STRATHCLYDE

DISCUSSION PAPERS IN ECONOMICS

THE POSITIVE CASUAL IMPACT OF FOREIGN DIRECT INVESTMENT ON PRODUCTIVITY: A NOT SO TYPICAL RELATIONSHIP

BY

RODOLPHE DESBORDES AND VINCENZO VERARDI

NO. 11-06

DEPARTMENT OF ECONOMICS
UNIVERSITY OF STRATHCLYDE
GLASGOW

The Positive Causal Impact of Foreign Direct Investment on Productivity: A Not So Typical Relationship

Rodolphe Desbordes[‡] Vincenzo Verardi[§]

February 2011

Abstract

Previous research has argued that foreign direct investment (FDI) exerts a positive and causal impact on the productivity of the recipient countries. However, we find that there is little macroeconomic evidence that FDI fosters productivity growth in recipient countries, including in those with high absorptive capacity, once we use an instrumental variables (IV) estimator robust to outliers.

JEL classification: C2, F2, O3, O4.

Keywords: Foreign Direct Investment, Productivity, Robust Regression.

[‡]University of Strathclyde. Corresponding author. Address: Department of Economics, Sir William Duncan Building, University of Strathclyde, 130 Rottenrow, Glasgow G4 0GE, Scotland, United Kingdom. Telephone/Fax number: (+44) (0)141 548 3961/4445. E-mail: rodolphe.desbordes@strath.ac.uk

[§]University of Namur (CRED) and Université Libre de Bruxelles (ECARES and CKE). E-mail:vverardi@fundp.ac.be. Vincenzo Verardi is Associated Researcher of the FNRS and gratefully acknowledges their financial support.

> *The overall benefits of FDI for developing country economies are well documented. Given the appropriate host-country policies and a basic level of development, a preponderance of studies shows that FDI triggers technology spillovers, assists human capital formation, contributes to international trade integration, helps create a more competitive business environment and enhances enterprise development. All of these contribute to higher economic growth, which is the most potent tool for alleviating poverty in developing countries* (OECD, 2002, p.5).

It has been progressively acknowledged that the theoretical growth gains associated with capital market liberalisation have often failed to materialise, especially in developing countries (Prasad et al., 2003). Nevertheless, as the above quote makes clear, much faith has remained in the ability of Foreign Direct Investment (FDI) flows to foster productivity[1] growth, thanks to their alleged contribution to the international diffusion of technology. This issue has been hotly debated empirically. The positive results of early studies were often dismissed on misspecification grounds (Carkovic and Levine, 2005; Wooster and Diebel, 2010). Endogeneity was particularly an issue. A confounding variable may have been omitted or the higher income associated with higher productivity could have been the cause of larger FDI flows, rather than the consequence. Researchers turned to instrumental variables (IV) approaches to address these concerns, and the most recent studies (Kose et al., 2009; Vua and Noyb, 2009; Kemeny, 2010) conclude that FDI does indeed foster productivity growth in recipient countries, even after controlling for several forms of endogeneity.

Given the actual state of the empirical literature, it could be inferred that the positive

[1] In this paper, when we mention productivity, we refer to total factor productivity, not labour productivity.

effects of FDI on productivity are no longer controversial. This paper argues that such a conclusion is premature as recent papers, despite the increasing sophistication of their econometric techniques, have neglected a key assumption underlying their IV estimators: the absence of outliers. If this assumption is not satisfied, even one outlier may cause an IV estimator to be heavily biased. In the jargon of the statistics literature, the classical IV estimator is said to be not a robust estimator. Unfortunately, outliers are likely to be present in FDI data. For instance, investments in the oil sector may correspond to a large share of GDP, or a small country may be a large recipient of FDI thanks to generous tax policies, leading to "roundtripping" and "trans-shipping" FDI.[2] It is also well-known that productivity measures are frequently distorted by the presence of natural resources; Hall and Jones (1999) report that in the absence of any correction, oil-rich Oman and Saudi Arabia would dominate their productivity ranking. Even though it must be acknowledged that some studies report having paid attention to outliers, it is unlikely that they have successfully dealt with this issue.[3] They used outlier diagnostics based on least-squares residuals. Given that the least-squares estimator is extremely non-robust to outliers, these diagnostics share the same fragility and very often fail to detect atypical observations. In addition, their approach did not take into account the combined influence of outliers in the first and second stages of their IV estimations.

[2]Roundtripping refers to the situation where different treatments of foreign and domestic investors encourage the latter to channel their funds into special purpose entities (SPEs) abroad in order to subsequently repatriate them in the form of incentive-eligible FDI. With trans-shipping, funds channeled into SPEs in offshore financial centres are redirected to other countries, leading to strong divergences between the source country of the FDI and the ultimate beneficiary owner.

[3]For instance Kose et al. (2009) report, p.575 "*We first eliminated all observations with financial openness values that were more than two standard deviations from their respective full sample means. [...] We also used the method proposed by Hadi (1994) for detecting outliers in multivariate regressions. Again, eliminating such outliers made little difference to the key results.*"

We remedy to this omission of the literature by refining a ‘robust’ (to outliers) IV estimator (a RIV estimator) initially proposed by Cohen-Freue and Zamar (2006), in order to estimate the ‘robust’ impact of FDI on productivity in a panel of 106 countries over the 1970-2005 period. We improve on Cohen-Freue and Zamar (2006)’s estimator in three different ways. First, we use a weighting scheme that makes our estimator more efficient and allows the computations of the usual identification and overidentifying restrictions tests. Second, we show how the asymptotic variance of their estimator can be made robust to heteroskedasticity and asymmetry. Finally, we exploit this new estimator of the asymptotic variance to implement a generalised Hausman test for the presence of outliers.

In our empirical application, we find that controlling for the existence of outliers make a profound difference to the results. Whereas an IV estimator suggests that a larger FDI stock to GDP ratio increases productivity, the exact opposite conclusion is reached when employing the RIV estimator. A graphical tool allows us to identify the outliers which are responsible for this divergence in parameter estimates. The most outlying observations correspond to a war-stricken resource-rich country (Liberia) and a tax haven (Luxembourg). Finally, we investigate whether a more positive impact can be detected in countries which are likely to have been better placed to absorb the foreign technology spillovers. Our RIV estimates support this hypothesis in the sense that the impact of FDI on productivity becomes statistically insignificant in countries with favourable attributes such a large stock of human capital or a well-developed financial system. In those countries, the negative and positive productivity spillovers possibly balance out.

The remainder of the paper is organised as follows: section 1 reviews the classical

IV estimator, presents the RIV estimator and describes our generalised test for outliers. Section 2 demonstrates the good behaviour and properties of the RIV estimator and the test for outliers via Monte-Carlo simulations. In section 3 we describe the data used in our empirical analysis and motivate our econometric approach. Section 4 presents and interprets our empirical results and section 5 concludes.

1 Instrumental variables estimation

1.1 Classical instrumental variables estimation

1.1.1 Classical instrumental variables estimator

The objective of linear regression analysis is to study how a dependent variable is linearly related to a set of regressors. The linear regression model is given by:

$$y_i = \mathbf{x}_i^t\theta + \varepsilon_i \tag{1}$$

where y_i is the scalar dependent variable and $\mathbf{x}_i$ is the $(p \times 1)$ vector of covariates observed for $i = 1, ...n$. Vectors and matrices will be denoted by boldface throughout. Vector θ of size $(p \times 1)$ contains the unknown regression parameters and needs to be estimated. On the basis of the estimated parameter $\hat{\theta}$, it is then possible to fit the dependent variable by $\hat{y}_i = \mathbf{x}_i^t\hat{\theta}$, and estimate the residuals $r_i(\theta) = y_i - \hat{y}_i$ for $i = 1 \leq i \leq n$. Although θ can be estimated in several ways, the common intuition is to try to get as close as possible to the true value of the parameters by reducing the total magnitude of the residuals, as measured by an aggregate prediction error. In the case of the ordinary least squares (LS) method, this aggregate prediction error is defined as the sum of squared residuals. The vector of parameters estimated by LS is then

$$\hat{\theta}_{LS} = \arg\min_{\theta} \sum_{i=1}^{n} r_i^2(\theta) \tag{2}$$

with $r_i(\theta) = y_i - \theta_0 - \theta_1 x_{i1} - ... - \theta_1 x_{ip-1}$ for $1 \leq i \leq n$. Calling $\mathbf{X}$ the $(n \times p)$ matrix containing the values for the p regressors (constant included) and $\mathbf{y}$ the $(n \times 1)$ vector containing the value of the dependent variable for all the individuals, the solution to this minimisation leads to the well-known formula

$$\hat{\theta}_{LS} = \underbrace{\left(\mathbf{X}^t\mathbf{X}\right)}_{n\mathbf{\Sigma_{XX}}}^{-1} \underbrace{\mathbf{X}^t\mathbf{y}}_{n\mathbf{\Sigma_{Xy}}} \tag{3}$$

which is simply the product of the ($p \times p$) covariance matrix of the explanatory variables $\mathbf{\Sigma_{XX}}$ and the ($p \times 1$) vector of the covariances of the explanatory variables and the dependent variable $\mathbf{\Sigma_{Xy}}$ (the n simplify).

The unbiasedness and consistency of the LS estimates crucially depend on the absence of correlation between $\mathbf{X}$ and ε. When this assumption is violated, instrumental variable estimators are generally used. The logic underlying this approach is to find some variables, known as instruments, which are strongly correlated with the troublesome explanatory variables, known as endogenous variables, but independent of the error term. This is equivalent to estimating the relationship between the response variable and the covariates by using only the part of the variability of the endogenous covariates that is uncorrelated with the error term.

