ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Babetskii, Ian; Campos, Nauro F.

Working Paper

Does reform work? An econometric examination of the reform-growth puzzle

IZA Discussion Papers, No. 2638

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Babetskii, Ian; Campos, Nauro F. (2007) : Does reform work? An econometric examination of the reform-growth puzzle, IZA Discussion Papers, No. 2638, Institute for the Study of Labor (IZA), Bonn,
https://nbn-resolving.de/urn:nbn:de:101:1-20080416247

This Version is available at:
https://hdl.handle.net/10419/33865

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

DISCUSSION PAPER SERIES

IZA DP No. 2638

Does Reform Work? An Econometric Examination of the Reform-Growth Puzzle

Ian Babetskii
Nauro F. Campos

February 2007

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Does Reform Work? An Econometric Examination of the Reform-Growth Puzzle

Ian Babetskii
Czech National Bank, CERGE-EI,
and CES-ROSES, University of Paris 1

Nauro F. Campos
Brunel University,
CEPR, WDI and IZA

Discussion Paper No. 2638
February 2007

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 2638
February 2007

ABSTRACT

Does Reform Work?
An Econometric Examination of the Reform-Growth Puzzle*

Why are socially beneficial reforms not implemented? One simple answer to this question (which has received little attention in the literature) is that this may be caused by generalized uncertainty about the effectiveness of reforms. If agents are unsure about whether a proposed reform will work, it will be less likely to be adopted. Despite the numerous benefits economists assign to structural reforms, the empirical literature has thus far failed to establish a positive and significant effect of reforms on economic performance. We collect data from 43 econometric studies (for more than 300 coefficients on the effects of reform on growth) and show that approximately one third of these coefficients is positive and significant, another third is negative and significant, and the final third is not statistically significant different from zero. In trying to understand this remarkable variation, we find that the measurement of reform and controlling for institutions and initial conditions are main factors in decreasing the probability of reporting a significant and positive effect of reform on growth.

JEL Classification: O11, P21, C49

Keywords: structural reforms, economic growth

Corresponding author:

Nauro F. Campos
Department of Economics and Finance
Brunel University
Uxbridge, Middlesex UB8 3PH
United Kingdom
E-mail: Nauro.Campos@brunel.ac.uk

* We thank Balazs Egert, Jan Fidrmuc, Jarko Fidrmuc, Laszlo Halpern, Jan Hanousek, Roman Horvath, Evzen Kocenda, Iikka Korhonen, Tomasz Mickiewicz, Bilin Neyapti, Karsten Staehr, Johan Swinnen, Jan Visek and seminar participants at the Czech National Bank, Czech Economic Society Conference (Prague), Western Economic Association Conference (San Diego) and European Econometric Society Meetings (Vienna) for valuable comments on previous drafts. Evgeny Plaksen, Dana Popa and Ekaterina Shironosova provided superb research assistance. Campos is grateful for financial support from the Economic and Social Research Council (ESRC), Research Grant RES-000-22-0550. All remaining errors are entirely our own.

1. Introduction

One of the most heated debates in recent times has been that on the macroeconomic implications of structural reforms. Since the late 1980s, a large number of reform programmes have been implemented with varying degrees of success across the world. The reasons for this variation are still at large and have raised a number of important questions. Did reform work? Did the expected growth and welfare pay-offs materialise? What do we know about the evaluation of those reform efforts? How robust are the econometric estimates of their effects? What are the main factors that help explain their variation? Is it driven by data quality issues, diversity of theoretical frameworks, or differences in econometric methodology? Does how reform is measured matter? This paper focuses on such questions.

We start by taking stock of the voluminous body of econometric evidence on the impact of structural reforms. We put together a unique data set based on more than 300 estimates of the effect of reforms on growth (from the 43 studies listed in Appendix 1). Our data set has quantitative information on both the econometric estimates of the impact and on a large set of characteristics that may potentially help explain the variation across these estimates. Remarkably, we observe that a third of the coefficients of reform on economic performance are positive and significant, another third is negative and significant, and the final third is not statistically significantly different from zero. Figures 1-3 plots these results. Notice that the variance changes little when differences in lag structures are taken into account (this is important because although the contemporaneous effect is generally found to be negative, the lagged effect tends to be positive.) The extent of such variation of the impact of reforms suggests to us that this is indeed fertile ground for meta-regression analysis.

Meta-regression analysis (hereafter, MRA) is a statistical method that gives a summary as well as a quantitative assessment of a given body of evidence (Stanley, 2001; Stanley and Jurrell, 1989). In a MRA study, the dependent variable is usually a summary statistic (for

instance, elasticities, t-values, etc.) while the independent variables often cover various features of the econometric strategy and data used in each study. MRA has been widely used despite it being a relatively new technique. In environmental economics, Florax (2002) reviews 40 meta-regression studies (mostly on pollution valuation) published since 1980. It has also been used extensively in labour economics: Card and Krueger (1995) use MRA to assess the evidence on minimum wages, Stanley and Jurrell (1998) use it to evaluate that on gender wage differentials in the United States, while Ashenfelter et al. (1999) use MRA to investigate the robustness of the evidence on returns to education.[1] In public finance, MRA has been used to assess the impact of tax policies (Phillips and Goss, 1995) and to evaluate econometric findings on the Ricardian equivalence (Stanley, 1998). In international macroeconomics, Rose (2004) uses MRA to evaluate the evidence on the effects of currency unions on international trade, Fidrmuc and Korhonen (2006) to assess that on business cycle synchronisation, Stiroh (forthcoming) to examine the evidence on the information technology-growth relationship, and Égert and Halpern (2006) to appraise that on equilibrium exchange rates. In this paper, we apply MRA to the evidence on reform and growth.[2]

The data set we put together for this paper is based on information collected from 43 econometric studies, generating a grand total of 321 coefficients, and includes a quantification of more than 40 features of those studies encompassing estimation method, measurement and specification (Appendix 2 provides a complete list). We investigate the sign, magnitude and significance of the (contemporaneous as well as cumulative) effect of reform on economic performance. Our findings suggest that the use of panel methods, controls for macroeconomic stabilisation and time coverage are important in explaining the variation in the estimated

[1] Jurrell and Stanley (1990) use MRA to evaluate the evidence on the union/non-union wage gap, Weichselbaumer and Winter-Ebmer (2001) to assess that on gender wage differentials across countries, and Doucouliagos (1995) uses it to take stock of the econometric evidence on worker participation.

[2] In comparative economics, Djankov and Murrell (2002) use MRA to assess the empirical evidence on enterprise restructuring. Havrylyshyn (2001) provides a review of the relationship between reform and growth but we are unaware of any MRA study of this issue.

effects of reform on economic performance. Of particular interest is that the measurement of reform and controlling for institutions and initial conditions significantly decreases the probability of finding a positive and significant effect of reform on growth.

We should also note at the outset that here we focus on a particular body of econometric evidence on the growth-reform nexus, namely the experience of the transition economies. This is so for at least four reasons: (1) this is a group of countries for which there is a sufficiently large number of published econometric studies;[3] (2) these economies provide an almost natural experiment setting for the question at hand as they started out with rather similar initial conditions but experienced very dissimilar reform and growth trajectories (with some of them implementing reform packages in an unprecedented scale); (3) this body of evidence therefore tends to use similar measures of reform[4] as well as growth figures which attenuates one potentially crucial source of bias; and (4) the studies tend to use similar (to a certain extent) econometric specifications, estimation strategies and sets of explanatory variables. Contrasting these four reasons with the variation we see in the results (Figures 1-3) supports our view that this is indeed fertile ground for meta-regression analysis.

The rest of the paper is organised as follows. The next section describes the methodological framework we use. Section 3 presents the data set we put together for this paper. Section 4 discusses our econometric approach and main findings, while Section 5 concludes with suggestions for future research.

[3] Acemoglu et al. (2006) argue that the empirical evidence on reform-growth is very scarce, with even that for the OECD countries being limited to a handful of papers.

[4] This simply means that the reform data used in the empirical literature is almost uniformly the same, and does not imply that we see these measures as error-free. Indeed, Campos and Coricelli conclude their survey by noting that: "more emphasis should be placed upon a better understanding of the role of economic reforms and reform strategies in dictating the path of the transition process (…) There are a number of theoretical models that stress the role of reform strategies. Yet the data for discriminating among these models is lacking. The few indicators available are unnecessarily subjective (…)" (2002, p. 831).

2. Methodological Framework

Meta-analysis refers to a set of statistical methods for reviewing and evaluating bodies of empirical evidence in a comprehensive and rigorous way. If a large number of studies have been carried out on a particular topic, then combining their results in a systematic manner can provide additional insights and greater explanatory power than can the more informal, narrative discussions of the individual results. MRA usually goes beyond what is often called vote-counting or head-counting (Light and Smith, 1997), in which the inference that a specific category occurs in a majority of cases is usually taken as evidence of the significance and magnitude of the "true" effect. Head-counting is neither systematic nor statistically powerful in drawing conclusions about a body of evidence. When the number of existing studies is very large, head-counting is even more likely to support misleading conclusions because the Type-II errors of the individual studies do not cancel out, but add up instead (Florax et al., 2002).

One first procedure to summarise a given body of evidence was developed by Fischer (1932). It assumes that the underlying p-values are uniformly distributed under the null hypothesis of no effect, and then proposes that minus twice the sum of the logs of the p-values follows a chi-square distribution. This approach assumes independence across studies and that each one of them is unbiased; this is clearly an important assumption which is usually addressed by estimating MRA equations with study fixed-effects so as to capture unobserved heterogeneity among findings.

It is well-established that the Fischer test is too generous in ascribing significance. Stanley and Jurrell (1998) discuss three main reasons for this. First, it does not distinguish between positive and negative statistically significant effects, as both are only counted as significant. Second, the null hypothesis of the Fischer test is that none of the observations reflects a genuine effect. A finding of significance therefore does not necessarily mean that the average effect is statistically significant. Third, the assumption of unbiased estimates is

often violated in the case of non-experimental evidence.

