

Chatelain, Jean-Bernard; Ralf, Kirsten

Article — Accepted Manuscript (Postprint)

Spurious regressions and near-multicollinearity, with an application to aid, policies and growth

Journal of macroeconomics

Suggested Citation: Chatelain, Jean-Bernard; Ralf, Kirsten (2014) : Spurious regressions and near-multicollinearity, with an application to aid, policies and growth, Journal of macroeconomics, ISSN 0164-0704, Elsevier, Amsterdam, Vol. 39, Iss. A, pp. 85-96, <https://doi.org/10.1016/j.jmacro.2013.11.003>

This Version is available at:

<https://hdl.handle.net/10419/96630>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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

Spurious regressions and near-multicollinearity, with an application to aid, policies and growth*

Jean-Bernard Chatelain[†] Kirsten Ralf[‡]

Post print: Journal of Macroeconomics (2014), vol. 39A, pp. 85-96.

May 9, 2014

Abstract

In multiple regressions, explanatory variables with simple correlation coefficients with the dependent variable below 0.1 in absolute value (such as aid/gross domestic product (GDP) with GDP growth) face a problem of parameter identification. They may have very large, statistically significant, estimated parameters which are unfortunately “outliers driven” and spurious. This is obtained by including another regressor which is highly correlated with the initial regressor, such as a lag, a square or interaction terms of this regressor. The analysis is applied on the “Gambia and Botswana outliers driven” Burnside and Dollar [2000] article which found that aid/GDP had an effect on growth only for countries achieving “good” macroeconomic policies.

Highlights:

*We thank for very helpful comments an anonymous referee, Michel Armatte, Robert Chirinko, Giacomo Corneo, Fabian Gouret, Christophe Hurlin, Antoine Parent, Xavier Ragot, Glenn Shafer, Jean-Philippe Touffut and Florian Wendelspiess, as well as participants to sessions at Probability and Statistical days in Mainz, AFSE conference in Paris, INFER conference in Orleans, the GDRE Money Banking and Finance conference in Poitiers, the Money Macro and Finance conference at Queen Mary University in London and to seminars at Cournot centre, Freie Universität Berlin and Reading universities. Preprint of a paper in Journal of Macroeconomics.

[†]Paris School of Economics, Université Paris 1 Panthéon Sorbonne, CES (Centre d’Economie de la Sorbonne), 106-112 Boulevard de l’Hôpital, 75647 Paris Cedex 13, France. Email: jean-bernard.chatelain@univ-paris1.fr

[‡]ESCE (Ecole Supérieure du Commerce Extérieur), 10 Rue Sextius Michel, 75015 Paris, France. Email: kirsten.ralf@esce.fr

- A classical suppressor has a zero simple correlation with a dependent variable.
- Classical suppressors often lead to spurious inference in multiple regressions.
- Highly correlated classical suppressors are very often statistically significant.
- Parameter inflation factor, pre-tests, outliers signal those spurious regressions.
- Aid has a widely quoted but spurious effect on growth conditional on good policy.

Keywords: Spurious regressions, Parameter identification problem, Hypothesis testing, Unstable conditional independence, Near- multicollinearity, Growth and Foreign aid.

JEL classification: **C12, O19, P45**

Résumé: Une variable explicative ayant un coefficient de corrélation simple avec la variable dépendante plus petit que 0,1 en valeur absolue (telle que l'aide au développement et la croissance) peut avoir des paramètres estimés très grands et statistiquement significatifs dans des régressions multiples, alors qu'il n'y a pas d'effet ou que cet effet provient seulement de quelques observations atypiques. Ce résultat peut être obtenu en ajoutant une autre variable explicative très corrélée avec la précédente telle que cette variable retardée ou mise au carré ou multipliée par une autre variable (terme d'interaction). Nous appliquons l'analyse ci-dessus à l'article de Burnside et Dollar (2000) qui a montré que l'aide aurait un effet sur la croissance seulement pour les pays ayant de "bonnes" politiques macroéconomiques, alors que ces résultats dépendent en grande partie des observations atypiques du Botswana et de la Gambie.

Mots clés: Régressions fallacieuses, problème d'identification d'un paramètre, test d'hypothèse, indépendance conditionnelle instable, quasi-colinéarité, aide au développement, croissance.

1 Introduction

This paper presents three results demonstrating that the statistical significance of an estimated parameter using the t -test may lead to misleading inference for the existence of a relationship between a dependent variable and a regressor in a *multiple* regression, when, by contrast, in a *simple* regression with the same dependent variable, the parameter of this regressor is not statistically significant. Such a regressor with a near-zero sample correlation with the dependent variable is sometimes called a “classical suppressor

variable”, see, for example, Friedman and Wall [2005].¹

(1) The first result is the existence of a parameter identification problem for identifying a spurious effect versus a non-spurious effect of a classical suppressor in a multiple regression. A parameter is not identified when it is not possible to distinguish between alternative values of the parameter related to alternative models which are observationally equivalent.

(2) A classical suppressor turns out to be statistically significant when including a second regressor in a multiple regression in three cases. In case 1, the second regressor is orthogonal to the classical suppressor and highly correlated with the dependent variable. In case 2, the second regressor is moderately correlated with the classical suppressor and with the dependent variable. In case 3, the second regressor has a simple correlation close to zero with the dependent variable (it is a second classical suppressor) and it is highly correlated with the classical suppressor. Case 3 includes dynamic models with highly correlated lags of the classical suppressor and non-linear models with quadratic and interactions terms including the classical suppressor. The publication bias for statistically significant parameters may foster the publication of regressions including a pair of highly correlated classical suppressors because they are so easily available to researchers (lags, square, interaction terms).

(3) When a model is not identifiable, it is not theoretically possible to learn the true value of this model’s underlying parameter with a large number of observations. The statistical significance of the estimated parameter of a classical suppressor in a multiple regression is not enough to distinguish between a spurious versus a non spurious effect. As an identification strategy, we first show that case 1 corresponds to a spurious effect, with an irrelevant Student t -test in multiple regression. We then show that case 3 has a potential for over-fitting. *If the regression is mostly fitting outliers, the t -test is not robust to these outliers: this allows the identification of a spurious effect.* This over-fitting leads to very large forecast errors on other samples than the estimation sample. Dummies for outliers, robust regression estimators, the evaluation of the root mean square errors on other samples than the estimation sample and other cross validation, resampling and bootstrapping techniques should be used.

The paper proceeds as follows. Section 2 investigates the parameter identification problem between a spurious effect versus a non-spurious effect of a classical suppressor in a multiple regression. Section 3 determines three cases

¹An irrelevant variable, by contrast, remains uncorrelated with the dependent variable also in the multiple regression. The focus of this paper is the case where the classical suppressor is not irrelevant in the multiple regression, whereas it is irrelevant in a simple regression.

where a classical suppressor turns out to be statistically significant when including a second regressor in a multiple regression. Section 4 proposes two identification strategies of spurious effect for two of these cases. Section 5 proposes tools for practitioners to avoid these spurious regressions. Section 6 applies these tools to Burnside and Dollar's paper on aid, policies and growth. Section 7 concludes.

2 A parameter identification problem

We consider the following trivariate regression on standardized variables. There is no constant in the model and all variables have mean zero and a variance of one. Bold letters correspond to matrices and vectors:

$$\mathbf{x}_1 = \beta_{12}\mathbf{x}_2 + \beta_{13}\mathbf{x}_3 + \varepsilon_{1,23} \quad (1)$$

where $\mathbf{x}_1$ is the vector of N observations of the dependent variable, $\mathbf{X}_{2,3} = (\mathbf{x}_2, \mathbf{x}_3)$ is the matrix where column i corresponds to the N observations of the regressor $\mathbf{x}_i$ for $2 \leq i \leq 3$, $\beta = (\beta_{12}, \beta_{13})$ is a vector of standardized parameters to be estimated, and $\varepsilon_{1,23}$ is a vector of random disturbances that follow a normal distribution with mean zero and variance σ^2 . In a linear regression model with standard assumptions on the error term, ($E(\varepsilon_t|\mathbf{X}_{2,3,t}) = \mathbf{0}$ and $E(\varepsilon_t^2|\mathbf{X}_{2,3,t}) = \sigma^2$), Spanos and McGuirk [2002] derive in their theorem 1 a relation between the model parameters (β, σ^2) and the primary parameters of the model defined by a vector of means and a covariance matrix (identical to the correlation matrix for standardized variables). In their theorem 2, Spanos and McGuirk [2002] state that the parameterization (β, σ) exists if and only if the determinant of the correlation matrix of the regressors, denoted $\mathbf{R}_2$, is strictly positive: $\det(\mathbf{R}_2) = 1 - r_{23}^2 > 0$ where r_{23} denotes the simple correlation coefficient of the regressors. Then, the ordinary least squares estimated parameters are given by linear functions of the simple correlation coefficients r_{12} and r_{13} of each regressor with the dependent variable indexed by one (variables with a hat denote estimated variables, except for sample simple correlation coefficients to simplify notation):

$$\begin{pmatrix} \hat{\beta}_{12} \\ \hat{\beta}_{13} \end{pmatrix} = \mathbf{R}_2^{-1} \begin{pmatrix} r_{12} \\ r_{13} \end{pmatrix} \Rightarrow \hat{\beta}_{12} = \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} \text{ and } \hat{\beta}_{13} = \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2}. \quad (2)$$

One can always orthogonalize the two regressor variables, using the residuals of the auxiliary regression of the regressor $\mathbf{x}_3$ as a linear function of the regressor $\mathbf{x}_2$ ($\varepsilon_{3,2} = \mathbf{x}_3 - r_{23}\mathbf{x}_2$):

$$\mathbf{x}_1 = \hat{\beta}_{12}\mathbf{x}_2 + \hat{\beta}_{13}\mathbf{x}_3 + \hat{\varepsilon}_{1.23} = r_{12}\mathbf{x}_2 + \hat{\beta}_{13}(\mathbf{x}_3 - r_{23}\mathbf{x}_2) + \varepsilon_{1.23}. \quad (3)$$

Let us assume that the null hypothesis for the true simple correlation coefficient $H_0: r_{12} = 0$ is not rejected. On the one hand, the effect of the classical suppressor may be *spurious* because the regression with orthogonal regressors becomes:

$$\mathbf{x}_1 = 0 \cdot \mathbf{x}_2 + \hat{\beta}_{13}(\mathbf{x}_3 - r_{23}\mathbf{x}_2) + \varepsilon_{1.23} \text{ and } \mathbf{x}_2 = \varepsilon_2 \text{ and } \varepsilon_{3.2} = \mathbf{x}_3 - r_{23}\mathbf{x}_2 \quad (4)$$

A shock ε_2 of the random variable $\mathbf{x}_2$ has *no direct effect* on the dependent variable (because $r_{12} = 0$) and *no indirect effect*, because the variance of the residuals $\mathbf{x}_3 - r_{23}\mathbf{x}_2$ corresponds to the variance of $\mathbf{x}_3$ which is not related at all to the variations of $\mathbf{x}_2$ (figure 1). The t -test which is relevant in this multiple regression with orthogonal regressors is the test of the null hypothesis $H_0: r_{12} = 0$.

