published (1995 = 1; 1996 = 2; 1997 = 3; …; 2008 = 14)	2003.4	3.1	106.8	67.7
NOINDSE	The number of independent variables used in regressions	4.0	3.3	49.6	62.3
PCC	The partial correlation coefficient between the dependent and institutional variable	0.4	0.2	–	–
SEINVpcc	The inverse standard error of the partial correlation coefficients	0.1	0.1	11.0	4.7
LNTIMEDSE	The observed time-horizon of the dependent variable (e.g. 1, 2, 5 years change/average), logarithmic transformation	10.4	28.5	9.6	14.8
LNTIMEINSSE	The observed time-horizon of the institutional variable (e.g. 1, 2, 5 years change/average), logarithmic transformation	5.1	7.8	8.8	14.1
SAMSE	Dummy, 1 if a sample is for transition countries, 0 otherwise	0.3	0.4	2.3	4.6
LEVELGDPSE	Dummy, 1 if a study uses the level of output as dependent variable, 0 otherwise	0.5	0.5	6.8	7.3
GRGDPSE	Dummy, 1 if a study uses the growth rate of output as dependent variable, 0 otherwise	0.5	0.5	4.2	5.1
SPINSSE	Dummy, 1 if a study uses only institutional variable(s) in the regression, 0 otherwise	0.2	0.4	2.3	4.7
SPPRSE	Dummy, 1 if a study uses institutional variable(s) in the modified extended production function , 0 otherwise	0.4	0.5	3.8	5.9
SPOTHERSE	Dummy, 1 if a study uses institutional and other variables but not the standard growth factors, 0 otherwise	0.4	0.5	4.8	6.6
COMPSE	Dummy, 1 if the institutional variable is a subcomponent, 0 is for aggregated index	0.6	0.5	5.5	5.8
INITIALSE	Dummy, 1 if the initial level of output is included in the specification, 0 otherwise	0.2	0.4	1.2	3.1
PANELSE	Dummy, 1 if authors use panel data, 0 cross-section	0.2	0.4	3.4	7.2
PROPRSE	Dummy, 1 if authors use property rights as the main institutional variable, 0 otherwise	0.1	0.3	0.8	2.6
RULESE	Dummy, 1 if authors use rule of law as the main institutional variable, 0 otherwise	0.1	0.4	1.7	4.7
LAGSE	Dummy, 1 if authors use the lagged institutional variable in the regressions (one-year lag), 0 otherwise	0.1	0.2	0.5	2.6
INSVARSE	Dummy, 1 if authors use instrumental variables (i.e. IV or 2SLS) to control for endogeneity, 0 otherwise	0.6	0.5	6.8	7.1
SLS3SE	Dummy, 1 if authors use the 3SLS procedure to control for endogeneity, 0 otherwise	0.1	0.2	0.6	2.5

Source: Authors. Note, in the preferred model reported in Table 3, the omitted categories are *LEVELGDPSE* for different studies; *SPOTHERSE* for different specifications; while with respect to endogeneity we omit the studies that do not control for the potential endogeneity of institutions. This is also explained in the text.

give the same weight to every study. This weighting procedure gives equal weight to each study; it is thus additional to – and differently motivated from – the WLS strategy.

The first weighting procedure – by the inverse of the standard error – is an essential step in the methodology of MRA: it addresses the heteroskedasticity inherent in effect sizes from heterogeneous primary studies; it gives greater weight to the more precise estimates; and the effect size is transformed into the t-statistic on the variable of interest in each regression in the sample. In contrast, the second weighting procedure is not essential. Rather it reflects a preference for giving all studies equal influence on the estimates rather than potentially allowing studies reporting multiple estimates to have more influence. Hence, we apply a second set of weights to all terms in Eqs. (2) and (3) (where the weight = 1/number of reported regressions in the MRA *from a particular study*) which has the effect of giving each study equal influence on the reported coefficients and standard errors. (For example, if Study one reports one regression then this regression is fully weighted; if Study two reports two regressions then each of these is weighted 0.5 so that, together, they have the same influence on the estimates as the single regression from Study one.) Of course, if our model is correctly specified such that all important heterogeneities are correctly controlled, then allowing some studies to exert more influence on the estimates than others should not matter, because weighted and unweighted estimates should be the same. Reassuringly, then, our regression results from estimating Eqs. (2) and (3) *without* the second set of weights are almost entirely consistent with the estimates reported below (i.e., from Eqs. (2) and (3) estimated with both sets of weights). Across the full and two sub-sample models there are no economically substantial changes in the sizes of estimated coefficients and the very few changes in statistical significance do not have any important implications for the interpretation of our results. Because these results add no further insight, we do not report them (however, they are available in an on-line appendix; see footnote 6).

4. Graphical investigation and initial testing for publication bias and authentic empirical effect

"Publication bias" in a literature arises when editors, reviewers, and/or researchers have a preference for results that are statistically significant (Stanley, 2005) and/or that satisfy theoretical expectations (Doucouliagos, 2005). Research based on smaller samples is usually at a disadvantage in finding statistically significant results, because limited degrees of freedom are associated with larger standard errors on estimated coefficients. Hence, authors who work with smaller samples may "search" more (e.g. across specifications; estimators; techniques; data sets; proxies/indices; etc.) in order to obtain more significant results, which *per se* makes publication bias more likely. Accordingly, publication bias can make empirical effects appear larger than they are in reality (Stanley et al., 2008).

Given the strong orientation of institutional literature towards a positive role of institutions in economic development, and the corresponding lack of competing hypotheses in the empirical literature under consideration, in this MRA evidence of publication bias may be detecting non-submission or, in the case of submission, non-publication of results suggesting that institutional

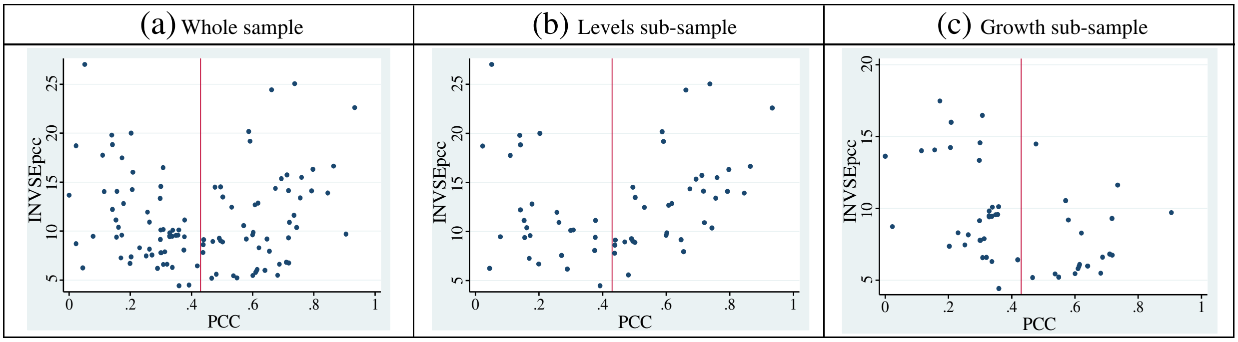

Fig. 1. Funnel plots: partial correlation coefficient (*PCC*) versus the inverse standard error of *PCC* (*INVSEpcc*) (whole sample mean *PCC* marked by the vertical line).

