INSTITUTIONAL EFFECTS ON ECONOMIC PERFORMANCE IN TRANSITION: A DYNAMIC PANEL ANALYSIS*

Adnan EFENDIC – Geoff PUGH

Journal: ***Acta Oeconomica 2015/4***

ON-LINE APPENDIX

Appendix 1

Model specification – importance of initial conditions

a) The economic theory of conditional convergence stresses the importance of controlling initial conditions in such economic models.
b) Per capita GDP on the eve of transition may be an aggregative indicator of relative capability for transformation and performance during transition (in which case, we anticipate a positive sign).
c) Specification with initial per capita income is the solution to a potential inconsistency between our dependent variable and our independent variable of interest: whereas specifying economic performance in terms of levels of per capita GDP captures the result of the entire history of time varying growth performance, our institutional variable of interest by definition measures changes confined to the transition period. However, initial (1989) per capita income controls for the influence of the entire process governing economic performance up to shortly before our sample period. Hence, effects of other independent variables are estimated net of pre-transition influences. This ensures that our dependent variable is subject to *further influences* predominantly in the period during which our independent variable of interest is able to exert influence.
d) Specification with initial per capita income helps to address potential endogeneity associated with omitted variables. Institutional quality may be correlated with time invariant unobservable influences on economic performance captured by the country-specific error terms (v_i). Consequently, specifying our model with initial per capita income is to displace an important but otherwise unobserved time invariant influence from the error term (the v_i) into the observed systematic part of the model (Roodman 2009a). Hence, there is less likely to be correlation between our independent variable of interest

and unobserved time invariant influences on our dependent variable. Conversely, omitting initial conditions may introduce unmodelled persistence into our model. If not included in the observable part of the model, the influence of initial conditions will be displaced into the group-specific error term where it may be a source of both autocorrelation and endogeneity, either of which might invalidate estimation of a dynamic model by the General Method of Moments (Bond 2002; Roodman 2009a, 2009b).[1]

[1] Evidence for the former was found when excluding our proxy for initial conditions from the model reported in Table 1; namely, the *m2* test yielded a rejection of the null of no second-order serial correlation in the differenced residuals. In contrast, the *m2* test reported in Table 1 supports non-rejection of the null.

Appendix 2

VARIABLES	DEFINITIONS AND SOURCES
gdppc ***(lngdppc)***	The level of GDP *per capita* in US Dollars. Source: EBRD, 2008; official web page; www.ebrd.com. Note, *lngdppc* is a logarithmic transformation of *gdppc* variable.
Inst ***(inst5)***	The EBRD index of structural and institutional reforms, published annually, includes the following areas: Governance and enterprise restructuring; Price liberalization; Trade and foreign exchange system; Competition policy; Banking reform and interest rate liberalization; Securities markets and nonbank financial institutions; Large-scale privatization; Small-scale privatization. Since the EBRD indices range from 1 to 4 + (where 4 + is approximation of an advanced market economy) we have linearized the scores, assigning the value of 0.33 to a '+' indicator (following Eicher and Schreiber 2010). Hence, all indices are divided by 4.33 in order to get the rank from 0 to 1, where 1 is the maximum value of the index. Source: EBRD, 2008. Note, *inst5* denotes five years difference of *inst* variable.
Cpi	Consumer price index, annual change in percentages. Source: EBRD, 2008, official web page; www.ebrd.com
budget	General government budget balance in percentages of GDP. Source: EBRD, 2008, official web page; www.ebrd.com
fdiper	Foreign direct investment, net inflow as percentage of GDP recorded in the balance of payment. Source: WB (2008) for data 1992-2006; for 2007, data are taken from EBRD, 2008; and for Montenegro, missing data for 2006 is also taken from EBRD, 2008. In 1992 Armenia recorded FDI inflow as 348.19% of GDP. Since this year was an obvious outlier in the transition sample, this observation is removed.
invest	Gross capital formation as apercentage of GDP. Source: WB, 2008a.
initial (lninitial)	Purchasing Power Parity Income per Capita in US Dollars in 1989. Source: IMF (2000). Missing data for Bosnia and Herzegovina, Serbia, and Montenegro were calculated by the authors from Savezni zavod za statistiku (1991). Note, *lninitial* is a logarithmic transformation of *initial* variable.
Eu	Dummy variable that takes a value of one if a country joined the European Union over the period 1992-2007, 0 otherwise. Source: Authors' calculations.
see	Dummy variable that takes a value of one if a country belongs to South East Europe, 0 otherwise. Source: Authors'calculations.
Cis	Dummy variable that takes a value of one if a country belongs to the Commonwealth and Independent Stages group of countries, 0 otherwise. Source: Authors' calculations.
commun	Number of years in which a particular transition country was under the communism. Source: IMF (2000). Authors' estimate for Bosnia and Herzegovina, which was missing in the sample.
war	Continuous variable that measures the number of years in which a particular transition country was involved in any kind of military conflict. The months are also calculated as part of year. Source: authors' calculations based on data obtained from www.en.wikipedia.org in January 2008.
distance	Distance in thousands of kilometres between capital city of a transition country and Brussels. 'Distance calculations are based on the WGS84 ellipsoid using *geod* (a part of