On the other hand, the effect of the classical suppressor may also be *not spurious*. One always has $r_{12} = \hat{\beta}_{12} + r_{23} \cdot \hat{\beta}_{13}$: the simple correlation effect r_{12} is the sum of the *direct effect* $\hat{\beta}_{12}$ of $\mathbf{x}_2$ upon $\mathbf{x}_1$ and of the *indirect effect* of $\mathbf{x}_2$ upon $\mathbf{x}_1$ mediated by the mediating variable $\mathbf{x}_3$ (figure 1). This indirect effect is the product of the effect r_{23} of $\mathbf{x}_2$ upon $\mathbf{x}_3$ time the effect $\hat{\beta}_{13}$ of $\mathbf{x}_3$ upon $\mathbf{x}_1$. If $r_{12} = 0$ is not rejected *and* if $r_{23} = 0$ is rejected, then the direct effect $\hat{\beta}_{12}$ of $\mathbf{x}_2$ upon $\mathbf{x}_1$ is exactly and fully offset by the indirect effect of $\mathbf{x}_2$ upon $\mathbf{x}_1$ mediated through $\mathbf{x}_3$: $\hat{\beta}_{13} = -\frac{1}{r_{23}}\hat{\beta}_{12}$. In this case, one may interpret the relationship between regressors $\mathbf{x}_3 = r_{23}\mathbf{x}_2 + \varepsilon_{2.3}$ as a immediate negative feedback relationship of the variable $\mathbf{x}_3$ upon $\mathbf{x}_2$ which maintains constant the dependent variable $\mathbf{x}_1$ whenever $\mathbf{x}_2$ deviates from its mean value (here equal to zero) by a value ε_2 . The multiple regression may then belong to a triangular system of equations:

$$\mathbf{x}_2 = \varepsilon_2 \text{ and } \mathbf{x}_3 = r_{23}\mathbf{x}_2 + \varepsilon_{3.2} \text{ and } \mathbf{x}_1 = \beta_{12}\mathbf{x}_2 - \frac{1}{r_{23}}\beta_{12}\mathbf{x}_3 + \varepsilon_{1.23} \quad (5)$$

where the relevant t -test is the test of the null hypothesis $H_0: \beta_{12} = 0$.

The above discussion highlights the existence of a parameter identification problem of a classical suppressor variable in a multiple regression (Hoover, 2001, pp. 45-49). The two observationally equivalent models are related to t -tests of two different null hypothesis with conflicting outcome, where the conditional independence of the variables $\mathbf{x}_1$ and $\mathbf{x}_2$ is not stable when one includes an additional regressor. To our knowledge, no identification strategy has been proposed to decide between these two models. For example,

Spanos [2006] concludes that the multiple regression including a statistically significant classical suppressor is *always* spurious using pre-tests on simple correlation coefficients. But he does not explicitly address the above parameter identification problem. Because the parameter could be either equal to zero (with a coefficient of determination equal to zero in simple regression) or even reach maximal feasible absolute values in trivariate regressions (with a coefficient of determination equal to one), there is no room for partial identification or for set identification. However, we found two identification strategies which are described in section 4 and are based on results found in the next section 3.

3 Testing the stability of conditional independence

In this section, we use Spanos and McGuirk [2002] framework to investigate the conditions for a classical suppressor to turn out to be statistically significant when including an additional regressor. This is the *opposite* of the concern of econometric textbooks when dealing with near-multicollinearity. Textbooks focus is when a regressor that is relevant in the simple regression has no longer a statistically significant effect when another highly correlated regressor is included. They focus on type I inference discordance defined as follows. Firstly, the t -test does not reject the null hypothesis of no effect ($H'_0 : \beta_{12} = 0$) between the dependent variable x_1 and a regressor x_2 , in the trivariate regression. Secondly, the t -test *rejects* the null hypothesis of no effect ($H_0 : r_{12} = 0$) between the dependent variable x_1 and a regressor x_2 , in the simple regression. Our concern is type II inference discordance defined as follows. Firstly, the t -test *rejects* the null hypothesis of no effect ($H_0 : \beta_{12} = 0$) in the trivariate regression. Secondly, the t -test *does not reject* the null hypothesis of no effect ($H_0 : r_{12} = 0$) between the dependent variable x_1 and a regressor x_2 in the simple regression, in the set of feasible correlation coefficients of the trivariate regression. A formerly not relevant variable becomes relevant when another variable is included.

Spanos and McGuirk [2002] compute the estimated standard deviations of estimated parameters $\widehat{\sigma}_{\beta_{12}}$ and their Student's t -statistics $\widehat{t}_{\beta_{12}}$ assuming the normality of the disturbances:

$$\begin{pmatrix} \widehat{\sigma}_{\beta_{12}} \\ \widehat{\sigma}_{\beta_{13}} \end{pmatrix} = \frac{\sqrt{\det(\mathbf{R}_3)}}{\sqrt{N-2}} \frac{1}{1-r_{23}^2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \widehat{t}_{\beta_{12}} \\ \widehat{t}_{\beta_{13}} \end{pmatrix} = \frac{\sqrt{N-2}}{\sqrt{\det(\mathbf{R}_3)}} \begin{pmatrix} r_{12} - r_{13}r_{23} \\ r_{13} - r_{12}r_{23} \end{pmatrix}$$

$$0 \leq \det(\mathbf{R}_3) = (1 - R_{1.23}^2) \det(\mathbf{R}_2) \leq \det(\mathbf{R}_2) = 1 - r_{23}^2 \leq 1. \quad (6)$$

Estimated standard errors and t-statistics depend on the sample correlation matrix $\mathbf{R}_3$ whose entries are the correlation coefficients of all pairs of variables denoted r_{ij} , including the dependent variable, indexed by 1, on the first row and first column. One has $r_{ij}^2 \leq 1$ for $1 \leq i \leq 3$ and $1 \leq j \leq 3$. The Schur property of the determinants of the correlation matrices relates the determinant of the matrix $\mathbf{R}_3$ to the determinant of the submatrix $\mathbf{R}_2$ of the regressors and to the coefficient of determination $R_{1,23}^2$ of the trivariate regression.

Feasible correlation coefficients when the correlation between regressors increases. The condition on correlation coefficients $\det(\mathbf{R}_3) \geq 0$ implies that the partial correlation coefficients are at most equal to one in the case of the exact regression, where all the residuals are equal to zero. It can be written as follows:

$$\begin{aligned} \det(\mathbf{R}_3) &= -(1 + r_{23}) \left(\frac{r_{12} - r_{13}}{\sqrt{2}} \right)^2 - (1 - r_{23}) \left(\frac{r_{12} + r_{13}}{\sqrt{2}} \right)^2 + 1 - r_{23}^2 \geq 0 \\ &\Rightarrow \frac{|r_{12} - r_{13}r_{23}|}{\sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2}} \leq 1. \end{aligned} \quad (7)$$

In the case where $\det(\mathbf{R}_3) = 0$, the Schur property implies that the coefficient of determination $R_{1,23}^2$ is equal to one. Then, all residuals are equal to zero. The regression is an exact linear relation between the three variables. The estimated standard errors of the estimated parameters are equal to zero. The t -statistics tends to infinity. For a given $|r_{23}| < 1$ and varying r_{13} and r_{12} , $\det(\mathbf{R}_3) = 0$ describes an ellipse centered at the origin ($r_{12} = r_{13} = 0$) in the plane (r_{13}, r_{12}) . All possible values of correlation coefficients of r_{12} and r_{13} have to be enclosed by the large ellipse or on its border so that $\det(\mathbf{R}_3) \geq 0$.

When the correlation between regressors increases, the set of feasible correlation coefficients with the dependent variable decreases and the area enclosed by the large ellipse shrinks. The values of r_{23} are chosen to be equal to 0, 0.5, 0.95 and 0.99, in figures 2 to 5 respectively. When $r_{23} = 0$, the ellipse is a circle centered at the origin ($r_{12} = r_{13} = 0$) with a radius of 1 (figure 2). When $r_{23} > 0$, the minor axis has a length of $2\sqrt{1 - r_{23}^2}$ and is on the line $r_{12} = -r_{13}$. As can be seen on figures 2 to 5, the width on both sides of the minor axis decreases when the correlation coefficient r_{23} increases from zero to one. By contrast, the length of the major axis $2\sqrt{1 + r_{23}^2}$ (which is on the line $r_{12} = r_{13}$, when $r_{23} > 0$) increases when r_{23} increases. In other words, *the higher two variables are positively correlated, the closer are their correlations with a third variable.* The area enclosed by the ellipse ($\pi\sqrt{1 - r_{23}^2}$)

shrinks and tends to zero when r_{23} increases from zero to one. In the limit case of exact positive collinearity between regressors ($r_{23} = 1$), the ellipse degenerates into the segment of the line $r_{12} = r_{13}$ defined on $[-1, +1]$. When $r_{23} < 0$, the major axis is $r_{12} = -r_{13}$ and the minor axis is $r_{12} = r_{13}$. In the other limit case $r_{23} = -1$, the ellipse degenerates into the segment of the line $r_{12} = -r_{13}$, defined on $[-1, +1]$. These two limit cases correspond to the singularity of the correlation matrix of the regressors (exact collinearity of the explanatory variables: $\det(\mathbf{R}_2) = 0$), the ordinary least squares estimators cannot be computed.

