Spurious Regressions and Near-Multicollinearity, with an Application to Aid, Policies and Growth

Chatelain, Jean-Bernard and Ralf, Kirsten

Centre d'Economie de la Sorbonne, Paris School of Economics, ESCE Ecole Superieure du Commerce Exterieur

8 November 2012

Spurious Regressions and Near-Multicollinearity, with an Application to Aid, Policies and Growth*

Jean-Bernard Chatelain† Kirsten Ralf‡

November 8, 2012

Abstract

In multiple regressions, explanatory variables with simple correlation coefficients with the dependent variable below 0.1 in absolute value (such as aid with economic growth) may have very large and 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 "Botswana outliers driven" Burnside and Dollar [2000] article which found that aid had an effect on growth only for countries achieving "good" macroeconomic policies.

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

Keywords: Near-Multicollinearity, Student t-Statistic, Spurious regressions, Ceteris paribus, Parameter Inflation Factor, Growth, Foreign Aid.

*We thank for helpful comments Glenn Shafer, Robert Chirinko, Michel Armatte, Xavier Ragot, Jean-Philippe Touffut, Chirstophe Hurlin and Giacomo Corneo as well as participants to sessions at AFSE conference in Paris and probability and statistical days in Mainz and to seminars at Cournot centre, Freie Universität Berlin and Reading universities.

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

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

1. Introduction

Should applied researchers include in multiple regressions explanatory variables which have a simple correlation close to zero with the dependent variable? Since Horst [1941], a regressor which is not correlated with the dependent variable is called a "classical suppressor variable" (Cohen *et al.* [2003], Friedman and Wall [2005], Christensen [2006])). For Yule [1897], a classical suppressor variable is not considered to be a problem in a multiple regression. Since Fisher [1925] introduced hypothesis testing in multiple regressions assuming the normality of the disturbances, the decision rule is that, if a "classical suppressor" variable is statistically significant, it should be included in the multiple regression. By contrast, this paper presents a number of results for being particularly cautious when including a statistically significant classical suppressor variable in multiple regressions.

Firstly, a classical suppressor variable, which is usually not statistically significant in a simple regression, may become statistically significant in a multiple regression, if the multiple regression includes *another "classical suppressor" which is highly correlated with the first one.* This fact remained unnoticed,as the literature tends to over-emphasize that highly correlated variables lead to a "low power of the t-test" (Silvey [1969], Belsley [1991], Mason and Perreault [1991], Hill and Adkins [2001]. Spanos and McGuirk [2002] is an exception with this respect.

The second point is that applied researchers can easily include another highly correlated suppressor variable, such as a lag, a square or an interaction term of the first classical suppressor. Those specifications are usually interesting dynamic or non-linear models which are frequently published.

Unfortunately, and this is the third point, a regression model including two highly correlated classical suppressor variables presents two interpretations which cannot be decided upon by the regression model only (Hoover [2001]). The first interpretation is that the regression is related to a homeostatic model with negative feedback between the two classical suppressors. In this case, the effect of a classical suppressor on the dependent variable is completely offset by a highly correlated negative feedback response of the other classical suppressor, so that the dependent variable remains invariant. The second interpretation is that at least one of the two classical suppressor variable has no effect at all on the dependent variable, which is shown after orthogonalization of the explanatory variables. In the second case, there is a spurious regression effect of the first classical suppressor variable in the regression, independently from its statistical significance.

The fourth point is that, even if the true model is the homeostatic model, and even if the estimated parameters are highly statistically significant (there is a high power of the t-test in the sample), the highly correlated classical suppressor regressors have large estimated parameters which are highly sensitive to a few observations with high leverage, that are observations which are far form the mean of those regressors. A first consequence is that these large parameters will be unstable and are likely to change signs including or excluding a few of these observations. A second consequence is related to the interpretation of the model. The applied researcher may claim that a large effect of the classical suppressor variable on the dependent is valid on the full population, whereas it is only driven by a few outliers.

The fifth point is related to the publication biases in top journals (Stanley [2005], Ioannidis [2008]). First, journals select papers which presents statistically significant effects. Second, top journals may favor unexpected, novel and controversial correlations. A spurious regression is likely to present unexpected correlations. Third, the instability of large parameters depending on a few observations may foster controversies, citations and the journal impact factor. Fourth, a pair of classical suppressors may appear in interesting specifications such as dynamic and non-linear models with interaction terms. For all these reasons, the publications of regressions including a pair of classical suppressors are likely to be widespread.

The sixth point provides an example of the above “winners’ curse" phenomenon (Ioannidis [2008]) in the field of aid, policies and growth, namely the paper by Burnside and Dollar [2000] (more than 3000 citations in Google Scholar database in 2012). Foreign aid is a classical suppressor variable for economic growth. Two interaction terms where included in the paper to reach statistical significance for a pair of classical suppressors. Dozens of papers followed their insight and controversies flourished. Over the last 15 years, the process was likely to divert researchers scarce resources from other lines of research and from more robust specifications.

The paper proceeds as follows. Section 2 investigates how spurious regressions may occur with classical suppressor variables. Section 3 presents graphical views of the critical regions of the t-test including or not an additional regressor and depending on the correlation among regressors. Section 4 proposes tools for practitioners to avoid these spurious regressions. Section 5 applies these tools to Burnside and Dollar’s paper. Section 6 concludes.

2. Spurious regression with classical suppressor variable

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:

$$\mathbf{x}_1 = \beta_{12}\mathbf{x}_2 + \beta_{13}\mathbf{x}_3 + \boldsymbol{\varepsilon}_{1.23} \tag{1}$$

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 $2 \leq i \leq 3$, $\boldsymbol{\beta} = (\beta_{12}, \beta_{13})$ is a vector of standardized parameters to be estimated, and $\boldsymbol{\varepsilon}_{1,23}$ 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, ($E(\boldsymbol{\varepsilon}_t|\mathbf{X}_{2,3,t}) = \mathbf{0}$ and $E(\boldsymbol{\varepsilon}_t^2|\mathbf{X}_{2,3,t}) = \boldsymbol{\sigma}^\mathbf{2}$), Spanos and McGuirk [2002] derive in their theorem 1 a relation between the model parameters $(\boldsymbol{\beta}, \boldsymbol{\sigma}^\mathbf{2})$ and the primary parameters of the model defined by a vector of means and a covariance matrix. In their theorem 2, Spanos and McGuirk [2002] state that the parameterization ($\boldsymbol{\beta}$,σ) exists if and only if the determinant of the covariance matrix is positive.