More precisely, let's define $\mathbf{Z}$ the $(n \times m)$ matrix (where $m \geq p$) containing the instruments. The instrumental variable estimator (generally called two stages least squares when $m > p$) can be conceptualised as a two stage estimator. In the first stage, each endogenous variable is regressed on the instruments and on the variables in $\mathbf{X}$ that are not correlated with the error term. The predicted value for each variable is then fitted. In this way, each variable is purged of the correlation with the error term. Exogenous explanatory variables are used as their own instruments. Technically speaking, the first stage consists in fitting

$$\hat{\mathbf{X}} = \mathbf{Z}\left(\mathbf{Z}^t\mathbf{Z}\right)^{-1}\mathbf{Z}^t\mathbf{X} \tag{4}$$

In the second stage, the standard LS formula (3) is used, but $\mathbf{X}$ matrix is replaced by $\hat{\mathbf{X}}$

$$\hat{\theta}_{IV} = \left(\hat{\mathbf{X}}^t\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}'\mathbf{y} \tag{5}$$

By replacing (4) in (5) we have that

$$\hat{\theta}_{IV} = \left(\mathbf{X}^t\mathbf{Z}\left(\mathbf{Z}^t\mathbf{Z}\right)^{-1}\mathbf{Z}^t\mathbf{Z}\left(\mathbf{Z}^t\mathbf{Z}\right)^{-1}\mathbf{Z}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{Z}\left(\mathbf{Z}^t\mathbf{Z}\right)^{-1}\mathbf{Z}^t\mathbf{y} \tag{6}$$

that simplifies to

$$\hat{\theta}_{IV} = (\underbrace{\mathbf{X}^t\mathbf{Z}}_{n\Sigma_{\mathbf{XZ}}}\underbrace{(\mathbf{Z}^t\mathbf{Z})}_{n\Sigma_{\mathbf{ZZ}}}^{-1}\underbrace{\mathbf{Z}^t\mathbf{X}}_{n\Sigma_{\mathbf{ZX}}})^{-1}\underbrace{\mathbf{X}^t\mathbf{Z}}_{n\Sigma_{\mathbf{XZ}}}\underbrace{(\mathbf{Z}^t\mathbf{Z})}_{n\Sigma_{\mathbf{ZZ}}}^{-1}\underbrace{\mathbf{Z}^t\mathbf{y}}_{n\Sigma_{\mathbf{Zy}}} \tag{7}$$

We finally have

$$\hat{\theta}_{IV} = \left(\boldsymbol{\Sigma}_{\mathbf{XZ}} \left(\boldsymbol{\Sigma}_{\mathbf{ZZ}}\right)^{-1} \boldsymbol{\Sigma}_{\mathbf{ZX}}\right)^{-1} \boldsymbol{\Sigma}_{\mathbf{XZ}} \left(\boldsymbol{\Sigma}_{\mathbf{ZZ}}\right)^{-1} \boldsymbol{\Sigma}_{\mathbf{Zy}} \tag{8}$$

where $\Sigma_{\mathbf{XZ}}$ is the covariance matrix of the original right-hand side variables and the instruments, $\boldsymbol{\Sigma}_{\mathbf{ZZ}}$ is the covariance matrix of the instruments and $\Sigma_{\mathbf{Zy}}$ is the vector of covariances of the instruments with the dependent variable. A drawback of the IV method is that if outliers are present, all the estimated covariances are biased. Cohen-Freue and Zamar (2006) therefore suggest to replace the classical covariance matrices in (8) by some robust counterparts that withstand the contamination of the sample by outliers.

1.1.2 Asymptotic variance

The asymptotic variance of the classical IV estimator (that withstands heteroskedasticity) is the standard Huber-White sandwich estimator based on $\hat{\mathbf{X}}$ rather than $\mathbf{X}$, i.e. $\mathbf{V}_{IV} = \left(\hat{\mathbf{X}}^t\hat{\mathbf{X}}\right)^{-1} \hat{\mathbf{X}}^t\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^t\hat{\mathbf{X}}^t \left(\hat{\mathbf{X}}^t\hat{\mathbf{X}}\right)^{-1}$. Note however that the residuals used to estimate the variance are $r_i = y_i - \mathbf{x}_i^t\hat{\theta}_{IV}$ and not $\tilde{r}_i = y_i - \hat{\mathbf{x}}_i^t\hat{\theta}_{IV}$. The formula of the estimated asymptotic variance is therefore

$$\hat{\mathbf{V}}_{IV} = \left(\hat{\mathbf{X}}^t\hat{\mathbf{X}}\right)^{-1} \hat{\mathbf{X}}^t\mathbf{r}\mathbf{r}^t\hat{\mathbf{X}}^t \left(\hat{\mathbf{X}}^t\hat{\mathbf{X}}\right)^{-1} \tag{9}$$

1.2 Robust instrumental variables estimation

1.2.1 Robust instrumental variables estimator

When outliers are present, the covariances in equation (8) need to be robust to outliers. We follow Cohen-Freue and Zamar (2006) by using what are called the S-estimators of scatter.

An useful preliminary introduction to these estimators is the notion of generalised variance. This measure, originally introduced by Wilks (1932), is a one-dimensional assessment of multidimensional spread. Without loss of generality, we explain this concept calling on a 2×2 covariance matrix. The generalisation to higher dimensions is straightforward.

Let's define a covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma^2_{\mathbf{x}_1} & \sigma_{\mathbf{x}_1\mathbf{x}_2} \\ \sigma_{\mathbf{x}_1\mathbf{x}_2} & \sigma^2_{\mathbf{x}_2} \end{pmatrix} \tag{10}$$

where $\sigma^2_{\mathbf{x}_1}$, $\sigma^2_{\mathbf{x}_2}$ and $\sigma_{\mathbf{x}_1\mathbf{x}_2}$ are respectively the variance of variable $\mathbf{x}_1$, the variance of variable $\mathbf{x}_2$ and the covariance between the two. The generalized variance is defined as the determinant of Σ: i.e. $\sigma^2_{\mathbf{x}_1}\sigma^2_{\mathbf{x}_2} - \sigma^2_{\mathbf{x}_1\mathbf{x}_2}$. This expression is composed of two elements: the product of $\sigma^2_{\mathbf{x}_1}$ and $\sigma^2_{\mathbf{x}_2}$ and the squared covariance $\sigma^2_{\mathbf{x}_1\mathbf{x}_2}$. The first term ($\sigma^2_{\mathbf{x}_1}\sigma^2_{\mathbf{x}_2}$) represents the raw bi-dimensional spread of the observations. However, if $\mathbf{x}_1$ and $\mathbf{x}_2$ are not independent, some of the variance in $\mathbf{x}_2$ is already accounted for by the variance in $\mathbf{x}_1$. When we look at the formula of the determinant, we see that this redundancy is dealt through the substraction of the second term ($\sigma^2_{\mathbf{x}_1\mathbf{x}_2}$). Hence, the generalised variance is a unidimensional assessment of the bi-dimensional spread once the covariation has been accounted for. Having defined the generalised variance, it is now easy to present the underlying principle of an S-estimator of scatter. For the sake of clarity, we start with the univariate case before introducing the multivariate case.

Consider the minimal linear model $y_i = \mu + \varepsilon_i$. the objective of parameter estimation is to find the estimate $\hat{\mu}$ such as the predicted values $\hat{y}_i$ are as close as possible to the observed values y_i. The LS objective is to minimize the sum of squared residuals $\sum_{i=1}^{n}(y_i - \mu)^2$, or equivalently minimise the variance of the residuals $\frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2 = \sigma^2$. If we rewrite the last expression

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{\mu}}{\hat{\sigma}}\right)^2 = 1 \tag{11}$$

we can say that $\hat{\mu}$, our measure of location, is the estimate that minimises the measure of dispersion $\hat{\sigma}$ under the constraint that equality (11) holds. The value of $\hat{\mu}$ which satisfies this condition is the sample mean.

A drawback is that the squared distance criterion is very sensitive to outliers as it attributes a huge importance to large (absolute values of) y. Thus, to increase robustness, another objective function ρ_0 can be chosen, which is less sensitive to extreme values of y. [4] However, in that case, the estimated $\hat{\sigma}$ and $\hat{\mu}$ will no longer be the standard deviation and the sample mean when data are Gaussian. A solution is simply to modify equality (11) such that the problem is now to find the smallest robust scale of the residuals $\hat{\sigma}^S$

[4] Function $\rho(\cdot)$ is even, non decreasing for positive values, less increasing than the square with a unique minimum at zero.

satisfying

$$\frac{1}{n}\sum_{i=1}^{n}\rho_0(\frac{y_i-\hat{\mu}}{\hat{\sigma}^S})=\delta \tag{12}$$

where $\delta = E[\rho_0(u)]$ with $u \sim N(0,1)$. This modification guarantees that the estimated $\hat{\sigma}^S$ is coherent with the standard deviation for Gaussian data, at least in large samples. The value of μ that minimizes $\hat{\sigma}^S$ is called an S-estimator of location. More formally, an S-estimator of location is defined as:

$$\hat{\mu}^S = \arg\min_{\mu}\ \hat{\sigma}^S(y_1-\hat{\mu},...,y_n-\hat{\mu}) \tag{13}$$

where $\hat{\sigma}^S$ is the robust estimator of scale as defined in (12). If we consider model (1), instead of the minimal model, the logic remains unchanged and the S-estimator of regression becomes:

$$\hat{\theta}^S = \arg\min_{\theta}\ \hat{\sigma}^S(r_1(\theta),...,r_n(\theta)) \tag{14}$$

under the equality constraint

$$\frac{1}{n}\sum_{i=1}^{n}\rho_0(\frac{r_i(\theta)}{\hat{\sigma}^S})=\delta \tag{15}$$

The choice of $\rho_0(\cdot)$ is crucial to have good robustness properties and a high Gaussian efficiency. The Tukey Biweight function defined as

$$\rho_0(u)=\begin{cases} 1-\left[1-\left(\frac{u}{k}\right)^2\right]^3 & \text{if } |u|\le k \\ 1 & \text{if } |u|>k \end{cases} \tag{16}$$

with first derivative

$$\rho_0'(u) = \begin{cases} \frac{6}{k^6} u \left(u^2 - k^2\right)^2 \text{ if } |u| \leq k \\ 0 \qquad\qquad\qquad\; \text{if } |u| > k \end{cases} \tag{17}$$

is a common choice. The LS and Tukey Biweight objective functions are plotted in figure 1.

[Figure 1 about here]

The tuning parameter k is the number of robust dispersion estimate from the mean at which ρ_0' (the first derivative of ρ_0) becomes zero. To guarantee a resistance up to 50% of outliers, tuning constant can be set to 1.546. The Gaussian efficiency of such an estimator is however rather low (approximately 28%), meaning that the S estimator needs more than three times as many observations as the LS estimator to achieve the same variance when data are Gaussian. Increasing the value of the tuning constant would increase efficiency but reduce the percentage of contamination the estimator can withstand.

In multivariate analysis, a similar logic can be applied and an S-estimator of location and scatter can be estimated by finding $\hat{\mu}_S$, the multivariate location parameter, that minimises the $\det(\hat{\mathbf{\Sigma}}_S)$ (i.e. a unidimensional assessment of multivariate spread) subject to:

$$\frac{1}{n}\sum_{i=1}^{n} \rho_0(\sqrt{(\mathbf{x}_i - \hat{\mu}_S)\hat{\mathbf{\Sigma}}_S^{-1}(\mathbf{x}_i - \hat{\mu}_S)'}) = \delta \tag{18}$$

where $\rho(\cdot)$ is a loss function which is even, non decreasing for positive values and less increasing than the square function. The term $d_i = \sqrt{(\mathbf{x}_i - \hat{\mu}_S)\hat{\mathbf{\Sigma}}_S^{-1}(\mathbf{x}_i - \hat{\mu}_S)'}$ (called a Robust Mahalanobis distance) is a unidimensional assessment of the standardised distance of each observation from the center of the multivariate data cloud (i.e. the multivariate equivalent of $\frac{x_i - \hat{\mu}}{\hat{\sigma}^S}$). It is distributed as $\sqrt{\chi_p^2}$ for Gaussian data. As for the univariate case, the $\rho_0(\cdot)$ function considered here is the Tukey Biweight defined in (16). In the multivariate case, as proposed by Campbell et al. (1998), the value of tuning parameter k is chosen by specifying a cut-off constant as the number of robust dispersion estimate from the mean on the univariate scale at which ρ_0' becomes zero (i.e. 1.546) and then converting it to a value on the chi-squared scale of d_i^2 by using the Wilson and Hilferty (1931)'s transformation. The constant δ is taken as the expected value of $\rho_0(d_i)$ assuming a multivariate normal distribution (see Campbell et al, 1998).