The technique that focuses on the magnitude of the effect was developed by Glass (1976), who suggests focusing on the "effect size" to compare and integrate findings. He defines the effect size as the average outcome of the treatment group minus the average outcome of the control group, divided by the standard deviation of the control group.[5]

A related development refers to estimating the "average" effect in order to assess its practical and statistical significance and to explain its variation among studies. Stanley (1998) and Rose (2004) compute the average (normalised) test statistic while Djankov and Murrell (2002) use aggregate t-statistics for this purpose. One important feature of the literature on the effects of reform on growth is that different studies use different measures of stabilisation, factor inputs, initial conditions and institutional development, and thus combining estimates has to be done carefully. One way to combine results from different studies is combining t-statistics. Djankov and Murrell (2002) suggest aggregating t-statistics as follows: $\bar{t} = \sum_{k=1}^{M} t_k / \sqrt{M}$, where M is the total number of studies, and $t_1,...,t_M$ are t-statistics on $\hat{\beta}$, with $\bar{t}$ taken as normally distributed given sufficiently large sample sizes in all analyses. Djankov and Murrell (2002) make the point that statistical tests based on this $\bar{t}$ are more powerful than tests on individual observations, as the former provides additional information on the statistical significance of the effect.

The typical study of reform on growth estimates an equation of the form:

$$Y = \alpha + \beta Lib + \gamma Stab + \varphi F + \delta IC + \phi Inst + \varepsilon \qquad (1)$$

where Y is the per capita GDP growth rate, Lib is a measure of reform, IC is a measure of

[5] One common criticism of Glass's effect size method is its estimate of the overall standard deviation. Hedges and Olkin (1985) argue that it is inappropriate and propose a pooled estimate in its stead, corrected for the degrees of freedom. They also argue that conventional methods, ANOVA in particular, are often unsuitable for analysing the effect size as heteroskedasticity across studies can mask design differences that the meta-analysis itself is intended to investigate.

initial conditions (say macroeconomic and structural distortions at the beginning of the period), *Stab* are different measures of stabilisation (inflation, exchange rate regime, government deficit, etc.), *Inst* is institutional development, *F* are factors of production (capital and labour) and ε is the error term. Coefficient β represents the partial effect of reform on growth, *ceteris paribus,* and its size, sign and level of statistical significance are of central interest here.

The meta-regression equations we estimate take the form:

$$Y_i = \beta_0 + \sum \beta_k Z_{ki} + \varepsilon_i \qquad (2)$$

where Y_i is the value of a summary statistic from analysis *i* and Z_{ki} are *k* meta-independent variables. It is common practice to use estimated coefficients or the results of statistical tests (e.g., t-values) as the summary measure. In light of the very large variation in the results from the body of evidence we are interested in (Figures 1-3), we report estimates from ordered logit models in which the dependent variable captures whether a result is positive and significant, negative and significant, or not significant. More details are provided below.

One major shortcoming of MRA concerns the so-called "file drawer" problem, namely the tendency of academic journals to favour studies that report statistically significant results. Card and Krueger (1995) and Ashenfelter et al. (1999) address publication bias in their studies of minimum wage and returns to schooling, respectively. One potential problem in this case is the implicit assumption that working papers are not published (and will not be) because they do not contain a sufficient number of statistically significant results. One important issue in our case is that the literature on reform and growth is more recent and less established (than, for instance, those on the minimum wage and returns to education). Therefore, we think it is better to leave the investigation of the severity of publication bias for future work. We do, however, try to deal with study heterogeneity by re-estimating all our results (below) with study fixed-effects.

3. Data Set

MRA must start with a careful search for the appropriate literature from which the observations (in this case, coefficients on the effects of reform) will be taken. We include papers in our data set if they investigate the effect of reform on growth in a transition context, if they report the coefficients on reform, if their estimates are from regression analysis, and if their t-values or standard errors are reported in full. We find 43 papers (listed in Appendix 1) which fulfil these criteria and use them as the basis for our data set.

We follow Weichselbaumer and Winter-Ebmer (2001) and include all available test statistics from each study in our meta-regression analysis. There is no clear consensus on whether to choose one estimate from each study or all of the reported estimates. Stanley (2001) proposes to choose only one estimate – the one which the author of the study indicates to be her preferred one. Alternatively, Ashenfelter et al. (1999) and Weichselbaumer and Winter-Ebmer (2001) include all the reported estimates to make full use of the existing information and to avoid arbitrary judgement on the authors' preferred results.[6] Here we choose to use all the reported coefficients, but as a robustness exercise we re-run all our results adding a set of study-specific dummy variables. This is in large part because in our case the authors seldom indicate which one is their preferred estimate.

We follow two complementary strategies for the construction of our dependent variable. The first strategy follows the rule of not having more than one coefficient from each reported regression. This implies that if both contemporaneous and lagged effects of reform on growth are in the original specification, then we only select the one on the contemporaneous effect. By the same token, if several alternative measures of reform are used, the default is to take the

[6] Djankov and Murrell (2002) and Stanley and Jarrell (1998) collect multiple estimates from the same study only if the estimates are derived from conceptually distinct analyses, i.e. different forms of the dependent variable from different countries or from different years.

most highly significant t-value.[7] Our second strategy is to combine all the relevant reform coefficients from a given specification, which gives rise to the matter of how to coalesce standard errors when we do not have access to all the original data sets.[8]

For each of the 43 papers, the estimates of reform on growth and their corresponding meta-independent variables were collected. This procedure gave one observation in our data set per estimated coefficient for a grand total of 321 observations. We chose to capture mainly two aspects of these effects: their sign and the magnitude of their statistical significance. The sign variable is coded as a dummy variable taking the value of 1 if the sign of the coefficient is positive, zero if its sign is negative (a similar coding was used to code the sign of the cumulative effect). About 50 percent of the coefficients are signed positive, with this increasing to about 80 percent in the case of the cumulative effect.

As for the statistical significance of the effect of reform on growth, we capture it by collecting the values of the t-statistics for each of the 321 coefficients. For about 50 percent of our sample, this is simply the t-statistic of the contemporaneous coefficient of reform on growth. The remaining estimates in our sample come from regressions in which the lagged values of reform are also included or reform is measured in first-differences, that is, if the study is interested in the "speed" of reform. When the t-statistics of the joint or cumulative effect (that is, from the combined contemporaneous and lagged coefficients) are not available in the original paper, we must compute it ourselves; one way of doing this is to use the standard errors of each coefficient and take the square root of the sum of their variances. Notice however that because we do not have access to all original datasets, this overestimates

[7] The least significant t-value is retained for the purpose of sensitivity checks. In our data, 22 out of 321 observations come from regressions with multiple estimates and are either the most or least significant t-values. Although the results below are with the 22 most significant, they do not change if we use the 22 least significant ones instead. These are available from the authors upon request.

[8] We compute, for one lag, $SE_cumul = \sqrt{(SE_{LIB(t)})^2 + (SE_{LIB(t-1)})^2}$ and, in the case of two lags, $SE_cumul = \sqrt{(SE_{LIB(t)})^2 + (SE_{LIB(t-1)})^2 + (SE_{LIB(t-2)})^2}$

the true cumulative effect, as it does not take into account the covariances.[9] The mean t-value for the direct effect of reform on growth is 0.06 and a Jarque-Bera test suggests that they are normally distributed (Figure 1), while for the cumulative effect the average t-value is 1.29 with the Jarque-Bera test also supporting normality (Figure 3).

For the sake of conciseness, we combine these two dependent variables (dummy variable for the sign of the coefficient and the value of the t-statistic) and report ordered logit results in which the dependent variable takes one of three possible values: positive and statistically significant, negative and statistically significant, and not statistically significant.[10]

The meta-independent variables capture the various characteristics of the studies so as to explain the large variation we observe in their findings. We focus on three main blocs of study characteristics: *method, measurement* and *specification*. Under method, we are interested in, *inter alia*, general modelling features (number of observations, explanatory variables, and degrees of freedom), choice of econometric technique and data features (panel or cross-section and the time period of the sample). Under measurement, we are mostly interested in the way reform is measured. And under specification, we try to capture the types and number of various control variables (Appendix 2 has a complete list of these variables). [11] Let us comment on each of these blocs in turn.

In terms of general model features, for each of the 321 regressions reported in the 43 studies, we collect information on the number of observations, the number of explanatory variables in each specification (including the reform variables) and on the degrees of freedom

[9] The distribution is shown in Figure 1. We removed atypical observations ("outliers"), defined as those for which t-statistics exceeded 3 standard deviations. There were very few of them, as we found only four such observations, all from paper 25 (Appendix 1), with t-statistics equal to 16.7, 17.2, 20.4 and 21.1.

[10] Our baseline results are for the 10% significance level. We also experimented with 1% and 5% levels and find no qualitative changes in our results. Further, in previous versions, we reported logit equations for the sign and OLS regressions for the magnitude of the effect; again we find no qualitative changes in our results. These are available from the authors upon request.

[11] Notice that Appendix 2 has the full list of variables in our data set. For the sake of space, we do not report results for all of them. We should also mention that most MRA studies we are aware of collect data on a relatively small number of study characteristics and surely none of them produces as detailed and comprehensive a data set as ours which contains 40 different potential explanatory variables.

available. The average number of explanatory variables from the regressions in our sample is almost 10 and the average degrees of freedom are slightly above 127 (with standard deviations of 8 and 80 respectively). The number of explanatory variables range from 2 to 58.

In terms of econometric modelling, we create dummy variables that: take the value of 1 if the estimates are based on panel data (zero if cross-section), if fixed (country) effects are present (zero otherwise), if fixed (time) effects are present (zero otherwise), and if reform is treated as an endogenous variable (and zero otherwise). The choice of econometric modelling reflects whether the possibility of endogeneity bias is addressed. This measure serves to answer whether the assumption of exogeneity of reforms is correct, since significantly different results from OLS and 2SLS or GMM would suggest the presence of two-way causation in the growth-reform relationship. A vast majority of specifications (almost 80 percent) are estimated on panel data, with just below a third of them addressing potential endogeneity bias and even fewer making allowances for fixed effects.