The ordinary least squares estimated parameters were computed by Yule [1897] (variables with a hat denote estimated variables, except for sample simple correlation coefficients to simplify notation):

$$\begin{pmatrix} \widehat{\beta}_{12} \\ \widehat{\beta}_{13} \end{pmatrix} = \frac{1}{1-r_{23}^2} \begin{bmatrix} 1 & -r_{23} \\ -r_{23} & 1 \end{bmatrix} \begin{pmatrix} r_{12} \\ r_{13} \end{pmatrix} = \frac{1}{1-r_{23}^2} \begin{pmatrix} r_{12} - r_{13}r_{23} \\ r_{13} - r_{12}r_{23} \end{pmatrix} \tag{2}$$

Let us assume that the null hypothesis for the true simple correlation coefficient $H0 : r_{12} = 0$ **is not rejected** A zero correlation coefficient $r_{12} = 0$ implies the following relation between the parameters of the multiple regression: $\widehat{\beta}_{12} + r_{23} \cdot \widehat{\beta}_{13} = r_{12} = 0$. Let us orthogonalize the two suppressor variables using the residuals of the intermediary regression of the regressor $\mathbf{x}_3$ as a function of the regressor $\mathbf{x}_2$ ($\boldsymbol{\varepsilon}_{3.2} = \mathbf{x}_3 - r_{23}\mathbf{x}_2$):

$$\mathbf{x}_1 = \widehat{\beta}_{12}\mathbf{x}_2 + \widehat{\beta}_{13}\mathbf{x}_3 + \widehat{\boldsymbol{\varepsilon}}_{1.23} = 0.\mathbf{x}_2 + \widehat{\beta}_{13}\left(\mathbf{x}_3 - r_{23}\mathbf{x}_2\right) + \widehat{\boldsymbol{\varepsilon}}_{1.23}. \tag{3}$$

In the second equality, the 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 $\mathbf{x}_2$. Hence, the orthogonalized regressors multiple regression shows that there is no effect of the variable $\mathbf{x}_2$ on the dependent variable $\mathbf{x}_1$

Let us now assume the particular case of a high correlation r_{23} between the two regressors. The higher the correlation among regressors, the lower the variance of the residual, which is much smaller than the variance of each standardized regressors when those ones are highly correlated: $Var\left(\mathbf{x}_3 - r_{23}\mathbf{x}_2\right) = 1 - r_{23}^2 < Var\left(\mathbf{x}_3\right)$. After orthogonalization, the problem of near-collinearity turns out to be the problem of the relatively small variance of the unique regressor $\mathbf{x}_3 - r_{23}\mathbf{x}_2$ with respect to the dependent variable. An explanatory variable which has most of its observations very close to its mean leads to a simple regression which is highly sensible 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).

A second drawback of a second regressor highly correlated with a classical suppressor is that it also turns to be a classical suppressor where the null hypothesis $H0 : r_{13} = 0$ is not rejected. This intuitive property is derived later on from the determinant of the correlation matrix which has to be positive. Hence, the orthogonalization can be done the other way round with the residual of the following regression ($\boldsymbol{\varepsilon}_{2.3} = \mathbf{x}_2 - r_{23}\mathbf{x}_3$), and both regressors may turn to have potentially spurious effects.

By contrast, the first equality leads to a potentially large "*ceteris paribus*" effect of $\mathbf{x}_2$ with $\widehat{\beta}_{12} = -r_{23} \cdot \widehat{\beta}_{13}$, with a non linear increasing effect of the correlation coefficient between regressors r_{23}. There is a potentially large effect $\widehat{\beta}_{13}$ of $\mathbf{x}_3$ on the dependent variable $\mathbf{x}_1$ with the opposite sign with respect to $\widehat{\beta}_{12}$ when $r_{23} > 0$. In this equation, the homeostatic model with a negative feedback of the variable $\mathbf{x}_2$ on $\mathbf{x}_3$ (or the reverse) is a possible interpretation. The "*ceteris paribus*" interpretation of the large estimated parameters $\widehat{\beta}_{12}$ and $\widehat{\beta}_{13}$ is meaningless, as a change of $\mathbf{x}_2$ leads to an immediate change of $\mathbf{x}_3$.

Hoover [2001] states that these two models (the spurious one versus the homeostatic one with a negative feedback among regressors) are ontologically different, and that the parameter estimation does not allow to decide which model is the correct one. However, if the t-test states that statistical significance is not obtained for $\widehat{\beta}_{12}$ and $\widehat{\beta}_{13}$ for highly correlated classical suppressors, then the issue of the interpretation of the estimated parameters does not matter in practice. By contrast, if the t-test states that statistical significance is easily obtained for $\widehat{\beta}_{12}$ and $\widehat{\beta}_{13}$ for highly correlated classical suppressors, then the applied researcher has no formal clue on how to decide between the two interpretation of the estimated parameters For this purpose, the next section investigates how the critical region of the t-test evolves when the correlation between regressors increases.

3. The Changes of the Critical Region of the t-Test when the Correlation among Regressors Increases

3.1. The shapes of the critical region

Let $\mathbf{R}_3$ be a sample correlation matrix whose entries are the correlation coefficients of all pairs of variables, including the dependent variable on the first row and column. The submatrix $\mathbf{R}_2$ corresponds to the correlation matrix of the regressors. 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 (valid for k variables, here $k = 3$) is related to the coefficient of determination $R_{1.23}^2$:

$$\begin{aligned} 0 &\leq \det(\mathbf{R}_3) = \left(1 - R_{1.23}^2\right) \det(\mathbf{R}_2) \leq \det(\mathbf{R}_2) = 1 - r_{23}^2 \leq 1 \\ R_{1.23}^2 &= 1 - \frac{\det(\mathbf{R}_3)}{\det(\mathbf{R}_2)} = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2} = 1 - \left(1 - r_{12}^2\right)\left(1 - r_{13.2}^2\right) \end{aligned}$$

Assuming the normality of the disturbances, Fisher [1925] computed the estimated standard deviations of estimated parameters $\widehat{\sigma}_{\widehat{\beta_{12}}}$ and their Student's t-statistics $t_{\widehat{\beta_{12}}}$:

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

Feasible correlation coefficients and exact regression boundary. The condition on correlation coefficients $\det(\mathbf{R}_3) \geq 0$ is:

$$\det(\mathbf{R}_3) = -(1 + r_{23}) \left(\frac{r_{12} - r_{13}}{\sqrt{2}}\right)^2 - (1 - r_{23}) \left(\frac{r_{12} + r_{13}}{\sqrt{2}}\right)^2 + 1 - r_{23}^2 \geq 0. \quad (4)$$

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. Hence, 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 $0 < r_{23} < 1$ and varying r_{13} and r_{12}, $\det(\mathbf{R}_3) = 0$ describes an ellipse centered at the origin ($r_{12} = r_{13} = 0$). When $r_{23} > 0$, the major axis has a length of $2\sqrt{1 + r_{23}}$ and is on the line $r_{12} = r_{13}$, i.e. it has a slope of one. The minor axis has a length of $2\sqrt{1 - r_{23}}$ and is on the line $r_{12} = -r_{13}$. When $r_{23} < 0$, the major axis is $r_{12} = -r_{13}$ and the minor axis is $r_{12} = r_{13}$.