The solution to this problem leads to a robust counterpart of the covariance matrix. It is then easy to robustly estimate $\mathbf{\Sigma_{XZ}}$, $\mathbf{\Sigma_{ZZ}}$ and $\mathbf{\Sigma_{Zy}}$ and replace these estimates in equation (8). The robust instrumental variable estimator can therefore be written as:

$$\hat{\theta}^S_{RIV} = \left(\mathbf{\Sigma}^S_{\mathbf{XZ}}\left(\mathbf{\Sigma}^S_{\mathbf{ZZ}}\right)^{-1}\mathbf{\Sigma}^S_{\mathbf{ZX}}\right)^{-1}\mathbf{\Sigma}^S_{\mathbf{XZ}}\left(\mathbf{\Sigma}^S_{\mathbf{ZZ}}\right)^{-1}\mathbf{\Sigma}^S_{\mathbf{Zy}} \tag{19}$$

An alternative estimator that would allow a substantial gain in efficiency is:

$$\hat{\theta}^W_{RIV} = \left(\mathbf{\Sigma}^W_{\mathbf{XZ}}\left(\mathbf{\Sigma}^W_{\mathbf{ZZ}}\right)^{-1}\mathbf{\Sigma}^W_{\mathbf{ZX}}\right)^{-1}\mathbf{\Sigma}^W_{\mathbf{XZ}}\left(\mathbf{\Sigma}^W_{\mathbf{ZZ}}\right)^{-1}\mathbf{\Sigma}^W_{\mathbf{Zy}} \tag{20}$$

where W stands for weights. The idea here is to estimate robust covariance $\mathbf{\Sigma_{XZy}}$ and calculate robust Mahalanobis distances. Relying on these, outliers are identified by looking at observations that have a d larger than $\sqrt{\chi^2_{m,0.99}}$. Observations that are associated with d_i larger than the cut-off point are downweighted and the classical (reweighted) covariance matrix is estimated. The weighting we adopt here is simply awarding a weight one for observations associated to a d smaller than a critical value and zero otherwise.

The advantage of this last estimator is that standard overidentification, underidentification and weak instruments tests can easily be obtained since this weighting scheme amounts to running a standard IV estimation on a sample free of outliers. The asymptotic variance of the estimator is also readily available. Furthermore, a substantial gain in efficiency with respect to the standard instrumental variable estimator proposed by Cohen-Freue and Zamar (2006) can be attained.[5]

As far as the estimator presented in (19) is concerned, we propose to improve on Cohen-Freue and Zamar (2006), by calculating an asymptotic variance that withstands heteroskedasticity and asymmetry. This is achieved by adapting the estimator proposed for the asymptotic variance of the S-estimator in Croux et al. (2003) to the case of a robust instrumental variable estimator. The main benefit of this approach is that it allows us, following Dehon et al. (2010), to implement a test to check if outliers distort classical instrumental variables estimations enough that robust methods are warranted.

[5]Unfortunately, it is not possible to know beforehand the reachable efficiency. However, in our simulations, the Gaussian efficiency is 97%.

1.2.2 Asymptotic variance

To calculate the asymptotic variance of the $\hat{\theta}^S_{RIV}$ estimator, we rely on the same logic of Croux et al. (2003). It can be shown that robust IV estimators are a special case of Method of Moments estimators for $\theta = (\theta^t, \sigma)^t$ with moment matrix $\mathbf{m}$ (for observation i)

$$\mathbf{m}_i(\theta) = \begin{pmatrix} \rho_0'(\frac{\mathbf{y}_i - \mathbf{x}_i^t\theta_0}{\sigma})\hat{\mathbf{x}}_i \\ \\ \rho_0(\frac{\mathbf{y}_i - \mathbf{x}_i^t\theta_0}{\sigma}) - \delta \end{pmatrix} = \begin{pmatrix} \rho_{0i}'\hat{\mathbf{x}}_i \\ \\ \rho_{0i} - \delta \end{pmatrix}$$

where $\rho_{0i} = \rho(\varepsilon_{0i})$, with $\varepsilon_{0i} = \frac{\mathbf{y}_i - \mathbf{x}_i^t\theta_0}{\sigma}$. The first line of $\mathbf{m}_i$ corresponds to the F.O.C. of the minimisation problem associated to the S-estimator while the second is related to the equality constraint. Note that as in the classical case, $\hat{\mathbf{x}}_i$ is used for the final estimator and not $\mathbf{x}_i$. Obviously, equation (19) guarantees that $\hat{\mathbf{x}}_i$ is robustly estimated. The estimated residuals $r_i = \mathbf{y}_i - \mathbf{x}_i^t\hat{\theta}_0$ are fitted, as in the classical case, relying on $\mathbf{x}_i^t$ rather than on $\hat{\mathbf{x}}_i^t$.

Following Hansen (1982), Croux et al. (2003) show that $\hat{\theta}$ has a limiting normal distribution given by

$$\sqrt{N}(\hat{\theta} - \theta) \longrightarrow N_p(\mathbf{0}, \mathbf{V})$$

where, defining the matrix of the derivatives $\mathbf{G}_S$ the matrix of the derivatives of $\mathbf{m}_i(\theta)$ with respect to θ (i.e.$\mathbf{G}_S = E\left[\frac{\partial \mathbf{m_i}(\theta)}{\partial \theta^{t\prime}}\right]$) and $\mathbf{\Omega}_S = E[\mathbf{m}_i(\theta)\mathbf{m}_i^t(\theta)]$, the asymptotic variance $\mathbf{V}$ is

$$\mathbf{V} = \mathbf{G}_S^t \mathbf{\Omega}_S^{-1} \mathbf{G}_S^{-1}$$

which, for the exactly identified case, is equivalent to

$$\mathbf{V} = \mathbf{G}_S^{-1}\boldsymbol{\Omega}_S(\mathbf{G}_S^t)^{-1} \tag{21}$$

Since $\boldsymbol{\Omega}_S = E\begin{pmatrix} (\rho'_{0i})^2\,\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^t & \rho'_{0i}\rho_{0i}\hat{\mathbf{x}}_i \\ \rho_{0i}\rho'_{0i}\hat{\mathbf{x}}_i^t & (\rho_{0i})^2 - \delta^2 \end{pmatrix}$

and $\mathbf{G}_S^{-1} =$

$$-\begin{pmatrix} \sigma[E(\rho''_{0i}\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^t)]^{-1} & -\sigma[E(\rho''_{0i}\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^t)]^{-1}E(\rho''_{0i}\hat{\mathbf{x}}_i\varepsilon_{0i})[E(\rho'_{0i}\varepsilon_{0i})]^{-1} \\ \mathbf{0} & \sigma[E(\rho'_{0i}\varepsilon_{0i})]^{-1} \end{pmatrix}$$

Defining $\mathbf{B} = \sigma[E(\rho''_{0i}\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^t)]^{-1}$ and $\mathbf{b} = \mathbf{B}E(\rho''_{0i}\hat{\mathbf{x}}_i\varepsilon_{0i})[E(\rho'_{0i}\varepsilon_{0i})]^{-1}$ and calling on (23) we have

$$G_S^{-1} = -\begin{pmatrix} \mathbf{B} & -\mathbf{b} \\ 0 & \sigma[E(\rho'_{0i}\varepsilon_{0i})]^{-1} \end{pmatrix}$$

And subsequently

$$Avar(\hat{\theta}^S) = \mathbf{B}E((\rho'_{0i})^2\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^t)\mathbf{B} - \mathbf{b}E(\rho'_{0i}\rho_{0i}\hat{\mathbf{x}}_i^t)\mathbf{B} - \mathbf{B}E(\rho'_{0i}\rho_{0i}\hat{\mathbf{x}}_i)\mathbf{b}^t + \mathbf{b}E((\rho_{0i})^2 - \delta^2)\mathbf{b}^t$$

This asymptotic variance is robust to heteroskedasticity and asymmetry. If we assume homoskedasticity (i.e. $\mathbf{x}_i$ and ε_{0i} are independent) and symmetry ($\mathbf{b}= 0$), the formula boils down to

$$Avar(\hat{\theta}^S) = \sigma^2 \frac{E\left((\rho'_{0i})^2\right)}{\left(E(\rho''_{0i})\right)^2} \left(E(\hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^t)\right)^{-1}$$

We however discourage the use of this simplified formula in practice since in addition to its being fragile to heteroskedasticity, it also lacks robustness with respect to outliers.

1.2.3 Testing for outliers

As previously mentioned, one of the drawbacks of S-estimators is their low Gaussian efficiency. To cope with this, Yohai (1987) introduced MM-estimators that combine a high resistance to outliers and a high efficiency. These estimators are two step estimators where the first step is a standard S-estimator as defined in (13) and the second step is

$$\hat{\theta}^{MM} = \arg\min_{\theta} \sum_{i=1}^{n} \rho\left(\frac{r_i(\theta)}{\hat{\sigma}^S}\right) \quad (22)$$

where the measure of scale is fixed at the value estimated by the S-estimator, $\hat{\sigma}^S$. The function $\rho(\cdot)$ (with first derivative $\psi(\cdot)$) is the same as for the S-estimator, except that the tuning parameter is set in such a way the Gaussian efficiency is higher. The preliminary S-estimator guarantees a high breakdown point, and the the final MM-estimate a high Gaussian efficiency. As illustrated by Croux et al. (2003) the MM-estimators are exactly identified Generalized Method of Moments estimators (GMM) for $\theta = (\theta^t, \sigma)^t$

with moment matrix $\mathbf{m}$ (for observation i) is

$$\mathbf{m}_i(\theta) = \begin{pmatrix} \psi(\frac{\mathbf{y}_i - \mathbf{x}_i^t\theta}{\sigma})\hat{\mathbf{x}}_i \\ \\ \rho_0'(\frac{\mathbf{y}_i - \mathbf{x}_i^t\theta_0}{\sigma})\hat{\mathbf{x}}_i \\ \\ \rho_0(\frac{\mathbf{y}_i - \mathbf{x}_i^t\theta_0}{\sigma}) - \delta \end{pmatrix} = \begin{pmatrix} \psi_i\hat{\mathbf{x}}_i \\ \\ \rho_{0i}'\hat{\mathbf{x}}_i \\ \\ \rho_{0i} - \delta \end{pmatrix}$$

where $\psi_i = \psi(\varepsilon_i)$, with $\varepsilon_i = \frac{\mathbf{y}_i - \mathbf{x}_i^t\theta}{\sigma}$. The last two lines correspond to the moment matrix of the S-estimator described above while the first line corresponds to the F.O.C. of the minimisation problem shown in (22).