The acceptance region of the test of the null hypothesis $H'_0 : \beta_{12} = 0$ against the alternative hypothesis $H'_1 : \beta_{12} \neq 0$ when the correlation between regressors increases. It is given by:

$$\frac{|r_{12} - r_{13}r_{23}|}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}} \leq \frac{t_{\widehat{\beta}_1}(\alpha)}{\sqrt{N - 2 + t_{\widehat{\beta}_{12}}^2(\alpha)^2}} = \gamma(\alpha, N, 2) < 1 \quad (8)$$

with a significance threshold $\alpha = 5\%$, related to the percentile $t_{\widehat{\beta}_{12}}(\alpha)$ of a Student distribution. The acceptance region of the null hypothesis ($H'_0 : \beta_{12} = 0$) is *reduced* when the above interval for r_{12} is *reduced*, that is when N increases and/or $t(\alpha)$ decreases (the applied researcher sets for example a significance threshold $\alpha = 10\%$ instead of 5%) and/or r_{13}^2 tends to one (the other explanatory variable is strongly correlated with the dependent variable) and/or *in particular when r_{23}^2 tends to one (highly correlated regressors)*. As seen in figure 2 to 5, the acceptance region is the area enclosed by the small ellipse.

When the correlation between regressors increases, the set of feasible correlation coefficients with the dependent variable in the acceptance region of the test $H'_0 : \beta_{12} = 0$ decreases and the area inside the small ellipse shrinks. The *major axis* of the small ellipse is related to the null hypothesis $H_0 : \beta_{12} = 0$, which is equivalent to $r_{12} = r_{23}r_{13}$. The locus of the null hypothesis $r_{12} = r_{23}r_{13}$ defines a line through the origin with a slope of r_{23} in the plane (r_{12}, r_{13}) . Hence, its slope in absolute value *increases* from 0 to the limit value of 1 when the correlation between regressor increases. It reaches the value $r_{13} = -1$ for $r_{12} = -r_{23}$ and $r_{13} = 1$ for $r_{12} = r_{23}$. Finally, the rejection region of the t test of the null hypothesis $H_0 : \beta_{12} = 0$ lies inside the large ellipse of feasible correlation coefficients and outside the small ellipse in figures 2 to 5. In the rejection region, we would conclude that the variable x_2 has an explanatory partial effect on the dependent variable x_1 .

The acceptance region of the test of the null hypothesis $H_0 : r_{12} = 0$ against $H_1 : r_{12} \neq 0$, taking into account that r_{12} is a feasible

correlation coefficient:

$$|r_{12}| \leq \frac{t_{r_{12}}(\alpha)}{\sqrt{N-1+t_{r_{12}}(\alpha)^2}} = \gamma(\alpha, N, 1) < 1 \text{ and } \frac{|r_{12} - r_{13}r_{23}|}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}} \leq 1 \quad (9)$$

Inside the large ellipse describing feasible correlation coefficients, the rejection region of the null hypothesis ($H_0 : r_{12} = 0$) *increases* when the above interval for r_{12} is *reduced*, when the number of observations N and/or the threshold α increase. In figures 2 to 5, the acceptance region of this test is between the two horizontal lines $r_{12} = \pm 0.196$ (for $N = 102$ and $\alpha = 5\%$) on both sides of the horizontal line related to the null hypothesis $H_0 : r_{12} = 0$ and inside the large ellipse. If the correlation coefficients are within the area between the two horizontal lines, we do not reject the null hypothesis of no effect in the simple regression of the variable x_2 on the variable x_1 .

The acceptance region of the test of the null hypothesis $H_0 : r_{13} = 0$ against $H_1 : r_{13} \neq 0$, taking into account that r_{13} is a feasible correlation coefficient is given by similar inequalities. In figures 3 to 5 (not shown on figure 2 for simplicity), the acceptance region of this test is between the two vertical lines $r_{13} = \pm 0.196$ (for $N = 102$ and $\alpha = 5\%$) on both sides of the vertical line related to the null hypothesis $H_0 : r_{13} = 0$ and inside the large ellipse.

Graphical representations of inference discordances. In what follows we discuss some results concerning type I and type II discordances. Our discussion here refers only to the graphical representation, the theoretical results can be found in the appendix. On figure 2 to 5, type II discordance occurs for (r_{13}, r_{12}) between the highest horizontal line and the lowest horizontal line $r_{12} = \pm 0.196$ (the acceptance region of the t -test of the null hypothesis $H_0 : r_{12} = 0$) *and* inside the large ellipse and outside the small ellipse (the rejection region of the t -test of the null hypothesis $H_0 : \beta_{12} = 0$). Conversely, type I discordance occurs for (r_{13}, r_{12}) in the rejection region of the t -test of the null hypothesis $H_0 : r_{12} = 0$ which is above the highest horizontal line $r_{12} > 0.196$ or below the lowest horizontal line $r_{12} < -0.196$ *and* inside the small ellipse (acceptance region of the t -test of the null hypothesis $H_0 : \beta_{12} = 0$).

A typology of type II inference discordance when the correlation between regressors increases.

Case 1. When the null hypothesis $H_0 : r_{23} = 0$ that the second regressor is orthogonal to the classical suppressor regressor is not rejected (for example: $0 \leq |r_{23}| \leq 0.2$, when $N \approx 102$ and $\alpha = 5\%$), the range of feasible values of the sample correlation coefficient r_{12} for type II discordance is larger when the second regressor is highly and increasingly correlated with the dependent

variable (for example: $0.8 \leq |r_{13}| \leq 1$). For example, in figure 2, with orthogonal regressors ($r_{23} = 0$), the estimated parameters are identical $\hat{\beta}_{12} = r_{12}$ and the range of feasible values of the sample correlation coefficient r_{12} for type II discordance is larger when the correlation between regressors r_{13}^2 tends to one:

$$\gamma(\alpha, N, 2) \sqrt{1 - r_{13}^2} \leq |r_{12}| \leq \gamma(\alpha = 5\%, N = 102, 1) = 0.196 \quad (10)$$

The acceptance region for the test of the null hypothesis $H'_0 : \beta_{12} = 0$ is *included* in the acceptance region for the test of the null hypothesis $H_0 : r_{12} = 0$ so that there is no type I inference discordance between the two tests. The coefficient of determination is equal to $R_{1,23}^2 = r_{13}^2 + r_{12}^2$ when $r_{23} = 0$. It tends to one when r_{13}^2 tends to one and when r_{12} is close to zero. In other words, when r_{13}^2 tends to one, the second regressor (indexed by 3) reduces on average the squares of the residuals so much that the root mean square error of the residuals tends to zero. By contrast, when r_{12} is zero, the first regressor does not reduce the squares of the residuals. But the standard error of the estimated parameter of first regressor $\hat{\beta}_{12}$ tends to zero in the multiple regression, because it is a linear increasing function of the root mean square error of the residuals. Hence, the parameter $\hat{\beta}_{12}$ turns out statistically significant although the variable x_2 , orthogonal to x_3 , is *unrelated to this reduction of the root mean square error*. For this reason, in regressions with orthogonal regressors, only the simple correlation rejection region provides meaningful inference (Chatelain and Ralf, 2010, present a similar argument for time invariant variables in panel data). Software programmers should change the computation of standard errors with incomplete principal component ordinary least square regressions (e.g. PROC REG, PCOMIT instruction in SAS).

Case 2. When the second regressor is moderately correlated with the classical suppressor regressor ($0.2 \leq |r_{23}| \leq 0.8$), the range of feasible values of the sample correlation coefficient r_{12} for type II discordance is larger when the second regressor is *moderately correlated* with the dependent variable, for example: $0.1 \leq |r_{13}| \leq 0.9$, as seen in figure 3, where $r_{23} = 0.5$. The major axis of the small ellipse has a positive slope equal to $r_{23} = 0.5$. It is the location of the null hypothesis $\beta_{12} = 0$. The area of type II discordances (where the classical suppressor reaches statistical significance in the multiple regression) increases with respect to the case of orthogonal regressors. There is now an area for type I discordances.

Case 3 (highly correlated pair of classical suppressors). When the second regressor is highly correlated with the classical suppressor regressor

($0.8 \leq |r_{23}| < 1$), the range of feasible values of the sample correlation coefficient r_{12} for type II inference discordance is larger when the second regressor is likely to be another classical suppressor poorly correlated with the dependent variable, so that the null hypothesis $H_0 : r_{13} = 0$ is not rejected, for example, $0 \leq |r_{13}| \leq 0.196 = \gamma(\alpha = 5\%, N = 102, 1)$ in figure 4 and 5. In particular, when $r_{13} = 0$, the range of feasible values of the sample correlation coefficient r_{12} for type II discordance is larger when the correlation between regressors r_{23}^2 tends to one:

$$\gamma(\alpha, N, 2) \sqrt{1 - r_{23}^2} \leq |r_{12}| \leq \gamma(\alpha = 5\%, N = 102, 1) = 0.196 \quad (11)$$

In figure 4, the correlation between the two regressors is equal to $r_{23} = 0.95$ (in figure 5, which is qualitatively equivalent to figure 4, the correlation between the two variable is equal to $r_{23} = 0.99$). The location of the null hypothesis $H_0 : \beta_{12} = 0$ (so that $r_{12} = 0.95r_{13}$) is the major axis on the small ellipse with a slope equal to 0.95: it is nearly identical to the slope of the major axis of the large ellipse equal to 1 (so that $r_{12} = r_{13}$). The large ellipse *shrinks* with respect to figure 2 because the high positive correlation between the two regressors implies that the correlation of each of the regressors with the dependent variable should be nearly the same. As a consequence, the acceptance region of the t -test of the null hypothesis $H_0 : \beta_{12} = 0$ included inside the small ellipse *also shrinks*.