The large blue ellipses in figures 1 to 4 are all combinations of r_{12} and r_{13} for which $\det(\mathbf{R}_3) = 0$. The values of r_{23} are chosen to be equal to 0, 0.5, 0.95 and 0.99, in figures 1 to 4 respectively. All possible values of correlation coefficients of r_{12} and r_{13} have to be inside the ellipse or on its border. When $r_{23} = 0$, the ellipse is a circle centered at the origin ($r_{12} = r_{13} = 0$) with a radius of 1 (figure 1). As seen on figures 1 to 4, when the correlation coefficient r_{23} increases from zero to one, the width of the minor axis decreases. Hence, for $r_{23} = 1$ (exact positive collinearity), the ellipse degenerates into the segment of the line $r_{12} = r_{13}$, defined on $[-1, +1]$. When $r_{23} = -1$ (exact negative collinearity), 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.

Critical region for the test of the null hypothesis $H_0 : r_{12} = 0$ against $H_1 : r_{12} \neq 0$. The sample distribution of the correlation coefficient has been found by Fisher [1921]. The complement of the critical region of the test with the null hypothesis $H_0 : r_{12} = 0$ can be stated as:

$$-1 \leq -\frac{t_{r_{12}}(\alpha)}{\sqrt{N - 1 + t_{r_{12}}(\alpha)^2}} \leq r_{12} \leq \frac{t_{r_{12}}(\alpha)}{\sqrt{N - 1 + t_{r_{12}}(\alpha)^2}} \leq 1.$$

The border of the critical region is depicted as two horizontal lines over and above the horizontal line where $H_0 : r_{12} = 0$ (see figures 1 to 4). A sample correlation coefficient which is between these two lines leads to accept the null hypothesis $H_0 : r_{12} = 0$ for a given probability of type I error. The area of the critical region for rejecting 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. The complement of the critical region of the test with the null hypothesis $H_0 : r_{13} = 0$ against the alternative $H_1 : r_{13} \neq 0$ is given by similar inequalities. The border of the critical region is depicted as two vertical lines on both sides of the vertical line where $H_0 : r_{13} = 0$.

Critical region for the test of the null hypothesis $H_0 : \beta_{12} = 0$ against the alternative hypothesis $H_1 : \beta_{12} \neq 0$. The null hypothesis H_0: $\beta_{12} = 0$ is equivalent to $r_{12} = r_{23}r_{13}$. This condition defines a line through the origin with a slope of r_{23} in the plane (r_{12}, r_{13}). It reaches the limits $r_{13} = -1$ for $r_{12} = -r_{23}$ and $r_{23} = 1$ for $r_{12} = r_{23}$. It is depicted in figures 1 to 4 as a green line. When the two regressors are orthogonal ($r_{23} = 0$), the segment describing the null hypothesis H_0: $\beta_{12} = 0$ is the horizontal line. When $r_{23} > 0.95$ (high correlation among regressors), the segment describing the null hypothesis H_0: $\beta_{12} = 0$ is close to the major axis of the blue ellipse, the slope of which is equal to one.

The critical region of the test of the null hypothesis $H_0 : \beta_{12} = 0$ against $H_1 : \beta_{12} \neq 0$ amounts to define a critical region for the test of the partial correlation coefficient $r_{12.3}$ with a given type I error, e.g. $\alpha = 5\%$, related to the percentile $t_{\widehat{\beta_{12}}}(\alpha)$ (the sample distribution of the partial correlation coefficients has been found by Fisher [1924]). The t-statistics is related to the partial correlation coefficient as follows:

$$-1 < r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} = \frac{t_{\widehat{\beta_{12}}}}{\sqrt{t_{\widehat{\beta_{12}}} + N - 2}} = \frac{\sqrt{1-r_{23}^2}}{\sqrt{1-r_{13}^2}}\widehat{\beta}_{12} < 1$$

The complement of the critical region to reject the null hypothesis is defined by:

$$\begin{aligned} -1 &\leq -\frac{t_{\widehat{\beta_{12}}}(\alpha)}{\sqrt{N-2+t^2_{\widehat{\beta_{12}}}(\alpha)^2}}\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2} \leq r_{12} - r_{13}r_{23} \leq \\ &\leq \frac{t_{\widehat{\beta_1}}(\alpha)}{\sqrt{N-2+t^2_{\widehat{\beta_{12}}}(\alpha)^2}}\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2} \leq 1. \end{aligned}$$

The border of the critical region is depicted as the small red ellipse in figures 1 to 4 with the major axis defined by $H_0 : \beta_{12} = 0$, the green line. The critical region of the t test of the null hypothesis $H_0 : \beta_{12} = 0$ lies outside this red ellipse, but inside the larger blue ellipse of feasible correlation coefficients (such that $\det(\mathbf{R}_3) = 0$). There, we have all combinations of r_{12} and r_{13} for a given r_{23} for which the null hypothesis $H_0 : \beta_{12} = 0$ is rejected and that are feasible in the sense that the determinant of the correlation matrix is positive or equal to zero. The area of the critical region to reject the null hypothesis $H_0 : \beta_{12} = 0$ *increases* when the above interval for r_{12} is *reduced*, that is when: N increases and/or $r_{23}^2 \to 1$ (there is high correlation among regressors) and/or $r_{13}^2 \to 1$ (the other explanatory variable is strongly correlated with the dependent variable) and/or $t(\alpha)$ decreases (the applied researcher sets for example a threshold for the probability of type I error of $\alpha = 10\%$ instead of 5%).