As for the time windows used in the different studies, we create variables for the first year of the sample, for the last year of the sample, and for its mid-point for each of the 321 regressions. Because output dynamics differ greatly across countries over time, we also create dummy variables for all end years of the samples in each specification (which range from years 1993 to 2004). In case the author did not disclose the exact end year for each specification, we assume all specifications in each paper are for data on the same time window. The median starting and ending years for our studies are 1990 and 1998, respectively. The variable coding the time period covered in a particular study (early, middle, or late) is used to try to uncover changing patterns of the significance of the effect over time.

Regarding different measures of reform and of reform dynamics, we create a series of dummy variables that take the value of 1 if the study used: the EBRD average reform index, the cumulative liberalisation index (De Melo et al., 1997), the World Bank's or EBRD's internal liberalisation index (prices and labour markets), the World Bank's or EBRD's

external liberalisation index (trade and capital flows), the World Bank's or EBRD's privatisation index (real and financial sectors), simple average of the three World Bank's or EBRD's indices, whether any of these individual reform components are used one at a time, a combination of the EBRD and World Bank indices (and zero otherwise for each one of these).[12] In terms of measuring reform dynamics, we generate dummy variables that take the value of 1 if both contemporaneous and lagged reforms are used, and if the reform measure is a measure of its "speed," or change over time. In addition, we capture whether the estimation has a lagged dependent variable (1 if it does, zero otherwise) and whether quadratic terms for reform are used (taking the value of 1 if they are used, zero otherwise). We find that about half of the specifications include both contemporaneous and lagged reforms and about a quarter use speed as the preferred measure of reform.[13]

Finally, regarding specification choice, we collect information on whether or not the reported specification includes variables for macroeconomic stabilisation (as well as their actual number), and in similar fashion for initial conditions, institutional development, and factors of production. We also construct measures of whether or not the results are reported separately for the former Soviet Union countries (split-sample analysis), for whether or not initial conditions are proxied by the De Melo et al. (1997) principal components indexes, for whether inflation is the stabilisation measure used, for whether the study measures underreported output,[14] and for whether the study separates the effect of reform on public and

[12] There are large literatures assessing the effect of specific reforms. These are excluded from our study because either they do not investigate more than one reform and/or they focus on individual countries. One excellent case in point is the literature on privatisation (authoritative reviews are provided by Megginson and Netter, 2001, and, for the specific case of the transition economies, by Zinnes, Eilat, and Sachs, 2001, and Hanousek, Kocenda, and Svejnar, 2005). See also Roland (2000) and Rozelle and Swinnen (2004) for a discussion of the related theoretical literature.

[13] Also note that all the studies in Appendix 1 focus on the so-called first generation reforms (stabilisation, liberalisation, and privatisation) but this is not because we do not believe that second generation reforms (e.g., institutional and regulatory changes) are important; we simply do not know of any econometric study that focuses on the latter type.

[14] Official GDP figures for the years immediately following 1989 are widely believed to be biased because statistical offices were not equipped to measure output from small private firms and because prices were liberalised at different speeds.

private sectors. Because approximately half of our coefficients come from authors whose main affiliations are not academia (and for multiple authors, in all cases they share the same type of affiliation), we also create a dummy variable for this characteristic.[15]

4. Econometric Results

Despite the large econometric literature on the effects of economic reform on economic growth during the transition from plan to market, the extent and depth of the divergence among results is almost bewildering. As noted, any casual or informal attempt to take stock of the lessons from this literature may be doomed from the start: a third of the large number of existing estimates is positive and significant, another third is negative and significant, and the final third is not statistically different from zero. It is our view that MRA can be very useful in this situation. In this section we present and discuss our results. We choose to organise this presentation in terms of three principal (potential) explanations we offer for the existing divergence, namely we investigate whether differences in (a) methods, (b) measurement, and (c) choice of specification are the main culprits. In what follows we report ordered logit results that differentiate between statistically significant and positive, statistically significant and negative, and not statistically significant effects.

From our sample of 43 studies (Appendix 1), the values of the 317 normalised t-test statistics range from –9.80 to 10.67 with mean 0.06 and standard deviation 3.64. As noted, the aggregate test statistic is $\bar{t} = \sum_{k=1}^{M} t_k / \sqrt{M} = 1.138$, which is not statistically significant at conventional levels (that is, the average effect of reform on growth in transition economies seems not to be significantly different from zero.) Yet this descriptive statistic cannot

[15] This is not uncommon in MRA. For instance, Fidrmuc and Korhonen (2006) find that central bankers' inflation forecasts tend to be significantly more conservative (lower) than academicians'. Our prior in this case is that academicians' estimates will likely be lower than those from non-academicians as suggested by the much more divided views among the former than among the latter.

represent such a diverse literature. Relying only on average test statistics, we should refrain from inferring that there is no strong relationship between growth and reform and, more importantly, this does not allow us to say much about the reasons for this divergence.

4.1 Determinants of the sign and statistical significance of the reform effect

Our meta-regression model takes the form of the following ordered logit model:

$$\Pr(S_i = j \mid Z_i, \beta_j) = e^{z_i'\beta_j} / (1 + \sum_{j=-1}^{1} e^{z_i'\beta_j}) \quad i = 1, 2 \ldots M \tag{3}$$

where S_i is a categorical variable capturing the sign and significance of the reform effect (it equals one if the effect is positive and significant, zero if insignificant, and minus one if the effect is negative and significant), Z_i is a vector of study characteristics (following the method-measurement-specification scheme as listed in Appendix 2), M is the number of estimates from the empirical literature (listed in Appendix 1) and β_i is a vector of meta-regression coefficients which reflect the effect of particular characteristics of the original study on the reform effect.

4.1.a The role of method

The results in Table 1 refer to our explanation of the divergence of results in terms of various aspects of the choice of econometric *method* in each study. They suggest that the higher the number of the degrees of freedom (DF) in the original study or the larger the number of observations, the less likely it will be that we will find a positive and significant relationship between reform and growth. We also find some evidence that studies conducted by academicians (AUTHAFF), as opposed to non-academicians, are less likely to support a positive and significant reform–growth relationship. The use of country-specific dummies (FIXED) is found to increase the probability of finding a positive reform-growth relationship, as opposed to the use of panel data (PANEL), which tend to decrease this probability.

A number of authors (among others, Heybey and Murrell, 1997; Kruger and Ciolko, 1997; Wolf, 1999; Berg et al., 1999; Fidrmuc, 2001; Falcetti et al., 2002; Staehr, 2003; and Merlevede, 2003) recognise the problem of the potential endogeneity of reform vis-à-vis economic growth and address this by instrumental variables, three-stage least squares, etc. Our meta-regression analysis reveals that those studies that treat reform as endogenous (ENDO=1) are more likely to yield a positive and statistically significant reform effect (Table 1, column 2). However, this result is not robust to the inclusion of other controls. The impact of reforms on growth is likely to increase over time as suggested by statistically significant coefficients on the LATE variable (Table 1, column 3). Nonetheless, the inclusion of other controls also makes this effect insignificant.

In summary, three method-related characteristics – AUTHAFF, PANEL, and FIXED – which appeared to be significant in determining the sign and significance of the reform effect, are taken as our first set of results.

4.1.b The role of measurement

The next set of factors we appraise is the way economic reform and its dynamics are measured. In this respect, we distinguish the origin of the index (whether it was developed by the World Bank, EBRD, or a combination of both), and the nature of the index (internal or external liberalisation, privatisation, their average, and the marginal effect of each of them if other measures are also accounted for).

The results in Table 2 show that measuring reform by the EBRD index does not seem to significantly affect the sign of the reform impact on growth performance. Yet, there is some evidence that use of the World Bank's Cumulative Liberalisation Index increases the probability of finding a positive and significant effect of reform on growth, although this finding is not robust to the inclusion of other controls.

Among three main types of reform, the only one that is statistically significant and

robust is that referring to external liberalisation (LIE), which systematically shows a positive impact on growth. Note that this effect holds even if the remaining components are included in the meta-regression. Further, and not entirely surprising, if reform is measured by the average of three main components (LII, LIE, and LIP), the resulting effect is likely to be positive (Table 2, column 3), although this effect becomes insignificant when other controls are added. Surprisingly, the privatisation measure (LIP) seems to be negatively associated with economic growth (but as can be seen in the table, this effect is not robust).

The inclusion of lagged values of reform (Table 2, column 6) is one common way of dealing with dynamics in this context. We find that it increases the probability of the effect of reform on growth to be negative and significant, and similarly with the use of speed as the measure of reform. This finding in a sense supports the "no pain, no gain" view. The effect of structural reforms seems to occur over a longer period of time, and reforms have initial costs which seem to be offset in subsequent years. Conversely, cross-sectional estimates based on average liberalisation measures over several years (or yearly estimates measuring only the contemporaneous effect of reform) tend to report a positive effect of reforms on economic performance.

Notice that when reform measures and controls for dynamics are considered together, the estimated impact of the use of panel data (as opposed to cross-section) changes the sign to positive and is significant (Table 2, columns 6-8). Also, control for fixed effects and author affiliation lose their impact in explaining the significance of the reform effect.

Column 8 in Table 2 shows our summary specification so far. Overall, five study characteristics (namely, PANEL, LAGS, SPEED, LIE, and MARGEFF) turn out to be important in explaining the variation in sign and significance of the reform effect on growth among the 43 studies we identified.

4.1.c The role of specification

The next set of variables we consider are aimed at investigating how the choice of specification in the original study, in particular the inclusion of specific sets of control variables, affect the reform effect. We thus pay special attention to (a) the inclusion of controls for initial conditions, stabilisation, institutional development and factor inputs, (b) correcting for the possibility of underreported output, (c) accounting for former Soviet Union membership, and (d) controlling for the differential public-private sector effect.

We find that simultaneously controlling for initial conditions, macroeconomic stabilisation, and institutions significantly decreases the probability of finding a positive and significant impact of reform (Table 3, column 9). Note that these effects remain when other control variables are included (Table 3, column 8). We find that controlling for being a former Soviet Union country (split sample) does not seem to change the estimated reform effect (Table 3, column 5). We also find that the effect does not seem to depend on accounting for traditional factors of production, for the public sector or underreported output (Table 3, columns 4, 6 and 7).