Following Hansen (1982), Croux et al. (2003) show that $\hat{\theta}$ has a limiting normal distribution given by

$$\sqrt{N}(\hat{\theta} - \theta) \longrightarrow N_p(0, \mathbf{V})$$

where, defining $\mathbf{G}_{MM}$ the matrix of the derivatives of $\mathbf{m}_i(\theta)$ with respect to theta (i.e.$\mathbf{G}_{MM}= E\left[\frac{\partial \mathbf{m_i}(\theta)}{\partial \theta^{t\prime}}\right]$) and $\mathbf{\Omega}_{MM}= E[\mathbf{m}_i(\theta)\mathbf{m}_i^t(\theta)]$, the asymptotic variance $\mathbf{V}$ is

$$\mathbf{V} = (\mathbf{G}_{MM}^t \mathbf{\Omega}_{MM}^{-1} \mathbf{G}_{MM})^{-1}$$

which, for the exactly identified case, is equivalent to

$$\mathbf{V} = \mathbf{G}_{MM}^{-1} \mathbf{\Omega}_{MM} (\mathbf{G}_{MM}^t)^{-1} \tag{23}$$

Since $\boldsymbol{\Omega}_{MM} = E \begin{pmatrix} \psi_i^2 \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^t & \psi_i \rho'_{0i} \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^t & \psi_i \rho_{0i} \hat{\mathbf{x}}_i \\ \rho'_{0i} \psi_i \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^t & (\rho'_{0i})^2 \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^t & \rho'_{0i} \rho_{0i} \hat{\mathbf{x}}_i \\ \rho_{0i} \psi_i \hat{\mathbf{x}}_i^t & \rho_{0i} \rho'_{0i} \hat{\mathbf{x}}_i^t & (\rho_{0i})^2 - \delta^2 \end{pmatrix}$

and $\mathbf{G}_{MM}^{-1} =$

$$- \begin{pmatrix} \sigma[E(\psi_i^t \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^t)]^{-1} & \mathbf{0} & -\sigma[E(\psi'_i \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^t)]^{-1} E(\psi'_i \hat{\mathbf{x}}_i \varepsilon_i)[E(\rho'_{0i} \varepsilon_{0i})]^{-1} \\ \mathbf{0} & \sigma[E(\rho''_{0i} \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^t)]^{-1} & -\sigma[E(\rho''_{0i} \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^t)]^{-1} E(\rho''_{0i} \hat{\mathbf{x}}_i \varepsilon_{0i})[E(\rho'_{0i} \varepsilon_{0i})]^{-1} \\ \mathbf{0} & \mathbf{0} & \sigma[E(\rho'_{0i} \varepsilon_{0i})]^{-1} \end{pmatrix}$$

$\mathbf{V} = \mathbf{G}_{MM}^{-1} \boldsymbol{\Omega}_{MM} (\mathbf{G}_{MM}^t)^{-1}$

Defining $\mathbf{A} = \sigma[E(\psi'_i \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^t)]^{-1}; \mathbf{a} = \mathbf{A} E(\psi'_i \hat{\mathbf{x}}_i \varepsilon_i)[E(\rho'_{0i} \varepsilon_{0i})]^{-1}; \mathbf{B} = \sigma[E(\rho''_{0i} \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^t)]^{-1}$ and $\mathbf{b} = \mathbf{B} E(\rho''_{0i} \hat{\mathbf{x}}_i \varepsilon_{0i})[E(\rho'_{0i} \varepsilon_{0i})]^{-1}$ and calling on (23) we have

$$G^{-1} = - \begin{pmatrix} \mathbf{A} & \mathbf{0} & -\mathbf{a} \\ \mathbf{0} & \mathbf{B} & -\mathbf{b} \\ 0 & 0 & \sigma[E(\rho'_{0i} \varepsilon_{0i})]^{-1} \end{pmatrix}$$

And subsequently

$$Avar(\hat{\theta}^{MM}) = \mathbf{A} E(\psi_i^2 \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^t) \mathbf{A} - \mathbf{a} E(\psi_i \rho_{0i} \hat{\mathbf{x}}_i^t) \mathbf{A} - \mathbf{A} E(\psi_i \hat{\mathbf{x}}_i \rho_{0i}) \mathbf{a}^t + \mathbf{a} E((\rho_{0i})^2 - \delta^2) \mathbf{a}^t$$

$$Avar(\hat{\theta}^S) = \mathbf{B}E((\rho'_{0i})^2\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^t)\mathbf{B} - \mathbf{b}E(\rho'_{0i}\rho_{0i}\hat{\mathbf{x}}_i^t)\mathbf{B} - \mathbf{B}E(\rho'_{0i}\rho_{0i}\hat{\mathbf{x}}_i)\mathbf{b}^t + \mathbf{b}E((\rho_{0i})^2 - \delta^2)\mathbf{b}^t$$

$$Acov(\hat{\theta}^{MM}, \hat{\theta}^S) = \mathbf{A}E(\psi_i\rho'_{0i}\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^t)\mathbf{B} - \mathbf{a}E(\rho'_{0i}\rho_{0i}\hat{\mathbf{x}}_i^t)\mathbf{B} - \mathbf{A}E(\psi_i\hat{\mathbf{x}}_i\rho_{0i})\mathbf{b}^t + \mathbf{a}E((\rho_{0i})^2 - \delta^2)\mathbf{b}^t$$

Having the asymptotic variance of $\hat{\theta}^S$, $\hat{\theta}^{MM}$ and the covariance between the two, it is now possible, following Dehon et al. (2010), to call on a test procedure that balances robustness against efficiency. The underlying idea is to compare an estimator which is very robust but highly inefficient ($\hat{\theta}^S$) with an estimator that is potentially less robust but more efficient ($\hat{\theta}^{MM}$). On the one hand, if the difference between the S estimate and MM estimate is small, it would be preferable to use the MM-estimator given its higher efficiency. On the other hand, if the difference between the two estimates becomes too large, the gain in efficiency is more than balanced by a loss in robustness, and it would be better to use the more robust estimator.

The probably most appropriate testing procedure to reach this aim is the generalised Hausman test defined as

$$W = (\hat{\theta}^{MM} - \hat{\theta}^S)[Var(\hat{\theta}^{MM}) + Var(\hat{\theta}^S) - 2Cov(\hat{\theta}^{MM}, \hat{\theta}^S)]^{-1}(\hat{\theta}^{MM} - \hat{\theta}^S)^t \quad (24)$$

Bearing in mind that this statistic is asymptotically distributed as a χ^2_p, where p is the number of covariates, it is possible to set an upper bound above which the estimated parameters can be considered as statistically different. Stated differently, if W is larger than $\chi^2_{p,(1-\alpha)}$ the difference between $\hat{\theta}^{MM}$ and $\hat{\theta}^{S}$ is too large, and the gain in efficiency cannot compensate the loss in robustness.

In our case, the robust but not efficient estimator will be the S-estimator defined in (19). For its efficient but not robust counterpart, we rely on the fact that LS is nothing else than a special case of the MM-estimator when k goes to infinity, since if $k \longrightarrow \infty$, $\rho(u) = \frac{u^2}{2}$, $\psi(u) = u$, $\psi'(u) = 1$ and MM boils down to LS. Hence, we will test for the presence of outliers in the dataset by contrasting the S-estimator with the LS estimator of the second stage of the IV estimator, as defined in (5), using the appropriate asymptotic variance in (9).

2 Monte-Carlo simulations

2.1 Behaviour of the RIV estimator

We use a setup that is similar to that of Cohen-Freue and Zamar (2006). We first generate 1000 observations for 5 random variables (x, u, v, w, z) drawn from a multivariate normal distribution with mean $\mu = (0, 0, 0, 0, 0)$ and covariance

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & 0 & 0 & 0.5 & 0 \\ 0 & 0.3 & 0.2 & 0 & 0 \\ 0 & 0.2 & 0.3 & 0 & 0 \\ 0.5 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

or equivalently, with correlation matrix

$$P = \begin{pmatrix} 1 & 0 & 0 & 0.5 & 0 \\ 0 & 1 & 0.67 & 0 & 0 \\ 0 & 0.67 & 1 & 0 & 0 \\ 0.5 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

We then consider the following data generating process (DGP) for $\mathbf{y}$:

$\mathbf{y} = \mathbf{1} + \mathbf{2x} + \mathbf{z} + \mathbf{u}$. We assume that $\mathbf{x}$ is measured with error and that only variable $\mathbf{X}$, that is generated as $\mathbf{X} = \mathbf{x} + \mathbf{v}$, is observed. Given that the correlation coefficient between $\mathbf{u}$ and $\mathbf{v}$ is about 0.7, we would obtain biased and inconsistent estimators if we simply regressed $\mathbf{Y}$ on $\mathbf{X}$ and $\mathbf{z}$, since $\mathbf{X}$ and $\mathbf{u}$ are not independent. We therefore have to use the instrumental variable estimator, exploiting the instrumental variable $\mathbf{w}$. The latter satisfies the two conditions required for a variable to be a good instrument: it is relevant, as it is correlated at 0.5 with $\mathbf{X}$, and it is independent of the error term $\mathbf{u}$. We reproduce this setup 1000 times under different contamination scenarii. In the 'mild' scenarii, we contaminate alternatively 5% of the observations of $\mathbf{x}$, $\mathbf{w}$, $\mathbf{z}$ or $\mathbf{y}$ by the value 5. In the 'heavy' scenarii, we contaminate alternatively 10% of the observations of $\mathbf{x}$, $\mathbf{w}$, $\mathbf{z}$ or $\mathbf{y}$ by the value 10. Finally we consider a setup with no contamination to simulate the efficiency of the RIV estimator relative to that of the classical IV estimator.

[Tables 1-3 about here]

It can be seen in tables 1-2 that the classical IV estimator is very sensitive to the contamination of the sample by outliers, even in the mild case. On the other hand, the RIV estimator is extremely stable, having a very low mean squared error whatever the scenario tested. Interestingly, outliers in the instrument only strongly influence the results of the classical IV estimator when heavy contamination is present. Intuitively, that is because outliers in the instrument (in our simulation setup) will never generate extreme values in the fitted value of the troublesome variable (X), as would be the case for direct contamination. Finally, table 3 shows that, in the absence of contamination, the classical and robust IV estimators have very similar performances. The variance of the RIV estimator at the normal is only 3% larger than that of the classical IV estimator.

2.2 Properties of the Hausman test for outliers

We now turn to the properties of the generalised Hausman test for outliers.

We first investigate the size of the test. We reproduce the sample 1000 times under the null, i.e. we do not generate any outliers, and calculate the percentage of rejection, with a degree of confidence of 95%. The estimated size of the test is 4.9%, very close to the nominal test size of 5%.