	the PROJ.4 Cartographic Projections library originally written by Gerald Evenden then of the USGS). The computation is for the great circle distance between points, and do not account for differences in elevation.' Source: http://www.infoplease.com/atlas/calculate-distance.html, visited in January 2008.
chrprob	The variable that measures the probability that a random chosen citizen is a Roman Catholic or Orthodox Christian. Authors' calculations using data obtained from www.en.wikipedia.org in January 2008
civil	The ratings of the Freedom House based on a scale of 1 to 7, with 1 representing the highest level of progress and 7 the lowest. It: 'Assesses the growth of nongovernmental organizations (NGOs), their organizational capacity and financial sustainability, and the legal and political environment in which they function; the development of free trade unions; and interest group participation in the policy process.' Source: The Freedom House, 2008.
industry	Share of industry in GDP, in percentages. Source: WB, 2008.
trade	Total trade as percentage of GDP. Source: WB, 2008.
popgr	Population growth in percentages. Source: WB, 2008.
current	The ratio of current account balance over GDP in percentages. Source: EBRD, 2008.
Source: Authors	

Appendix 3

Choice of dynamic panel model

Our preference for a dynamic panel model is based on the following arguments:

a) Static panel estimates omit dynamics causing dynamic panel bias (Baum 2006; Bond 2002; Greene 2008) and do not allow us to study the dynamics of adjustment (Baltagi 2008).

b) In our panel dataset there are 29 countries (N) that are analyzed over a period of 16 years (T). The dynamic panel model is designed for a situation where 'T' is smaller than 'N' in order to control for dynamic panel bias (Baltagi 2008).

c) The problem of the potential endogeneity of variables other than the lagged dependent variable is also much easier to address. An advantage of the dynamic GMM estimation is that all variables from the regression that are not correlated with the error term (including lagged and differenced variables) can be potentially used as valid instruments (Greene 2008).

d) Finally, static panel estimates do not allow a separate analysis of the short and long-run effects of institutions on economic performance. Hence, an additional advantage of the dynamic panel model is its ability to identify both short-run impact and long-run institutional effects (Baltagi 2008).

In addition, we had to decide which dynamic panel approach to apply. Notwithstanding that the GMM is the method of estimation of dynamic panel models that provides consistent estimates (Baum 2006), one still has to decide whether to use: difference-GMM (DGMM) developed by Arrelano and Bond (1991); or, system-GMM (SGMM) estimation established by Arrelano and Bover (1995) and Blundell and Bond (1998). Without going deeply into an investigation of differences/similarities between those two GMM approaches, we identify the main advantages of SGMM over DGMM:

a) SGMM has an advantage with respect to variables with a statistical generating mechanism that is a 'random-walk' or close to a random-walk (Baum 2006).

b) Initial conditions, the level of GDP per capita in 1989 for each country, is a time invariant variable and so would be differenced out if we were to use the DGMM approach (Roodman 2009a). We would also lose other such variables to be used later in our sensitivity analysis.

c) The SGMM approach generally produces more efficient estimates compared to DGMM by improving precision and reducing finite sample bias (Baltagi 2008).

d) If one works with an unbalanced panel then it is better to avoid DGMM estimation, which has the weakness of magnifying gaps (Roodman 2009a). Our panel is close to balance. Moreover, in some cases, so called ‘orthogonal deviations’ can be used to fill gaps (Roodman 2009a). However, estimation of our model with orthogonal deviations does not yield better model diagnostics in comparison to SGMM.