improvement does not have a significant effect and/or has a negative effect on economic performance.[5] Doucouliagos and Stanley (2008) conclude that publication selection could be strongest in literatures characterised by consensus. By implication, this may apply to the literature investigating the relationship between institutional quality and economic performance. Indeed, our sample reports only positive effects of institutions on economic performance, with differences only in the magnitude and statistical significance of these findings. In addition, institutional literature is consistent in reporting a positive effect while being very heterogeneous with respect to methodologies, data-sets, model specifications, estimation procedures, and endogeneity (non) modelling. Such heterogeneity giving rise to consensual results entails the possibility that authors modify their research strategy in different dimensions until the expected "correct" results emerge. Hence, we investigate the hypothesis that publication selection may inflate the average reported effect of institutions on economic outcomes.

To address these possibilities, MRA researchers have developed checking procedures as well as formal statistical tests both for detection and for correction of the effect of publication selection (Stanley, 2005, 2008). Following common practice, we begin with preliminary graphical analysis and continue with bivariate MRA.

The most common graphical approach to detecting publication bias in empirical literature is the funnel plot (Doucouliagos, 2005; Stanley, 2005), which compares the effect size for each regression (in our case, the *PCC*) against some measure of its precision (the inverse standard error of the *PCC*, *INVSEpcc*). The plot should appear as an inverted funnel, reflecting the heteroskedasticity induced by different sample sizes: the smaller the sample, the lower the precision and, hence, the more spread out the base of the plot; conversely, the larger the sample the more precise the estimate and the more concentrated the estimates around the best estimate of the true value of the *PCC* at the peak of the plot. The absence of publication bias requires symmetry around the peak values; a right skew, for example, suggests a selection process favouring larger effect sizes. For a good illustration of the symmetric "ideal type" of funnel plot that suggests the absence of publication selection in a research literature, see Stanley (2005, pp. 314–315; Figs. 1 and 2).

Fig. 1 displays the funnel plots for the whole sample and for two sub-samples with, respectively, dependent variables measuring growth and levels of economic performance.

Panel (c) is the easiest to interpret: the funnel plot for the sub-sample of studies with some measure of economic growth as the dependent variable displays the highest precision for a *PCC*, a little less than 0.2, while the plot is severely right skewed. This suggests publication bias of the hypothesised (positive) form. However, matters are not as straightforward for the whole sample (a) or for the sub-sample of studies with some measure of economic level as the dependent variable (b). According to Stanley (2005, p. 315) "it is the graph's symmetry (or asymmetry) that is crucial to assessing publication bias". From this perspective, neither of these plots suggests publication bias. However, their "U-shaped" appearance – almost the opposite of a funnel shape – may indicate more complex influences at work. For example, a plausible interpretation might be that we have a bimodal plot, reflecting either different research designs (heterogeneity created by methodological choice) or separate populations (genuine heterogeneity in the data). In either case, in both panels (a) and (b), the data can be partitioned into two groups: one group, to the left of the mean, has a peak *PCC* a little lower than 0.1 and displays a right skew; and another group, to the right of the mean, has a peak *PCC* a little lower than 0.8 and displays a left skew.

To explore these possibilities, we investigated those variables that are our main suspects for introducing heterogeneity into our sample to assess whether one or more of these may be driving the possibly bimodal character of these plots; namely: heterogeneity created by methodological choices (model specification, the use/non-use of IV techniques to address the potential endogeneity of institutions and the use of aggregated indices as the dependent variable); and a potential source of genuine heterogeneity (transition/non-transition samples in the primary studies).[6] From Fig. 1, it seems likely that the whole sample distribution is driven by the levels sub-sample. In turn, we found that these distributions can best be replicated by partitioning the

[5] Researchers' unwillingness to submit results with lack of statistical significance and/or the "wrong sign", or their inability to get them accepted for publication, is sometimes referred to as the "file drawer" problem. The consequent consignment of studies to the "file drawer" means that published studies may be a biased sample of the whole population of studies.

[6] For reasons of space, we do not report these; they are available at http://www.efsa.unsa.ba/~adnan.efendic/EJPE_-_Appendix_-_Efendic_et_al_2010.pdf.

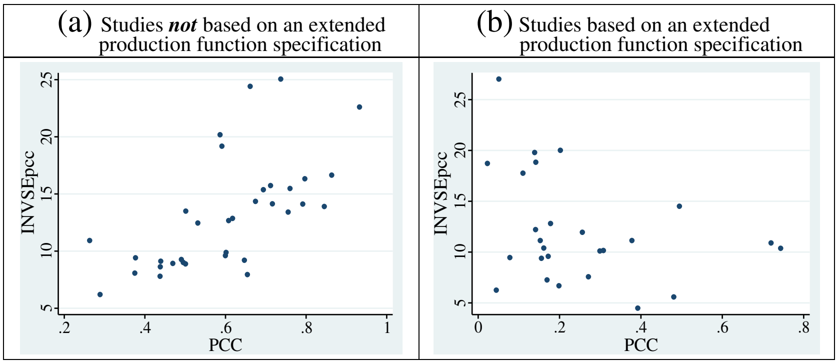

Fig. 2. Funnel plots: the levels sub-sample partitioned according to specification.

Table 2
Testing for publication bias and authentic empirical effect (Bivariate MRA FAT-PET model): OLS with cluster-robust standard errors; observations weighted to give each study equal weight.

Dependent variable: TSTAT, which is the t-statistic on the institutional variable of interest									
VARIABLES (short description)	All studies			Growth studies			Output-level studies		
	Coeff.	t-stat.	p-value	Coeff.	t-stat.	p-value	Coeff.	t-stat.	p-value
CONS (intercept)	0.11	0.08	0.940	2.51	3.97	0.001	−0.79	−0.31	0.757
SEINVpcc (inverse SE of PCC)	0.41	2.53	0.016	0.10	1.37	0.186	0.52	2.13	0.046
Model diagnostics									
Number of observations	No = 112			No = 52			No = 60		
R-squared	R-squared = 0.26			R-squared = 0.04			R-squared = 0.29		
F-test	F(1, 39) = 6.40			F(1, 19) = 1.88			F(1, 20) = 4.52		
Ho: *SEINVpcc* = 0	Prob > F = 0.016			Prob > F = 0.186			Prob > F = 0.05		
Ramsey RESET test	F(3, 107) = 0.14			F(3, 47) = 1.43			F(3, 55) = 0.23		
Ho: No omitted variables	Prob > F = 0.94			Prob > F = 0.25			Prob > F = 0.87		

Notes: The coefficient on *SEINVpcc* measures the magnitude of the effects of institutions on economic performance, corrected for publication selection. In columns 2 and 3 it measures the same effect in the output-growth and output-level sub-samples respectively. Reported t-statistics are based on heteroskedasticity cluster-robust standard errors.

levels sub-sample into studies that do not use an extended production function specification (34 observations) and studies that do (26 observations). The funnel plots for each group are displayed in Fig. 2.