[Figure 2 about here]

We then simulate the "power" of the test under some specific circumstances. We consider 4 contamination scenarii where 5%, 10%, 15% and 20% of the observations in $\mathbf{x}$, $\mathbf{w}$, $\mathbf{z}$ or $\mathbf{y}$ successively become outliers. For variables $\mathbf{z}$,and $\mathbf{x}$ the outliers are sequentially generated from a Normal distribution with unit variance and centered in values ranging from 0 to 3, with an increment of 0.1. For each of these contaminations we reproduce the sample 100 times and calculate the percentage of rejection. For variable $\mathbf{y}$ the outliers are generated from a Normal distribution with unit variance and centered in values ranging from 0 to 5, while for $\mathbf{w}$, the outliers are generated from a Normal distribution with unit variance and centered in values ranging from 0 to 30. As for $\mathbf{x}$ and $\mathbf{z}$, we reproduce the sample 100 times and calculate the percentage of rejection. The simulation results will give us an idea of the power of the test according to different setups.

Figure 2 suggests that the test is fairly powerful. Indeed, when the distance with respect to the null increases (horizontal axis), the percentage of rejection of the null (vertical axis) increases rapidly for contamination in $\mathbf{z}$, $\mathbf{x}$ and, though slightly less so, in $\mathbf{y}$. Furthermore this high rejection occurs faster when the percentage of contaminated observations increases. For variable $\mathbf{w}$, the effect of outliers on the rejection of the null is smaller, for the same reasons as given above; indirect contamination of the second stage has less impact on the estimation of the second stage than direct contamination of the variables involved in the second stage.[6]

3 The effects of FDI on productivity: empirical model and data

Having developed the necessary tools to take into account potential outliers in our variables, we investigate the impact of FDI on productivity.

We assume that output Y in country i at time t is produced according to a Cobb-

[6]Computer code and programs are available upon request to the authors.

Douglas production function, using physical capital K, and effective human capital-augmented labour AH. The latter is related to schooling, such as $H_{it} = e^{\phi S_{it}} L_{it}$, and A is a labor-augmenting measure of productivity.

$$\begin{aligned} Y_{it} &= K_{it}^{\alpha}(A_{it}H_{it})^{1-\alpha} \\ \frac{Y_{it}}{L_{it}} &= A_{it}(\frac{K_{it}}{Y_{it}})^{\frac{\alpha}{1-\alpha}}\frac{H_{it}}{L_{it}} \\ Ln(\frac{Y_{it}}{L_{it}}) &= Ln(A_{it}) + \frac{\alpha}{1-\alpha}Ln(\frac{K_{it}}{Y_{it}}) + \phi S_{it} \end{aligned} \quad (25)$$

Following Hall and Jones (1999), we have rearranged the production function so that the capital-output ratio appears on the right-hand side of equation. Expressing output per worker in this way ensures that all the long-run effects of an increase in education or productivity are attributed to these variables.

In line with previous research on the determinants of productivity (Hall and Jones, 1999; Rodrik et al., 2004; Kose et al., 2009), the equilibrium value of productivity is expected to depend on institutional quality ($INST$), the FDI stock to GDP ratio ($\frac{FDI}{GDP}$), time-invariant country-specific factors (C_i) and country-invariant time effects (T_t)[7]:

$$Ln(A_{it}^{*}) = \beta_1 INST + \beta_2 \frac{FDI}{GDP}_{it} + C_i + T_t + \epsilon_{it} \quad (26)$$

where ϵ_{it} is a not serially correlated error term.

[7] In unreported regressions, we allowed for country-specific time trends. Results were qualitatively unchanged.

Adjustment of the actual value of productivity to its equilibrium value is not instantaneous:

$$\frac{dLn(A_{it})}{dt} = \lambda[Ln(A^*_{it}) - Ln(A_t)]$$

Solving this first-order difference equation in $Ln(A_{it})$ and plugging equation 27 in 26 we obtain a partial adjustment model, in which the current value of productivity depends on its past value and the determinants of its (time-changing) equilibrium value:

$$Ln(A_{it}) = e^{-\lambda\tau}Ln(A_{it-1}) + \theta INST_{it} + \gamma\frac{FDI}{GDP}_{it} + (1 - e^{-\lambda\tau})T_t + \varepsilon_{it} \quad (27)$$

where $\tau = t - t_1$, $\theta = (1 - e^{-\lambda\tau})\beta_1$, $\gamma = (1 - e^{-\lambda\tau})\beta_2$, and $\varepsilon_{it} = (1 - e^{-\lambda\tau})C_i + (1 - e^{-\lambda\tau})\epsilon_{it}$.

Using this expression for productivity in equation 25 we get:

$$Ln(\frac{Y_{it}}{L_{it}}) = e^{-\lambda\tau}Ln(A_{it-1}) + \theta INST_{it} + \gamma\frac{FDI}{GDP}_{it} + \frac{\alpha}{1-\alpha}Ln(\frac{K_{it}}{Y_{it}}) + \phi S_{it} + (1 - e^{-\lambda\tau})T_t + \varepsilon_{it}$$

$$Ln(\frac{Y_{it}}{L_{it}}) - [\frac{\alpha}{1-\alpha}Ln(\frac{K_{it}}{Y_{it}}) + \phi S_{it} + e^{-\lambda\tau}Ln(A_{it-1}) + \theta INST_{it}] = \gamma\frac{FDI}{GDP}_{it} + (1 - e^{-\lambda\tau})T_t + \varepsilon_{it} \quad (28)$$

We are solely interested in γ, the coefficient on the FDI to GDP ratio. Hence, we constrain the coefficients in the brackets to values that are commonly used in the literature.[8] Following Gollin (2002) and Psacharopoulos and Patrinos (2004), we assume that α, the

[8]Results are not qualitatively sensitive to changes in the values that we have used.

physical capital's share in the production, is equal to $\frac{1}{3}$, and that, ϕ the return to one extra year of education, is 10%. The estimates of Bernard and Jones (1996) and Kose et al. (2009) suggest that a reasonable estimate for the speed of productivity convergence is 4% per year. Given that we use a five-year period panel, $e^{-0.04*5} \simeq 0.82$. Finally, we do not have a readily available estimate for the long-run effect of institutional quality on productivity. Our approach is to report our results using a range of values for β_1, such as institutional quality typically explains between 50 and 90% of the productivity gap between the countries whose productivity values are above the upper quartile and the countries whose productivity values are below the lower quartile in 2005.

Data on income and labour force come from Heston et al. (2009).[9] The capital stock is calculated using the perpetual inventory method,[10] while data on schooling come from Barro and Lee (2010), and correspond to the average years of total schooling for the population aged 15 and over. Our measure of FDI is the ratio of financial FDI stock (liabilities) to GDP, which come from Lane and Milesi-Ferretti (2007), and is fairly standard in the literature (see for instance Carkovic and Levine (2005) Prasad et al. (2007)). We use stocks in order to capture the cumulative effects of foreign presence (Bitzer and Görg, 2009). This financial measure cannot be expected to provide a perfect picture of foreign presence in a given country, but it is well correlated ($r \simeq 0.70$) with indicators of real foreign

[9]In unreported regressions, we tried to adjust the labour force for unemployment. Results are qualitatively very similar to our main results.

[10]Capital is assumed to be accumulated according to the following equation of motion $K_{it} = I_{it} + (1 - \delta)K_{it-1}$, where a depreciation rate δ of 6% is chosen. The initial capital stock is calculated on the basis of the expression for the steady-state capital stock in the Solow model: $K_0 = \frac{I_0}{g+\delta}$, where g is the average geometric growth rate for the investment series between the first year with available data and the tenth year with available data. In order to minimise the impact of those assumptions on the initial capital stock, data on estimated capital stocks are discarded as long as twenty years from he first year with available data have not elapsed.

activity such as the the total numbers of majority-owned foreign affiliates, as reported by the UNCTAD on the *Investment Map* website.[11] The institutional quality measure comes from Teorell et al. (2010) and corresponds to the mean value of the ICRG (International Country risk Guide-PRS Group) variables "Corruption", "Law and Order" and "Bureaucracy Quality", scaled 0-1. Higher values indicate higher quality of government.

Our panel consists of data for 106 countries over the period 1970-2005. Following Caselli et al. (1996), we use a five-year period panel ($\tau = 5$), such as the productivity values are five years apart and the values for $INST$ and $\frac{FDI}{GDP}$ have been averaged over non-overlapping five-year periods (1970-1974...2000-2004). Potential endogeneity of the $\frac{FDI}{GDP}$ ratio warrants an IV approach. In line with recent panel literature, we use as internal instruments the lagged values of the troublesome variable. More specifically, given the strong persistence of the FDI series,[12] we instrument the level values of the $\frac{FDI}{GDP}$ ratio, with its once- or twice-lagged differences. Lagged differences are valid instruments under the assumptions that (1) there is no correlation between the differences of these variables and the country-specific effects and, (2) the idiosyncratic part of the error term is not serially correlated (Blundell and Bond, 1998). These hypotheses can be tested through an Arellano and Bond (1991) test of serial correlation of the differenced error term and a Hansen (1982) test of over-identifying restrictions. Given its good properties in terms of bias and coverage rate (Angrist and Jörn-Steffen, 2009), we focus on the results obtained

[11] `http://www.investmentmap.org/invmap/index.aspx?prg=1`

[12] In a simple AR(1) model estimated by OLS, the coefficient on the lagged FDI stock is slightly larger than 0.80. Blundell and Bond (1998) show that lagged levels are weak instruments for subsequent first-differences when the autoregressive coefficient is high. They suggest using instead suitably lagged differences as instruments for the equations in levels. Indeed, Blundell et al. (2001)'s Monte-Carlo simulations indicate that a 'levels-GMM' estimator performs well when the series are highly persistent.

via a just-identified IV estimator, using the once-lagged difference of $\frac{FDI}{GDP}$ as an instrument for the latter. The test of overidentifying restrictions is computed using once- and twice-lagged differences of $\frac{FDI}{GDP}$ as instruments.

4 The effects of FDI on productivity: empirical results

Figure 3 reports the IV estimate of γ, the short-run coefficient on $\frac{FDI}{GDP}$ for a range of a range of values for β_1, the coefficient on $INST$. FDI appears to have a positive and statistically significant impact on productivity, whichever the value assumed for β_1. For $\gamma \simeq 0.15$, $\beta_2 = \frac{\gamma}{(1-e^{-\lambda\tau})} \simeq 0.83$, suggesting that a 10 percentage points rise in the FDI to GDP ratio would increase productivity in the long-run by about 8%. This finding is in line with the recent literature investigating the impact of international financial flows on productivity. For instance, Kose et al. (2009) find in table 4 of their paper that a 10 percentage point increase in the ratio of FDI and equity liabilities to GDP would increase productivity in the long-run by about $(\frac{0.00379}{0.40691} * 10) * 100 \simeq 9\%$. Such an effect is not dramatic but is nevertheless equivalent to increasing average years of schooling by 1 year. Figures 5 to 7 suggest that the once-lagged difference of $\frac{FDI}{GDP}$ is a strong and valid instrument.

In the absence of concerns relating to outliers, we would conclude that openness to FDI is likely to enhance productivity in the recipient countries. However, figure 4, which reports the RIV estimate of γ, tells a very different story.[13] If we assume that differences in institutional quality explain less than 65% of differences in productivity, the impact of

[13]Figures 5 to 7 suggest that the once-lagged difference of $\frac{FDI}{GDP}$ remains a strong and valid instrument, despite the omission of outliers.