In summary, eight study characteristics are deemed important in explaining the sign and significance of the reform effect (see Table 3, column 9), namely, the use of: panel data, lagged reform, speed of reform, external liberalisation, multiple reforms, controls for institutions, initial conditions, and stabilisation. We checked all our results for multicollinearity. The highest Variance Inflation Factor (VIFs) we found was 5.18, which is considerably below the standard cut-off value of 10, suggesting that multicollinearity is not a severe problem in our estimations.

4.1.d Sensitivity checks

Our estimation strategy, which is based on the sequential principle *method-measure-specification*, could be potentially sensitive to the order of inclusion of the blocs of variables. For example, one may argue that the final equation of the first bloc –*method*– is mis-specified by construction since it does not account for other (significant) explanatory variables added in the next two stages, *measurement* and *specification*. To address this issue, we perform an additional sensitivity check which consists of testing which starts from a general specification encompassing all explanatory variables, and then excluding the least significant variables one by one.

The final specification obtained from such a general-to-specific selection strategy is shown in Table 4, column 1. We end up with essentially the same specification as above. The only exception is that now two more explanatory variables are significant: EARLY and LATE, suggesting that the impact of reforms on growth is likely to increase over time. Overall, our final specification includes 10 significant characteristics, explaining 36% of variation in the data (these are as follows: EARLY, LATE, PANEL, LAGS, SPEED, LIE, MARGEFF, IC, INST, STAB).

Since multiple estimates from the same study are used in our meta-regression equations, we try to deal with the potential problem of biased sampling by including dummy variables for each study. In the resulting specification, we find that 22 study dummies are significant, indicating higher than conditional average probability to report a positive reform effect in 12 studies and a lower probability in 10 studies. More importantly, study dummies do not affect our main meta-regression results as none of the 10 study characteristics loses significance (Table 4, column 2).[16] We thus conclude that our results do not seem affected by biased sampling and that the use of multiple estimates from the same study is valid in this

[16] The coefficients of the study dummies are not shown but are available from the authors upon request.

respect.

4.2 Determinants of the sign and statistical significance of the cumulative reform effect

Now we turn to the discussion of the cumulative reform effect. Similar to the analysis above of the determinants of the reform effect, the factors responsible for the variance in the sign and significance of the cumulative reform are assessed using the same blocs of study characteristics, namely, *method-measurement-specification*. Then, an alternative *general-to-specific strategy* is applied. The results from ordered logit estimations with heteroskedasticity-consistent standard errors are reported in Table 5.

As noted, a number of studies consider reform dynamics, which prompted us to incorporate this aspect in our MRA. Note that in the case of the cumulative effect, the regression sample contains a lower number of observations (220) and fewer variables are present in the final specifications for the (cumulative) effect of reform on growth. The determinants we find significant are (with their impact on the probability of finding a positive and significant reform effect in parentheses): controls for endogeneity (negative), average of internal and external liberalisation and privatisation (positive), external liberalisation (positive), marginal effects if internal, external liberalisation and privatisation are used one at a time (negative), control for lags (positive), initial conditions (negative), institutions (negative), and macroeconomic stabilisation (negative).

Like the results presented above for the non-cumulative effect, controlling for institutions, initial conditions, and stabilisation is associated with a lower probability of finding a positive and significant impact of reforms on growth. Also, the control for internal and external liberalisation and privatisation's marginal effects leads to a lower impact of reforms, while the use of external liberalisation increases the probability of a positive effect.

Unlike the case of the non-cumulative effect, the inclusion of lagged values of reform

decreases the probability of the cumulative effect to be negative. In other words, the longer the period over which the cumulative effect is calculated, the higher is the probability of finding a positive impact of reforms on growth. Yet, in the case of the cumulative effect, there seems to be little evidence that controlling for the time coverage matters for a relationship between reform and growth. Column 1 in Table 5 presents our baseline model, where eight study characteristics are found to be significant, explaining 20% of variation in test statistics.

An assessment of the problem of biased sampling was carried out by estimating the above meta-regressions with study dummy variables (Table 5, column 2). We find that 14 study dummies are significant, indicating a higher than conditional average sign and significance level of the reform effect in eight studies and a similarly lower effect in six studies. Overall, the study dummies have a minor effect on our meta-regression results. Out of eight variables, only one (STAB) loses its significance, being just slightly below the 10% significance level (significant at 10.7%). The remaining seven study characteristics retain their significance. On this basis, therefore, biased sampling should not be a serious problem here.

4.3 Discussion of additional sensitivity checks

The results above try to explain the variance of 317 coefficients from regressions of growth on reform, using the 10% cut-off level to define the dependent variable, the sign and significance of the reform effect. One first concern is that out of these 317, 22 belong to regressions with multiple reform terms. As noted, the results presented above refer to the case in which those 22 most significant t-values were selected. One first sensitivity check was to see whether our results change when the 22 *least* significant t-values are employed instead.[17] Overall, the results we obtain are very similar to those reported here.

With respect to the choice of the significance level, we investigate whether a 5% cut-

[17] These results are available upon request from the authors.

off level would make a difference. Again, we experimented with the 22 *most* and *least* significant t-values, and with contemporaneous versus cumulative reform effects. The results for the reform effect and the cumulative reform effect are qualitatively similar to those reported.

Finally, in previous versions of the paper we experimented with an alternative coding of the dependent variable, namely using separately the absolute values of the t-statistics and the sign of the reform effect (either negative or positive, not distinguishing the insignificant estimates). Similarly, we tried the 22 *most* and *least* significant t-values and the cumulative effect. The results are in line with the findings discussed above and are not reported for the sake of space (they are available upon request).

5. Conclusions

The objective of this paper is to summarise and evaluate the existing econometric evidence on the effect of structural reforms on economic growth. The quantitative analysis is carried out through recently developed meta-regression analysis techniques. We put together a unique data set covering more than 300 estimates of the effect of reform on growth from more than 40 different econometric studies. Overall, the direction of the effect of structural reforms on economic performance and its statistical significance is sensitive to the specification, modelling choice, as well as various important sources of omitted variable bias (notably, on the latter, institutions and initial conditions). The use of lagged reform measures shows that reforms have negative contemporaneous effects which can be offset by positive effects in subsequent periods, after some level of reform is in place. The estimated relationship between reform and growth seems to change over time. We also find that the existing results are sensitive to the choice of the measure of reform used.

The results of our meta-regression analysis illustrate that ignoring the problem of endogeneity of reforms vis-à-vis growth seems to lead to severely biased results. Our findings

suggest that the use of panel methods and the time coverage are important in explaining the variation in the estimated effects. Further, the one aspect of reform packages that seems to receive overwhelming support in our data is the liberalisation of trade and capital flows (that is, external liberalisation). Of particular interest is the finding that controlling for institutions and initial conditions appears to be very effective in decreasing the probability of finding a large and positive effect of reform on growth.

The findings of this paper also help us identify suggestions for future research. We focus on the following three: (1) Considerably more attention should be paid to measurement issues. There are well-known and severe measurement problems both with respect to GDP and with respect to economic reforms. As for GDP growth, the official data seems to underestimate the participation of the nascent private sector (in some cases because of the large informal sectors) and overestimate that of the public sector (mostly, if not entirely, formal activities). With respect to reform, the existing measures are mostly subjective, difficult to replicate and tend not to reflect reform reversals (Campos and Horvath, 2006). In more concrete terms, it is somewhat surprising that we were not able to find a single study that pays explicit attention to the problem of errors-in-variables. Therefore, studies that try to deal with this matter in the future will certainly make an important contribution. (2) Our findings suggest that the use of measures of external liberalisation is central in understanding growth rates, yet almost no study we examine attempts to investigate how this reform interacts (or fails to interact) with other reforms such as privatisation, and labour market and price liberalisation. Recall that the backdrop for this failure is a theoretical literature in which the issue of the sequencing of reforms looms large and a policy debate in which the big-bang versus gradualism options are discussed, as this paper demonstrates, without much robust underlying econometric evidence. Therefore more attention to the issue of reform sequencing and interactions among reforms should also generate a genuine contribution in future work. (3) Efforts could also be made in terms of making explicit the theoretical framework guiding

the econometric analysis. In very few of the studies reviewed above can one identify concerns in this respect. This translates into econometric analyses that are seldom structural, a feature that can be seen by the fact that the majority of the studies reviewed discuss economic growth while presenting evidence based on annual data. Because we now have more than 15 years of data available, it is perhaps high time to improve upon this aspect.

Several policy implications also follow from our study. First, as highlighted in Figures 1-3, there is high variation in both the sign and the significance of the effect of structural reforms on growth. Since reforms are ultimately in the hands of policymakers, it is vital to improve understanding of how structural reforms influence economic growth. In particular, a better understanding of the reasons why the long-run impact of reforms on growth tends to be positive while in the short-run it seems to have non-negligible costs, and the role institutions play in this asymmetry, would be particularly welcome. Second, our meta-regression results suggest that the initial costs of reforms depend on the type of reform. For example, considerably higher costs are associated with privatisation and internal liberalisation (i.e., prices and labour markets) than with external liberalisation (of trade and capital flows) and this seems to be irrespective of whether the World Bank's or EBRD's indexes are used. Lastly, those building conventional and dynamic macro models should try to account for structural reforms, though there is already some research in this direction, namely Ernst et al. (2006) who explore the impact of structural reforms on the short-term macroeconomic performance in 15 OECD countries using a dynamic intertemporal model with nominal and real rigidities.

References

Acemoglu, D., Johnson, S. and Robinson, J.A., 2006. "Institutions as the Fundamental Cause of Long-Run Growth," in P. Aghion and S. Durlauf (eds.) *Handbook of Economic Growth*, Elsevier.

Ashenfelter, O., Harmon, C. and Oosterbeek, H., 1999. "A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias," *Labour Economics*, 6, pp. 453–70.

Campos, N. and Coricelli, F., 2002. "Growth in Transition: What We Know, What We Don't, and What We Should," *Journal of Economic Literature*, 40 (3), pp. 793–836.

Campos, N. and Horvath, R., 2006. "Reform Redux: Measurement, Determinants and Reversals," *IZA Discussion Paper*, No. 2093: Bonn.