Fig. 2 suggests that different model specifications might account for the bimodal appearance of the funnel plots for the whole sample (Fig. 1a) and for the levels sub-sample (Fig. 1b). However, this interpretation depends on only a few relatively high precision effect sizes. In general, funnel plots can be instructive for potential relationships and heterogeneities in the data. Yet, as Stanley (2005, p. 316) warns, "visual inspections are inherently subjective and somewhat ambiguous … publication selection need not be the only source of asymmetry". To proceed further, we use multivariate MRA to investigate the effect of institutional quality on economic performance conditional upon both publication bias and heterogeneities – whether real or methodological – in the data. Common procedure is to approach multivariate MRA by way of first estimating Eq. (2).

Stanley (2008, p. 108) argues that Eq. (2) may be used to test both for publication bias and for an authentic empirical effect "beyond publication bias". The Funnel Asymmetry Test (FAT) for detection of publication bias tests the hypothesis H_0: $\hat{\beta}_0 = 0$, where non-rejection is consistent with lack of publication bias (see Section 3 above). Alternatively, if the intercept term is significantly positive (or negative) then the effect size is subject to an upward (or downward) bias across its whole range, which is evidence of publication bias in the literature under investigation (Doucouliagos, 2005; Rose and Stanley, 2005; Stanley and Doucouliagos, 2006). The Precision-Effect Test (PET) tests whether the slope coefficient in Eq. (2) is zero: i.e. H_0: $\hat{\alpha} = 0$ (where rejection is consistent with an authentic empirical effect). Simulations by Stanley (2008) suggest that the estimate of $\hat{\alpha}$ serves as a test for "authentic" or "genuine" empirical effect corrected for publication bias. Results from the estimated bi-variate MRA model (2) are summarized in Table 2.

The regression for the whole sample suggests that publication selection may not exist and provides apparently strong evidence of an authentic empirical effect of institutions on economic performance; namely, a highly significant estimate of the partial correlation between institutions and economic performance. The mean value of the institutional effect is estimated to be 0.41 (compared to the unconditional mean of 0.43),[7] which is "large" in terms of the typical size of partial correlation coefficients

[7] According to Doucouliagos and Stanley (2008) this similarity suggests "little or modest" publication bias.

reported in applied economics (Doucouliagos, 2010). This suggests the importance of institutions in explaining economic outcomes, whether in terms of growth or income levels.

The FAT-PET results for the sub-sample with output-level as the dependent variable are consistent with the whole sample results. Yet this is not the case for the growth sub-sample, in which there is evidence of positive publication bias as well as lack of authentic empirical effect. Precisely the same conclusions arise from both non-weighted non-cluster robust and non-weighted cluster-robust estimates. Similar results are obtained by Doucouliagos (2005) for the literature focused on the economic freedom and economic growth relationship. The highly significant and positive $\hat{\beta}_0$ indicates publication bias in favour of a positive effect of institutions on economic growth. Moreover, $|\hat{\beta}_0|\rangle 2$ indicates "severe selectivity" (for more on "guidelines regarding the practical significance of publication selectivity" see Doucouliagos and Stanley, 2008, p. 13). Given that institutional research focused on economic growth is based on strong theoretical foundations, coming usually from extensions of the Solow growth model, this finding is consistent with our previous discussion on publication bias in consensual literatures. Overall, there is no clear evidence of an authentic effect of institutions on economic growth.

The bivariate MRA estimates reported in Table 2 may be biased when important explanatory variables are omitted (Doucouliagos and Stanley, 2009). Accordingly, we embed our bivariate MRA – hence, our FAT-PET procedure (Stanley, 2008) – within a fully specified multivariate meta-regression model. This enables us to control for sources of heterogeneity in the literature as well as to explore their consequences for the estimated effects of institutions on economic performance. To this end, the moderator variables listed in Table 1 model the impact on the estimated effect of institutional quality on economic performance of not only the choice of dependent variable but also of different specifications, samples, time-horizons observed, and methodologies for dealing with potential endogeneity. After controlling for these influences, we are able to test with greater confidence for publication bias as well as for the presence of an authentic empirical effect in the literature. In this case, if an empirical effect is identified in the literature, it is measured conditional on the sources of heterogeneity identified in the MRA as well as net of publication bias.

5. Multivariate meta-regression analysis

We estimate model (3) where $k=1, \ldots, 12$ moderator variables. The final model was chosen by the "general-to-specific" approach to econometric modelling, which is common practice in MRA (Stanley, 2005, p. 332). Hence, some of our initial set of potential moderator variables are excluded from the final model. The need to implement a testing down procedure was dictated by constraints on the available degrees of freedom, which reflects a relatively small sample and an initial list of 20 moderator variables (Table 1). Our testing down procedure was not conducted mechanically, purely according to threshold levels of statistical significance or fit, but was influenced also by model diagnostics and theoretical judgements concerning the importance of each variable and the corresponding dangers of omitted variable bias. For example, although the variable *COMPSE* is not statistically significant in the full model (Table 3), its retention substantially improves the model diagnostics (for the Ramsey RESET test, the p-value drops to 0.10 upon the exclusion of *COMPSE*). Moreover, *COMPSE* proves to be a significant and important variable in our later sub-sample estimates (Table 4).

Our results for the whole sample are reported in Table 3. Because individual regressions are clustered with studies, we report cluster-robust standard errors, which allow the error terms within each cluster (study) to be correlated, requiring only that they are not correlated across clusters (studies) (Baum, 2006). Consequently, we adopt the most conservative available criteria for statistical inference.

The model *F-statistic* indicates that the estimated MRA coefficients are jointly significant, while the overall fit of the regression is quite high for a meta regression ($R^2=0.77$). Accordingly, the model has largely captured the main sources of heterogeneity in the literature as reflected in deviations from the overall empirical effect (estimated by the coefficient on *SEINVpcc*). In addition, the Ramsey RESET test confirms that the assumption of linear functional form (no omitted variables) cannot be rejected at conventional levels of significance, while the maximum and mean variance inflation factors suggest no undue effects from multicollinearity.

In addition to the highly significant coefficient on the inverse SE (*SEINVpcc*), eight of the twelve moderator variables are estimated with coefficients statistically significant at the ten percent level (or, in two cases, borderline). Positive and significant coefficients suggest that a certain study characteristic represented by that moderator variable typically increases the reported partial correlation between institutions and economic performance; while a negative and significant coefficient suggests that a particular characteristic typically reduces the reported partial correlation coefficient.

