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

January 17, 2011

Abstract

High correlation among regressors leads to two problems when testing the null hypothesis of no effect between two variables. In some situations, t-tests incorrectly sustain the null hypothesis (type I). In other situations, t-tests incorrectly reject the null hypothesis (type II). Three tools may limit type II spurious regressions: (1) a parameter inflation factor metric, (2) tests on correlation coefficients, and (3) using the *ceteris paribus* interpretation of estimated parameters conditional on the outcome of the tests. These tools are applied on Burnside and Dollar (2000)'s article "Aid, Policies and Growth".

JEL classification: **C12 O19 P45**

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

1. Introduction

For theoretical and applied researchers, it is crucial to know under which conditions high correlation among regressors leads to t-tests that *incorrectly* reject the null hypothesis of no effect between two variables. Without clearly stated theoretical guidelines, empirical research is likely to emphasize for several years the existence of spurious effects between variables. This may lead to a misallocation of researchers' resources, while fostering empirical controversies until several meta-analysis conclude that the effect is spurious.

However, the commonly held view emphasizes that high correlation among regressors may lead to "imprecise" regressions, with "low power of the t-test" (Silvey [1969], Belsley [1991], Mason and Perreault [1991], Hill and Adkins [2001]) so that t-tests incorrectly sustain the null hypothesis of no effect between two variables.[1] This view emphasizes a "type I inference discordance" based on the following two conditions. Firstly, the t-test *rejects* the null hypothesis of no effect (denoted $r_{12} = 0$) between the dependent variable x_1 and a regressor x_2, when the regression *excludes* other re-

[1] The term "multicollinearity" (introduced by Frisch [1934]) is often used for "high correlation among regressors" in a multiple regression. However, Belsley [1991] and Spanos and McGuirk [2002] stress that multicollinearity is a numerical issue that concerns the potential volatility of the estimated coefficients. It stems from the ill-conditioning of the $X'X$ matrix. High correlation among regressors and conditioning are neither necessary nor sufficient to one another (Belsley [1991] and Isotalo, Puntanen and Styan [2006]). To avoid misleading statements, we use the term "high correlation among regressors" in the rest of the paper.

gressors highly correlated with the first one, such as x_3. Secondly, the t-test *sustains* the null hypothesis of no effect (denoted $\beta_{12} = 0$) between the dependent variable x_1 and a regressor x_2, when the regression *includes* at least another regressor x_3 highly correlated with the first one. But this view neglects the alternative type II discordance of the outcome of the tests of these null hypothesis (table 1):

Table 1: Inference discordances

	Accept $r_{12} = 0$	Reject $r_{12} = 0$
Accept $\beta_{12} = 0$, regression includes x_3	No effect	Type I discordance
Reject $\beta_{12} = 0$, regression includes x_3	Type II discordance (spurious)	Effect

The "type II inference discordance" is based on the following two conditions. Firstly, the t-test *does not reject* the null hypothesis of no effect (denoted $r_{12} = 0$) between the dependent variable x_1 and a regressor x_2, when the regression *excludes* other regressors highly correlated with the first one, such as x_3. A regressor which is not correlated with the dependent variable is called a "classical suppressor variable" (Horst [1941], Cohen *et al.* [2003], Friedman and Wall [2005], Christensen [2006]). Secondly, the t-test *rejects* the null hypothesis of no effect (denoted $\beta_{12} = 0$) between the dependent variable x_1 and a regressor x_2, when the regression *includes* at least another regressor x_3 highly correlated with the first one. Since Yule [1897], a "classical suppressor" variable is not considered to be a problem in a multiple regression.[2] Hence, the commonly held view is that type II discordances do not matter.