3.2. Testing the violations of the hypothesis of the stability of conditional independence

Let us define a "type I inference discordance" based on the following two conditions. Firstly, 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. Secondly, the t-test does not reject the null hypothesis of no effect $(H_0\prime : \beta_{12} = 0)$ between the dependent variable x_1 and a regressor x_2, in the trivariate regression. This leads to the following formal condition:

Proposition 1: When $r_{12} > 0$ and when $r_{13}r_{23} > 0$, a type I inference discordance in the trivariate regression occurs for:

$$\begin{aligned}
\gamma(\alpha, N, 1) &\leq r_{12} \leq r_{13}r_{23} + \gamma(\alpha, N, 2)\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2} \\
\text{where } \gamma(\alpha, N, k) &= \frac{t_{r_{12}}(\alpha)}{\sqrt{N-k+t_{r_{12}}(\alpha)^2}}
\end{aligned}$$

The usual concern of econometric textbooks with near-multicollinearity is related to type I inference discordance which may occur when r_{23} is relatively large.

Let us now define a type II discordance where a "classical suppressor" turns to be statistically significant when including an additional regressor. Firstly, 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 Secondly, the t-test *rejects* the null hypothesis of no effect $(H_0 : \beta_{12} = 0)$ in the trivariate regression:

Proposition 2: When $r_{12} > 0$ and when $r_{13}r_{23} > 0$, a type II inference discordance in the trivariate regression occurs for:

$$\begin{aligned}
r_{13}r_{23} + \gamma(\alpha, N, 2)\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2} &\leq r_{12} \leq \gamma(\alpha, N, 1) \\
\text{where } \gamma(\alpha, N, k) &= \frac{t_{r_{12}}(\alpha)}{\sqrt{N-k+t_{r_{12}}(\alpha)^2}}
\end{aligned}$$

The left hand side of the above inequality has to be sufficiently small. As a consequence, when the second regressor is highly correlated with the dependent variables (r_{13} close to one in absolute value), type II inference discordance are likely to occur when the regressors are close to be orthogonal (r_{23} close to zero). Conversely, when the regressors are highly correlated (r_{23} close to one in absolute value), type II inference discordance is likely to occur when the second regressor is a classical suppressor (r_{13} close to zero).

We now discuss the occurrence of type I versus type II discordances when the correlation between regressors r_{23} increases (figures 1 to 4, for $N = 102$). These figures show the critical regions of both t-tests in bivariate and trivariate regressions, in the plane (r_{13}, r_{12}), with the the correlation coefficient r_{12} on the vertical axis as a function of the correlation coefficient r_{13} on the horizontal axis. On figure 1 to 4, type I discordance occurs for $(\widehat{r_{13}}, \widehat{r_{12}})$ above the highest red horizontal line (or below the lowest red horizontal line) and inside the red ellipse. Type II discordance occurs for $(\widehat{r_{13}}, \widehat{r_{12}})$ between the highest red horizontal line and the lowest red horizontal line, inside the blue ellipse, and outside the red ellipse.

In figure 1, with orthogonal regressors ($\widehat{r_{23}} = 0$), the estimated parameters are identical $\widehat{\beta}_{12} = \widehat{r_{12}}$. Orthogonal regressors are found in incomplete principal component regressions or with Gram-Schmidt orthogonalized regressors. The critical region for the test of the null hypothesis $H_0 : \beta_{12} = 0$ is *included* in the critical region for the test of the null hypothesis $H_0 : r_{12} = 0$. There is no type I discordance between the two tests. Type II discordance occurs when the correlation of the other regressor with the dependent variable get closer to one ($\widehat{r_{13}} \to 1$). The R^2 increases, the root mean square error decreases, so that the standard error of $\widehat{\beta}_{12}$ decreases, although the variable x_2, orthogonal to x_3, is **totally unrelated to the reduction of the root mean square error**. A limit case of type II discordance is when the second regressor is perfectly correlated with the dependent variable: $\widehat{r_{13}} = 1$. Because the residuals are all equal to zero, $R^2_{1.23} = 1$. The root mean square error component of the estimated standard error of $\widehat{\beta}_{12}$ is zero. The t-statistic tends to infinity. The difference $R^2_{1.23} - r^2_{12}$ is equal to one in the limit case $\widehat{r_{12}} = 0$. The larger is $R^2_{1.23} - r^2_{12}$, the larger is the power of the t-test of the null hypothesis $\beta_{12} = 0$ in a multiple regression (Cohen [1988]). For the sample parameters $\widehat{r_{12}} = \widehat{\beta}_{12}$ close to zero but not exactly equal to zero, the t-test will reject the null hypothesis $H_0 : \beta_{12} = 0$ with an extremely high power of the t-test when $R^2_{1.23} = 1$. Another example is to include a large number of orthogonal regressors (up to all the principal components in principal component regressions) that increase mechanically the coefficient of determination so that the root mean square error shrinks. As a consequence, the estimated standard error of the classical suppressor shrinks as well. Then, its estimated parameter turns to be statistically significant although its contribution to the explained variance in the $R^2_{1.23}$ is close to zero.

To conclude on the particular case of orthogonal regressors, we suggest that only the simple correlation critical region makes sense for doing inference (Chatelain and Ralf [2010] present a similar argument for time invariant variables in panel data). This suggestion rules out type II discordance, and amounts to exclude classical suppressors in regressions with orthogonal regressors. Software programmers (e.g. PROC REG, PCOMIT instruction in SAS) should change the computation of standard errors with incomplete principal component ordinary least square regressions.

In figure 2, the sample correlation between the two variable is equal to $\widehat{r_{23}} = 0.5$ The green line with a positive slope equal to $\widehat{r_{23}} = 0.5$ is the location of the null hypothesis $\beta_{12} = 0$. There is now an area for type I discordances The area of type II discordances increases with respect to the case of orthogonal regressors. There is little overlap of area of the type II discordance for the other parameter (reject $H_0 : \beta_{13} = 0$ and do not reject $H_0 : r_{13} = 0$) which is not represented on figure 2 for the sake of clarity. For the classical suppressor to reach statistical significance in the trivariate regression, the second regressor has to be neither highly nor weakly correlated with the dependent variable: figure 2 suggests these figures: $0.1 < \widehat{r_{13}} < 0.9$. In the orthogonal regression (equation 3), its residuals ($\boldsymbol{\varepsilon}_{2.3} = \mathbf{x}_2 - r_{23}\mathbf{x}_3$) still have a relatively large variance (0.75), so that its coefficient may be relatively precise.