Based on the key features of SGMM and DGMM, our decision is to use SGMM as the preferred method of estimation.

Model diagnostics

Specification 1 is estimated by SGMM and implemented by *xtabond2*, a user-written programme for STATA 10 and later versions (Roodman 2009a). The estimated model is for the period 1992-2007 and covers the set of 29 TCs according to the EBRD definition (Kosovo and Turkey are not part of this sample). The results are reported in Table 1, while discussion of model diagnostics and how we address potential endogeneity in the model is available on-line.

The validity of the obtained results in SGMM depends on the model diagnostics. Compared to Ordinary Least Squares (OLS), SGMM does not assume normality and it allows for heteroskedasticity in the data. Dynamic panel models are known for endemic heteroskedasticity of the data, which can be addressed (Baltagi 2008). Accordingly, we report two-step estimates that yield theoretically robust results (Roodman 2009a). Moreover, we apply the two-step estimator to obtain the robust *Sargan test*, in other words, the (robust) *Hansen J-test*, which is not available in one-step estimation. A small panel sample may produce ‘downward bias of the estimated asymptotic standard errors’ in the two-step procedure (Baltagi 2008, p. 154). As a remedy we report corrected results by implementing the Windmeijer correction (Windmeijer 2005).

The SGMM approach assumes that the applied instruments in the model are exogenous. Consequently, an important procedure in testing the statistical properties of this model is testing the validity of instruments, which requires testing for the presence of first- and, in particular, second-order autocorrelation in the error term. Moreover, SGMM requires ‘the steady state' assumption throughout the analyzed period (Roodman 2009a), which also needs to be investigated. The results of the relevant statistical tests and checks are available on-line.

Considering together the various diagnostic tests and checks that have been conducted, there is sufficient evidence to satisfy the key assumptions of SGMM estimation and to conclude that this model is an appropriate statistical generating mechanism.

a) According to Arrelano – Bond (1991), the GMM estimator requires that there is first-order serial correlation (*m_1 test*) but that there is no second-order serial correlation (*m_2 test*) in the differenced residuals. As we see from Table 1, these tests support the validity of the model specification.

b) *The Hansen J-statistic* tests the null hypothesis of valid overidentifying restrictions, or in other words, validity of instruments (Baum 2006). According to Baum (2006, p. 201), the *Hansen J- test* is the most commonly used diagnostic in GMM estimation for assessment of the suitability of the model. T*he Hansen test* of overidentifying restrictions does not reject the null at any conventional level of significance (*p-value*=0.885); hence, it is an indication that the model has valid instrumentation.

c) *The Hansen J-test* evaluates the entire set of overidentfying restrictions/instruments. It is also important to test the validity of subsets of instruments (levels, differenced, and standard IV instruments). For this purpose, one can use a *difference-in-Sargan/Hansen test*, also known as the *C-test* (Baum 2006). The null hypothesis of the *C-test* is that the specified variables are valid instruments. As we see from Table 1, we cannot reject the null hypothesis of the exogeneity of any GMM-instruments used, or of the validity of the standard IV instruments.

d) Sarafidis et al. (2009) utilize a combination of the m_2 and *difference-in-Hansen tests* for testing cross-section dependence. This approach examines 'whether any error cross section dependence remains after including time dummies' in the model (p.149). The null hypothesis of this test is that the cross section dependence is homogenous across pairs of cross section units. In the reported model diagnostics, the m_2 statistic is satisfactory with respect to the null, while the difference between the Hansen statistics for the full set of instruments available and for each of the various subsets of instruments is not sufficiently large to reject the null of homogenous cross-section dependence (De Hoyos – Sarafidis 2006, p. 484). Conversely, if we run the same regression without time dummies the model diagnostics are much worse (particularly noteworthy is the deterioration of the m_2 test), suggesting the presence of unmodelled cross-section dependence (Sarafidis et al. 2009). Hence, inclusion of time-dummies in our specification improves the model diagnostics by removing universal time-related shocks from the error term.