Card, D. and Krueger, A.B., 1995. "Time-Series Minimum Wage Studies: a Meta-Analysis," *American Economic Review*, 85, pp. 238–43.

Djankov S. and Murrell, P., 2002. "Enterprise Restructuring in Transition: A Quantitative Survey," *Journal of Economic Literature*, 40(3), pp. 736–92.

Doucouliagos, C., 1995. "Worker Participation and Productivity in Labor-Managed and Participatory Capitalist Firms: A Meta-Analysis," *Industrial and Labor Relations Review*, 49/1, pp. 58–77.

Égert, B. and Halpern, L. 2006. "Equilibrium Exchange Rates in Central and Eastern Europe: A Meta-Regression Analysis," *Journal of Banking and Finance*, 30 (5), pp. 1359–74.

Ernst, E., Gong, G., Semmler, W. and Bukeviciute, L., 2006. "Quantifying the Impact of Structural Reforms," *ECB Working Paper*, No. 666, August.

Fidrmuc, J. and Korhonen, I., 2006. "Meta-Analysis of the Business Cycle between the Euro Area and the CEECs," *Journal of Comparative Economics*, 34 (3), pp. 518–37.

Fischer R.A., 1932. "Statistical Methods for Research Workers," 4th ed. London: Oliver and Boyd.

Florax, R.J.G.M., 2002. "Meta-Analysis in Environmental and Natural Resource Economics: The Need for Strategy, Tools and Protocol," Department of Spatial Economics, Free University, Amsterdam.

Florax, R.J.G.M., de Groot, H. L.F. and De Mooij, R. A., 2002. "Meta-Analysis: A Tool for Upgrading Inputs of Macroeconomic Policy Models," *Tinbergen Institute Discussion Paper*, No. TI 041/3.

Glass, G.V., 1976. "Primary, Secondary, and Meta-Analysis of Research," *The Educational Researcher,* 10, pp. 3–8.

Hanousek, J., Kocenda, E. and Svejnar, J., 2005. "Privatization and Performance over The Transition: A Reassessment," CERGE-EI, mimeo.

Havrylyshyn, O., 2001. "Recovery and Growth in Transition: A Decade of Evidence," *IMF Staff Papers*, 48, Special Issue.

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

Jurrell, S. B. and Stanley, T.D., 1990. "A Meta-Analysis of the Union-Non-Union Wage Gap," *Industrial and Labor Relations Review*, 44/1, pp. 54–67.

Light, R.J. and Smith, P.V., 1997. "Accumulating Evidence: Procedures for Resolving Contradictions Among Different Research Studies," *Harvard Educational Review*, 41, pp. 429–71.

Megginson, W. and Netter, J., 2001. "From State to Market: A Survey of Empirical Studies on Privatization," *Journal of Economic Literature*, 39(2), pp. 321–89.

Phillips, J. M. and Goss, E. P., 1995. "The Effect of State and Local Taxes on Economic Development: A Meta-Analysis," *Southern Economic Journal*, 62/2, pp. 320–33.

Roland, G., 2000. *Transition and Economics: Politics, Markets and Firms.* Cambridge, MA: MIT Press.

Rose, A. K., 2004. "A Meta-Analysis of the Effect of Common Currencies on International Trade," *NBER Working Paper*, No. 10373.

Rozelle S. and Swinnen J., 2004. "Success and Failure of Reforms: Insights from Transition Agriculture," *Journal of Economic Literature*, 42(2), pp. 404 – 56.

Stanley T.D. and Jurrell, S. B., 1998. "Gender Wage Discrimination Bias? A Meta-Regression Analysis," *Journal of Human Resources,* 33/4, pp. 947–73.

Stanley T.D., 1998. "New Wine in Old Bottles: A Meta-Analysis of Ricardian Equivalence," *Southern Economic Journal*, 64, pp. 713–27.

Stanley T.D., 2001. "Wheat From Chaff: Meta-Analysis as Quantitative Literature Review," *Journal of Economic Perspectives*, 15/3, pp.131–50.

Stanley, T.D. and Jurrell, S. B., 1989. "Meta-Regression Analysis: A Quantitative Method of Literature Surveys," *Journal of Economic Surveys*, 3, pp.54–67.

Stiroh, K., "Reassessing the Role of IT in the Production Function: A Meta-Analysis," *Annales d'Economie et de Statistique*, forthcoming.

Weichselbaumer, D. and Winter-Ebmer, R., 2001. "The Effects of Markets, Politics, and Society on the Gender Wage Differential: A Meta-Analysis," Dept. of Economics, University of Linz, Austria.

Zinnes, C., Eilat, Y. and Sachs, J., 2001. "The Gains from Privatization in Transition Economies: Is 'Change of Ownership' Enough?" *IMF Staff Papers*, 48, Special Issue.

Table 1.

The determinants of the reform-growth effect: The role of method

	(1)	(2)	(3)	(4)	(5)
DF	-0.00245*	-0.0031	-0.00402***	-0.00319	
	[0.0013]	[0.0023]	[0.0013]	[0.0023]	
AUTHAFF		-0.636***	-0.559**	-0.618**	-0.588***
		[0.23]	[0.24]	[0.26]	[0.22]
PANEL		-0.57		-0.6	-0.888***
		[0.37]		[0.37]	[0.20]
FIXED		1.106***		1.142***	0.944**
		[0.40]		[0.42]	[0.38]
ENDO		0.473*		0.514*	
		[0.26]		[0.27]	
MID			-0.101	-0.145**	
			[0.079]	[0.073]	
EARLY			-0.329	-0.673**	
			[0.32]	[0.32]	
LATE			0.678**	0.461	
			[0.33]	[0.34]	
Observations	317	317	317	317	317
Chi2	3.564	25.93	20.96	31.89	26.25
Pseudo R2	0.00526	0.0403	0.0242	0.047	0.0339

Notes:

Estimation was carried out using ordered logit, with heteroskedasticity-consistent standard errors reported in brackets. Dependent variable takes the value of -1 if the coefficient of reform on growth is negative and statistically significant at 10%, 0 if the coefficient is not statistically significant, and 1 if the coefficient is positive and statistically significant at 10%.

DF is degrees of freedom, AUTHAFF is author's affiliation (1 if non-academic, zero otherwise), PANEL is a dummy variable that takes the value of 1 if the reform-growth coefficient is from panel data, FIXED is a dummy variable that takes the value of 1 if country-specific dummy variables are included, ENDO is 1 if there is an attempt to deal with endogeneity bias (zero otherwise), MID refers to the mid-point of the time window, EARLY and LATE refer to the time windows used for estimation (1989-1993 and 1999-2004, respectively).

***, **, and * denote significance at 1%, 5% and 10% levels respectively.

Table 2.

The role of the measurement of reform and reform dynamics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AUTHAFF	-0.66***	-0.54**	-0.56**	-0.351	-0.39	-0.32	-0.46*	
	[0.22]	[0.23]	[0.22]	[0.24]	[0.24]	[0.26]	[0.27]	
PANEL	-0.60***	-0.88***	-0.81***	-0.83***	-0.81***	0.81**	1.11**	1.23***
	[0.22]	[0.21]	[0.20]	[0.21]	[0.21]	[0.40]	[0.44]	[0.35]
FIXED	0.89**	0.95**	0.92**	1.07***	1.04***	0.34	0.3	
	[0.39]	[0.38]	[0.39]	[0.40]	[0.39]	[0.38]	[0.36]	
CLI	1.23***						0.56	
	[0.33]						[0.44]	
EBRD		-0.13					0.48*	
		[0.22]					[0.27]	
AV			0.44*				0.81	
			[0.26]				[0.54]	
LIE				1.51**	1.86**		4.83***	4.42***
				[0.61]	[0.84]		[1.45]	[1.41]
LII				-0.85**	-0.69		0.093	
				[0.42]	[0.43]		[0.91]	
LIP				-1.03***	-0.88**		0.26	
				[0.37]	[0.40]		[0.73]	
MARGEFF					-0.491		-1.35	-1.85**
					[0.56]		[0.86]	[0.78]
SPEED						-1.19***	-1.12**	-1.12***
						[0.39]	[0.44]	[0.42]
LAGS						-3.15***	-3.67***	-3.73***
						[0.40]	[0.42]	[0.40]
Observations	317	317	317	317	317	317	317	317
Chi2	44.38	27.58	31.23	48.34	49.3	147.8	146.8	134.8
Pseudo R2	0.0472	0.0344	0.0386	0.0652	0.0665	0.265	0.323	0.311

Notes:

Estimation was carried out using ordered logit, with heteroskedasticity-consistent standard errors reported in brackets. Dependent variable takes the value of -1 if the coefficient of reform on growth is negative and statistically significant at 10%, 0 if the coefficient is not statistically significant, and 1 if the coefficient is positive and statistically significant at 10%.

AUTHAFF is author's affiliation (1 if non-academic, zero otherwise), PANEL=1 if the reform-growth coefficient is from panel data, FIXED=1 if country-specific dummy variables are included, CLI=1 if the cumulative liberalisation index from the World Bank is used as a reform measure, EBRD=1 if the reform index originates from the EBRD, AV=1 if average (simple or weighted, or simple sum) of reform indices LIE, LII or LIP was used; LIE, LII and LIP refer to external liberalisation; internal and/or price liberalisation, and privatisation and banking reform, respectively; MARGEFF=1 if LII, LIE, LIP are used jointly; SPEED=1 if speed is the measure of reform, LAGS=1 if both contemporaneous and lagged reform variables are used.

***, **, and * denote significance at 1%, 5% and 10% levels respectively.

Table 3.