The multivariate-MRA results are consistent with those from the bivariate-MRA, again suggesting a lack of publication bias. These results also suggest a positive authentic empirical association between institutions and economic performance. However, in the multivariate MRA the "authentic effect" is captured by the combination of all the moderator variables (Doucouliagos and Stanley, 2009). Changes in the inverse of the standard error (*SEINVpcc*) also affect all moderator variables interacted with *SEINVpcc*. Consequently, the estimated effect size is not, as in the bivariate model, simply the coefficient on *SEINVpcc*; instead, the estimated effect size depends both on all the moderator variables interacted with *SEINVpcc* and on the corresponding choices of omitted (reference) categories. For example, in Table 3 the estimated coefficient on *SEINVpcc* is 0.90, while the coefficient on the moderator variable for those regressions with growth as the dependent variable (*GRGDPSE*) is −0.30. Conversely, when we estimate with the moderator variable for those regressions with output-level as the dependent variable (*LEVELGDPSE*), hence with *GRGDPSE* as the omitted category, the coefficient on *SEINVpcc* is 0.60 and the coefficient on *LEVELGDPSE* is 0.30. In the first case, the partial correlation between economic performance and the institutional variable of interest in the sub-sample of studies with economic growth as the dependent variable is 0.60 (=0.90 − 0.30); whereas in the sub-sample investigating output level it is 0.90 (=0.60 + 0.30), the same as reported in Table 3. In both cases, the remaining coefficient estimates are unchanged.

Table 3
Multivariate MRA FAT-PET model, testing for publication bias and authentic empirical effect (OLS with cluster-robust SEs; observations weighted to give each study equal weight).

Dependent variable: TSTAT, which is the t-statistic on the institutional variable of interest				
Moderator variables used in MRA (short description of moderator variable)	Coeff.	Cluster-robust SE	t-stat.	P>|t|
CONS (intercept term)	−0.51	0.70	−0.72	0.475
SEINVpcc (the inverse standard error of the partial correlation coefficients)	0.90	0.15	6.11	0.000
SPPRSE (modified extended production function specification)	−0.38	0.12	−3.24	0.002
SPINSSE ("institutional specification")	0.13	0.08	1.64	0.110
GRGDPSE (dependent variable is growth of output)	−0.30	0.14	−2.08	0.044
INSVARSE (instrumental variable methodology used in research)	−0.24	0.07	−3.55	0.001
SLS3SE (three-stage-least squares methodology used in research)	−0.52	0.11	−4.74	0.000
LAGSE (institutional variable is lagged once)	−0.13	0.07	−1.88	0.067
INITIALSE (initial level of output included in specification)	0.19	0.09	2.17	0.036
LNTIMEINSSE (natural log of the time-horizon of the institutional variable)	−0.04	0.03	−1.38	0.175
LNTIMEDSE (natural log of the time-horizon of the dependent variable)	0.04	0.03	1.17	0.250
COMPSE (institutional variable is focused on certain components of an index)	0.03	0.09	0.33	0.741
SAMSE (sample covered by research; transition = 1, "non-transition" = 0)	−0.12	0.11	−1.09	0.284
NOINDSE (number of independent variables in the regression)	−0.01	0.01	−1.61	0.116

Model diagnostics	
Number of observations	No. obs. = 112
Number of clusters (studies)	No. of clusters = 40
R-squared	R-squared = 0.77
F-test	F(13, 39) = 15.41
Ho: independent variables are jointly equal to zero	Prob>F = 0.00
Ramsey RESET test	F(3, 95) = 1.96
Ho: No omitted variables	Prob>F = 0.12
Variance Inflation Factor (VIF)	Max: 5.44; Mean: 3.02

Notes: The coefficient on *SEINVpcc* measures the magnitude of the effects of institutions on economic performance, corrected for publication selection. Reported t-statistics are calculated using heteroskedasticity cluster-robust standard errors. *LNTIMEINSSE* and *LNTIMEDSE* are continuous variables measured in years. Both were transformed into logarithms to obtain a substantial improvement in the functional form of our estimated model: the Ramsey RESET test for the model including the original variables yielded a p-value of 0.00; while after logarithmic transformation the p-value is 0.12, which by conventional standards supports non-rejection of the null of linearity.

Table 4
Comparative multivariate MRA FAT-PET estimates for the output-growth and output-level studies (OLS with cluster-robust SEs; observations weighted to give each study equal weight).

Dependent variable: TSTAT, which is the t-statistic on the institutional variable of interest						
Moderator variables used in MRA (short description of moderator variable)	Output-growth studies			Output-level studies		
	Coeff.	t-stat.	p-value	Coeff.	t-stat.	p-value
CONS (intercept term)	−1.07	−0.94	0.360	−1.99	−1.34	0.194
SEINVpcc (inverse SE of the partial correlation coeff.)	0.27	1.55	0.137	0.67	4.06	0.001
SPPRSE (modified extended production function specification)	0.03	0.21	0.833	–	–	–
SPINSSE ("institutional specification")	0.39	3.31	0.004	0.31	3.65	0.002
INSVARSE (IV methodology used in research)	−0.13	−1.44	0.167	−0.23	−2.96	0.008
SLS3SE (three-stage-least squares used in research)	−0.37	−2.02	0.058	–	–	–
LAGSE (institutional variable is lagged once)	−0.11	−1.36	0.191	–	–	–
INITIALSE (initial level of output included in specification)	0.17	2.05	0.055	–	–	–
COMPSE (institutional variable is component of an index)	0.16	1.72	0.101	0.22	1.91	0.070
SAMSE (sample covered; transition = 1, "non-transition" = 0)	0.19	2.20	0.040	−0.26	−2.36	0.029

Model diagnostics		
Number of observations	No = 52	No = 60
Number of clusters (studies)	No. of clusters = 20	No. of clusters = 21
R-squared	R-squared = 0.50	R-squared = 0.58
F-test	F(9, 19) = 3.62	F(5, 20) = 34.40
Ho: independent variables are jointly = 0	Prob>F = 0.01	Prob>F = 0.00
Ramsey RESET test	F(3, 39) = 1.51	F(3, 51) = 0.47
Ho: No omitted variables	Prob>F = 0.23	Prob>F = 0.71
Variance Inflation Factor (VIF)	Max: 2.85; Mean: 1.98	Max: 1.90; Mean: 1.45

Notes: The coefficient on *SEINVpcc* measures the magnitude of the effects of institutions on output growth (Columns 2–4) and on output level (Columns 5–7) corrected for publication selection. Reported t-statistics are based on heteroskedasticity cluster-robust standard errors.

Multivariate MRA of the whole sample reveals sources of heterogeneity in the literature that are both economically substantial and statistically significant at the five percent level: (non)use of an extended production function specification (*SPPRSE*); growth/levels definition of the dependent variable (*GRGDPSE*); (not)addressing potential institutional endogeneity (*INSVARSE* and *SLS3SE*); and (non)inclusion of the initial level of output (*INITIALSE*). In the remainder of this section we first comment on each variable significant at or close to the 10 percent level; we then conclude by focussing on these four main effects to calculate the authentic institutional effect for the whole sample.