e) The check for the 'steady state' assumption suggested by Roodman (2009a) can be also used to investigate the validity of instruments in SGMM. This assumption requires a kind of steady-state in the sense that deviations from long-term values are not systematically related to the group-specific effects in the error term (v_i). This assumption requires that the estimated coefficient on the lagged dependent variable in the model should indicate convergence by having a value less than (absolute) unity (Roodman 2009a, p. 114), otherwise SGMM is invalid. The estimated coefficient on the lagged dependent variable is 0.9, which is consistent with the steady-state assumption. The second condition that Roodman (2009a) suggests is that the convergence process must not be correlated with the fixed effects (the v_i), which has been addressed by controlling for initial conditions in the model.

f) Bond (2002) suggests additional investigation of the dynamic panel estimates' validity by checking whether the estimated coefficient on the lagged dependent variable lies between the values obtained from OLS and FE estimators, which is confirmed in our model (the following values are obtained: OLS=0.98 > GMM=0.91 > FE=0.60).

g) Roodman (2009b) strongly suggests that one should report the number of instruments used, since dynamic panel models can generate an enormous number of potentially 'weak' instruments that can cause biased estimates. There are no clear rules concerning how many instruments is 'too many' (Roodman 2009b), but some rules of thumb and tell-tale signs may be used. First of all, the number of instruments should not exceed the number of observations, which is the case here (41 instruments < 325 observations). Second, a tell-tale sign is a perfect *Hansen J-statistic* with the *p-value*=1.00. At the same time, the *p-value* should have a higher value than the conventional 0.05 or 0.10 levels; at least 0.25 is suggested by Roodman (2009b, p. 142). In our model, *the Hansen J-test* reports a *p-value*=0.88, which satisfies both rules. We estimated a number of other regressions by increasing or decreasing the number of instruments, using Roodman's (2009b) *collapse* command for decreasing the number of instruments, but any other restrictions worsen the model diagnostics.

h) *The F-test* of joint significance reports that we may reject the null hypothesis that the estimated coefficients on the independent variables are jointly equal to zero (p=0.000).

Addressing endogeneity

As we explain above, in a dynamic panel specification endogeneity potentially arises from omitted and typically unobservable time-invariant variables captured in the country-specific component of the model error term (v_i). Taking a differenced value of our institutional variable of interest may help to reduce such correlation and so minimize the effect of endogeneity. However, to investigate the potential endogeneity of our institutional variable of interest, we instrument *inst5* in the same manner as the endogenous lagged dependent variable. We continue to instrument the lagged dependent variable minimally; however, we find that a larger than minimum, but fewer than maximum, number of available instruments are necessary to estimate the effect of *inst5* with acceptable precision.

Considering the model diagnostics, the m_1 and m_2 tests are consistent with instrument validity (respectively, p=0.013 and p=0.085). However, because we have a relatively small sample at our disposal, the Hansen test is being used to assess more overidentifying restrictions (37, compared to 22 when *inst5* is assumed exogenous) without any increase in information, and thus has reduced power to reject the null of instrument validity. Accordingly, the overall Hansen test with p=1.00 may indicate a problem of 'too many instruments' (Roodman 2009b). To investigate this possibility, Roodman (2009b) recommends using the difference-in-Hansen tests, in order to gain statistical power by focusing on the instruments of greatest concern. First, we consider the test on the differenced instruments for the levels equation, which also constitutes a test of the validity of system versus difference GMM estimation as well as of the 'steady-state' assumption of system GMM; there is a clear non rejection of the validity of this group of instruments (p=0.976). In addition, the Hansen test excluding this group –a test of the validity of all the other instruments – fails to reject the null of instrument validity while reducing the *p-value* to a little below the tell-tale value of 1.00 (p=0.996). Secondly, a similar result is obtained from the Hansen test excluding the instruments on the lagged dependent variable; in other words, a non rejection of the validity of all other instruments (p=0.998). Finally, the joint test of the validity of the instruments on both the lagged dependent variable (endogenous by definition) and on *inst5* (potentially endogenous) yields a non rejection with p=0.802. These difference-in-Hansen tests yield *p-value*s in the range suggested by Roodman (2009b) ($0.25 \leq p < 1$) which, together with the results of the m_1 and m_2 tests, suggest non rejection of the null of instrument validity.