By contrast, the key idea of this paper is to emphasize that type II discordance does matter when both regressors are highly correlated classical suppressor variables. This occurs when the t-tests do not reject the joint null hypothesis $r_{12} = 0$ and $r_{13} = 0$. These outcomes of the t-tests may occur when there is a high correlation among these two regressors (for example, such that the sample correlation is $|r_{23}| > 0.8$). Because the regressors are highly correlated, a ceteris paribus interpretation of β_{12} is not feasible in the multiple regression. Because the simple correlation coefficients with the dependent variable are not statistically different from zero, one cannot exclude the hypothesis that the multiple regression with highly correlated classical suppressors is spurious. For this reason, even when the t-test rejects the null hypothesis $\beta_{12} = 0$ with a power of the t-test reaching 100%, the statistical significance of β_{12} has no substantive significance. Hence, *a multiple regression including highly correlated classical suppressors which are statistically significant is a particular type of spurious regression.*[3] A pair of highly correlated classical suppressors may be related to a violation of the "stability condition of conditional independence" (Pearl [2009, p.48]) also known as the "faithfulness condition" (Spirtes *et al.* [2000], Hoover [2001]) in the context of graphical causal modeling, where x_1 and x_2 are unconditionally independent

[2] Yule [1897] initial claim is "*if* $r_{12} = 0$*, one cannot conclude that the third variable is of no use, for* β_{12} *will not be zero unless [the correlation coefficient between the regressors* x_2 *and* x_3 *denoted]* r_{23} *also* $= 0$".

[3] Aldrich [1995] presents a historical overview of other categories of spurious regressions along with their posterity.

of one another, but dependent, conditional upon x_3.

The first contribution of this paper is to show precisely how type II discordance occurs. This amounts to show how the critical regions of the test of the hypothesis $r_{12} = 0$ and of the test of the hypothesis $\beta_{12} = 0$ do not overlap, depending on the degree of correlation among regressors. This provides additional information to Spanos and McGuirk [2002] and Weber and Monarchi [1977] demonstration that increasing the correlation among regressors does not necessarily decrease the precision of the coefficient estimators or worsen the statistical significance of the coefficients and the power of the t-test. A regression in which the explaining variables are individually uncorrelated with the dependent variable is not automatically "spurious", it can be the case that two variables have no individual effect, because the second regressor varies according to an immediate response to shocks on the first regressor in order to stabilize instantly the dependent variable (perfectly homeostatic model). For Hoover (2001), p. 44-46, these two models are "ontologically" different, but one cannot decide between the two interpretations (spurious regression versus perfectly homeostatic model) using the regression method. The researcher needs a prior to decide. But because the prior makes the decision about the true effect, the empirical test following the prior is useless.

The second contribution of this paper is to propose three tools for practitioners to avoid spurious regressions with highly correlated classical suppressor variables. First, the paper proposes a simple parameter inflation metric, for journal editors and referees, as a reasonable starting point that highlights the potential problem of spurious regressions, that the variance inflation factor or collinearity index is unable to capture. Secondly, preliminary tests on simple correlation coefficients should be used in order to eliminate highly correlated suppressor variables. This extends Spanos [2006] typology of six cases to be tested in case of omitted variable bias, to a seventh case related to highly correlated classical suppressors, where $r_{12} = 0$, $r_{13} = 0$ and $r_{23} \neq 0$. This proposal amounts to a pre-test estimator for multiple regression. It is also related to Bryant et al. [2009] tests for disproving of causal relationship. Thirdly, in the case where the tests reject the null hypothesis $r_{12} = 0$, $r_{13} = 0$, and also $r_{23} = 0$, then the "ceteris paribus" interpretation of estimated parameters (other variables being unchanged) should not be used.

The third contribution is to apply these tools on the most quoted paper of the *American Economic Review* for the year 2000, Burnside and Dollar [2000] "Aid, Policies and Growth". It shows that its key results were based on spurious regressions using highly correlated classical suppressors. Then, the academic literature which followed this paper in this field is mostly based on spurious regressions. The field of aid and growth suggests that these spurious regressions are likely to be endemic; Roodman [2008] and Doucouliagos and Paldam [2008] found large and statistically significant effects when (correct) priors are that there is no correlation among these variables.