FDI is negative and statistically insignificant. On the other hand, once this productivity threshold is exceeded, the impact of FDI remains negative but is substantially larger and statistically significant. For 75% of the productivity gap explained, a 10 percentage points rise in the FDI to GDP ratio would decrease productivity in the long-run by about 37%. A plausible explanation for this finding is that any positive FDI-related effects are out-weighted by a 'market-stealing- effect', in the sense that the entry of foreign competitors causes less-competitive domestic producers to cut production to such an extent that they experience an overall productivity decline (Aitken and Harrison, 1999).

[Figures 3-8 about here]

Figure 8 shows that our Hausman test always rejects the absence of outliers in the sample. Identification of the outliers can be achieved by plotting robust Mahalanobis distances against residuals of the second stage standardised by a robust estimate of their standard deviation.[14] Outliers are usually classified as vertical outliers, good/horizontal outliers (good leverage points) or bad outliers (bad leverage points). A vertical outlier is an observation outlying in the vertical dimension only; in the context of our empirical application, that means that the predicted value of the productivity is very different from the actual value. Its presence mostly affects the value of the intercept parameter by shifting the regression line upwards or downwards, even though it can also affect the slope estimates. A good outlier is an observation outlying in the horizontal dimension only; the

[14]We exploit the residuals of the second stage, with the underlying assumption that our methodology guarantees that the estimates are robust to outliers in the first stage. As a corollary, it is important to note that it would not enough to check for outliers in the second stage, as this would not guarantee robustness to outliers in both stages.

FDI to GDP ratio is very different from the rest of the observations. It has little effect on the estimated coefficient since it lies in the continuity of the regression line. Finally, a bad outlier is outlying both in the vertical and horizontal dimensions; the predicted value of the productivity is very different from the actual value **and** the FDI to GDP ratio is very different from the rest of the observations. The presence of this type of outliers in the data is considered particularly harmful, as their remoteness from the rest of the data strongly influences the slope estimates, given the attempt of the 2SLS estimator to minimise both at the first and second stages the squared distance between these observations and the regression line. In figure 9, we facilitate the identification of each type of outliers by setting vertical and horizontal cut-off points. The vertical cut-off points are 2.25 and -2.25. Assuming that the data are Gaussian, residuals are normally distributed, and values above or below these cut-off points are strongly atypical since they are 2.25 standard deviations away from the mean, with a probability of occurrence of 0.025. In line with our downweighting scheme, the horizontal cut-off point is $\sqrt{\chi^2_{p,0.99}}$. Vertical outliers are in Section (S) 1, good outliers are in S3 and bad outliers are in S2.

[Figures 9-11 about here]

It is obvious that Liberia (LBR) and Luxembourg (LUX) are excessively bad outliers. That is not surprising given that their FDI stocks are 5 and 22 times greater than their GDP respectively. Once the observations related to these two countries are omitted (figure 10) a large number of outliers remain, even though the bulk of them appear to be good outliers. Figure 11 illustrates the influence of each kind of outlier, by reporting the estimate of γ

and its associated confidence interval once each category is removed. As expected, the omission of vertical and good outliers has little impact on the IV estimates. On the other hand, eliminating bad outliers shows how the classical IV estimator can totally breakdown in the presence of bad outliers and generates very misleading results.

Studies investigating the impact of FDI on economic growth frequently argue that the existence and diffusion of positive productivity spillovers ought to be conditional on a country's absorptive capacity, as measured by its level of income per worker, human capital, trade openness, financial development or institutional quality (Carkovic and Levine, 2005). We investigate this possibility by looking at the robust effect of FDI on productivity in samples of countries for which a given measure of absorptive capacity is above the sample median in period 2000-2004. Our measures of income per worker, human capital and institutional quality have already been defined. Trade openness corresponds to the trade openness ratio $\frac{(X+M)}{GDP}$, as reported in Heston et al. (2009). Finally, financial development is the value of credits by financial intermediaries to the private sector divided by GDP; this variable can be found in the updated database of (Beck et al., 2000).

Figures 12 to 16 suggest that absorptive capacity may indeed mediate the effects of FDI on productivity. In comparison to our previous results, we never find a scenario in which higher FDI would lead to a fall in productivity. However, we can also never reject the null hypothesis that FDI has no impact on productivity. These disappointing findings may reflect a rough balance between the negative and positive foreign spillovers in countries well-endowed enough to profit from them.

[Figures 12-16 about here]

5 Conclusion

The application of a robust instrumental variable (RIV) approach to investigate the impact of foreign direct investment (FDI) on productivity in a large panel of countries has allowed us to demonstrate that outliers need to be taken seriously. We find that the positive and statistically significant impact of FDI on productivity suggested by the classical IV estimator is an artefact stemming from the presence of several atypical observations in the sample. Once the influence of the latter is downweighted, there is little macroeconomic evidence that suggests that FDI fosters productivity growth in recipient countries, even those with high absorptive capacity. Hence, the more optimistic results of previous studies should be treated with caution. These earlier results may not be robust to the presence of outliers in their data. Fortunately, our RIV estimator, and its associated test for outliers, will allow future research resorting to IV estimations to control for outliers in a simple and systematic way.

References

Aitken B, Harrison A. 1999. Do domestic firms benefit from direct foreign investment? evidence from venezuela. *American Economic Review* **89**: 605–618.

Angrist JD, Jörn-Steffen P. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Arellano M, Bond S. 1991. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies* **58**: 277–97.

Barro RJ, Lee JW. 2010. A new data set of educational attainment in the world, 1950-2010. NBER Working Paper, n°15902.

Beck T, Demirgüç-Kunt A, Levine R. 2000. A new database on the structure and development of the financial sector. *World Bank Economic Review* **14**: 597–605.

Bernard AB, Jones CI. 1996. Productivity across industries and countries: Time series theory and evidence. *The Review of Economics and Statistics* **78**: 135–46.

Bitzer J, Görg H. 2009. Foreign direct investment, competition and industry performance. *The World Economy* **32**: 221–233.

Blundell R, Bond S. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* **87**: 115–143.

Blundell R, Bond S, Windmeijer F. 2001. Estimation in dynamic panel data models: Improving on the performance of the standard gmm estimator. In Baltagi BH (ed.) *Advances in Econometrics: Nonstationary Panels, Panel Cointegration and Dynamic Panels*.

Campbell NA, Lopuhaä HP, Rousseeuw PJ. 1998. On the calculation of a robust s-estimator of a covariance matrix. *Statistics in Medicine* **17**: 2685–2695.

Carkovic M, Levine R. 2005. Does foreign direct investment accelerate economic growth? In Moran TH, Graham EM, Blomstrom M (eds.) *Does Foreign Direct Investment Promote Development?* Washington , DC : Institute for International Economics.

Caselli F, Esquivel G, Lefort F. 1996. Reopening the convergence debate: A new look at cross-country growth empirics. *Journal of Economic Growth* **1**: 363–89.

Cohen-Freue G, Zamar R. 2006. A robust instrumental variables estimator. Mimeo.

Croux C, Dhaene G, Hoorelbeke D. 2003. Robust standard errors for robust estimators. K.U. Leuven Discussion Papers Series 03.16.

Dehon C, Gassner M, Verardi V. 2010. Testing for the presence of outliers in a cross-section regression. Mimeo, ULB.

Gollin D. 2002. Getting income shares right. *Journal of Political Economy* **110**: 458–474.

Hadi AS. 1994. A modification of a method for the detection of outliers in multivariate samples. *Journal of the Royal Statistical Society. Series B (Methodological)* **56**: pp. 393–396.

Hall R, Jones C. 1999. Why do some countries produce so much more output per worker than others? *The Quarterly Journal of Economics* **114**: 83–116.

Hansen L. 1982. Large sample properties of generalizes method of moments estimators. *Econometrica* **50**: 1029–1054.

Heston A, Summers R, Aten B. 2009. *Penn World Table Version 6.3*. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Kemeny T. 2010. Does foreign direct investment drive technological upgrading? *World Development* **38**: 1543–1554.

Kose AM, Prasad ES, Terrones ME. 2009. Does openness to international financial flows raise productivity growth? *Journal of International Money and Finance* **28**: 554–580.

Lane PR, Milesi-Ferretti GM. 2007. The external wealth of nations mark ii: Revised and extended estimates of foreign assets and liabilities, 1970-2004. *Journal of International Economics* **73**: 223–250.

OECD. 2002. *Foreign Direct Investment for Development*. Paris: OECD.

Prasad E, Rogoff K, Wei SJ, Kose AM. 2003. Effects of financial globalisation on developing countries: Some empirical evidence. *Economic and Political Weekly* **38**: 4319 – 4330.

Prasad ES, Rajan RG, Subramanian A. 2007. Foreign capital and economic growth. *Brookings Papers on Economic Activity* **38**: 153–230.

Psacharopoulos G, Patrinos HA. 2004. Returns to investment in education: A further update. *Education Economics* **12**: 111–134.

Rodrik D, Subramanian A, Trebbi A. 2004. Institutions rule: The primacy of institutions over geography and integration in economic development. *Journal of Economic Growth* **9**: 131–165.

Teorell J, Charron N, Samanni M, Holmberg S, Rothstein B. 2010. *The Quality of Government Dataset*. University of Gothenburg: The Quality of Government Institute.

Vua TB, Noyb I. 2009. Sectoral analysis of foreign direct investment and growth in the developed countries. *Journal of International Financial Markets, Institutions and Money* **19**: 402 – 413.

Wilks S. 1932. Certain generalizations in the analysis of variance. *Biometrika* **24**: 471–494.

Wilson EB, Hilferty MM. 1931. The distribution of chi-square. *Proceedings of the National Academy of Sciences of the United States of America* **17**: pp. 684–688.

Wooster RB, Diebel DS. 2010. Productivity spillovers from foreign direct investment in developing countries: A meta-regression analysis. *Review of Development Economics* **14**: 640–655.

Yohai VJ. 1987. High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics* **15**: pp. 642–656.