The estimates with *inst5* instrumented to address potential endogeneity are not radically different from those with *inst5* treated as exogenous: the coefficient on the lagged dependent variable is a little smaller (0.880 compared with 0.913); and the coefficient on our institutional variable of interest is around 10 percent larger (0.438 compared with 0.403). Given (1) that *inst5* is defined so as to obviate potential endogeneity, (2) the similarity of the estimated effects when remaining concerns about potential endogeneity are addressed, and (3) that the effect of *inst5* is less precisely estimated when instrumented (although significant at the 5% level), we focus further discussion and interpretation on the results obtained by assuming *inst5* to be exogenous.

Appendix 4

Sensitivity analysis

If we use a four-year difference of institutional quality as the explanatory variable the results are quite consistent, but the model diagnostics are weaker. If we further decrease the difference to three years, the variable of interest becomes insignificant while the model diagnostics weaken further. Similarly, increasing the difference to six or seven years likewise resulted in unacceptable model diagnostics, while the institutional proxy proved to be insignificant. All in all, the most valid results are obtained in the preferred model, suggesting that *the time-horizon over which institutional performance is measured greatly affects conclusions concerning the determinants of economic performance in transition*.

Countries' status with respect to the process of EU integration may be also an important explanatory variable in explaining economic performance and institutional effects in transition (Chousa et al. 2005; Di Tommaso et al. 2007). After including an (exogenous) EU dummy variable (the base category is non-EU TCs) the model diagnostics were a little worse than those of the base model, while the estimated effects of other variables remain much the same regarding sign, magnitude, and significance. The EU dummy has a positive sign but was not significant at conventional levels of significance. Because those TCs that entered the EU tend to have the best economic and institutional performance, this dummy variable may potentially be endogenous. Accordingly, we treat the EU dummy as a predetermined variable, using the standard SGMM instruments. In addition, following Di Tommaso et al. (2007), we instrument the EU dummy using the geographical distance from Brussels as an external instrument. However, the effect of EU membership still proves insignificant, while model diagnostics weaken further.

If we control in our model specification for different clusters of countries (EU, SEE, and CIS transition economies), the model diagnostics worsen compared to those of the base-line model, while none of these dummy variables proves to be statistically significant. Hence, we do not identify differences in the model between different clusters of TCs.

As part of our investigation, we estimated the baseline model augmented by interactions between institutions and domestic/foreign investment. We did not find any significant interaction effects, while the model diagnostics in these cases worsened.

We estimated a number of other regressions with the institutional variable from the current period as well as with lags, in each case instrumenting them with additional external instruments to be found in the literature: years under communism (*commun*); war (*war*); distance (*distance*); EU membership (*eu*); and fractionalisation by religion (*chrprob*). However, in all cases the model diagnostics were inappropriate, while the institutional variable did not appear as significant. Accordingly, we could not identify precisely the current or lagged influence of the institutional variable on economic performance, which again confirms the key findings on the importance of how we measure the timing of institutional effects.

Institutional theory is not explicit about what variables constitute the essential core of an appropriate empirical specification. Facing this challenge, authors often use the specific-to-general approach in specifying their institutional models (Klomp – De Haan 2009) or investigate more specifications with sets (vectors) of different variables (Blume – Voigt 2011). In our case, specifying a model with all potentially important variables identified in the transition research is incompatible with estimating a SGMM model on the size of sample available. Yet estimation on different groups of variables yields widely varying results; as reported above, Efendic et al. (2011) find that the link between institutions and economic performance was conditional on, amongst other sources of heterogeneity, empirical specifications. Accordingly, our preferred specification must be further investigated in order to check its robustness.

In order to assess our specification we conducted a variant of Extreme Bounds Analysis (EBA), which we adapted to take account of model diagnostics (more on EBA can be found in: Leamer 1985). EBA is an econometric methodology to assess whether minor changes in the list of independent variables alter the main conclusions of the model (Leamer 1985). It is also a test of whether some 'doubtful' omitted variables truly do not belong to the model; if so, then the base specification should produce better estimates (Leamer 1983).