In the central square, $|r_{1j}| < 0.196$, $j = 2, 3$, both correlation coefficients are in the acceptance region of the t -test of the null hypothesis $H_0 : r_{1j} = 0$, $j = 2, 3$: one does not reject the null joint hypothesis that *both* regressors are classical suppressors. This central square is now within the rejection region of the t -test of the null hypothesis $H_0 : \beta_{12} = 0$, except inside the *thin* diagonal area around the null hypothesis $H_0 : \beta_{12} = 0$ which is the locus of the major axis $r_{12} = 0.95r_{13}$ of the small ellipse. Then, in order to reach statistical significance of $\hat{\beta}_{12}$ for a p -value at least equal to 5%, when $r_{23} = 0.95$ and $r_{12} = 0$, a difference of only a few percentage points between the correlation coefficient r_{12} and r_{13} is sufficient so that $|r_{12} - 0.95r_{13}|$ is not too small, knowing that $|r_{13}|$ cannot exceed 0.312 for the exact regression case $\det(\mathbf{R}_3) = 0$. For examples, if $N = 402$, it is sufficient that $|r_{13}| > 0.031$, if $N = 102$, it is sufficient that $|r_{13}| > 0.061$ and if $N = 22$, it is sufficient that $|r_{13}| > 0.127$.

The usual concern of textbooks for near-multicollinearity related to type I discordances is also found in figure 4. For $0.196 < |r_{12}| \leq 1$, even including values of $|r_{12}|$ close to one with a very high power (close to one) of the t -test of the null hypothesis: $H_0 : r_{12} = 0$, $|r_{12}|$ may belong to the thin area

inside the small ellipse which is the acceptance region of the t -test of the null hypothesis $H_0 : \beta_{12} = 0$.

Increasing the number of observations may foster spurious inference of highly correlated classical suppressors. We conducted Monte Carlo simulations on samples of multi-normal laws with zero mean, standard deviations of one, $r_{12} = 0$ and $r_{13} = 0.03$ for 1000 replications. In case 2, when $r_{23} = 0.5$, the proportion of type II discordance is equal to 2.1% of the replications of samples of $N = 102$ observations; it is equal to 4.9% of the replications of samples of $N = 1002$ observations. In case 3, when $r_{23} = 0.99$ (cf. figure 5), the proportion of type II discordance is equal to 52.3% of the replications of samples of $N = 102$ observations; it is equal to 95.5% of the replications of samples of $N = 1002$ observations. By contrast, in the four types of simulations, the proportion of type I discordances, which is the usual focus of textbooks on near-multicollinearity issues, remained between 0% and 3.3% (Chatelain, 2010).

On the effect of a pair of highly correlated classical suppressors on a third regressor: Let us now consider this regression with normalized variables when the following null hypotheses are not rejected: $r_{12} = r_{13} = 0$:

$$\mathbf{x}_1 = \beta_{12}\mathbf{x}_2 + \beta_{13}\mathbf{x}_3 + \beta_{14}\mathbf{x}_4 + \varepsilon_{1.234} \quad (12)$$

To investigate the effect of the pair of classical suppressors, one can analyse the variance of the third regressor with the following orthogonal regressors:

$$\mathbf{x}_4 = r_{14}\mathbf{x}_1 + r_{24}\mathbf{x}_2 + \beta_{24}(\mathbf{x}_3 - r_{23}\mathbf{x}_2) + \varepsilon_{4.123} \quad (13)$$

If the third regressor is orthogonal to the two classical suppressors (the following hypothesis are not rejected: $r_{24} = \beta_{24} = 0$), its parameter $\beta_{14} = r_{14}$ does not change when adding orthogonal regressors. But its estimated standard error may benefit from the decrease of the root mean square error, if for example $\mathbf{x}_3 - r_{23}\mathbf{x}_2$ fits with a large outlier. As mentioned before, in the case of orthogonal regressors, this outcome of the t -test is questionable.

Else, if the third regressor is not orthogonal to the two classical suppressors, its parameter $\beta_{14} \neq r_{14}$ will change when including these two classical suppressors and will include: $cov(\mathbf{x}_4 - r_{14}\mathbf{x}_1, \mathbf{x}_2)$ and $cov(\mathbf{x}_4 - r_{14}\mathbf{x}_1, \mathbf{x}_3 - r_{23}\mathbf{x}_2)$. However, the variance component $var(\mathbf{x}_4 - r_{14}\mathbf{x}_1)$ is unrelated to the variance that needs to be explained: $var(\mathbf{x}_1)$ in the regression where the dependent variable is $\mathbf{x}_1$ (they are related to two distinct orthogonal subspace). Hence, it is not obvious that introducing $cov(\mathbf{x}_4 - r_{14}\mathbf{x}_1, \mathbf{x}_2)$ and $cov(\mathbf{x}_4 - r_{14}\mathbf{x}_1, \mathbf{x}_3 - r_{23}\mathbf{x}_2)$ in β_{14} provides meaningful additional information with respect to r_{14} when the purpose at hand is to explain the effect

of the variance of $\mathbf{x}_4$ on the variance of $\mathbf{x}_1$. This suggests that excluding the two classical suppressors does not add noise in the interpretation of the parameter of the third regressor on the dependent variable $\mathbf{x}_1$.

On the distribution of true correlation coefficients: The frequency of spurious regressions depends upon the distribution on true correlation coefficients (r_{12}^T, r_{13}^T) on the planes depicted in the figures 2 to 5 given r_{23}^T for any available data set of three variables in any field of research. A researcher may consider that a Laplacian prior is possible: the distribution is uniform. But, in practice, researchers may select variables while doing exploratory regressions and data mining, in order to reach statistical significance (reject $H_0 : \beta_{12} = 0$), because of the publication bias for rejecting null hypothesis. When the simple correlation of the dependent variable with the explanatory regressor of interest is zero, it is easy to find highly correlated regressors, such as lags, powers and interaction terms of the classical suppressor or finding a control variable which has a common unobservable factor with the classical suppressor. Then statistical significance is likely to follow suit.

4 The identification of statistically significant spurious effects

A type II inference discordance may lead a researcher to present as statistically significant a spurious effect for regressors such that the null hypothesis $H_0 : r_{12} = 0$ is not rejected. Hopefully, we found two strategies to identify spurious effect related to case 1 and case 3 of the typology of type II inference discordance of the previous section.

In case 1, the null hypothesis $H_0 : r_{23} = 0$ that the second regressor is orthogonal to the classical suppressor regressor is not rejected. This implies that the existence of a negative feedback relation of the control variable x_3 on the variable x_2 is rejected. As a consequence, the non-spurious effect model is rejected. What is more, the above section highlighted that the statistical significance of β_{12} is meaningless: it should not be taken into account.

In case 3, the null hypothesis $H_0 : r_{13} = 0$ that the second regressor is another classical suppressor is not rejected, and it needs to be highly correlated ($|r_{23}| \geq 0.8$) with the first classical suppressor in order that both regressors have statistically significant parameters. This particular case often occurs when the second regressor is a lag or the square or an interaction term of the classical suppressor. On the one hand, for the spurious effect model, it adds another spurious effect of the second regressor $\mathbf{x}_3$ on the dependent variable $\mathbf{x}_1$ so that the model includes two spurious effects instead of one! On

the other hand, for the non-spurious effect model, the variable $\mathbf{x}_3$ is highly correlated to the other variable $\mathbf{x}_2$ in the negative feedback control auxiliary equation: $\mathbf{x}_3 = r_{23}\mathbf{x}_2 + \varepsilon_{2.3}$.

When a pair of highly correlated classical suppressors faces over-fitting of the dependent variable, *the spurious effect can be identified because its statistical significance is not robust to outliers*. The higher the correlation among regressors, the lower the variance of the residual $\mathbf{x}_3 - r_{23}\mathbf{x}_2$, which turns out to be much smaller than the variance of the dependent variable: $Var(\mathbf{x}_3 - r_{23}\mathbf{x}_2) = 1 - r_{23}^2 < Var(\mathbf{x}_1) = 1$. An explanatory variable which has most of its observations very close to its mean has a very large estimated parameter equal to $\hat{\beta}_{13}$, which is a linear function of the ratio of standard errors $\sigma(\mathbf{x}_1)/\sigma(\mathbf{x}_3 - r_{23}\mathbf{x}_2) = 1/\sqrt{1 - r_{23}^2}$. This parameter $\hat{\beta}_{13}$ is highly sensitive to the omission or the inclusion of a few observations which are the farthest from the mean of the variable in the sample (observations with high leverage). Hence, the residual $\mathbf{x}_3 - r_{23}\mathbf{x}_2$ may be highly correlated with dummies for a few outliers. Then, the multiple regression is mostly fitting outliers. This suggests that the t -test rejects the null hypothesis $H_0: \beta_{13} = 0$ but that it is not robust to outliers. By the same argument, as $\hat{\beta}_{12} = -r_{23} \cdot \hat{\beta}_{13}$, the t -test rejects the null hypothesis $H_0: \beta_{12} = 0$ but it is not robust to outliers. If, when removing outliers or when decreasing their weights using robust regressions, t -tests do not reject any longer the null hypothesis $H_0: \beta_{13} = 0$ and the null hypothesis $H_0: \beta_{12} = 0$, one can conclude that the correct interpretation of the first regression on the full sample was the one of a spurious effect. An additional hint is that when including both classical suppressors leads to a very small increase of the coefficient of determination of a multiple regression, the model is less likely to be robust to outliers and estimations on subsamples than when they contribute to a large increase of the coefficient of determination (i.e. a reduction of the root mean square error of the residuals).

But, if, when removing outliers, the t -test of the null hypothesis rejects $H_0: \beta_{12} = 0$, whereas the t -test does not reject the null hypothesis $H_0: r_{12} = 0$, the researcher is still facing conflicting outcomes between the two t -tests to decide between a spurious effect or a non-spurious effect model. So far, we leave to further research an identification strategy for case 2 (where both $H_0: r_{23} = 0$ and $H_0: r_{13} = 0$ are rejected), which is more likely to be robust to outliers as the variance $Var(\mathbf{x}_3 - r_{23}\mathbf{x}_2) = 1 - r_{23}^2$ is relatively larger than in case 3.