With "*other specifications*" (*SPOTHERSE*) as the reference category, those studies that use institutional specifications (institution(s) as the only independent variable; *SPINSSE*) are more likely to report higher correlations between institutions and economic performance. Conversely, the extended production function specification (*SPPRSE*), which combines institutional variables and some standard growth factors, tends to lower the estimated effect. Together, these estimated coefficients suggest that in comparison with the more fully specified institutional models, which use some form of the extended production function, the other two specifications tend to inflate the average "authentic" effect of institutions.

Specifications in which output growth is the dependent variable (*GRGDPSE*) (with *LEVELGDPSE* – output level – as the reference category) typically yield lower estimated institutional effects. The results reported in Table 3 also suggest that if a study includes initial level of output in the specification (*INITIALSE*) then it is more likely to yield a higher estimated effect of institutions on economic performance.

The number of explanatory variables (*NOINDSE*) in different specifications is also a (borderline) significant determinant; namely, the more explanatory variables in the specification the lower the effect of institutional quality on macroeconomic performance. This is consistent with the findings on different specifications; recall, institutional specification with only one explanatory variable tends to yield higher institutional effects. A significant and negative coefficient is estimated on *LAGSE*, which captures the effect of lagged institutional variables (one year lag) on economic performance. Accordingly, specifications including the lagged effect of institutions on economic performance tend to yield a lower institutions–economic performance correlation, on average. This evidence suggests that the long-run effect of institutions might be smaller than the short-run or contemporaneous effect. However, some caution is needed in advancing this interpretation: first, this moderator variable is estimated with only four observations coming from three different studies; and, secondly, our present sample gives us no basis for investigating the context of potentially rich dynamic effects (for example, non-linear effects over time).

Endogeneity has been accounted for in many studies by using either an instrumental variables methodology (IV or 2SLS, denoted *INSVARSE*) or, in a few studies, the three-stage-least squares methodology (*SLS3SE*). The base category is those studies that do not address potential institutional endogeneity. Both coefficients of interest are highly significant and negative. We expected this variable to be a significant and a negative influence: first, because institutional endogeneity is generally recognized in institutional research as a potential problem; and, secondly, because in the presence of endogeneity, estimation by ordinary least squares should inflate estimates relative to instrumental variables approaches. This negative effect is an especially important finding; it suggests that studies not addressing potential endogeneity may overestimate the effect of institutions.

We construct the authentic institutional effect in this multivariate MRA from coefficients estimated at the conventional five percent significance level, which are also the economically most substantial effects. This puts the focus on three types of influence: definition of economic performance in terms of, respectively, output levels or growth; model specification; and (non)approach to potential endogeneity. Accordingly, studies that investigate the effects of institutions on economic outcomes defined in terms of levels (GDP, GDP per capita, GNP per capita and GDP per worker; the reference category), that are specified as some form of extended production function (-0.38), that include initial conditions (0.19), and that address the potential endogeneity of institutions by applying either IV or 2SLS techniques (-0.24) yield a statistically significant effect size of 0.47 ($=0.90-0.38+0.19-0.24$; $t=3.75$, $p=0.001$). (There are no studies specified with lagged institutional effects or using 3SLS in the levels sub-sample.) This is not far removed from the "authentic" effect of institutions obtained in the simple bivariate MRA for the levels sub-sample; namely, a *PCC* of 0.52 (see Table 2). Conversely, in the growth literature sub-sample the estimated effect for studies with these characteristics is substantially lower (0.18; the result of 0.17 when calculated directly from Table 3 reflects cumulative rounding error) and not statistically significant at any conventional level ($t=0.96$, $p=0.344$). However, if we calculate the representative effect size according to the same statistical significance criterion then we must add to the calculation the large negative effect on reported effect size of estimating by three-stage least squares (-0.52). In this case, the calculated effect in the growth subsample is not only negative (-0.34) but also significant at the five percent level ($t=-2.11$, $p=0.042$). It may be argued that this result should be disregarded, because there are only two studies – accounting for 6 observations – that use 3SLS. Yet such a caveat does not change the basic finding of pronounced heterogeneity between the levels and the growth subsamples. In the case of the studies investigating institutional effects on output levels, the multivariate finding is consistent with the bivariate estimate reported in Table 2 above (a statistically significant effect of 0.52). In contrast, in the case of the studies investigating institutional effects on output growth the effect is at best consistent with the much smaller and non-significant bivariate estimate reported in Table 2 (0.10). The next section focuses on this particular source of heterogeneity in the literature.

6. Are there differences between output-growth and output-level studies?

We divide the sample into the output-growth and output-level studies in order to investigate further the effects of dependent variable heterogeneity.[8] Recall, in the bivariate MRA the results for publication bias (FAT) and authentic empirical effect (PET) are

[8] We cannot achieve the same effect by adding a full set of slope dummy variables to the full model, because this requires the inclusion of an intercept shift variable that is – by construction – collinear with the corresponding moderator variable interacting with the standard error.

Table 5
Comparative multivariate MRA FAT-PET estimates for the output-level studies that, respectively, use and do not use an extended production function specification (OLS with cluster-robust SEs; observations weighted to give each study equal weight).

Dependent variable: TSTAT, which is the t-statistic on the institutional variable of interest						
Moderator variables used in MRA	Output-level studies — *use* an extended production function specification			Output-level studies — *do not use* an extended production function specification		
	Coeff.	t-stat.	p-value	Coeff.	t-stat.	p-value
CONS (intercept term)	0.59	0.40	0.701	−3.62	−4.03	0.002
SEINVpcc (inverse SE of the partial correlation coeff.)	0.44	1.76	0.113	1.02	12.35	0.000
SPINSSE ("institutional specification")	–	–	–	0.02	0.34	0.739
INSVARSE (IV methodology used in research)	−0.33	−1.82	0.102	−0.04	−0.83	0.424
COMPSE (Institutional variable is component of an index)	0.09	0.78	0.455	−0.11	−1.41	0.186
SAMSE (sample covered; transition = 1, "non-transition" = 0)	−0.23	−1.26	0.240	–	–	–
Model diagnostics						
Number of observations	No = 26			No = 34		
Number of clusters (studies)	No. of clusters = 10			No. of clusters = 12		
R-squared	R-squared = 0.24			R-squared = 0.91		
F-test	F(4, 9) = 1.04			F(4, 11) = 97.32		
Ho: independent variables are jointly = 0	Prob > F = 0.44			Prob > F = 0.00		
Ramsey RESET test	F(3, 18) = 1.15			F(3, 26) = 1.08		
Ho: No omitted variables	Prob > F = 0.36			Prob > F = 0.37		
Variance Inflation Factor (VIF)	Max: 1.88; Mean: 1.53			Max: 1.62; Mean: 1.31		

Notes: The coefficient on *SEINVpcc* measures the magnitude of the effects of institutions on economic performance, corrected for publication selection. Reported t-statistics are calculated using heteroskedasticity cluster-robust standard errors.

different between those two subsamples, with the growth literature affected by positive publication bias and revealing no authentic empirical effect. Accordingly, we investigate whether or not these sub-sample findings are robust to a full MRA specification.