The key step in the EBA procedure is to estimate the base specification extended for all possible combinations of up to three variables (henceforth, EBA models). For each estimated EBA model, one investigates whether the coefficient on the variable of interest, $\hat{\omega}_{mj}$, where m indexes the coefficient estimated on the variable of interest (*inst5*) in j regressions, remains statistically significant and of the theoretically predicted sign. The extreme bounds refer to the highest and the lowest values of $\hat{\omega}_{mj}$ obtained from its standard error $\hat{\delta}_{mj}$ in the EBA model, according to the following formulas: Lower bound = $\hat{\omega}_{mj} - 2 \cdot \hat{\delta}_{mj}$; and Upper bound = $\hat{\omega}_{mj} + 2 \cdot \hat{\delta}_{mj}$. Researchers(s) report the lower and upper bounds of $\hat{\omega}_{mj}$ and assess whether the coefficient of interest is likely to be zero (EBA test 'is not passed') or not (EBA test 'passed') (McAleer et al. 1985).

One shortcoming of EBA is lack of clear guidelines about the diagnostics that should be investigated for the EBA models. Different authors have considered some model diagnostics issues, but without achieving consensus. Accordingly, in this EBA we use an important additional criterion. Namely, we refer to the model diagnostics, since SGMM 'works only under arguably special circumstances' (Roodman 2009b, p. 156). Hence, the final judgment on EBA models will be based on the standard EBA tests suggested by Leamer (1985) but augmented by the model diagnostic tests, establishing in that way the more 'systematic approach' to EBA evaluation suggested by McAleer et al. (1985, p. 306).

There is no clear rule as to which variables should be considered in EBA. Rather, variable selection depends on theory but also on researchers' judgments of the potentially important variables (McAleer et al. 1985) as well as on data availability (Sala-i-Martin 1997). Initially, as relevant variables we consider those that are already exploited in the empirical institutional research on TCs. Using this criterion, we come to a set of more than 20 variables that were used in different studies focused on either output growth or output levels. However, some of the variables provide similar information to the existing variables in the base specification, and so are not interesting for our EBA. Considering all the potentially interesting 'transition' variables, our final list includes nine variables (with very short descriptions): *commun* (number of years under communism); *war* (number of years during the last two decades in which a particular country endured military conflicts); *industry* (percentage share of industry in GDP); *distance* (distance of the capital city from Brussels); *trade* (total trade share as a percentage of GDP); *eu* (dummy variables for TCs that are EU members); *chrprob* (probability that a randomly chosen citizen is Christian); *popgr* (average annual population growth in percentages); and *current* (current account deficit as a percentage of GDP). Importantly, none of these examined variables is highly correlated with either our variable of interest (*inst5*) or with the other variables. This suggests that multicollinearity should not be a significant problem when using these variables together in the same regression.

Using those nine variables in different combinations we estimated 129 EBA regressions. Table 3 reports results from the twenty-four EBA regressions that yield adequate model diagnostics.

Table 3. Summary findings from the Extreme Bounds Analysis (EBA)

Sets of up to three variables used in the EBA models	Coeff. (*inst5*) $\hat{\omega}_{mj}$	Lower bound	Upper bound	EBA test (+/-)	Significance test	Sign test (+/-)	Robust/ fragile (+/-)
Trade	0.36	0.06	0.66	+	+	+	+
Chrprob	0.47	0.13	0.81	+	+	+	+
Popgr	0.42	0.04	0.80	+	+	+	+
trade chrprob	0.44	0.08	0.80	+	+	+	+
trade popgr	0.43	0.07	0.79	+	+	+	+
trade current	0.41	0.05	0.77	+	+	+	+
eu popgr	0.51	0.13	0.89	+	+	+	+
chrprob popgr	0.47	0.11	0.83	+	+	+	+
popgr current	0.49	0.29	0.69	+	+	+	+
commun war trade	0.51	0.07	0.95	+	+	+	+
commun distance eu	0.58	0.12	1.04	+	+	+	+
war industry trade	0.61	0.05	1.17	+	+	+	+
war distance trade	0.49	0.11	0.85	+	+	+	+
war distance popgr	0.33	- 0.85	1.51	-	-	+	-
war trade current	0.59	0.10	0.99	+	+	+	+
war popgr current	0.61	0.23	0.99	+	+	+	+
industry trade eu	0.45	0.05	0.85	+	+	+	+
industry eu current	0.46	0.10	0.82	+	+	+	+
distance eu popgr	0.53	0.11	0.95	+	+	+	+
trade chrprob popgr	0.42	0.02	0.82	+	+	+	+
trade chrprob current	0.53	0.13	0.93	+	+	+	+
trade chrprob popgr	0.42	0.02	0.82	+	+	+	+
eu chrprob popgr	0.49	0.09	0.89	+	+	+	+
eu chrprob current	0.53	0.15	0.91	+	+	+	+
eu popgr current	0.45	0.11	0.79	+	+	+	+

Source: Authors' calculations using EBA regression results from STATA 10.