5 How to prevent spurious regressions

We first suggest that editors require the systematic disclosure of the correlation matrix including the simple correlation coefficients with the dependent variable of all regressors, including interaction terms, lags and squares. Following Spanos (2006), we suggest that applied researchers perform preliminary tests on simple correlation coefficients of regressors with the dependent variable. A spurious regression may occur when at least one of the tests of null hypothesis of a negligible effect of each of the regressors indexed by j ($2 \leq j \leq k+1$) on the dependent variable ($H_0 : r_{1j} < 0.1$ when $r_{1j} > 0$, or $H_0 : r_{1j} > -0.1$ when $r_{1j} < 0$) is not rejected (say for the regressor indexed by j'), and such that the test of a null effect of this regressor in a multiple regression is rejected ($H_0 : \beta_{1j'} = 0$). The threshold 0.1 implies that the true correlation coefficient should explain at least 1% of the variance of the dependent variable in a simple correlation model (the coefficient of determination is such that: $R_{1,j}^2 > 1\%$). It refers to Cohen's [1988] (pp.79-81) classification of effects in his evaluation of the power of tests for cross sections where at least a correlation of $r_{1j} = 0.1$ or $r_{1j}^2 = 1\%$ is required to consider a meaningful small effect on the dependent variable. These tests are available in statistical softwares, such as SAS 9.3 (the instruction is: `proc corr data=database fisher (rho0=0.1 lower);`). The test of a null effect ($H_0 : \beta_{1j'}^T = 0$) is also feasible. But, for very large samples ($N > 1000$), it will reject the null hypothesis for very small sample values of the correlation coefficient. One may also use a rule of thumb excluding regressors with sample correlation coefficient such $|r_{12}| < 0.1$.

Our paper proposes additionally to calculate the parameter inflation factor *PIF* defined below. Three indicators have been proposed in the literature to detect near-collinearity: the determinant of the correlation matrix between regressors $\det(\mathbf{R}_2)$, the variance inflation factor (*VIF*) and the condition index *CI*. In the trivariate case, the traditional indicators depend only on the correlation coefficients between the two explanatory variables, r_{23} :

$$\det(\mathbf{R}_2) = \lambda_{\max}\lambda_{\min} = 1 - r_{23}^2, \quad VIF = \frac{1}{\det(\mathbf{R}_2)} = \frac{1}{1 - r_{23}^2}, \quad CI = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} = \sqrt{\frac{1 + r_{23}}{1 - r_{23}}} \quad (14)$$

where $\lambda_{\max} = 1 + r_{23}$ and $\lambda_{\min} = 1 - r_{23}$ are the two eigenvalues of the correlation matrix of the regressors $\mathbf{R}_2$. For example, a high correlation among regressors can be defined by a unique rule of thumb such as $r_{23} \geq 0.8$ so that $\det(\mathbf{R}_2) < 0.36$ or $VIF > 2.7$ or $CI > 3$. A measure able to evaluate whether coefficients are oversized or not, is also needed for referees and journal editors. A reasonable starting point for highlighting this problem is what

we call the *parameter inflation factor* (or PIF_{12}) as the ratio of the multiple correlation standardized parameter β_{12} and the correlation coefficient r_{12} . It is also equal to the ratio of the non standardized multiple regression parameter $\beta_{1.2/3\dots k}^{NS}$ and the non standardized parameter of the simple regression $\beta_{1.2}^{NS}$.

$$PIF_{12} = \frac{\beta_{1.2/3\dots k}}{r_{12}} = \frac{\beta_{1.2/3\dots k}^{NS}}{\beta_{1.2}^{NS}}. \text{ For } k = 2 : PIF_{12} = \left(1 - \frac{r_{13}}{r_{12}}r_{23}\right) VIF_{12}. \quad (15)$$

The second equation is the PIF -formula for the trivariate case. As compared to the variance inflation factor (VIF) which depends only on correlation coefficients between the regressors, the PIF takes also into account the vector r_{1j} of correlations of the regressors with the dependent variable. It measures if the numerator of the multiple correlation coefficient is *sufficiently large to benefit from the multiplier effect* of the denominator (the VIF) in case of high correlation among regressors. Highly correlated classical suppressors are obvious candidates for the highest values of the PIF .

6 The identification of spurious regressions: aid, policies and growth

The following example shows how the use of the PIF , tests on correlation coefficients and robust regressions would have helped referees to evaluate the Burnside and Dollar (denoted BD) (2000) paper. BD use ordinary least squares estimates on an unbalanced panel of 56 countries over 6 four-years periods between 1970 and 1993 ($N = 275$ observations). In BD first regression, real per capita gross domestic product (GDP) growth rate is a function of three macroeconomic policy variables (budget surplus, inflation and openness) and other control variables. Then, BD decide to constrain the estimated parameters found in this first regression in order to define a macroeconomic policy index:

$$\text{Policy} = 1.28 + 6.85 * \text{Budget Surplus} - 1.40 * \text{Inflation} + 2.16 * \text{Openness}. \quad (16)$$

Table 1 reports three regressions (BD1, BD2, BD3) from BD tables and two additional regressions (CR1, CR2) that we run using their data set. In regression BD1, the regressors are the policy variable (instead of the separate variables budget surplus, inflation, openness) with exactly the same other control variables than in their initial regression: by construction, the estimated parameter for the policy index is equal to one. The Aid/GDP

estimated parameter is close to zero (0.034) with a very large standard error. Then in regression BD2, their main result is that real per capita gross domestic product (GDP) growth depends significantly at the 5-percent level on two highly correlated interactions terms: $(\text{Aid/GDP}) \times \text{Policy}$ and $(\text{Aid/GDP})^2 \times \text{Policy}$, leading to the recommendation that foreign aid should be linked to good macroeconomic policy.

But firstly, both regressors are weakly correlated with the dependent variable but highly correlated among themselves: $r_{12} = 0.128$, $r_{13} = 0.058$, $r_{23} = 0.92$, with $PIF_{12} = 0.203/0.095 = 2.1$ and $PIF_{13} = -0.019/0.00458 = -4.2$ with sign reversal. This indicates that there might be a problem of a spurious regression. Secondly, the auxiliary regression of $(\text{Aid/GDP})^2 \times \text{Policy}$ on $(\text{Aid/GDP}) \times \text{Policy}$ has an extremely large studentized residual which exceeds 10.7 for Gambia 1986-1989. This indicates that there might be an outlier problem. According to Belsley, Kuh, and Welsch (1980), observations with studentized residual larger than 1.96 in absolute value should be investigated. The residuals variable orthogonal to $(\text{Aid/GDP}) \times \text{Policy}$ (denoted $\varepsilon_{3.2} = \mathbf{x}_3 - r_{23}\mathbf{x}_2$ in section 2) has a simple correlation coefficient with a dummy for Gambia 1986-1989 equal to 0.57. The Gambia 1986-1989 dummy variance explains one third ($R^2 = 0.57^2 = 32.5\%$) of the variance of this residual variable. When we run the growth regression excluding the observation of Gambia 1986-1989 (CR1, table1), the two regressors $(\text{Aid/GDP}) \times \text{Policy}$ and $(\text{Aid/GDP})^2 \times \text{Policy}$ are no longer both statistically significant. Thirdly, BD used the White heteroskedasticity consistent standard errors in regression BD2 including Gambia 1986-1989: this is not useful when heteroskedasticity is driven by a large outlier. With ordinary least square standard errors, both regressors $(\text{Aid/GDP}) \times \text{Policy}$ and $(\text{Aid/GDP})^2 \times \text{Policy}$ are no longer statistically significant with the specification BD2, including the observation Gambia 1986-1989. Fourthly, the marginal contribution to the coefficient of determination R^2 of including $(\text{Aid/GDP}) \times \text{Policy}$ and $(\text{Aid/GDP})^2 \times \text{Policy}$ is very small in regression BD2 with respect to regression BD1 which excludes these regressors: $0.3981 - 0.3918 = 0.0063$. This contribution to R^2 enters into the computation of the power of the Fisher test for including $(\text{Aid/GDP}) \times \text{Policy}$ and $(\text{Aid/GDP})^2 \times \text{Policy}$ which is then *only* 30% (hence, a p -value of type II error of 70%) for a threshold of 5% for the type I error, whereas at least 80% is usually required according to Cohen (1988) (the estimated power is computed using the procedure *power* with SAS). Finally, removing sequentially up to 14 additional outliers ranked by their leverage on $(\text{Aid/GDP}) \times \text{Policy}$ (which are distant from the mean of $(\text{Aid/GDP}) \times \text{Policy}$ from at least 1.48 times its standard error), we were never able to obtain the simultaneous statistical significance of each of the regressors $(\text{Aid/GDP}) \times \text{Policy}$ and $(\text{Aid/GDP})^2 \times \text{Policy}$ (those regressions

are available upon request) except when the outlier Gambia 1986-1989 is included. All these tests confirmed that the statistical significance of $(\text{Aid/GDP}) \times \text{Policy}$ and $(\text{Aid/GDP})^2 \times \text{Policy}$ in regression BD2 is not robust to outliers. This quadratic interaction term can be identified as a spurious effect.

BD also run a regression including $(\text{Aid/GDP}) \times \text{Policy}$ which is statistically significant at the 5% level (0.19*) excluding the variable $(\text{Aid/GDP})^2 \times \text{Policy}$ and 5 observations: Gambia 1986-1989, Gambia 1990-1993, Nicaragua 1990-1993, Guyana 1990-1993, Nicaragua 1986-1988 (regression BD3).² However, we run a regression excluding the next outliers with high leverage for $(\text{Aid/GDP}) \times \text{Policy}$, which are distant from the mean of $(\text{Aid/GDP}) \times \text{Policy}$ by more than 3.2 its standard error: Bostwana 1986-1989, 1982-1985 and 1978-1981. Then, the variable $(\text{Aid/GDP}) \times \text{Policy}$ is no longer statistically significant with the remaining 267 observations. When removing up to 7 remaining high leverage outliers (samples with 267 to 260 observations), the coefficient of $(\text{Aid/GDP}) \times \text{Policy}$ remains *negative* and not statistically significant.