In figure 3, the sample correlation between the two variables is equal to $\widehat{r_{23}} = 0.95$. The green line has a slope equal to 0.95. For type I discordances, even very large values of $\widehat{r_{12}}$ close to one, with a power of the t-test of the simple correlation effect $H_0 : r_{12} = 0$ close to one are compatible with the complement of the critical region of the t-test of coefficient of multiple correlation $H_0 : \beta_{12} = 0$, with a low power of this t-test. This is the usual concern of textbooks. This time, the area of type II discordance of the first regressor overlaps much more with the area of the type II discordance for the other regressor (near-multi-collinear pairs) than in the case of figure 2. This means that both regressors are statistically significant and highly correlated classical suppressors. In this case, they may be related to spurious inference or to an unstable estimation with outliers driven inference, even if the power of the t-test $H_0 : \beta_{12} = 0$ is close to one, for example, in the cases close to the blue ellipse.

Let us provide numerical insights on minimal values of r_{13} for reaching the critical region of statistical significance ($p = 0.05$) when $r_{23} = 0.95$ and $r_{12} = 0$ depending on the number of observations. For $\widehat{\beta}_{12}$ to be statistically significant ($p = 0.05$), it is sufficient that $|r_{13}| > 0.031$ for $N = 402$ with $\widehat{\beta}_{12} = 0.3 > r_{12} = 0$, $|r_{13}| > 0.061$ for $N = 102$ (figure 3) with $\widehat{\beta}_{12} = 0.59 > r_{12} = 0$), $|r_{13}| > 0.127$ for $N = 22$ with $\widehat{\beta}_{12} = 1.23 > r_{12} = 0$, knowing that $|r_{13}|$ cannot exceed 0.312 and $\widehat{\beta}_{12}$ cannot exceed 3.04 (blue ellipse border). In order to reach statistical significance of $\widehat{\beta}_{12}$ for a p-value at least equal to 5%, a difference of a few percentage points between the correlation coefficient r_{12} and r_{13} is sufficient, so that $|r_{12} - r_{23}r_{13}|$ is not too small.

In figure 4, the sample correlation between the two variable is equal to $\widehat{r_{23}} = 0.99$. The figure is qualitatively equivalent to figure 3. But this time, the overlap of the type II discordance for the pair of regressors is nearly complete.

Monte Carlo simulations knowing true correlation coefficients. An important issue is the sample distribution of (r_{12}, r_{13}) on the areas depicted in the figures 1 to 4 given r_{23}. This joint distribution is related to the sample distribution of the correlations coefficients (Fisher [1921]) and of the partial correlation coefficients (Fisher [1924]). Some insights of this sample distribution of type I and type II discordances are easily obtained by Monte Carlo simulations. Samples have been drawn 1000 times from a multivariate normal distribution with the true correlation coefficients given first as $r_{23}^T = 0.50$, $r_{12}^T = 0$, $r_{13}^T = -0.03$ and a sample size of $N = 102$ (case of figure 3) and then of $N = 1002$. The following tables 1 and 2 report the proportion of outcomes of tests of null hypothesis of parameters for the simple regression and for the trivariate regression at the 5% threshold for 1000 replications. The proportion of type II discordance (spurious) remains very small for a correlation coefficient of 0.5 between the two regressors.

We report in tables 3 and 4 the replications of the two previous simulations, changing only the true correlation coefficient between the two regressors to $r_{23}^T = 0.99$ (figure 4, for $N = 102$) The key result is that the proportion of spurious regressions (equal to $52,1\%$) is not negligible for the finite sample distribution of coefficients. In table 4, With a larger sample size of 1002 observations, spurious regressions with statistical significance in the trivariate regression occur in 95.5% of the simulations. This suggests that the usual recommendation of increasing the sample size in order to alleviate near-multicollinearity problems may foster spurious regressions.

Table 1: Inference discordances $r_{12}^T = 0$, $r_{13}^T = -0.03$, $r_{23}^T = 0.50$ $N = 102, 1000$ replications

	Do not reject $r_{12} = 0$	Reject $r_{12} = 0$
Do not reject $\beta_{12} = 0$	No effect: 92.3%	Type I: 3.3%
Reject $\beta_{12} = 0$,	Type II (spurious): 2.1%	Effect: 2.3%

Table 2: Inference discordances $r_{12}^T = 0$, $r_{13}^T = -0.03$, $r_{23}^T = 0.50$ $N = 1002, 1000$ replications

	Do not reject $r_{12} = 0$	Reject $r_{12} = 0$
Do not reject $\beta_{12} = 0$	No effect: 90.6%	Type I: 2.4%
Reject $\beta_{12} = 0$,	Type II (spurious): 4.9%	Effect: 2.1%

Table 3: Inference discordances $r_{12}^T = 0$, $r_{13}^T = -0.03$, $r_{23}^T = 0.99$ $N = 102, 1000$ replications

	Do not reject $r_{12} = 0$	Reject $r_{12} = 0$
Do not reject $\beta_{12} = 0$	No effect: 42.3%	Type I: 2.8%
Reject $\beta_{12} = 0$,	Type II (spurious): 52.1%	Effect: 2.8%

Table 4: Inference discordances $r_{12}^T = 0$, $r_{13}^T = -0.03$, $r_{23}^T = 0.99$ $N = 1002, 1000$ replications

	Do not reject $r_{12} = 0$	Reject $r_{12} = 0$
Do not reject $\beta_{12} = 0$	No effect: 0%	Type I: 0%
Reject $\beta_{12} = 0$,	Type II (spurious): 95.5%	Effect: 4.5%

On the distribution of true correlation coefficients: The frequency of spurious regressions depends also upon the distribution on true correlation coefficients (r_{12}^T, r_{13}^T) on the planes depicted in the figures 1 to 4 given r_{23}^T for any available data set of three variables in any field. This may depend on prior distributions. A researcher may choose a Laplacian prior: the distribution is uniform. For researchers such as Pearl [2009], p.62, the area of the discordances related to spurious regressions, which violates the assumption of the stability of conditional independence, has a prior measure equal to zero. The argument put forward is that strict equalities among product of parameters (for example, leading to a zero simple correlation coefficient with the dependent variable: $\beta_{12} + r_{23} \cdot \beta_{13} = 0$) have zero Lebesgue measure in any probability space in which parameters can vary independently (Spirtes et al. [2000]). Freedman [1997] by contrast, claimed that there is no reason to assume the prior that parameters are not in fact tied together by equality constraints of this sort.