Notes: Coeff. (*Inst5*) $\hat{\omega}_{mj}$ - coefficient on the variable of interest (*inst5*) estimated in EBA model
Lower bound – lower bounds in the EBA model
Upper bound –upper bounds in the EBA model
EBA test – whether the EBA model satisfies (+) or not (-) the EBA decision rules for the bounds.
Significance test – whether *inst5* in EBA model is significant (+) or not (-)
Sign test - whether *inst5* in EBA model is of expected positive (+) or negative (-) sign
Robust/fragile – whether EBA model has satisfied all tests – robust (+) or failed at least one - fragile (-)

Our variable of interest always had the anticipated sign; hence, it is robust in its positive influence on economic performance. Regarding significance and EBA tests, we identify only one EBA model (that includes *war*, *distance* and *popgr* as additional independent variables) which is 'fragile'. Conversely, in 23 of the 24 estimated EBA regressions *with acceptable model diagnostics* our variable of interest was 'robust' to changes in specification. Accordingly, by applying Sala-i-Martin's (1997) suggestion to look at the entire distribution of the EBA results, we conclude that the institutional variable was robust to changes in the base specification when estimated in a statistically well specified model.

The most demanding robustness check will be to estimate the model with an extended data set that includes the period of the global financial crisis and its aftermath. However, as yet, some of the key data are not available for 2008-2012. Moreover, the accuracy of the data for this period is still questionable and subject to revision. This suggests that it may be still too early to expect a reliable robustness check by applying our model to post-crisis data.

References (used in this appendix but not included in the paper)

Arrelano, M. – Bover, O. (1995): Another look at the instrumental variables estimation of error components models. *Journal of Econometrics*, 68: 29-51.

Baltagi, B. H. (2008): *Econometric analysis of panel data*. Chichester: John Wiley & Sons Ltd..

Baum, F. C. (2006): *An introduction to modern econometrics using Stata*. Texas: Stata Press.

Blume, L. – Voigt, S. (2011): Does organizational design of supreme audit institutions matter? A cross-country assessment. *European Journal of Political Economy*, 27: 215-229.

Bond, S. (2002): Dynamic panel models: a guide to micro data methods and practice. *Portugees Economic Journal*, 1: 141–162.

De Hoyos, R. E. – Sarafidis, Y. (2006): Testing for cross-sectional dependence in panel-data models. *The Stata Journal*, 6: 482-496.

European Bank for Reconstruction and Development (EBRD) (2008): Selected economic indicators. Downloaded from: www.ebrd.com in July 2008.

Freedom House (2008) *Nations in transition 2007: democratization from central Europe to Eurasia*. Budapest: Freedom House Europe Kht.

International Monetary Fund (IMF) (2000): *World economic outlook focus on transition economies*. Washington, IMF.

Klomp, J. – De Haan, J. (2009): Political institutions and economic volatility. *European Journal of Political Economy*, 25: 311-326.

Leamer, E. E. (1985): Sensitivity analysis would help. *The American Economic Review*, 75: 308-313.

McAleer, M. – Pagan, A. A. – Volker, P. A. (1985): What will take the con out of econometrics?. *The American Economic Review*, 75: 293-307.

Roodman, D. (2009a.): How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal*, 9: 86–136.

Roodman, D. (2009b): A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71: 135-158.

Sala-i-Martin, X. X. (1997): I just run two million regressions. *AEA Papers and Proceedings*, 87: 178-183.

Sarafidis, V. – Yamagata, T. – Robertson, R. (2009): A test of cross-section dependence for a linear dynamic panel model with regressors. *Journal of Econometrics*, 148: 149-161.

Savezni zavod za statistiku (1991): *Statistički godišnjak Jugoslavije 1991*. Beograd: Savezni zavod za statistiku.

The World Bank (WB) (2008): *World development indicators, (CD-Rom)*. Washington D. C.: World Bank.

Windmeijer, F. (2005): A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, 126: 25-51.