The paper proceeds as follows. Section 2 presents graphical views of the critical regions of the t-test with or without an omitted regressor, depending on the correlation among regressors. Section 3 proposes three tools for practitioners in order to avoid spurious regressions with high correlation among classical suppressor variables. Section 4 applies these tools to Burnside and Dollar's paper. Section 5 concludes. An appendix proposes a simple general proof that high correlation among regressors is compatible with precise regressions for $k \geq 3$ explanatory variables.

2. Spurious Regressions with High Correlation among Regressors

2.1. The trivariate regression

The problem of a high correlation among regressors is fully highlighted with only a pair of regressors (see appendix two for a general case with $k > 2$ regressors). In what follows, we consider regressions on standardized variables. Then, 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:

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

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. This defines the set of possible values for the primary parameters (see also Spanos [1995]). In what follows we take up the same approach, considering correlations coefficients for N observations. The main concern of this paper is to fully use the information on the covariance matrix (for standardized variables, the correlation matrix) in order to deal with spurious regressions and with the ceteris paribus interpretations of parameters when there is a high correlation among regressors.

Let $\mathbf{R}_3$ be a block 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 $1 \leq j \leq 3$. They are related to two simple regressions with the dependent variable, and to an auxiliary regression of one of the explanatory variable with the other one, where $\boldsymbol{\varepsilon}_{1,2}$, $\boldsymbol{\varepsilon}_{1,2}$ and $\boldsymbol{\varepsilon}_{2,3}$ are vectors of random disturbances that follow a normal distribution with mean zero and variances $\sigma_{1.2}^2$, $\sigma_{1.3}^2$ and $\sigma_{2.3}^2$ for each of the regressions, as follows:

$$\mathbf{x}_1 = r_{12}\mathbf{x}_2 + \boldsymbol{\varepsilon}_{1.2},\ \mathbf{x}_1 = r_{13}\mathbf{x}_3 + \boldsymbol{\varepsilon}_{1.3,}\ \mathbf{x}_2 = r_{23}\mathbf{x}_3 + \boldsymbol{\varepsilon}_{2.3}.$$

The ordinary least squares estimated parameters, the coefficient of determination, the related inequalities on the determinants of the correlation matrices, and the coefficient of partial correlation were computed by Yule [1897]:[4]

$$\begin{aligned}
\begin{pmatrix} \widehat{\beta}_{12} \\ \widehat{\beta}_{13} \end{pmatrix} &= \frac{1}{1-r_{23}^2}\begin{bmatrix} 1 & -r_{23} \\ -r_{23} & 1 \end{bmatrix}\begin{pmatrix} r_{12} \\ r_{13} \end{pmatrix} = \frac{1}{1-r_{23}^2}\begin{pmatrix} r_{12}-r_{13}r_{23} \\ r_{13}-r_{12}r_{23} \end{pmatrix} \\
R_{1.23}^2 &= 1-\frac{\det(\mathbf{R}_3)}{\det(\mathbf{R}_2)} = \frac{r_{12}^2+r_{13}^2-2r_{12}r_{13}r_{23}}{1-r_{23}^2} = 1-\left(1-r_{12}^2\right)\left(1-r_{13.2}^2\right) \\
0 &\leq \det(\mathbf{R}_3) = 1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}r_{13}r_{23} \leq \det(\mathbf{R}_2) = 1-r_{23}^2 \leq 1 \\
-1 &< r_{12.3} = \frac{r_{12}-r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} = \frac{t_{\widehat{\beta_{12}}}}{\sqrt{t_{\widehat{\beta_{12}}}+N-2}} = \frac{\sqrt{1-r_{23}^2}}{\sqrt{1-r_{13}^2}}\widehat{\beta}_{12} < 1.
\end{aligned}$$

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

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

[4] In what follows,variables with a hat denote estimated variables, except correlation coefficients to simplify notation.

Three indicators have been proposed in order to detect collinearity: the determinant $\det(\mathbf{R}_2)$, 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 $\mathbf{R}_2$. We assume from now on $r_{23} \geq 0$, the alternative leading to symmetric results easy to find. 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 $CI > 3$.