Some study characteristics are not present in both samples (e.g. *SLS3SE* and *INITIALSE* in the output-level literature) and those moderator variables cannot be estimated in both samples. In addition, some variables are included in the full model (Table 3) yet are omitted from the sub-sample models (Table 4) not only on grounds of statistical significance but also because they prove incompatible with a linear functional form in one or both of the sub-sample model(s) (*SPPRSE* in the output-level sample and *NOINDSE*, *LNTIMEDSE* and *LNTIMEINSSE* in both sub-samples).

Multivariate MRA is consistent across both sub-samples in finding no clear evidence for publication selection; the null hypothesis of "lack of publication bias" cannot be rejected at conventional significance levels. In the case of the growth sub-sample, both the funnel plot and bivariate (FAT-PET) MRA suggest substantial publication bias in the hypothesised positive direction. Conversely, in the multivariate sub-sample regression reported in Table 4 lack of statistical significance suggests that the publication selection term is washed out when the moderator variables are included. However, this does not necessarily mean that there is no publication bias. Rather, when taken together it is likely that our results indicate that the process of publication selection is complex, involving multiple specification choices as well as sample selection (e.g., transition/non-transition).[9]

The multivariate MRA reported in Table 4 suggests that the output levels studies and the output growth studies display rather different characteristics with respect to the strength of evidence for an authentic empirical effect. Firstly, the unmoderated estimates of the authentic empirical effects differ with respect to statistical significance: respectively, 0.67 ($p = 0.001$) in the output levels subsample; and 0.27 ($p = 0.137$) in the output growth subsample. Secondly, linear combinations of all the estimated coefficients in the respective sub-sample regressions yield similar results: for the levels sub-sample, the combined or moderated effect is 0.71 ($t = 4.40$, $p = 0.000$); and for the growth sub-sample, the combined effect is 0.60 ($t = 1.80$, $p = 0.087$ and the confidence interval ranges from −0.10 to 1.30).

The sub-sample results in Table 4 may exaggerate the contrast between the growth and levels sub-samples with respect to the robustness of their respective authentic empirical effects. When the levels sample is estimated with *SPPRSE* – the (non)use of an extended production function specification – there is clear rejection of the diagnostic null that the linear functional form is appropriate; the Ramsey RESET test yields $F(3, 50) = 7.27$, $p = 0.0004$.[10] This is consistent with evidence reported above: namely, the evidence of the funnel plots in Fig. 2 that this specification choice may drive the apparent heterogeneity of the levels sub-sample; and the evidence in Table 3 that the estimated effect of *SPPRSE* is both economically substantial and statistically significant. Accordingly, the sub-sample results reported in Table 4 may suffer from serious omitted variables bias. To investigate this possibility, we estimated a multivariate MRA model for the same sub-samples investigated by the funnel plots in Fig. 2:

[9] We owe the ideas in the preceding three sentences to an anonymous referee.
[10] *NOINDSE*, *LNTIMEDSE* and *LNTIMEINSSE* are also omitted from the levels sub-sample model. In contrast to *SPPRSE*, in the full sample the estimated effects of these variables are neither economically substantial nor statistically significant.

Table 6
Summary of estimates of representative authentic empirical effect.

	Bivariate MRA (Table 2)			Multivariate MRA (Table 3)		Multivariate MRA (Table 4)		Multivariate MRA Levels sub-sample by specification (Table 5)	
	Sample of studies								
	Full	Growth	Levels	Growth	Levels	Growth	Levels	Not production function	Production function
	1.	2.	3.	4.	5.	6.	7.	8.	9.
Estimated representative size (*PCC*)[a]	0.41	0.10	0.52	−0.34	0.47	0.60	0.71	1.02	0.11
Significant at the 5% level?	Yes	No	Yes	Yes	Yes	No	Yes	Yes	No

[a] Bivariate MRA — directly estimated; Multivariate MRA — linear combination of the unmoderated estimated effect and the estimated moderating influences.

namely, studies of levels effects that, respectively, do not use an extended production function specification (34 observations) and that do (26 observations). The results are reported in Table 5.

The diagnostics tests and checks for both sub-sample models are satisfactory. However, because these models are estimated with relatively small samples, the corresponding lack of precision in estimation suggests that it may be reasonable to discuss results significant at – or close to – the 10% level of significance.

In essentials, these results confirm the impressions created by the funnel plots in Fig. 2. MRA for the levels studies *not* specified as an extended production function find a "large" and significant – but unmoderated – authentic empirical effect (1.02) in the presence of "severe" negative bias, for which there is no clear explanation. For those levels studies that are specified as an extended production function, we find evidence of a smaller but still "large" effect (0.44), although the 95% confidence interval ranges across zero (from −0.127 to 1.012). Although not statistically significant, the positive intercept term is consistent with the positive bias indicated by Fig. 2(b) and, hence, with the hypothesised direction of publication bias. In the former case, there is no other variable estimated with a sufficient degree of statistical significance to justify forming a linear combination with the estimated effect size. In the latter case, however, the linear combination of the precision effect with *INSVARSE* – moderating the impact of IV estimation to address the potential endogeneity of institutions – yields a non-significant but positive effect, which is somewhat smaller than the norm for empirical literatures in economics (0.11, $t = 1.14$, $p = 0.285$).[11]

From the discussion in Sections 5 and 6 it could plausibly be argued that the most valid estimates arise from those studies focused on output levels, using an extended production function specification and addressing endogeneity. Yet, for this specific group the finding of an authentic effect appears to be weak. One explanation might be omitted variables bias. For example, in those studies that investigate transition countries, we find a complete absence of the theoretically important human capital variable.

7. Conclusion

This paper seeks to identify whether or not there is a representative empirical effect of institutional quality on economic performance and, if so, to measure the size of this effect. The results obtained from both preliminary graphical analysis and bivariate MRA are mixed. Investigation of both the full sample and the sub-sample of studies concerned with levels of economic performance reveals no evidence of publication bias while suggesting the possible presence of an authentic, positive and large empirical effect. In contrast, investigation of the sub-sample of studies concerned with economic growth reveals evidence of substantial, positive publication bias but finds no evidence suggesting an authentic empirical effect.

The multivariate results for both the full sample and the output-levels sub-sample again reveal no evidence of publication bias but do suggest a statistically significant and positive authentic empirical effect of institutional quality on economic performance. Yet the results reported for the growth sub-sample show that the publication selection revealed by the bivariate plot and regression is washed out when the moderator variables are included.

From the estimates reported for the full sample, we take into account the main sources of heterogeneity in the literature to calculate that the subsample of studies investigating the effects of institutional quality on economic performance levels yields a statistically significant representative partial correlation coefficient of 0.47. By the standards of empirical literatures in economics, the estimated representative effect of institutional quality on economic performance levels is large. In contrast, the effect of institutional quality on economic growth is not robustly estimated.

For convenience, Table 6 summarizes all the estimates of authentic empirical effect "beyond" – i.e., controlling and corrected for – publication bias and the main sources of heterogeneity in this literature.