Another assessment of the aid and policy interaction effects on GDP growth is to observe the effect of outliers on the marginal effect of policy on growth as shown in figure 6. Without interaction term, the parameter of policy is equal to one. When introducing a quadratic interaction term with Aid/GDP, the Gambia 1986-1989 dummy in regression BD2 is shifting *downwards* the effect of policy for large values of Aid/GDP (Aid/GDP has a minimum of near 0 units and a maximum of 12 units). When this observation is excluded (CR1), the quadratic parameter remains quite close to the horizontal line with value one. When removing 5 outliers (regression BD3), the Bostwana observations are shifting *upwards* the effect of policy for large values of Aid/GDP, in the *opposite direction* of the Gambia 1986-1989 outlier in regression BD2. When the three Bostwana observations are excluded, the increasing marginal effect has a slope close to zero. For the maximal value of Aid/GDP (12), Bostwana observations are driving the marginal effect of Policy up to 3 instead of 1 in regression BD3, whereas the Gambia 1986-1989 is driving the marginal effect of policy down to 0.3 instead of 1 in regression BD2. Conversely, on figure 7, the marginal effect of aid/GDP as a function of policy in the linear interaction term model (BD3) is also sky rocketing for extreme values of policy (-4.5 units as a min and $+4.5$ units as a max), which is no longer the case when the three Bostwana data are excluded (CR2).

The test of the simple correlation with the dependent variable GDP growth with $(\text{Aid/GDP})^2 \times \text{Policy}$ do not reject the null hypothesis: H_0 :

² p is the p -value of the t -test of the parameter, * indicates statistical significance at the 5-percent level.

$r_{1j}^T = 0$ at the 5-percent for $N = 275$ observations and for smaller sets of observations excluding outliers. The test: $H_0 : r_{1j}^T < 0.1$ when $r_{1j}^T > 0$ does not reject the null hypothesis at the 5-percent level for (Aid/GDP)×Policy: $r_{16} = 0.128$, p -value = 0.318 for $N = 275$ observations. The test: $H_0 : r_{1j}^T > -0.1$ when $r_{1j}^T < 0$ does not reject the null hypothesis at the 5-percent level for Aid/GDP (with negative sign): $r_{16} = -0.173$, p -value = 0.109, for $N = 275$ observations, but, for $N = 267$ observations (regression CR2), $r_{16} = -0.224$, p -value = 0.019 so that the null hypothesis is rejected at the 5-percent level, but with a negative sign. However, because of missing observations for their other regressors, BD excluded 230 (45%) observations of their data set, which in fact includes 505 observations of the pair of variables (Aid/GDP, GDP growth) with an overall simple correlation of $r_{16} = 0.032$ close to the one found in the updated sample of Doucouliagos and Paldam (2009). Hence, for $N = 505$, The t -test does not reject at the 5-percent level the hypothesis that Aid/GDP is a classical suppressor for GDP growth: $H_0 : r_{16} = 0$, p -value = 0.46. The value of the standardized parameter estimate of AID/GDP changes from the negative effect value -0.173 ($N = 275$) in a simple regression to, for example, $+0.034$ ($N = 275$) in regression BD1 (table 1, column 1). This change is mostly explained by the inclusion of the Sub-Saharan Africa and East Asia dummies in the multiple regressions of table 1.

The pre-tests of simple correlation coefficients with the dependent variable $H_0 : r_{1j} = 0$ do not reject the null hypothesis at the 5-percent level for 5 remaining regressors: Log(GDP) at the beginning of each period, Ethnic fractionalization, Assassinations, Ethnic fractionalization × Assassinations, M2/GDP lagged. In particular, there is another highly collinear pair of classical suppressors with *spurious* statistical significance at the 10-percent level in all regressions in table 1, for Assassinations and the interaction term Ethnic fractionalization × Assassinations: $r_{14} = 0.063$, $r_{15} = 0.039$, $r_{45} = 0.86$ with $PIF_{14} = 2.3$ and $PIF_{15} = -3.3$ (with sign reversal).

Finally, the conclusion that we should draw from the data and the pooled OLS regression model is that GDP growth increases with the macroeconomic policy index, the institutional quality and the East Asian countries dummy and decreases with the Sub-Saharan countries dummy. Burnside and Dollar’s paper involves further work using a larger data set and panel data estimators dealing with endogeneity and with time invariant regressors (such as initial GDP and regional dummies) instead of pooled OLS, with estimators which were already available in the mid-1990s. With respect to the endogeneity of Aid/GDP, Burnside and Dollar’s tests led them to favor pooled OLS instead of their pooled 2SLS instrumental variables estimations. Bazzi and Clemens’ (2013) critique focuses on BD’s choice of the log of population as an in-

strument for Aid/GDP in their 2SLS estimations. Although it is a strong instrument, they emphasize that the log of population has a statistically significant partial relationship with several variables other than Aid/GDP that are plausible growth determinants, as done in other studies than BD study. Bazzi and Clemens (2013) claim that the log of population cannot be used to instrument more than one relevant regressor in growth regressions.

An ideal tool for winner’s curse papers. The winner’s curse in auctions means that the person who makes the highest bid and eventually wins the auction has overpaid. Like bidding in auctions, finding unexpected and large effects in science leads to higher probability not only to be published in top journals but also to be far from the truth, fostering controversies and citations. Ioannidis [2005] finds support of winner’s curse for papers in several fields of clinical research and epidemiology. The spurious effects analyzed in this article provide *the ideal tool of winner’s curse papers*: they allow to find effects that are statistically significant (because of highly correlated classical suppressors), large (because of near-multicollinearity), unexpected (because the effect is spurious), interesting (because of interaction terms, dynamic or non-linear and quadratic models) and fostering controversies and citations (because the results are outliers driven). The fate of winner’s curse papers is to be contradicted in replications on different samples published in second-tier journals, so that, a decade later, meta-analysis concludes that there is no effect, in accordance with the saying of classical antiquity: *veritas filia temporis*. BD’s paper followed this predicted sequence. BD’s paper is the most cited paper published in the *American Economic Review* in 2000: more than 3100 citations in Google Scholar database in May 2013. Easterly, Levine, and Roodman [2004] showed that the sign on $\text{aid/GDP} \times \text{Policy}$ is not stable when including 80 observations to the BD’s sample (this section shows that it is not stable neither when excluding only 3 observations related to Botswana). A meta-analysis by Doucouliagos and Paldam [2009] includes up to 355 estimates from 31 articles of this literature including a quadratic term $(\text{Aid/GDP})^2$ or interaction terms of Aid/GDP with various policies measures. They conclude that “*the aggregate coefficient to the interaction between foreign aid and policy proves to be very close to zero*”. This paper states that this information is directly available in the near-zero simple correlation of Aid/GDP with GDP growth.

7 Conclusion

Highly correlated regressors weakly correlated with the dependent variable may foster the publication of spurious regressions over-fitting outliers with

statistically significant large estimated parameters. *PIF and tests on simple correlation coefficients with the dependent variable are useful tools to prevent this issue.*

References

- [1] Bazzi S. and Clemens M.A. [2013]. Blunt Instruments: Avoiding Common Pitfalls in Identifying the Causes of Economic Growth. *American Economic Journal: Macroeconomics*, 5(2), 152-186.
- [2] Belsley, D. A., Kuh, E. and Welsch, R.E. [1980]. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley and Sons.
- [3] Burnside C. and Dollar D. [2000]. Aid, policies and growth. *American Economic Review*, 90, 847-68.
- [4] Chatelain J.B. [2010]. Can statistics do without artefacts. *Cournot Centre Prisme*, 19, 1-39.
- [5] Chatelain J.B. and Ralf K. [2010]. "Inference on Time-Invariant Variables using Panel Data: A Pre-Test Estimator with an Application to the Returns to Schooling," PSE Working Papers hal-00492039, HAL.
- [6] Cohen J. [1988]. *Statistical Power Analysis for the Behavioral Sciences*. 2nd edition. Lawrence Erlbaum Associates, Hillsdale.
- [7] Doucouliagos H. and Paldam M. [2009]. Conditional aid effectiveness: A meta study. *Journal of International Development*, 22, 391-410.
- [8] Easterly, W., R. Levine, and D. Roodman. [2004]. New Data, New Doubts: A Comment on Burnside and Dollars Aid, Policies, and Growth (2000). *American Economic Review*, 94, 774-780.
- [9] Friedman, L. and Wall, M. [2005]. Graphical views of suppression and multicollinearity in multiple linear regression. *The American Statistician*, 59, 127-136.
- [10] Hoover K.D. [2001]. *Causality in Macroeconomics*. Cambridge University Press, Cambridge.
- [11] Ioannidis J.P.A. [2005]. Contradicted and initially stronger effects in highly cited clinical research. *Journal of the American Medical Association*. 294, 218-228.

- [12] Spanos A. [2006]. Revisiting the omitted variables argument: substantive vs. statistical adequacy. *Journal of Economic Methodology*, 13:2, 179-218.
- [13] Spanos A. and McGuirk A. [2002]. The Problem of high correlation among regressors revisited: erratic vs systematic volatility. *Journal of Econometrics*, 108, 365-393.