In practice, researchers may select variables while doing exploratory regressions and data mining, in order to reach statistical significance (reject $H_0 : \beta_{12} = 0$) for publication (Stanley [2005]). 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. How to Prevent Spurious Regressions

Three indicators have been proposed in order 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, these indicators depend only on the correlation coefficients between the two explanatory variables, r_{23}:

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

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}, also equal to the ratio of the non standardized multiple regression parameter $\beta^{NS}_{1.2/3...k}$ and the non standardized parameter of the simple regression $\beta^{NS}_{1.2}$.

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

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 high values of the PIF.

When the correlation matrix is not reported in the paper, one may use Ioannidis' [2008] "vibration ratio" as the ratio of the largest versus the smallest effect on the same association in the paper and related papers. According to Ioannidis [2008], "*the vibration ratio will be larger in small datasets and in those with hazy definitions of variables, unclear eligibility criteria, large numbers of covariates, and no consensus in the field about what analysis should be the default. In most discovery research, this explosive mix is the rule.*"

Once the problem is highlighted, 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: $R^2_{1.j} > 1\%$). 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 $r^2_{1j} = 1\%$ is required to consider a meaningful "small" effect on the dependent variable. These tests are based on Fisher's [1921] Z transformation. They 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 ($H_0 : \beta^T_{1j'} = 0$) is also feasible. But, for very large sample ($N > 1000$), 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 $|\widehat{r}_{12}| < 0.1$.

5. Classical suppressor regressions: Aid, Policies and Growth

This example shows using the PIF and tests on correlation coefficients would have helped referees of Burnside and Dollar's [2000] paper. This paper states that real per capita gross domestic product (GDP) growth depends significantly (at the 5-percent level) on (Aid/GDP)×Policy, where the policy index is found by doing an auxiliary regression:

$$\text{Policy} = 1.28 + 6.85 \cdot \text{Budget Surplus} - 1.40 \cdot \text{Inflation} + 2.16 \cdot \text{Openness}.$$

They use an unbalanced panel including 56 countries over six four-years periods between 1970 and 1993 ($N = 275$ observations). However, the paper faces two main problems: It presents a spurious regression and its results depend on including or excluding a few outliers. To the first problem: In regression 4 (table 4), there are spurious regression effects for (Aid/GDP)×Policy and (Aid/GDP)^2×Policy, with statistical significance at the 5-percent level. 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 $PIF_{12} = 0.203/0.095 = 2.13$ and $PIF_{13} = -0.019/0.00458 = -4.15$ (with sign reversal). With respect to the control variables, there is another spurious effects with statistical significance at the 10-percent level, for Assassinations and the interaction term Ethnic fractionalization × Assassinations: $r_{14} = 0.063$, $r_{15} = 0.039$, $r_{45} = 0.86$ with $PIF_{14} = -0.45/-0.06296 = 7.15$ and $PIF_{15} = 0.80/-0.03934 = -20.3$.

The tests of the correlation with the dependent variable $H_0 : r_{1j} = 0$ do not reject the null hypothesis at the 5-percent level for 6 regressors: Log(GDP) at the beginning of each period, Ethnic fractionalization, Assassinations, Ethnic fractionalization × Assassinations, M2/GDP lagged, (Aid/GDP)^2×Policy. The more restrictive 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 $= 0.318$. The more restrictive 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 and growth negative correlation coefficient: $r_{12} = -0.173$, p-value $= 0.109$.

These results on spurious regressions do not change when using the larger data set available online by the authors (the number of observations is smaller in their reported regressors because observations are missing for some other regressors). The Aid/GDP and growth correlation is $r_{16} = -0.032$ for $N = 505$ observations and $H_0 : r_{16} = 0$ is not rejected (p-value $= 0.466$). The (Aid/GDP)×Policy and (Aid/GDP)^2×Policy coefficients are $r_{12} = 0.077$ and $r_{13} = 0.028$, respectively, for $N = 348$ observations. The null hypothesis, $r_{12} = 0$, is not rejected at the 5-percent level (p-value $= 0.148$).

With respect to the second problem, one may increase a too small correlation with the dependent variable by omitting or adding observations. In their regression (5) reported in table 4, Burnside and Dollar [2000] suppress 5 outliers which simultaneously are some of the extreme values of the (Aid/GDP)$\times$Policy and have a large negative or positive influence on the slope of (Aid/GDP)$\times$Policy (their figure 1): Gambia 86-89, 90-93, Guyana 90-93, Nicaragua 86-89, 90-93. On this sample of 270 observations, the correlation coefficients with the dependent variable increases: $r_{12} = 0.148$, $r_{13} = 0.113$, $r_{23} = 0.92$. Burnside and Dollar [2000] figure 1 also reveals that there remain four other outliers (Botswana 1978-1981, 82-85, 86-89 and Mali 86-89) which were not removed from the regression (4). When removing these four outliers, (266 observations), the correlation coefficients decreases: $r_{12} = 0.014$, $r_{13} = -0.065$, $r_{23} = 0.88$. When one removes one, then 2, 3 and 4 each of these outliers starting with Botswana 1978-1981, the coefficient of (Aid/GDP)$\times$Policy falls gradually from 0.19^{*} to 0.17^{*} ($p = 0.04$), 0.10 ($p = 0.19$), 0.05 ($p = 0.60$) and -0.02 ($p = 0.77$) (p is the p-value of the t-test of the parameter, * indicates statistical significance at the 5-percent level) The 0.19^{*} statistically significant parameter of regression (5) is driven by Botswana data. Botswana, a high growth country in Africa which received relatively more aid/GDP than other countries. It is an outlier in the regression because its value is one for the Sub-Saharan dummy regressor, whereas nearly all the other Sub-Saharan countries grew less, so that this dummy has a robust negative correlation with growth. The authors' choice to remove 5 outliers out of 9 in regression (5) corresponds to the largest value of the parameter of Aid/GDP$\times$Policy.

Regression (5B) replicates regression (4) on this data set with 270 observations including the highly correlated regressor (Aid/GDP)$^2\times$Policy. This regression is not reported in the published article. The signs are reverted for Aid/GDP, (Aid/GDP)$\times$Policy, (Aid/GDP)$^2\times$Policy as compared to regression (4). The (Aid/GDP)$\times$Policy parameter is now negative (-0.13) and no longer significant. This shows that the statistical significance of (Aid/GDP)$\times$Policy of the spurious effect found in regression (4) is not robust to the removal of five outliers.

Once spurious effects are removed in the regressions, 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.