2.2. Inference

For given values of r_{23}, the critical regions of the tests $H_0 : \beta_{12} = 0$ (no effect of the regressor x_2 on the dependent variable in multiple regression) against $H_1 : \beta_{12} \neq 0$ and $H_0 : r_{12} = 0$ (no effect of the regressor x_2 on the dependent variable in simple regression) against $H_1 : r_{12} \neq 0$ can be depicted in the $(\mathbf{r_{13}}, \mathbf{r_{12}})$-plane. Conditions for "type I inference discordances" (the tests accept the null hypothesis, i.e. no effect, in multiple regression and reject the null hypothesis, i.e. existence of an effect, in simple regression) and for "type II inference discordances" (the tests reject the null hypothesis, i.e. existence of an effect, in multiple regression and accept the null hypothesis, i.e. no effect, in simple regression) are then discussed.

Feasible correlation coefficients and exact regression boundary. The condition $\det(\mathbf{R}_3) \geq 0$ can be restated as:

$$\det(\mathbf{R}_3) = -(1 + r_{23}) \left(\frac{r_{12} - r_{13}}{\sqrt{2}} \right)^2 - (1 - r_{23}) \left(\frac{r_{12} + r_{13}}{\sqrt{2}} \right)^2 + 1 - r_{23}^2 \geq 0. \quad (2.1)$$

For fixed $0 < r_{23} < 1$ and varying r_{13} and r_{12}, for $\det(\mathbf{R}_3) = 0$ describes an ellipse centered at the origin ($r_{12} = r_{13} = 0$). 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 curves 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. When $r_{23} = 1$ (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, however, to the singularity of the correlation matrix of the regressors (perfect correlation of the explanatory variables), hence, the ordinary least squares estimators cannot be computed.

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 null hypothesis related to the other parameter can be derived analogously.

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

$$\begin{aligned}
0 &\leq r_{12.3}^2 \leq \frac{t_{\widehat{\beta_{12}}}^2(\alpha)}{N - 2 + t_{\widehat{\beta_{12}}}^2(\alpha)} < 1 \\
-1 &\leq -\frac{t_{\widehat{\beta_{12}}}(\alpha)}{\sqrt{N - 2 + t_{\widehat{\beta_{12}}}^2(\alpha)^2}}\sqrt{(1 - r_{13}^2)}\sqrt{(1 - r_{23}^2)} \leq r_{12} - r_{13}r_{23} \leq \\
&\leq \frac{t_{\widehat{\beta_1}}(\alpha)}{\sqrt{N - 2 + t_{\widehat{\beta_{12}}}^2(\alpha)^2}}\sqrt{(1 - r_{13}^2)}\sqrt{(1 - r_{23}^2)} \leq 1.
\end{aligned}$$

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 correlation matrix is positive or equal to zero.[5] 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.
- $r_{23}^2 \to 1$: There is high correlation among regressors.
- $r_{13}^2 \to 1$: The other explanatory variable is strongly correlated with the dependent variable.
- $t(\alpha)$ decreases (the applied researcher sets for example a threshold for the probability of type I error of $\alpha = 10\%$).

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) and the sample distribution of the partial correlation coefficients has been found by Fisher (1924). The complement of the critical region of the test with

[5] There is always room for precise (reject the null hypothesis) and imprecise estimation because $1 > \frac{t^2}{N-2+t^2}$. On figure 1 to 4, the ratio of the distance of the case of the null hypothesis $r_{12} - r_{13}r_{23}$ on a vertical slice of diagram to the frontier of the critical region divided by the distance of the case of the null parameter to the frontier of possible values of r_{12} (the ellipse) is the same proportion: $\frac{t^2_{\widehat{\beta_1}}(\alpha)}{N-2+t^2_{12}(\alpha)}$ for any value of r_{13} (for fixed r_{23}).

the null hypothesis $H_0 : r_{12} = 0$ can be stated as:

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

The border of the critical region is depicted as two horizontal lines over and above the horizontal line where $H_0 : r_{12} = 0$. 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 interval for r_{12} is *reduced*, when the number of observations N and/or r_{12} 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 $H_0 : r_{13} = 0$.