8 Appendix A

Proposition 1: Conditions for type II inference discordance: Firstly, the t -test *rejects* the null hypothesis of no effect ($H_0 : \beta_{12} = 0$) in the trivariate regression. Secondly, the t -test *does not reject* the null hypothesis of no effect ($H_0 : r_{12} = 0$) between the dependent variable x_1 and a regressor x_2 in the simple regression, in the set of feasible correlation coefficients of the trivariate regression. This corresponds to the following conditions:

$$r_{13}r_{23} + \gamma(\alpha, N, 2) \sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2} \leq r_{12} \leq \min \left(\gamma(\alpha, N, 1), r_{13}r_{23} + \sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2} \right) \text{ or} \quad (17)$$

$$\max \left(-\gamma(\alpha, N, 1), r_{13}r_{23} - \sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2} \right) \leq r_{12} \leq r_{13}r_{23} - \gamma(\alpha, N, 2) \sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2}. \quad (18)$$

Proposition 2: Conditions for type I inference discordance: Firstly, the t -test does not reject the null hypothesis of no effect ($H'_0 : \beta_{12} = 0$) between the dependent variable x_1 and a regressor x_2 , in the trivariate regression. Secondly, the t -test *rejects* the null hypothesis of no effect ($H_0 : r_{12} = 0$) between the dependent variable x_1 and a regressor x_2 , in the bivariate regression. This leads to the conditions:

$$\gamma(\alpha, N, 1) \leq r_{12} \leq r_{13}r_{23} + \gamma(\alpha, N, 2) \sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2} \quad (19)$$

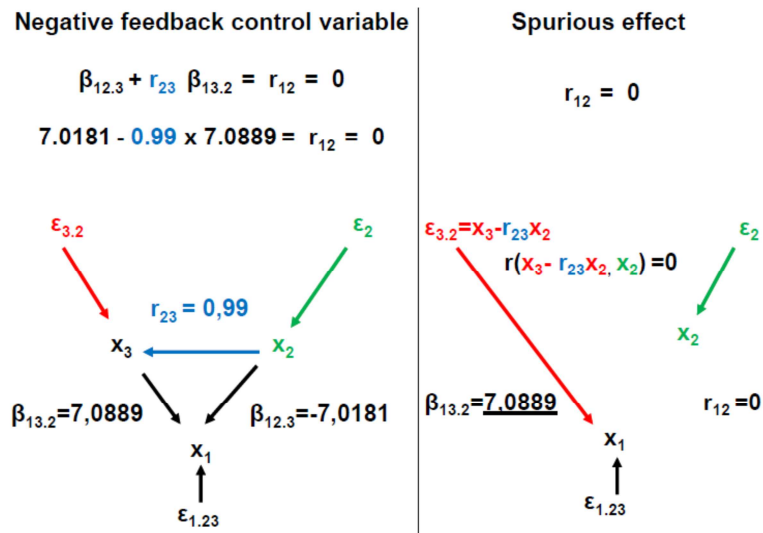
$$r_{13}r_{23} - \gamma(\alpha, N, 2) \sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2} \leq r_{12} \leq -\gamma(\alpha, N, 1). \quad (20)$$

Table 1 OLS growth regressions using data of 56 developing countries.

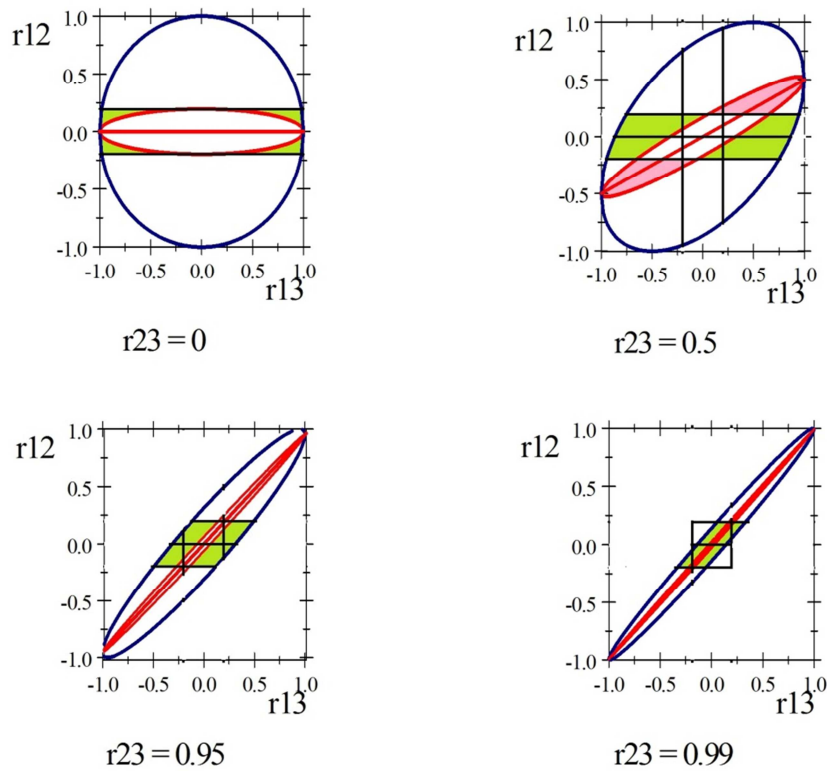
corr(x1,x(i)) for 275 and 267 obs.	BD1	BD2	CR1	BD3	CR2
Aid/GDP -0.173 and -0.224	0.034 (0.12) -0.1	0.049 (0.12) -0.2	0.054 (0.13) -0.2	-0.021 (0.16) 0.1	0.026 (0.16) -0.1
(Aid/GDP)×Policy 0.128 and 0.037	—	0.20* (0.09) 2.1	0.16 (0.11) 1.2	0.19* (0.07) 1.0	0.05 (0.10) 0.8
(Aid/GDP) ² ×Policy 0.058 and -0.014	—	-0.019* (0.0084) -4.2	-0.013 (0.013) -1.2	—	—
Policy Index 0.451 and 0.420	1.00* (0.14) 0.8	0.78* (0.20) 0.6	0.81* (0.20) 0.6	0.71* (0.19) 0.5	0.80* (0.19) 0.6
Institutional quality 0.271 and 0.241	0.64* (0.17) 0.8	0.67* (0.17) 0.8	0.67* (0.17) 0.8	0.69* (0.17) 0.9	0.63* (0.17) 0.9
Sub-Saharan Africa -0.210 and -0.256	-1.60* (0.73) 1.0	-1.84* (0.74) 1.1	-1.83* (0.74) 1.1	-1.87* (0.75) 1.1	-1.91* (0.75) 1.0
East Asia 0.320 and 0.333	0.91 (0.54) 0.2	1.20* (0.58) 0.3	1.18* (0.58) 0.3	1.31* (0.58) 0.4	1.23* (0.57) 0.3
Initial GDP 0.115 and 0.111	-0.61 (0.56) -1.0	-0.56 (0.56) -0.9	-0.56 (0.56) -1.0	-0.60 (0.57) -1.1	-0.66 (0.57) -1.2
M2/GDP (lagged) 0.017 and 0.027	0.014 (0.013) 3.0	0.016 (0.014) 3.5	0.014 (0.014) 3.2	0.012 (0.014) 1.5	0.008 (0.014) 1.1
EF: Ethnic fractionalization -0.083 and -0.099	-0.54 (0.72) 0.5	-0.42 (0.73) 0.4	-0.44 (0.72) 0.5	-0.42 (0.72) 0.4	-0.41 (0.71) 0.4
Assassinations (A) -0.063 and -0.061	-0.44 (0.26) 2.4	-0.45 (0.26) 2.4	-0.44 (0.26) 2.4	-0.45 (0.26) 2.3	-0.43 (0.26) 2.5
EF × A -0.039 and -0.038	0.82 (0.44) -3.5	0.80 (0.44) -3.4	0.80 (0.44) -3.4	0.79 (0.44) -3.1	0.73 (0.44) -3.3
Observations	275	275	274	270	267
R^2	0.3918	0.3981	0.3985	0.3944	0.3797

Notes: The dependent variable is real per capita GDP growth. Correlation coefficients of each regressor with GDP growth rates for N=275 and N=267 observations are below the names of the regressors. Below estimated parameters, White heteroskedasticity consistent standard errors are in parentheses and PIF values below standard errors. PIF absolute values over 2 are in bold. Regressions BD1, BD2 and BD3 are in Burnside and Dollar [2000] article (there is a typo in BD's article for regression BD2: the parameter is 0.049 for the variable Aid/GDP). Regression CR1 exclude the outlier Gambia 1986-89 with respect to BD2. Regression CR2 exclude three outliers related to Bostwana with respect to BD3. * statistical significance at the 5-percent level.

Figure 1 : Parameter identification problem

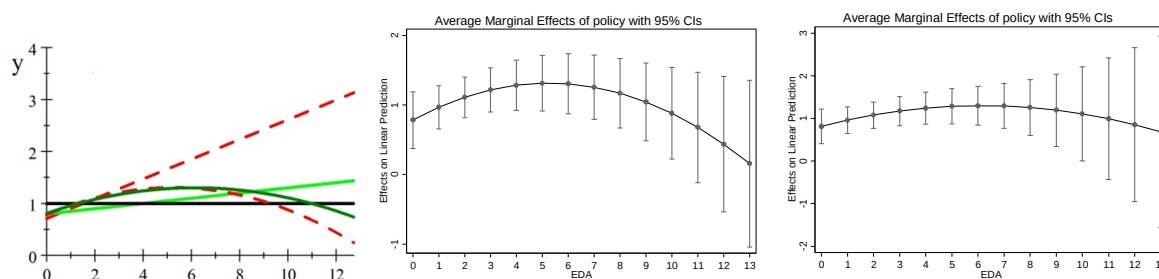


Figures 2, 3, 4 and 5: The rejection regions for the tests $H_0: r_{12}=0$, $H_0: r_{13}=0$, $H_0: \beta_{13}=0$ for $N=102$, $t=2$, $r_{12}=f(r_{13})$, $r_{23}=0$ (graph 1) then 0.5 (graph 2), 0.95 (graph 3) and 0.99 (graph 4).



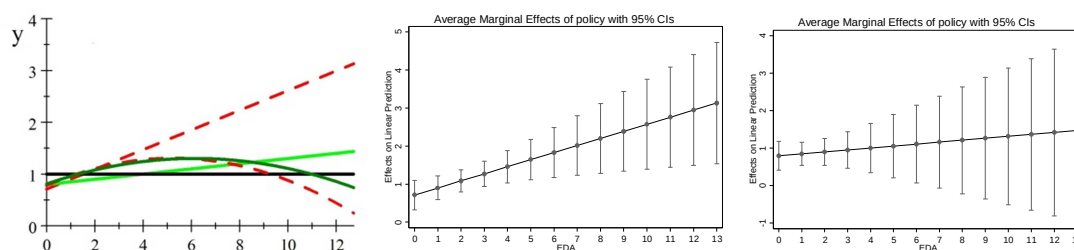
Feasible correlations coefficients are inside the large ellipse. The major axis of the small ellipse corresponds to the null hypothesis for the multiple regression $H_0: \beta_{12}=0$. The rejection region of the t-test of $H_0: \beta_{12}=0$ lies inside the large ellipse and outside the red ellipse. The region of acceptance of the t-tests of $H_0: r_{12}=0$ and $H_0: r_{13}=0$ is the central square $[-0.196, 0.196]^2$.