The role of the choice of econometric specification

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PANEL	1.26***	2.08***	1.10***	1.23***	1.24***	1.24***	1.234***	2.19***	2.04***
	[0.35]	[0.43]	[0.36]	[0.35]	[0.35]	[0.35]	[0.35]	[0.48]	[0.48]
LIE	4.49***	4.32***	4.27***	4.42***	4.42***	4.42***	4.406***	4.24***	4.27***
	[1.44]	[1.39]	[1.37]	[1.43]	[1.41]	[1.41]	[1.56]	[1.60]	[1.39]
MARGEFF	-1.97**	-1.84**	-1.92***	-1.85**	-1.85**	-1.85**	-1.859**	-2.24***	-2.06***
	[0.78]	[0.77]	[0.73]	[0.79]	[0.78]	[0.78]	[0.77]	[0.77]	[0.74]
LAGS	-3.68***	-3.81***	-3.53***	-3.73***	-3.73***	-3.73***	-3.73***	-3.62***	-3.59***
	[0.40]	[0.39]	[0.42]	[0.40]	[0.40]	[0.41]	[0.40]	[0.42]	[0.41]
SPEED	-1.09***	-1.16***	-1.30***	-1.13***	-1.12***	-1.12***	-1.12***	-1.34***	-1.29***
	[0.42]	[0.43]	[0.41]	[0.41]	[0.42]	[0.43]	[0.43]	[0.42]	[0.42]
IC	-0.45							-0.62**	-0.54*
	[0.30]							[0.29]	[0.29]
STABIL		-1.16***						-1.36***	-1.21***
		[0.43]						[0.47]	[0.47]
INST			-0.87***					-0.82***	-0.77**
			[0.30]					[0.31]	[0.30]
FACT				0.021				0.31	
				[0.36]				[0.38]	
FSU					-0.050			-0.57	
					[0.71]			[0.89]	
UNDERREP						-0.272		-0.70	
						[1.16]		[1.09]	
PUBPR							0.0365	0.24	
							[1.00]	[0.99]	
Obs.	317	317	317	317	317	317	317	317	317
Chi2	134	151.6	154.4	135.1	135.2	138.2	135.7	169.7	164.7
Pseudo R2	0.314	0.323	0.321	0.311	0.311	0.311	0.311	0.338	0.335

Notes:

Estimation was carried out using ordered logit, with heteroskedasticity-consistent standard errors reported in brackets. Dependent variable takes the value of -1 if the coefficient of reform on growth is negative and statistically significant at 10%, 0 if the coefficient is not statistically significant, and 1 if the coefficient is positive and statistically significant at 10%.

PANEL takes the value of 1 if the reform-growth coefficient is from panel data (zero otherwise); LIE=1 if the external liberalisation index was used as a reform measure, MARGEFF=1 if LII, LIE, LIP are used jointly; LAGS=1 if both contemporaneous and lagged reform variables are used; SPEED=1 if speed is the measure of reform; IC, STABIL, INST refer to controls for initial conditions, stabilisation and institutional development, respectively; FACT=1 if controlled for factors of production, FSU=1 if the results are reported separately for FSU countries, UNDERREP=1 if the study deals with underreported output.

***, **, and * denote significance at 1%, 5% and 10% levels respectively.

Table 4.

The determinants of the reform effect: sensitivity checks

	(1)[a]	(2)[b]
Method		
PANEL	1.896***	2.737***
	[0.47]	[0.62]
EARLY	-0.946***	-0.926**
	[0.36]	[0.44]
LATE	0.916**	0.708*
	[0.36]	[0.40]
Measurement		
LIE	4.723***	3.624***
	[1.46]	[1.37]
MARGEFF	-2.175**	-1.192*
	[0.85]	[0.72]
LAGS	-3.592***	-3.216***
	[0.41]	[0.48]
SPEED	-1.557***	-1.819***
	[0.49]	[0.49]
Specification		
IC	-0.629**	-0.725*
	[0.30]	[0.43]
STABIL	-1.225***	-1.832***
	[0.45]	[0.66]
INST	-1.219***	1.090*
	[0.30]	[0.61]
Obs.	317	317
Chi2	170.2	42064
Pseudo R2	0.359	0.496

Notes:

[a] *Method-measurement-specification* estimated at once, final specification.

[b] Estimated with 42 study dummies (not shown). 22 dummies are significant, out of which 12 are positive and 10 are negative.

Estimation was carried out using ordered logit, with heteroskedasticity-consistent standard errors reported in brackets. Dependent variable takes the value of -1 if the coefficient of reform on growth is negative and statistically significant at 10%, 0 if the coefficient is not statistically significant, and 1 if the coefficient is positive and statistically significant at 10%.

PANEL is a dummy variable that takes the value of 1 if the reform coefficient is from panel data (zero otherwise), EARLY and LATE refer to the time windows used for estimation (1989-1993 and 1999-2004, respectively), LIE=1 if external liberalisation components were used as a reform measure, MARGEFF=1 if LII, LIE, LIP are used jointly, LAGS=1 if both contemporaneous and lagged reform variables are used, SPEED=1 if speed is the measure of reform, IC, STABIL, INST refer to the controls for initial conditions, stabilisation and institutional development, respectively.

***, **, and * denote significance at 1%, 5% and 10% levels respectively.

Table 5.

The determinants of the cumulative reform effect

	(1)[a]	(2)[b]
Method		
ENDO	-1.819***	-2.256***
	[0.44]	[0.71]
Measurement		
LIE	1.688**	3.353**
	[0.75]	[1.58]
MARGEFF	-1.909***	-1.861**
	[0.64]	[0.92]
AV	1.896***	4.386***
	[0.48]	[1.08]
LAGS	1.612***	2.331***
	[0.52]	[0.78]
Specification		
IC	-1.624***	-1.670*
	[0.54]	[0.99]
STABIL	-0.990**	-1.21
	[0.48]	[0.75]
INST	-0.802**	-1.184***
	[0.36]	[0.45]
Obs.	220	220
Chi2	45.45	18673
Pseudo R2	0.199	0.467

Notes:

[a] *Method-measurement-specification* estimated at once, final specification.

[b] Estimated with 30 study dummies (not shown). 14 dummies are significant, out of which 8 are positive and 6 are negative.

Estimation was carried out using ordered logit, with heteroskedasticity-consistent standard errors reported in brackets. Dependent variable takes the value of -1 if the cumulative effect of reform on growth is negative and statistically significant at 10%, 0 if it is not statistically significant, and 1 if it is positive and statistically significant at 10%.

ENDO is a dummy variable that takes the value of 1 if the model used is 2SLS, 3SLS or GMM (zero otherwise), LIE=1 if external liberalisation components were used as a reform measure, MARGEFF=1 if LII, LIE, LIP are used jointly, AV=1 if average (simple or weighted, or simple sum) of reform measures LIE, LII or LIP was used, LAGS=1 if both contemporaneous and lagged reform variables are used, IC=1 if controlled for initial conditions, STABIL=1 if controlled for stabilisation, INST=1 if controlled for institutional development.

***, **, and * denote significance at 1%, 5% and 10% levels respectively.

APPENDIX 1.
Econometric studies of the effect of structural reforms on economic growth

(1) Abed,George, T. and Hamid R. Davoodi, 2000. "Corruption, Structural Reforms, and Economic Performance in the Transition Economies," *IMF WP*, No. 00/132.

(2) Ahrens, Joachim and Martin Meurers, 2002. "How Governance Affects the Quality of Policy Reform and Economic Performance: New Evidence for Economies in Transition," *Journal of Institutional Innovation, Development and Transition*, 6, p.35–56.

(3) Åslund, Anders, Peter Boone and Simon Johnson, 1996. "How to Stabilize: Lessons from Post-Communist Countries," *Brookings Papers on Economic Activity* 1, pp. 217–91.

(4) Aziz, Jahangir and Robert F. Wescott, 1997. "Policy Complementarities and the Washington Consensus," *IMF WP*, No. 97/118.

(5) Berg, Andrew, Eduardo Borenzstein, Ratna Sahay and Jeromin Zettelmeyer, 1999. "The Evolution of Output in Transition Economies: Explaining the Differences," *IMF WP*, No. 99/73.

(6) Braga de Macedo, Jorge and Joaquim Oliveira Martins, 2005. "Growth, Reform Indicators and Policy Complementarities," OECD, mimeo.

(7) Christoffersen, Peter and Peter Doyle, 2000. "From Inflation to Growth: Eight Years of Transition," *Economics of Transition*, 8(2), pp. 421–51.

(8) Cungu, Azeta and Johan Swinnen, 2003. "The Impact of Aid on Economic Growth in Transition Economies: An Empirical Study," *LICOS Discussion Paper*, No. 128/2003.

(9) De Melo, Martha, Cevdet Denizer and Alan Gelb, 1997. "From Plan to Market: Patterns of Transition," in: Blejer, Mario I. and Škreb, Marko (eds.), *Macroeconomic Stabilization in Transition Economies*, pp. 17–72. Cambridge: Cambridge University Press.

(10) De Melo, Martha, Cevdet Denizer, Alan Gelb and Stoyan Tenev, 2001. "Circumstance and Choice: The Role of Initial Conditions and Policies in Transition Economies," *World Bank Economic Review*, 15(1), pp. 1–31.

(11) Denizer, Cevdet, 1997. "Stabilization, Adjustment, and Growth Prospects in Transition Economies," *World Bank Policy Research Working Paper*, No. 1855.

(12) Eschenbach, Felix and Bernard Hoekman, 2005. "Services Policy Reform and Economic Growth in Transition Economies, 1990-2004," *World Bank WPS*, 3663 (July).

(13) Falcetti, Elisabetta, Tatiana Lysenko and Peter Sanfey, 2005. "Reforms and Growth in Transition: Re-examining the Evidence," *EBRD WP*, No 90 (March).

(14) Falcetti, Elisabetta, Martin Raiser, and Peter Sanfey, 2002. "Defying the Odds:Initial Conditions, Reforms, and Growth in the First Decade of Transition," *Journal of Comparative Economics*, 30, pp. 229–50.

(15) Fidrmuc, Jan, 2003. "Economic Reform, Democracy and Growth during Post-communist Transition," *European Journal of Political Economy*, 19 (3), pp. 583–604.

(16) Fidrmuc, Jan, 2001. "Forecasting Growth in Transition Economies: A Reassessment," Bonn, Center for European Integration Studies, mimeo.

(17) Fidrmuc, Jan and Ariane Tichit, 2004. "Mind the Break! Accounting for Structural Breaks in Growth during Transition," *William Davidson Institute Working Paper*, No. 643.