Table 1: Bias, Variance and MSE for a 5% contamination of the sample

	Classical			Robust		
	X	Z	_cons	X	Z	_cons
Contamination of: x						
Bias	-5.7440	-0.0017	-1.1284	-0.0004	0.0005	-0.0014
Variance	0.0754	0.0146	0.0230	0.0035	0.0009	0.0008
MSE	33.0690	0.0146	1.2963	0.0035	0.0009	0.0008
Contamination of: w						
Bias	0.0067	0.0000	-0.0003	0.0018	0.0006	-0.0002
Variance	0.0126	0.0008	0.0007	0.0036	0.0009	0.0008
MSE	0.0126	0.0008	0.0007	0.0036	0.0009	0.0008
Contamination of: z						
Bias	0.0006	-0.7226	-0.1382	0.0016	-0.0018	-0.0011
Variance	0.0060	0.0006	0.0013	0.0034	0.0009	0.0008
MSE	0.0060	0.5227	0.0204	0.0034	0.0009	0.0008
Contamination of: y						
Bias	-0.2000	-0.0984	0.4001	0.0189	0.0093	0.0264
Variance	0.0103	0.0026	0.0011	0.0037	0.0009	0.0009
MSE	0.0503	0.0123	0.1612	0.0041	0.0010	0.0016

Table 2: Bias, Variance and MSE for a 10% contamination of the sample

	Classical			Robust		
	X	Z	_cons	X	Z	_cons
Contamination of: x						
Bias	-5.6434	-0.0109	-2.3564	0.0019	-0.0006	-0.0038
Variance	0.4435	0.0600	0.4664	0.0039	0.0010	0.0093
MSE	32.2914	0.0601	6.0188	0.0039	0.0010	0.0093
Contamination of: w						
Bias	0.0462	-0.0013	-0.0003	0.0018	-0.0006	0.0003
Variance	0.2551	0.0014	0.0012	0.0035	0.0009	0.0008
MSE	0.2572	0.0014	0.0012	0.0035	0.0009	0.0008
Contamination of: z						
Bias	0.0023	-0.9095	-0.0890	0.0018	-0.0008	0.0003
Variance	0.0064	0.0002	0.0017	0.0035	0.0009	0.0008
MSE	0.0064	0.8273	0.0097	0.0035	0.0009	0.0008
Contamination of: y						
Bias	-0.1981	-0.0974	0.8989	0.0019	-0.0005	0.0004
Variance	0.0368	0.0088	0.0012	0.0035	0.0009	0.0008
MSE	0.0761	0.0183	0.8093	0.0035	0.0009	0.0008

Table 3: Bias, Variance and MSE when there is no contamination of the sample

	Classical			Robust		
	X	Z	_cons	X	Z	_cons
Bias	6.91E-05	0.001945	-0.00015	-0.0002	0.001675	-0.00039
Variance	0.002875	0.000693	0.000704	0.003304	0.00078	0.000728
MSE	0.0029	0.0007	0.0007	0.0033	0.0008	0.0007

Figure 1: Sensitivity of the LS and Tukey Biweight functions to outliers

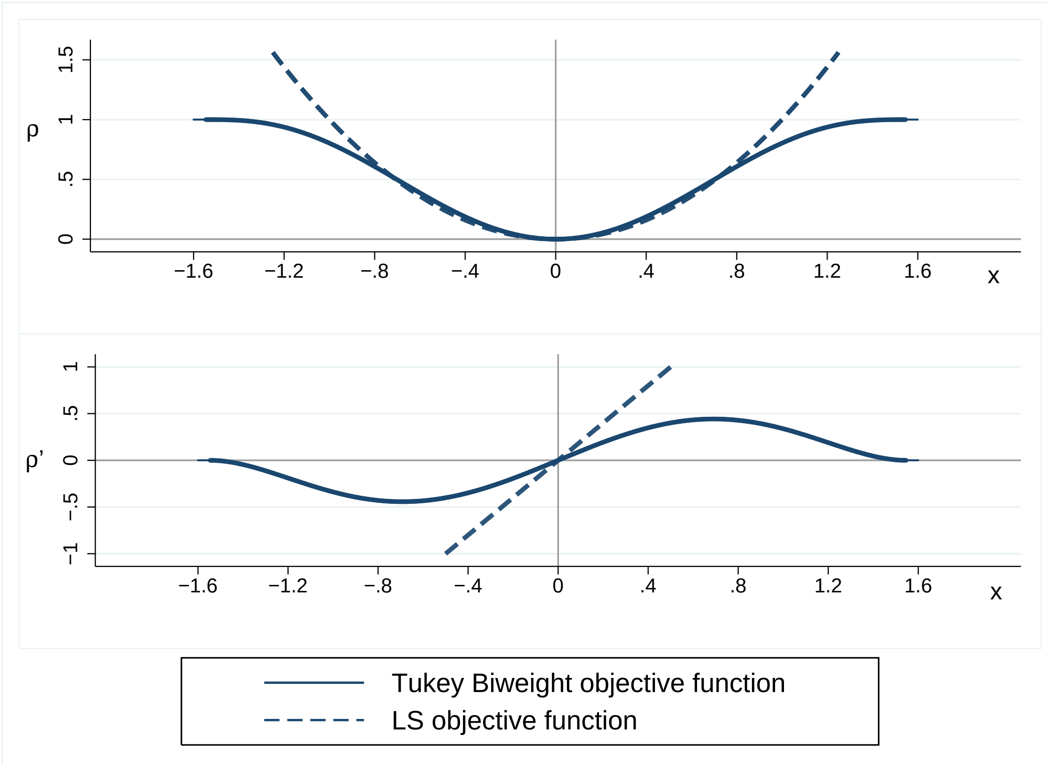


Figure 2: Power of the test for outliers

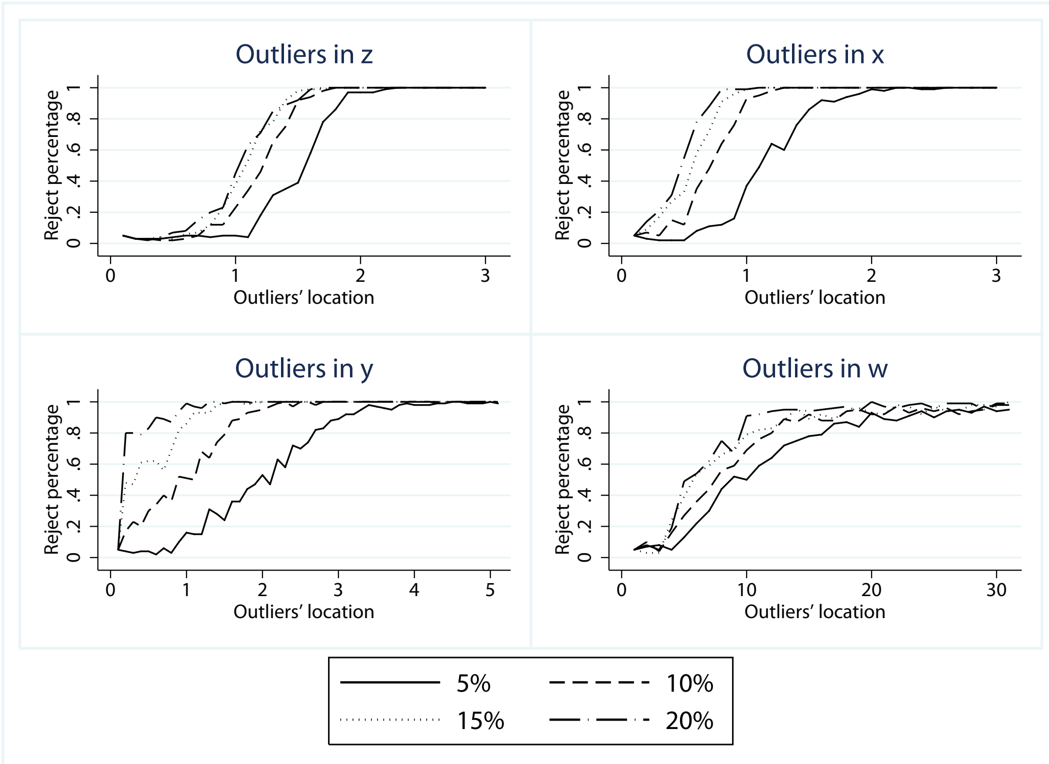
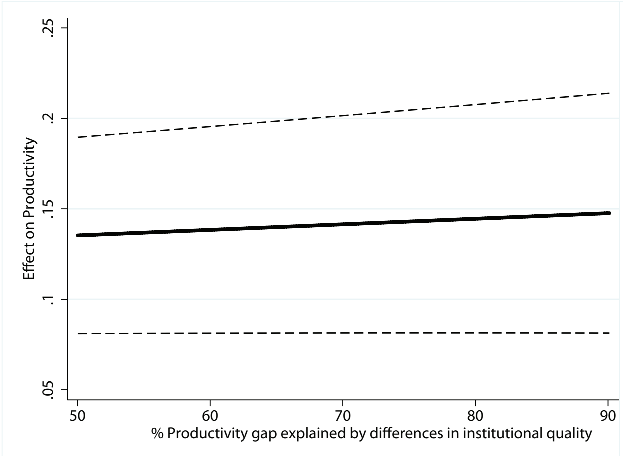
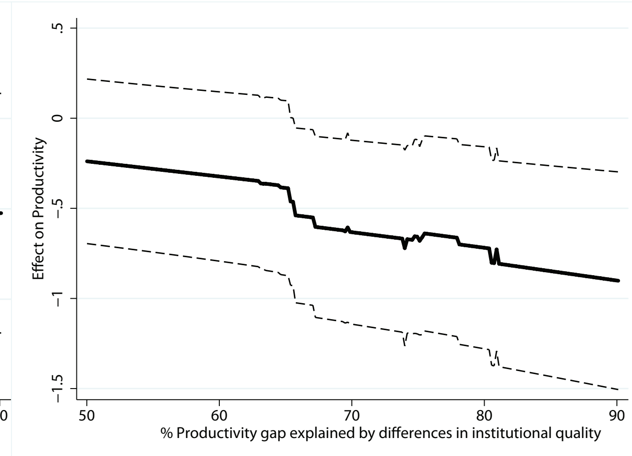


Figure 3: IV estimate of γ



Note: Dashed lines correspond to a 95% confidence interval.

Figure 4: RIV estimate of γ



Note: Dashed lines correspond to a 95% confidence interval.