Figure 6: Coefficient on policy as a function of Aid/GDP without interaction term (horizontal line with value 1) or with a linear interaction term including Bostwana observations (dash line) or not (solid line) or with a quadratic interaction terms including Gambia 86-89 with (dash curve) or without (solid curve). The range of the horizontal axis is the min and max of aid/GDP.



Quadratic interaction terms: Including Gambia 86-89 Excluding Gambia 86-89

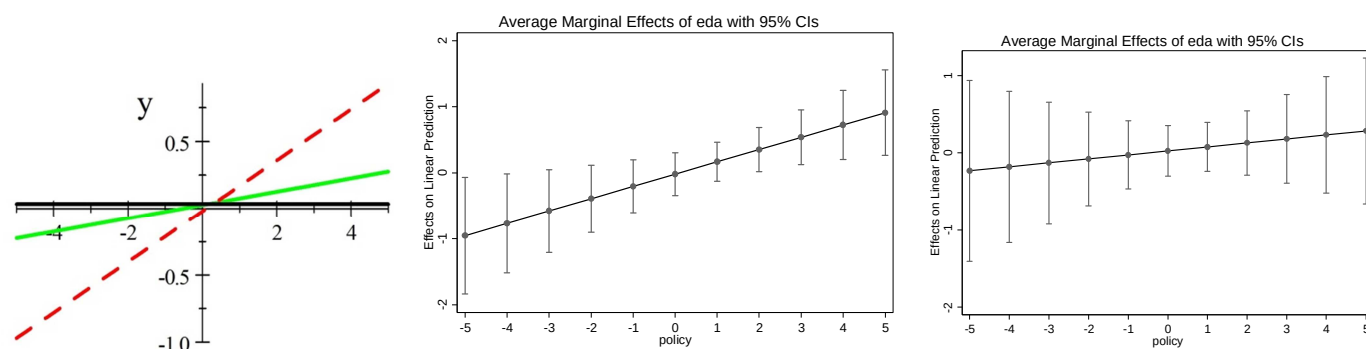
With quadratic interaction, the sensitivity equal to one is included in the 95% confidence interval even when including Gambia 86-89 observation: there is no stastical significance of the quadratic interaction term.



Linear interaction terms: dash red line including Bostwana green line excluding Bostwana

With linear interaction term, the sensitivity equal to one is included in the 95% confidence interval except when excluding 5 outliers but including 3 Bostwana outliers, when Aid/GDP exceeds 5 units.

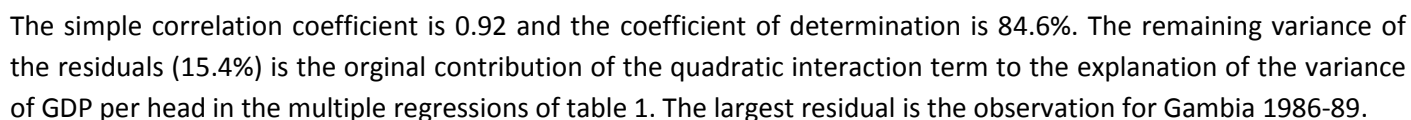
Figure 7: Coefficient on Aid/GDP as a function of policy without (horizontal line, value 0.034) interaction term or with for linear interaction term including Bostwana observations (dash line) or not (solid line). The range of the horizontal axis is the min and max of policy.



Excluding 5 outliers and including Bostwana excluding Bostwana

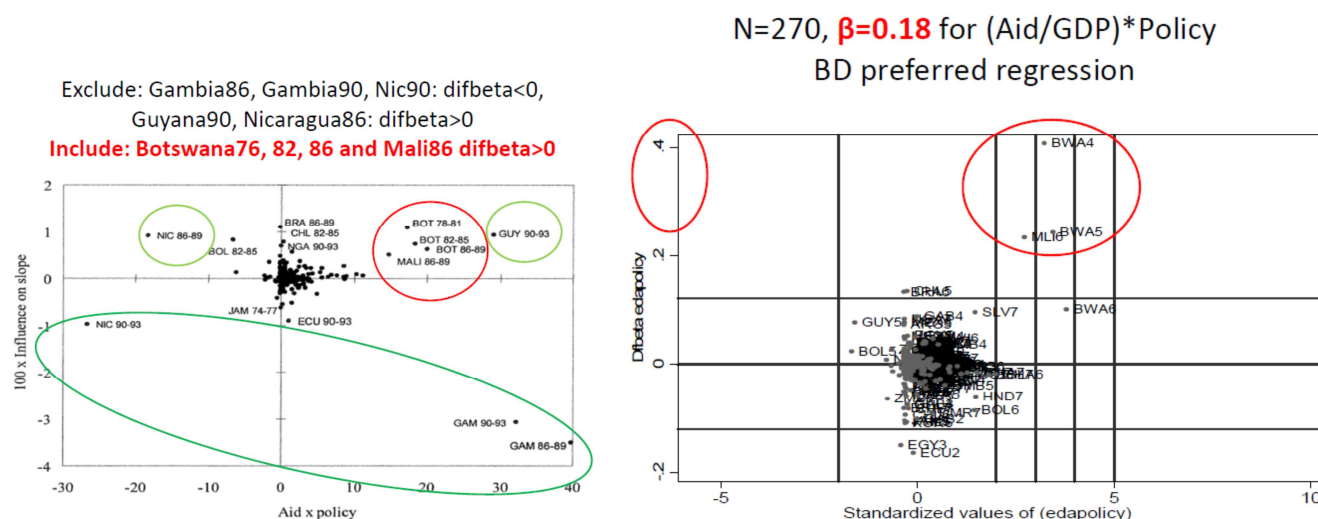
With linear interaction term, the sensitivity equal to one is included in the 95% confidence interval except when excluding 5 outliers but including 3 Botswana outliers for Policy < -2 or for Policy > 2.

Figure 8: Quadratic interaction term correlation with the simple interaction term.



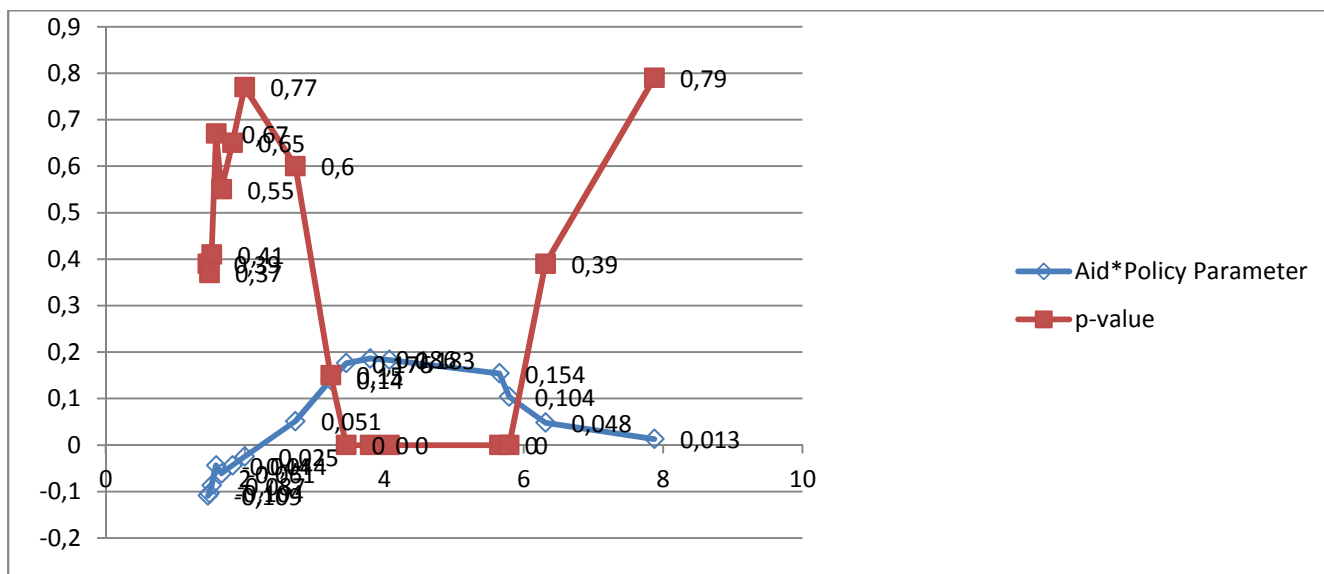
The correlation coefficient is 0.57 and the coefficient of determination of this simple regression 32.5%. 32.5% of the variance of the quadratic interaction term is brought by a dummy for Gambia 1986-89. An alternative specification in table 1 consists of substituting the quadratic interaction term by a dummy for Gambia 1986-89.

Figure 10 : Linear interaction term: Difbeta of the parameter of aid x policy as a function of aid x policy



When excluding 5 outliers, Botswana 78-81, Botswana 82-85 and Mali 86-89 outliers have still a very large positive influence on the estimated parameter of Aid x Policy according to Belsley, Kuh and Welsch (1980) criterion.

Figure 11 : Linear interaction term: estimated parameter of the Aid x Policy parameter and its p-value as a function of the absolute value of the difference of aid x policy observations to the mean of aid x policy divided by the standard error of aid x policy for the full sample (N=275 observations).



The largest aid x policy value (Gambia86-89) is nearly 8 standard errors far from the mean of aid x policy. Dots corresponds to a cumulative removal of high leverage observations (observations far from the mean) of aid x policy, up to 15 observations. The maximal values of the aid x policy parameters and its statistical significance obtain in removing the 5 outliers with highest leverage but including the next 3 outliers with highest leverage (3 observations related to Botswana). Once Botswana observations are excluded, the parameter is no longer statistically significant with values close to zero and finally negative.