Winner's curse paper. Ioannidis [2008] and Ioannidis and Trikalinos [2006] find support of winner's curse for papers in many fields of clinical research and epidemiology. Finding unexpected and large effects leads to higher probability to be published in top journals. These effects are then contradicted in lower ranked journals accepting replications of initial ideas.

The first step which followed Burnside and Dollar's paper was a controversy on the estimated effect. Easterly, Levine, and Roodman [2004] showed that, for example, the sign on aid/GDP×Policy is not stable when including 80 observations to the Burnside and Dollar's sample. In the previous section, we showed that it is not stable neither when excluding only 4 observations, in particular 3 from Botswana (table 4, equation 5B).

In a second step, a number of researchers followed Burnside and Dollar [2000]. They interacted aid with other terms, (e.g. with the fraction of a countries area that is in the tropics), split aid into subcomponents (bilateral versus multilateral flows, technical assistance and non-technical assistance, project aid and program aid, productive and unproductive aid, and so on), or introduced new terms that are inherently correlated with aid/GDP, e.g. a measure of aid instability or unpredictability that tends to scale with aid/GDP. They include in the regression at least another classical suppressor that is highly correlated with aid. Roodman [2008] mentions that for these statistically significant pairs of variables: "*Some of the coefficient magnitudes stretch credulity. Few of the studies report testing the variables of interest individually.*"

In a third step, a meta-analysis by Doucouliagos and Paldam [2009] includes up to 355 estimates from 31 articles of this literature dealing with "conditional" aid effectiveness, using non linear models with quadratic and/or interaction terms. They conclude that *"the aggregate coefficient to the interaction between foreign aid and policy proves to be very close to zero"*. They found similar results for aid-growth studies dealing with diminishing returns to aid, with a high correlation among the regressors aid/GDP and (aid/GDP)^2, and for the overall literature aid/growth literature including up to 543 estimates of the partial effects found in 68 published papers (Doucouliagos and Paldam [2008]). *Using meta-analysis, the average partial correlation coefficient converges to zero, which is also the value of the simple regression coefficient between aid and growth.*

The overall record in the field of aid and growth in the last 15 years is that a large amount of researchers scarce resources were diverted into spurious and/or outliers driven regressions including highly correlated classical suppressors with statistically significant oversized estimated parameters.

6. Conclusion

Highly correlated classical suppressors may foster the publication of spurious regressions and/or unstable regressions which presents statistically significant, large parameters, which are highly sensible to a few outliers. We advocate a very cautious use of classical suppressors in multiple regressions. If ever they are included in a regression, applied researchers and journal editors should provide a particularly detailed analysis of outliers.

References

[1] Belsley [1991]. *Conditioning Diagnostics: Collinearity and Weak Data in Regression*. Wiley, New York.

[2] Burnside C. and Dollar D. [2000]. Aid, policies and growth. *American Economic Review,* 90, 847-68.

[3] 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.

[4] Cohen J. [1988]. *Statistical Power Analysis for the Behavioral Sciences*. 2nd edition. Lawrence Erlbaum Associates, Hillsdale.

[5] Cohen J., Cohen, P., West S.G., Aiken L.S. [2003]. *Applied multiple regression/correlation for the behavioral sciences.* Wiley, New York.

[6] Christensen, R. [2006]. Comment on Friedman and Wall (2005). Letter to the Editor. *The American Statistician*, 60, 101-102.

[7] Doucouliagos H. and Paldam M. [2008]. Aid effectiveness on growth: A meta study. *European Journal of Political Economy*, 24, 1-24.

[8] Doucouliagos H. and Paldam M. [2009]. Conditional aid effectiveness: A meta study. *Journal of International Development*, 22, 391-410.

[9] 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.

[10] Freedman D. (1997). From Association to Causation via Regression. in V.R. McKim and S.P. Turner (eds), *Causality in Crisis?,* University of Notre Dame Press, Notre Dame, 113-61.

[11] Friedman, L. and Wall, M. [2005]. Graphical views of suppression and multicollinearity in multiple linear regression. *The American Statistician*, 59, 127-136.

[12] Fisher R. [1921]. On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron*, 1, 3-32.

[13] Fisher R. [1924]. The Distribution of the Partial Correlation Coefficient. *Metron*, 3, 329–332.

[14] Fisher R. [1925]. Applications of "Student's" distribution. *Metron*, 5, 90-104.

[15] Hill and Adkins [2001]. Collinearity. In: Baltagi B.H. (Ed.). *A Companion to Theoretical Econometrics*, Blackwell, Oxford (Chapter 12).

[16] Hoover K.D. [2001]. *Causality in Macroeconomics.* Cambrige University Press, Cambridge.

[17] Horst P. [1941]. The role of prediction variables which are independent of the criterion. in *The prediction of personal adjustment.* Social Sciences Reserach Council, New York, pp. 431-436.

[18] Ioannidis J.P.A. [2005]. Contradicted and initially stronger effects in highly cited clinical research. *Journal of the American Medical Association*. 294, 218-228.

[19] Ioannidis J.P.A., Trikalinos T.A. [2006]. Early extreme contradictory estimates may appear in published research: The Proteus phenomenon in molecular genetics research and randomized trials. *Journal of Clinical Epidemiology*, 58, 543-549.

[20] Ioannidis J.P.A. [2008]. Why Most Discovered True Associations Are Inflated. *Epidemiology*, 19, 640-648.

[21] Mason C.H. and Perreault W.D. [1991]. Collinearity, Power and Interpretation of Multiple Regression Analysis, *Journal of Marketing Research,* 28, 268-80.

[22] Pearl J. [2009]. *Causality: Models, Reasoning and Inference* (2nd edition). Cambrige University Press, Cambridge.

[23] Roodman D. [2008]. Through the Looking Glass, and What OLS Found There: On Growth, Foreign Aid, and Reverse Causality. Center for Global Development Working Paper Number 137.

[24] Spanos A. and McGuirk A. [2002]. The Problem of high correlation among regressors revisited: erratic vs systematic volatility. *Journal of Econometrics,* 108, 365-393.