Figure 5: Weak instrument F-statistic

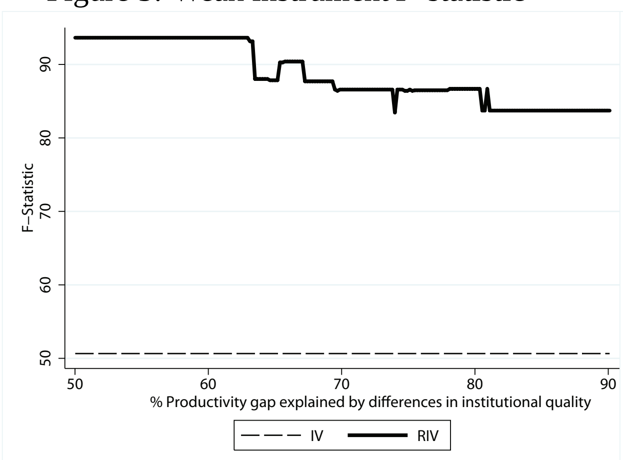


Figure 6: Autocorrelation test

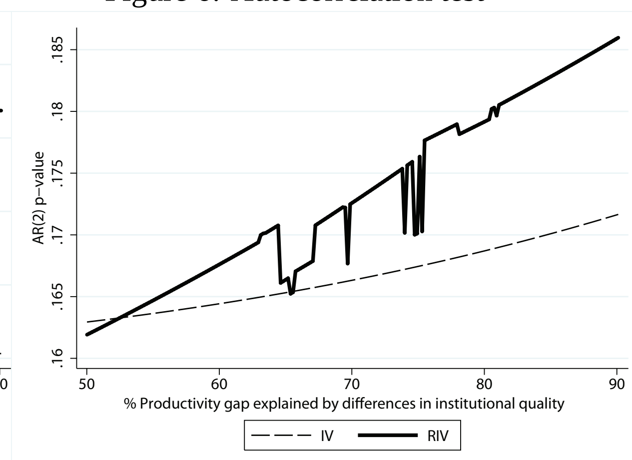


Figure 7: Test of overidentifying restrictions

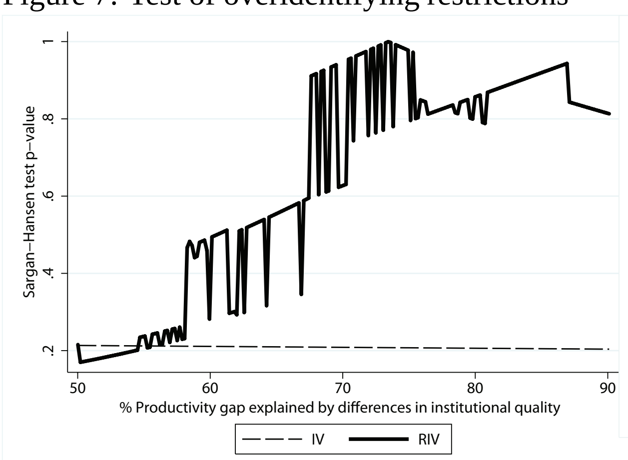
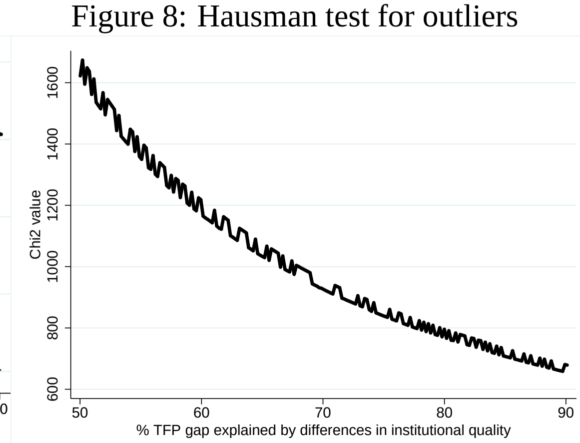
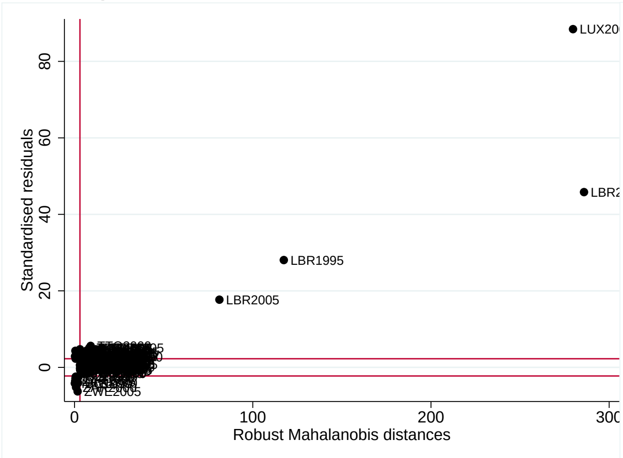


Figure 8: Hausman test for outliers



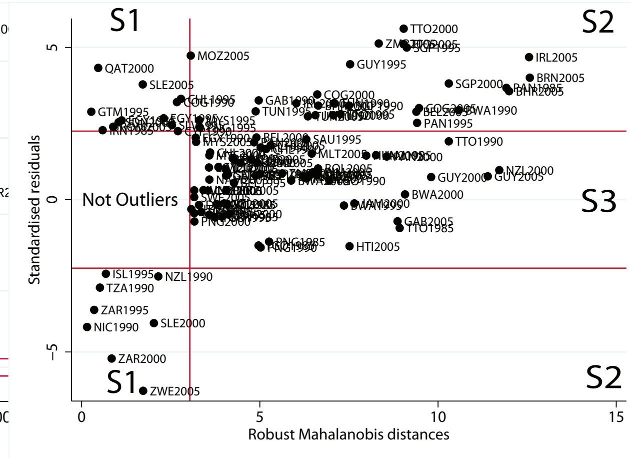
Note: The 5% critical value is 18.31.

Figure 9: Identification of outliers



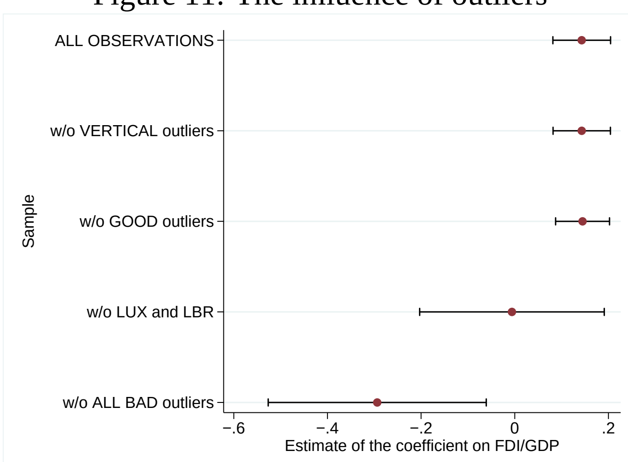
Note: A 75% productivity gap explained by differences in institutional quality is assumed.

Figure 10: Outliers, without LBR and LUX



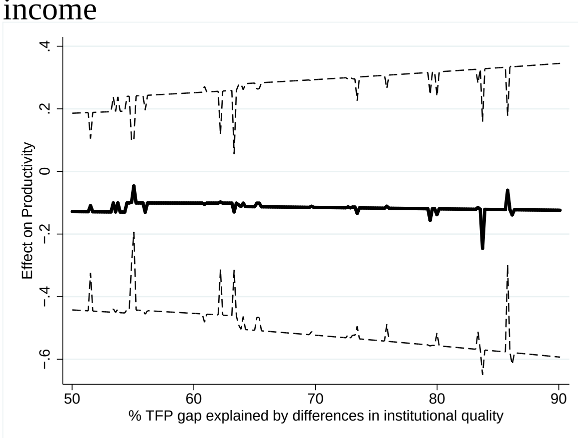
Note: A 75% productivity gap explained by differences in institutional quality is assumed.

Figure 11: The influence of outliers



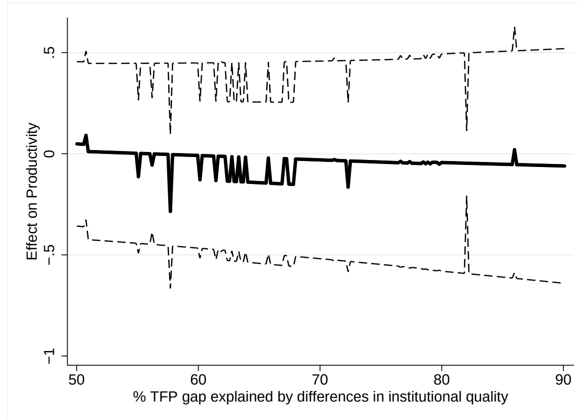
Note: A 75% productivity gap explained by differences in institutional quality is assumed.

Figure 12: RIV estimate of γ, above median income



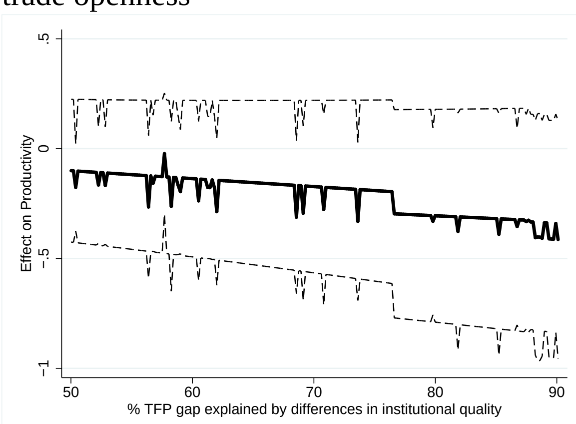
Note: Dashed lines correspond to a 95% confidence interval. Median F-statistic: 73. Median AR(2) p-value: 0.39. Median Hansen test p-value: 0.66.

Figure 13: RIV estimate of γ, above median schooling



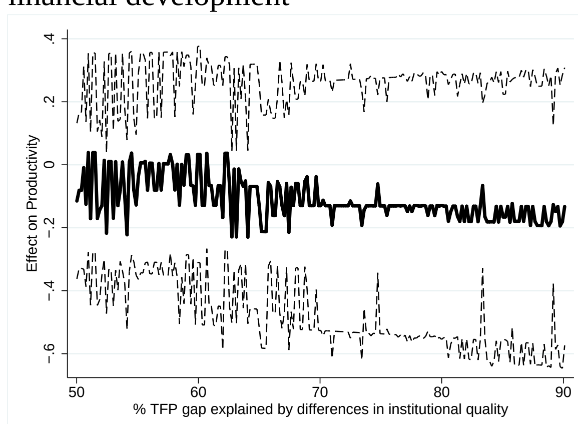
Note: Dashed lines correspond to a 95% confidence interval. Median F-statistic: 66. Median AR(2) p-value: 0.74. Median Hansen test p-value: 0.35.

Figure 14: RIV estimate of γ, above median trade openness



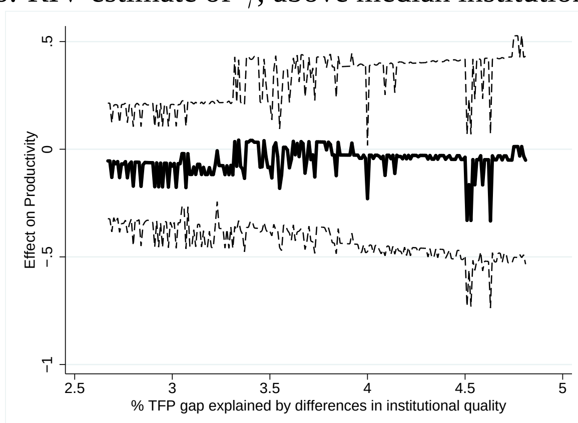
Note: Dashed lines correspond to a 95% confidence interval. Median F-statistic: 45. Median AR(2) p-value: 0.90. Median Hansen test p-value: 0.82.

Figure 15: RIV estimate of γ, above median financial development



Note: Dashed lines correspond to a 95% confidence interval. Median F-statistic: 74. Median AR(2) p-value: 0.36. Median Hansen test p-value: 0.77.

Figure 16: RIV estimate of γ, above median institutional quality



Note: Dashed lines correspond to a 95% confidence interval. Median F-statistic: 72. Median AR(2) p-value: 0.88. Median Hansen test p-value: 0.33.