Table 6 highlights our finding that this empirical literature is characterised by pronounced heterogeneity. Hence, it would be misleading to report a single estimated authentic empirical effect size as representative of the whole literature. Yet the weight of the evidence does support the qualitative conclusion that *the effect of institutional quality on economic performance is positive*.

The primary source of heterogeneity that we identify in this literature is between studies that investigate institutional effects on output growth and those that focus on output levels. On the whole, *the levels studies reveal a more robust representative effect*

[11] In the sample of empirical literatures from economics analyzed by Doucouliagos (2010, p. 5), "half of the observed partial correlations are smaller than ±0.16".

than do growth studies. All the effects for this sub-sample are positive; and all but one is large and statistically significant at the conventional five percent level.

Table 6 (column 9) also reveals that the one statistically insignificant and non-large result arises from taking into account *model specification as a further major source of heterogeneity*, which particularly affects the levels sub-sample. By isolating those studies investigating institutional effects on levels of economic performance by estimating an extended production function, we find that even the levels sub-sample does not uniformly yield robustly large and significant effects. Yet, in comparison, the growth studies yield less robust effects: not all of the estimated effects are positive; and of the two that are, neither is statistically significant.

The third major source of heterogeneity affects both the growth and levels samples; namely, the divide between those *studies that address the potential endogeneity of institutions* and those that do not. The moderating effects of methodologies for dealing with institutional endogeneity are particularly significant and substantial in almost all models. Specifically, *studies that account for endogeneity tend to report a substantially smaller effect of institutional quality on economic performance than do other studies*. Since theory recognizes institutions as potentially endogenous, and MRA suggests that studies ignoring this guidance may overestimate the association between institutions and economic performance, we conclude that the results from studies not addressing this issue should be treated with great caution.

Although the positive effect of institutions on economic performance is on its way to becoming conventional wisdom, this *MRA suggests that the evidence base is not as robust as it should be*. For researchers investigating institutional effects on economic performance, the main findings reported above suggest that future studies should provide a clear rationale for: whether or not and, if so, how to address the potential endogeneity of institutions; the specification of the model to be estimated; and, above all, the choice of dependent variable.[12] One result raised to the surface by this MRA is that the studies with growth-theoretic foundations (i.e., the output-growth models) do not give robust evidence for an authentic empirical effect of institutions. Conversely, the somewhat *more robust findings of positive, statistically significant and large institutional effects on output levels suggest this relationship to be a fruitful focus for research* (Basu, 2008; Knowles and Weatherston, 2006). Because national differences in per capita output levels reflect entire histories of time-varying growth performance, analyzing the determinants of differing per capita levels helps to avoid non-robust and, hence, spurious explanations that arise from potentially unrepresentative samples of impermanent growth processes (reflecting, according to Easterly, 2009, the "possibility that if you get a result associating high growth with a particular country … in one period, it is likely to vanish in the following period"). This approach is consistent with the evidence advanced in this paper.

Acknowledgments

This paper has been hugely improved by advice from Jean Mangan, participants in the 2008 colloquium of the Meta-Analysis in Economic Analysis network and two anonymous referees.

References

Acemoglu, D., Johnson, S., 2005. Unbundling institutions. Journal of Political Economy 113, 949–995.
Ahlerup, P., Olsson, O., Yanagizawa, D., 2009. Social capital vs institutions in the growth process. European Journal of Political Economy 25, 1–14.
Aidis, R., Estrin, S., Mickiewicz, T., 2009. Entrepreneurial entry: which institutions matter? IZA Discussion Paper Series 4123.
Aixala, J., Fabro, G., 2008. Does the impact of institutional quality on economic growth depend on initial income level? Economic Affairs 28, 45–49.
Alcala, F., Ciccone, A., 2004. Trade and productivity. The Quarterly Journal of Economics 119, 613–646.
Baliamoune-Lutz, M., Ndikumana, L., 2007. The growth effects of openness to trade and the role of institutions: new evidence from African countries. University of Massachusetts Working Paper 2007-05.
Basu, S.R., 2008. A new way to link development to institutions, policies and geography. UNCTAD Policy Issues in International Trade and Commodities Study Series 37.
Bates, R.H., 1995. Social dilemmas and rational individuals: an assessment of the new institutionalism. In: Harriss, J., Hunter, J., Lewis, C.M. (Eds.), The New Institutional Economics and Third World Development. Routledge, London, pp. 27–48.
Baum, F.C., 2006. An Introduction to Modern Econometrics Using Stata. Stata Press, Texas.
Beck, T., Laeven, L., 2006. Institution building and growth in transition economies. Journal of Economic Growth 11, 157–186.
Brousseau, E., Glachant, J.M. (Eds.), 2008. New Institutional Economics — A Guidebook. Cambridge University Press, Cambridge.
Chang, H.J., 2006. Understanding the relationship between institutions and economic development: some key theoretical issues. Revista de Economía Institucional 8, 125–136.
Chousa, J.P., Khan, H.A., Melikyan, D., Tamazian, A., 2005. Assessing institutional efficiency, growth and integration. Emerging Markets Review 6, 69–84.
De la Croix, D., Delavallade, C., 2009. Growth, public investment and corruption with failing institutions. Economics of Governance 10, 187–219.
Doucouliagos, C., 2005. Publication bias in the economic freedom and economic growth literature. Journal of Economic Survey 19, 367–387.
Doucouliagos, C., 2010. How large is large? Preliminary guidelines for interpreting partial correlations in economics. Deakin University, School of Accounting, Economics and Finance (pre-publication manuscript shared by the author).
Doucouliagos, C., Laroche, P., 2009. Unions and profits: a meta-regression analysis. Industrial Relations 48, 146–184.
Doucouliagos, C., Stanley, T.D., 2008. Theory competition and selectivity: are all economic facts greatly exaggerated? Deakin University Economic Series 2008-06.
Doucouliagos, C., Stanley, T.D., 2009. Publication selection bias in minimum-wage research? A meta-regression analysis. British Journal of Industrial Relations 47, 406–428.
Dreher, A., Meon, P.G., Shneider, F., 2007. The devil is in the shadow: do institutions affect income and productivity or only official income and official productivity? CESifo Working Paper Series 2150.
Easterly, W., 2009. The anarchy of success. The New York Review of Books LVI (October 8-21), pp. 28–30.
Easterly, W., Levine, R., 2003. Tropics, germs, and crops: how endowments influence economic development. Journal of Monetary Economics 50, 41–47.

[12] Additional sources of heterogeneity that may help to explain the variation in the estimates reported in this literature include different samples, different observed time-horizons, different definitions of the institutional variable of interest and (non)inclusion of the initial level of output. However, none of these effects are estimated as robustly as the sources of heterogeneity that we identify as being of primary interest.