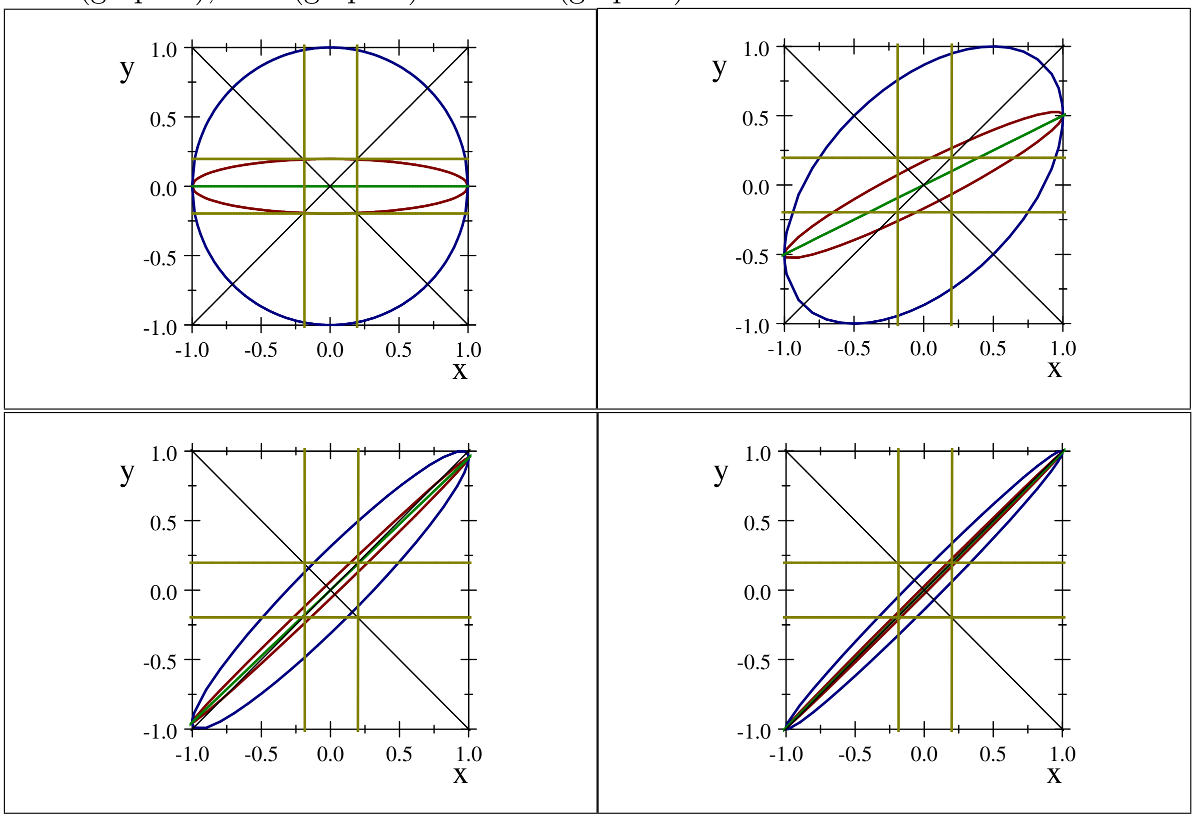
[25] Spirtes P., Glymour C.N. and Scheines R. [2000]. *Causation, Prediction, and Search* (2nd edition). Cambrige University Press, Cambridge.

[26] Silvey [1969]. Multicollinearity and Imprecise Estimation. *Journal of the Royal Statistical Society B*, 31, 539-552.

[27] Stanley, T.D., [2005]. Beyond publication bias. *Journal of Economic Surveys.* 19, 309-45.

[28] Yule G.U. [1897]. On the theory of correlation. *Journal of the Royal Statistical Society*, 60, 812-854.

Figure 1, 2, 3 and 4: The critical 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 trivariate regressions are inside the blue ellipse. The green line corresponds to the null hypothesis for the multiple regression H_0: $\beta_{12} = 0$. The critical region of the t-test related to H_0: $\beta_{12} = 0$ lies outside the red ellipse and inside the blue one. Inside the square delimited by red line around $r_{12} = r_{13} = 0$ is where both hypothesis of classical suppressors $H_0 : r_{12} = 0$, $H_0 : r_{13} = 0$ are not rejected.

Table 4 - OLS Growth Regressions: Using All Countries and the Policy Index.

	BD (3)	(4B)	BD (4)	BD (5)	(5B)	(5C)
Initial GDP	-0.61 (0.56)	-0.62 (0.56)	-0.56 (0.56)	-0.60 (0.57)	-0.68 (0.57)	-0.70 (0.57)
EF: Ethnic fractionalization	-0.54 (0.72)	-0.56 (0.72)	-0.42 (0.73)	-0.42 (0.72)	-0.52 (0.71)	-0.47 (0.72)
Assassinations (A)	-0.44 (0.26)	-0.44 (0.26)	-0.45 (0.26)	-0.45 (0.26)	-0.45 (0.26)	-0.41 (0.26)
EF $\times$ A	0.82 (0.44)	0.80 (0.44)	0.80 (0.44)	0.79 (0.44)	0.79 (0.43)	0.72 (0.44)
Institutional quality	0.64* (0.17)	0.64* (0.17)	0.67* (0.17)	0.69* (0.17)	0.63* (0.17)	0.64* (0.17)
M2/GDP (lagged)	0.014 (0.013)	0.014 (0.013)	0.016 (0.014)	0.012 (0.014)	0.014 (0.013)	0.008 (0.014)
Sub-Saharan Africa	-1.60* (0.73)	-1.60* (0.73)	-1.84* (0.74)	-1.87* (0.75)	-1.72* (0.74)	-1.85* (0.74)
East Asia	0.91 (0.54)	0.96 (0.56)	1.20* (0.58)	1.31* (0.58)	1.11* (0.56)	1.14* (0.56)
Policy Index	1.00* (0.14)	0.97* (0.19)	0.78* (0.20)	0.71* (0.19)	0.87* (0.18)	0.85* (0.18)
Aid/GDP	0.034 (0.12)	0.015 (0.012)	0.049 (0.12)	-0.021 (0.16)	-0.11 (0.17)	0.026 (0.16)
(Aid/GDP)$\times$Policy	–	0.013 (0.049)	0.20* (0.09)	0.19* (0.07)	-0.13 (0.15)	-0.025 (0.09)
(Aid/GDP)$^2\times$Policy	–	–	-0.019* (0.0084)	–	0.065* (0.028)	–
Observations	275	275	275	270	270	266
$\overline{R}^2$	0.36	0.35	0.36	0.36	0.36	0.38

Notes: The dependent variable is real per capita GDP growth. White heteroskedasticity consistent standard errors are in parentheses. Regressions (3), (4) and (5) are in Burnside and Dollar [2000] article (there is a typo in BD's article for regressions 4: the parameter is 0.049 for the variable Aid/GDP). * statistical significance at the 5-percent level.