(18) Fischer, Stanley and Ratna Sahay, 2004. "Transition Economies: The Role of Institutions and Initial Conditions," IMF Conference in Honour of Guillermo Calvo.

(19) Fischer, Stanley and Ratna Sahay, 2001. "The Transition Economies After Ten Years," in: Orlowski, Lucjan T. (ed.) *Transition and Growth in Post-communist Countries: The Ten-year Experience*, pp. 3–47. Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

(20) Fischer, Stanley, Ratna Sahay and Carlos A. Végh, 1998. "From Transition to Market: Evidence and Growth Prospects," *IMF Working Paper*, No. 98/52.

(21) Fischer, Stanley, Ratna Sahay and Carlos Végh, 1996a. "Economies in Transition: The Beginnings of Growth," *American Economic Review*, 86(2), pp. 229–33.

(22) Fischer, Stanley, Ratna Sahay and Carlos Végh, 1996b. "Stabilization and Growth in Transition Economies: The Early Experience," *Journal of Economic Perspectives* 10(2), pp. 45–66.

(23) Godoy, Sergio and Joseph E. Stiglitz, 2004. "Growth, Initial Conditions, Law and Speed of Privatization in Transition Countries: 11 Years Later," 2004 AEA meetings.

(24) Havrylyshyn, Oleh and Ron van Rooden, 2003. "Institutions Matter in Transition, But So Do Policies," *Comparative Economic Studies*, 45(1), pp. 2–24.

(25) Havrylyshyn, Oleh, Ivailo Izvorski and Ron van Rooden, 2001. "Recovery and Growth in Transition Economies 1990–97: A Stylised Regression Analysis," in: Hoen, Herman W. (ed.) *Good governance in Central and Eastern Europe: The Puzzle of Capitalism by Design*, pp. 26–53. Economies and Societies in Transition series. Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

(26) Hernandez-Cata, Ernesto, 1997. "Liberalization and the Behavior of Output During the Transition from Plan to Market," *IMF Staff Papers,* No. 44(4), pp. 405–29.

(27) Heybey, Berta and Peter Murrell, 1999. "The Relationship between Economic Growth and the Speed of Liberalization," *Journal of Policy Reform*, 3(2), pp. 121–37.

(28) Kim, Byung-Yeon and Jukka Pirttila, 2003. "The Political Economy of Reforms: Empirical Evidence from Post-Communist Transition in the 1990s," Bank of Finland, Institute for Economies in Transition Discussion Paper, No. 4.

(29) Krueger, Gary and Marek Ciolko, 1998. "A Note on Initial Conditions and Liberalization during Transition," *Journal of Comparative Economics,* 26(4), pp. 718–34.

(30) Lawson, Colin and Haifeng Wang, 2005. "Economic Transition in Central and Eastern Europe and the Former Soviet Union: Which Policies Worked?" University of Bath, Department of Economics and International Development, Working Paper, No. 01-05.

(31) Lougani, Parkash and Nathan Sheets, 1997. "Central Bank Independence, Inflation and Growth in Transition Economies," *Journal of Money, Credit and Banking*, 29(3), pp. 381–99.

(32) Merlevede, Bruno, 2003. "Reform Reversals and Output Growth in Transition Economies," *Economics of Transition*, 11(4), pp. 597–751.

(33) Mickiewicz, Tomasz M., 2005. "Post-Communist Recessions Re-examined," in: Mickiewicz, T.M., *Economic Transition in Central Europe and the Commonwealth of Independent States*, Ch.6, pp.99–118. Houndsmill: Palgrave Macmillan.

(34) Mickiewicz, Tomasz M., 2005. "Is The Link Between Reforms and Growth Spurious? A Comment," in: Mickiewicz, T.M., *Economic Transition in Central Europe and the Commonwealth of Independent States*, Ch.10, pp.181–96. Houndsmil: Palgrave Macmillan.

(35) Neyapti, Bilin and Nergiz Dincer, 2005. "Measuring Quality of Bank Regulation and Supervision, with an Application to Transition Economies," *Economic Inquiry,* 43, pp. 79–99.

(36) Piculescu, Violeta, 2003, "Direct and Feedback Effects on Economic and Institutional Developments in Transition: A Path Analysis Approach," Goteborg University, mimeo.

(37) Radulescu, Roxana and Barlow, David, 2002. "The Relationship between Policies and Growth in Transition Countries," *Economics of Transition*, 10(3), pp. 719–45.

(38) Sachs, Jeffrey D., 1996. "The Transition at Mid-Decade," *American Economic Review, Papers and Proceedings,* 86 (2), pp. 128–33.

(39) Saso Polanec, 2003. "Convergence at Last? Evidence from Transition Countries," LICOS Discussion Paper, No. 144/2004.

(40) Selowsky, Marcelo and Ricardo Martin, 1997. "Policy Performance and Output Growth in the Transition Economies," *American Economic Review* 87(2), pp. 349–53.

(41) Staehr, Karsten, 2003. "Reforms and Economic Growth in Transition Economies: Complementarity, Sequencing and Speed," Bank of Finland, BOFIT DP, No.1/2003.

(42) Thorsten Beck and Luc Laeven, 2006. "Institution Building and Growth in Transition Economies," *Journal of Economic Growth*, 11(2), pp. 157–86.

(43) Wolf, Holger C., 1999. "Transition Strategies: Choices and Outcomes," *Princeton Studies in International Finance*, 85 (June), pp. 1–30.

APPENDIX 2.
Description of variables in MRA data set

Dependent variables:

REF= -1 if the coefficient of reform is negative and statistically significant at 10%, 0 if the coefficient is insignificant, and 1 if the coefficient is positive and statistically significant at 10%.

REF_CUM = -1, 0, and 1 if the coefficient on cumulative reform is negative and significant at 10%, insignificant, and positive and significant at 10%, respectively.

Meta-independent variables:

I) Method

N = the number of observations in the individual study

K = the number of explanatory variables

DF = N–K is the degrees of freedom reflecting power of the test statistics

AUTHAFF = 1 if all authors are from academia, =0 if at least one of the authors is not

Econometric technique

PANEL = 1 if the model uses panel data, =0 if cross-section

ENDO =1 if model used is 2SLS, 3SLS, GMM; =0, otherwise

FIXED = 1 if fixed effects estimation is used (or country dummies)

Time period

START = first year of the sample

END = end year of the sample

MID = (START+END)/2 is the average year of the sample

EARLY =1 if the sample is 1989-1993

MIDDLE =1 if the sample is 1994-1998

LATE = 1 if the sample is 1999-2004

E93 – E04 = dummy for the end year of the data sample (1993 – 2004 respectively)

II) Measurement

a) Origin of the index. Each index belongs to one of the following four types:

WB = 1 if the reform index originates from the World Bank only

EBRD = 1 if the reform index originates from the EBRD only

COMB = 1 if a combination of EBRD and WB indices is used

OTHER = 1 if the share of private sector in GDP or openness is used

b) Type of index. (applies to WB / EBRD / COMB)

LII = 1 if internal and/or price liberalisation components are used as a reform measure

LIE = 1 if external liberalisation components are used

LIP = 1 if privatisation and banking reform components are used

MARGEFF = 1 if LII,LIE,LIP are used in the same specification

AV = 1 if average (simple or weighted, or simple sum) of LII,LIE,LIP is used

CLI = 1 if CLI (Cumulative Liberalisation Index from the World Bank, see De Melo et al., 1997 for details) is used

c) Measure of dynamics

LAGDEP = 1 if lagged dependent variable is used in the regression

SPEED = 1 if speed is the measure of reform

LAGS = 1 if both contemporaneous and lagged reform variables are used

LAGLIB = 1 if either LAGS=1 or SPEED = 1

TIME = 1 if time dynamics is controlled for

III) Specification

IC = 1 if controlled for initial conditions

IC12 =1 if first cluster and/or second cluster of initial conditions from de Melo et al. (1997) is used

NIC = the number of types controls for IC

STABIL = 1 if controlled for stabilisation

NSTAB = the number of types of controls for stabilisation

INFL = 1 if inflation is controlled for

INST = 1 if controlled for institutional development

NINST = the number of types controls for institutional development

FACT = 1 if controlled for factors of production

NFACT = the number of types controls for factors of production

FSU = 1 if the results are reported separately for FSU countries

PUBPR = 1 if the study separates the effect of reform on public and private sector (effect of public sector is reported)

UNDERREP = 1 if study measures underreported output and corrects for it

D2 – D43 = study dummies

Figure 1.
Histogram of the t-statistics of coefficients of structural reforms on economic growth
(317 coefficients from the 43 papers listed in Appendix 1)