Economides, G., Egger, P.H., 2009. The role of institutions in economic outcomes: editorial introduction. European Journal of Political Economy 25, 277–279.
Eggertsson, T., 1990. Economic Behaviour and Institutions. Cambridge University Press, Cambridge.
Eggertsson, T., 2005. Imperfect Institutions, Possibilities and Limits of Reform. University of Michigan Press, Ann Arbor.
Eicher, T.S., Garcia-Penalosa, C., Teksoz, U., 2006. How do institutions lead some countries to produce so much more output per worker than other. In: Eicher, S.T., Garcia-Penalosa, C. (Eds.), Institutions, Development, and Economic growth. The MIT Press, Cambridge, Massachusetts, pp. 65–80.
Eicher, T., Schreiber, T., 2010. Structural policies and growth: time series evidence from natural experiments. Journal of Development Economics 91, 169–179.
Fisher, R., 1954. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh.
Furubotn, E.G., Richter, R., 2005. Institutions and Economic Theory — The Contribution of the New Institutional Economics. The University of Michigan Press, Ann Arbor.
Glaeser, E.L., La Porta, R., Lopez-de-Silanes, F., Shleifer, A., 2004. Do institutions cause growth? Journal of Economic Growth 9, 271–303.
Greene, W.H., 2008. Econometric Analysis, Sixth Edition. Pearson International, Upper Saddle River, New Jersey.
Gwartney, J.D., Holcombe, R.G., Lawson, R.A., 2004. Economic freedom, institutional quality, and cross-country differences in income and growth. Cato Journal 24, 205–233.
Gwartney, J.D., Holcombe, R.G., Lawson, R.A., 2006. Institutions and the impact of investment on growth. KYKLOS 59, 255–273.
Hall, R.E., Jones, C.I., 1999. Why do some countries produce so much more output per worker than others? Quarterly Journal of Economics 114, 83–116.
Jacob, J., Osang, T., 2007. Institutions, geography and trade: a panel study. Southern Methodist University Departmental Working Paper 0706.
Jones, C.I., Romer, P.M., 2009. The new kaldor facts: ideas, institutions, population, and human capital. American Economic Journal: Macroeconomics 2, 224–245.
Joskow, P.L., 2008. Introduction to new institutional economics: a report card. In: Brousseau, E., Glachant, Jean-Michel (Eds.), New Institutional Economics — A Guidebook. Cambridge University Press, Cambridge, pp. 1–19.
Khan, M., 1995. State failure in weak states: a critique of new institutionalists explanations. In: Harriss, J., Hunter, J., Lewis, C.M. (Eds.), The New Institutional Economics and Third World Development. Routledge, London, pp. 71–86.
Kirman, A.P., 2007. Individual and aggregate behaviour: of ants and men. In: Hodgson, G.M. (Ed.), The Evolution of Economic Institutions — A Critical Reader. Edward Elgar Publishing Ltd., Cheltenham, Northampton, pp. 34–52.
Klomp, J., de Haan, J., 2009. Political institutions and economic volatility. European Journal of Political Economy 25, 311–326.
Knack, S., Keefer, P., 1995. Institutions and economic performance: cross-country tests using alternative institutional measures. Economics and Politics 7, 207–227.
Knowles, S., Weatherston, C., 2006. Informal institutions and cross-country income differences. University of Otago Economics Discussion Papers 0604.
Mauro, P., 1995. Corruption and growth. The Quarterly Journal of Economics 110, 681–712.
Moers, L., 1999. How important are institutions for growth in transition countries? Tinbergen Institute Discussion Papers 99-004/2.
North, D.C., 1990. Institutions, Institutional Change and Economic Performance. Cambridge University Press, Cambridge.
North, D.C., 2005. Understanding the Process of Economic Change, Princeton University Press, Princeton.
Nye, J., 2008. Institutions and the institutional environment. In: Brousseau, E., Glachant, J.M. (Eds.), New Institutional Economics — A Guidebook. Cambridge University Press, Cambridge, pp. 67–80.
Opper, S., 2008. New institutional economics and its application on transition and developing economies. In: Brousseau, E., Glachant, J.M. (Eds.), New Institutional Economics — A Guidebook. Cambridge University Press, Cambridge, pp. 389–405.
Pelikan, P., 2003. Bringing institutions into evolutionary economics: another view with links to changes in physical and social technologies. Journal of Evolutionary Economics 13, 237–258.
Potts, J., 2007. Evolutionary institutional economics. Journal of Economic Issues XLI, 341–350.
Raiser, M., Di Tommaso, M.L., Weeks, M., 2001. The measurement and determinants of institutional change: evidence from transition economies. EBRD Working Paper 60.
Redek, T., Susjan, A., 2005. The impact of institutions on economic growth: the case of transition economies. Journal of Economic Issues XXXIX, 995–1027.
Richter, B.K., Timmons, J.F., 2008. Why not adopt better institutions? SSRN Working Paper Series, p. 1213138.
Rodrik, D., 2004a. Getting institutions right. Harvard University. downloaded in November 2007 http://ksghome.harvard.edu/~drodrik/ifoinstitutions%20article%20_April%202004.pdf.
Rodrik, D., Subramanian, A., Trebbi, F., 2004. Institutions rule: the primacy of institutions over integration and geography in economic development. Journal of Economic Growth 9, 131–165.
Rose, A.K., Stanley, T.D., 2005. A meta-analysis of the effect of common currencies on international trade. Journal of Economic Surveys 19, 347–365.
Sachs, J.D., 1996. The transition at mid decade. American Economic Review 86, 128–133.
Sachs, J.D., 2003. Institutions don't rule: direct effects of geography on per capita income. NBER Working Paper 9490.
Shirley, M.M., 2008. Institutions and Development. Edward Elgar Publishing Limited, Cheltenham and Northampton.
Stanley, T.D., 2001. Wheat from chaff: meta-analysis as quantitative literature review. Journal of Economic Perspectives 15, 131–150.
Stanley, T.D., 2005. Beyond publication bias. Journal of Economic Surveys 19, 309–345.
Stanley, T.D., 2008. Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. Oxford Bulleting of Economics and Statistics 70, 103–127.
Stanley, T.D., Doucouliagos, C., 2006. Identifying and correcting publication selection bias in the efficiency-wage literature: Heckman meta-regression. Deakin University Economics Working Paper Series 29.
Stanley, T.D., Doucouliagos, C., Jarell, S.B., 2008. Meta-regression analysis as the socio-economics of economic research. The Journal of Socio-Economics 37, 276–292.
Stanley, T.D., Jarrell, S.B., 1989. Meta-regression analysis: a quantitative method of literature surveys. Journal of Economic Surveys 3, 161–170.
Stein, H., 1995. Institutional theories and structural adjustment in Africa. In: Harriss, J., Hunter, J., Lewis, C.M. (Eds.), The New Institutional Economics and Third World Development. Routledge, London, pp. 109–132.
Toye, J., 1995. The new institutional economics and its implications for development theory. In: Harriss, J., Hunter, J., Lewis, C.M. (Eds.), The New Institutional Economics and Third World Development. Routledge, London, pp. 49–68.
Ulubasoglu, M.A., Doucouliagos, C., 2004. Institutions and economic growth: a systems approach. Econometric Society 2004 Australasian Meetings Paper 63.
Vijayarghavan, M., Ward, W.A., 2009. Clemson University Working Paper 11.