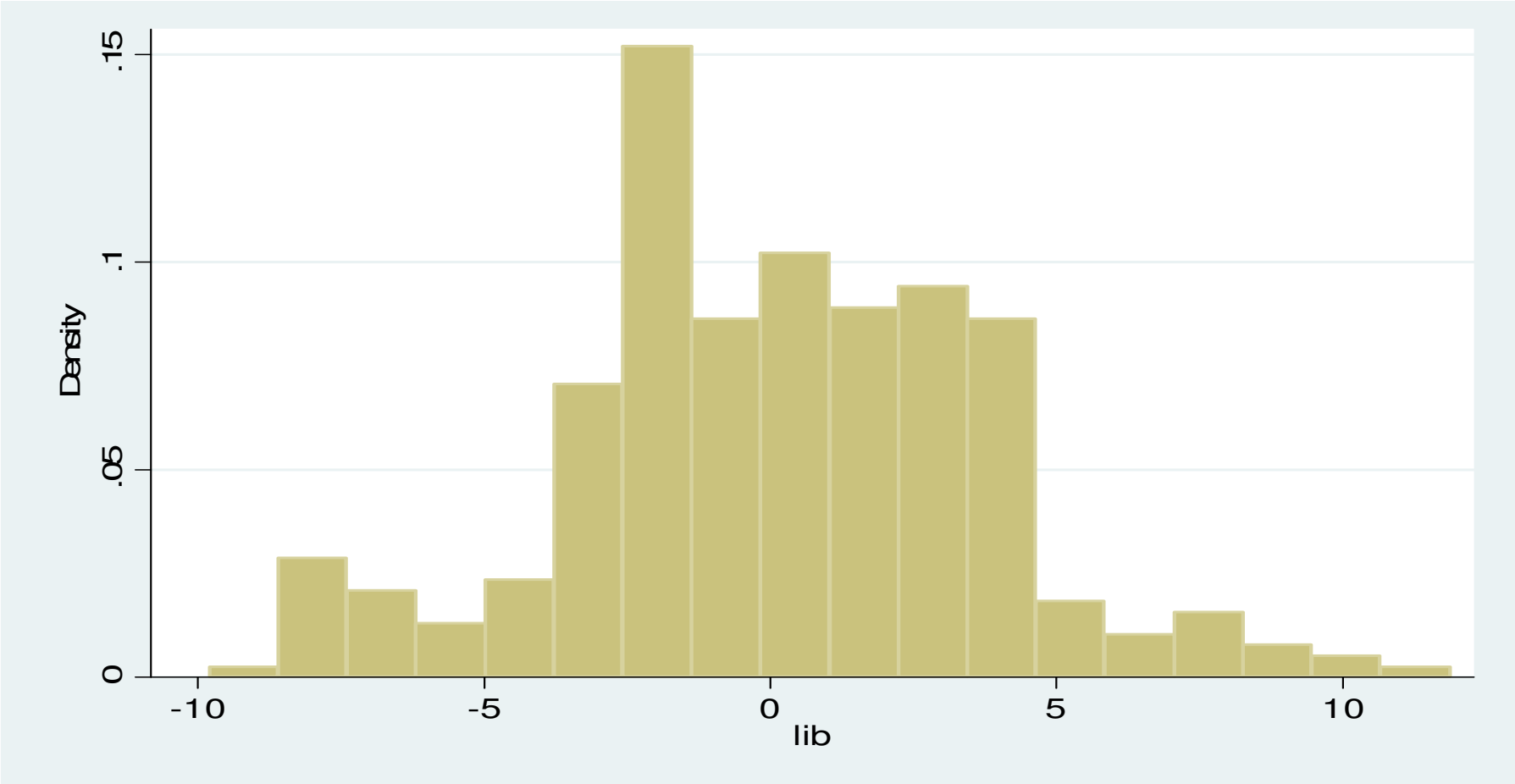

Figure 2.
Histogram of the t-statistics of coefficients of contemporaneous structural reforms on economic growth (excludes those from specifications with lagged reform and speed of reform; 157 coefficients from papers listed in Appendix 1)

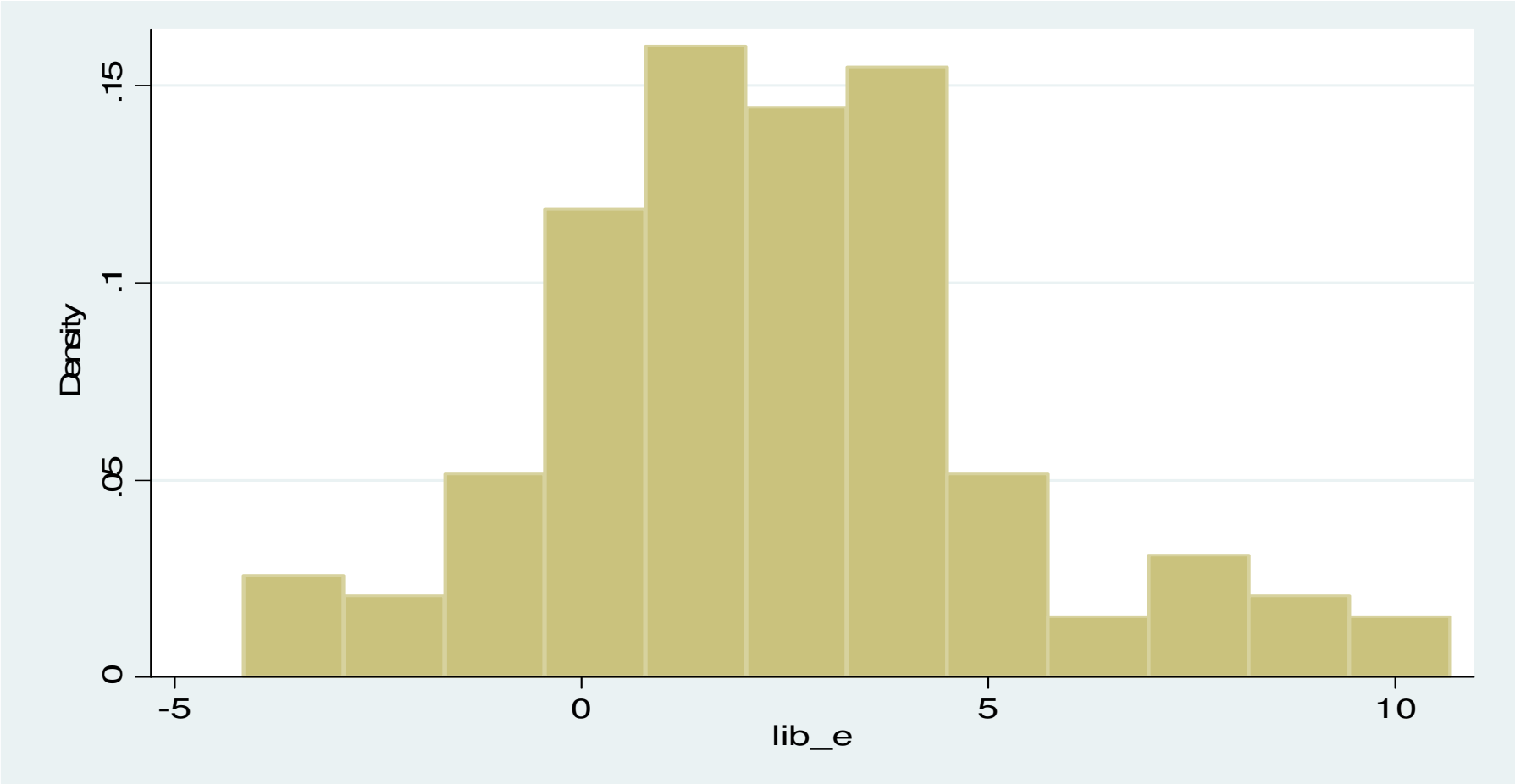
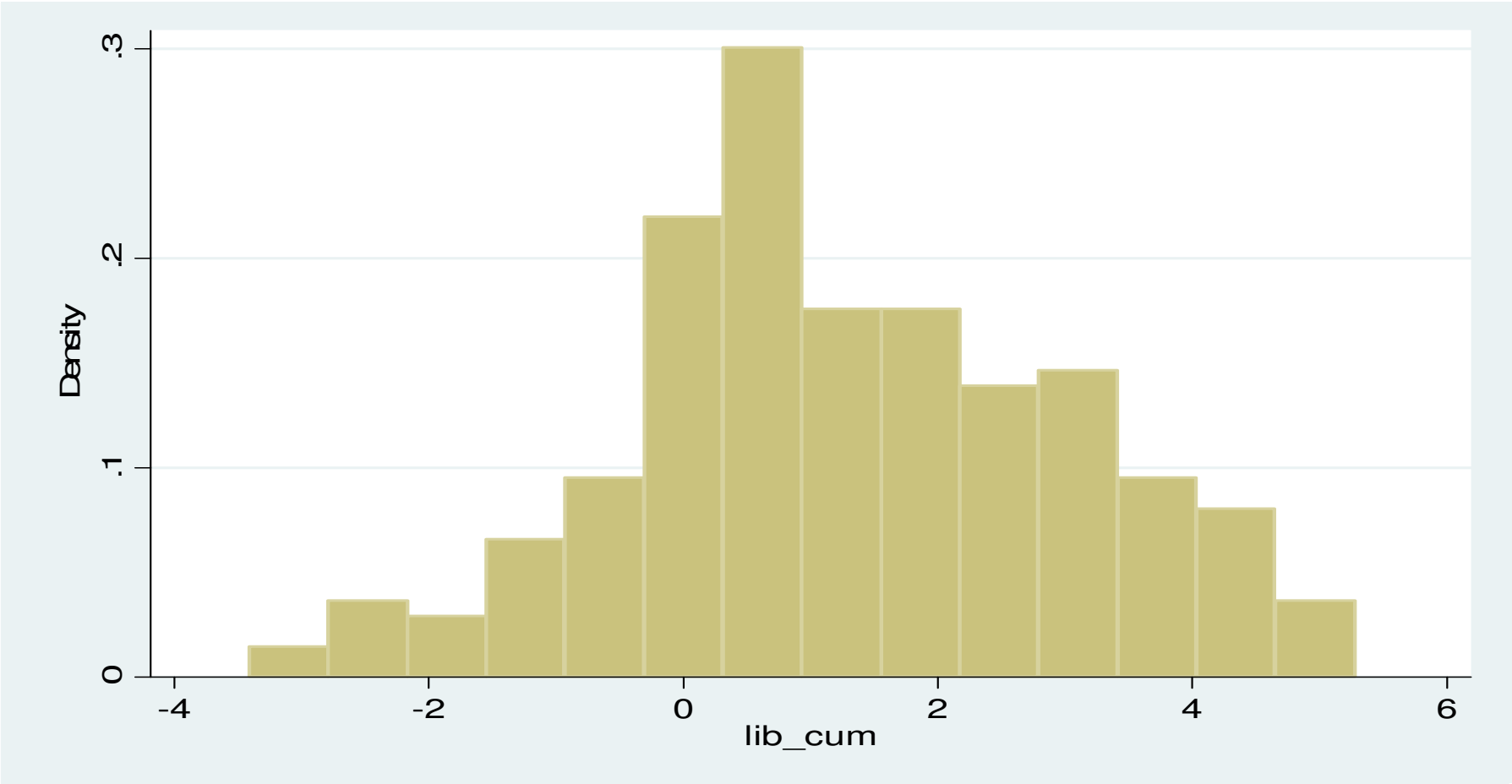
Figure 3.
Histogram of the t-statistics of coefficients of cumulative effect of structural reforms on economic growth
(221 coefficients from papers listed in Appendix 1)