2.3. Conditions for type II inference discordances (spurious regression).

In what follows we will discuss the occurrence of type I versus type II discordances when the correlation between regressors increases (figures 1 to 4). These figures show the critical regions of both t-tests in bivariate and trivariate regressions, in the plane (r_{13}, r_{12}), with the correlation coefficient r_{13} on the horizontal axis and the correlation coefficient r_{12} on the vertical axis, with a relatively small sample. $N = 102$ was chosen for the lines depicting the critical regions.

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. However, 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. For example, adding a large number of additional orthogonal regressors (up to all the principal components in principal component regressions) will turn as statistically significant the effect of the variable x_2 on the dependent variable x_1 in multiple regression, even though the null hypothesis $H_0 : r_{12} = 0$ is accepted in the simple regression case. Hence, in the particular case of orthogonal regressors (such as incomplete principal components ordinary least square regressions (IPC-OLS) or Gram-Schmidt orthogonalized regressors), we suggest that only the simple correlation critical region makes sense for doing inference. This suggestion rules out type II discordance with orthogonal regressors. Then, softwares programmers (e.g. PROC REG, PCOMIT instruction in SAS) should change the computation of standard errors with incomplete principal component ordinary least square regressions.

Type II discordance includes the limit case where $(\widehat{r_{13}}, \widehat{r_{12}})$ is on the limit blue circle and between the two horizontal red lines. Then, the regression is exact ($R^2 = 1$), while the power of the simple regression t-test $r_{12} = 0$ can be close to zero. The difference $R^2_{1.23} - r^2_{12}$ is very large. The larger it is, the larger is the power of the t-test of the null hypothesis $\beta_{12} = 0$ in a multiple regression (Cohen [1988]). For $\widehat{r_{12}}$ extremely close to zero but not exactly equal to zero, the power of the t-test $H_0 : \beta_{12} = 0$ can be extremely close to one when $R^2_{1.23} = 1$, and it will reject the null hypothesis while $\widehat{\beta}_{12} = \widehat{r_{12}} \approx 0$.

In figure 2, the sample correlation between the two variable is equal to $\widehat{r_{23}} = 0.5$ (a "strong" effect according to Cohen [1988]). The set of (true or sample) feasible correlation coefficients $(\widehat{r_{13}}, \widehat{r_{12}})$ lies within a blue ellipse. The green line with a positive slope equal to $\widehat{r_{23}} = 0.5$ is the location of the null hypothesis $\beta_{12} = 0$. The critical region of the t-test of the null hypothesis $\beta_{12} = 0$ lies outside the red ellipse. 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 complement of the critical region. 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 red ellipse. The area of type II discordance 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 $H_0 : r_{13} = 0$).

In figure 3, the sample correlation between the two variables is equal to $\widehat{r_{23}} = 0.95$. The set of (true and sample) feasible correlation coefficients $(\widehat{r_{13}}, \widehat{r_{12}})$ lies within a smaller blue ellipse with respect to figure 2. The green line has a slope equal to 0.95.

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 complement of the critical region. This time, 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 unity 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 very lower power of this t-test. This is the usual concern of textbooks.

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. 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-multicollinear pairs) than in the case of figure 2. This means that both regressors are related to spurious inference. This power of the t-test $H_0 : \beta_{12} = 0$ is still close to one on the blue ellipse for type II inference discordance.

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.

Table 2 states the practical requirements on r_{13} for reaching the critical region of statistical significance ($p = 0.05$) in case of high correlation among regressors ($r_{23} = 0.95$, $r_{23} = 0.99$) when there is a zero direct effect in a simple regression of the regressor x_2 on the dependent variable : $r_{12} = 0$, as a function of the number of observations ($N = 22$ or 102 or 402). The null hypothesis corresponds to the value

$r_{13} = r_{12}/r_{23} = 0$. The limit feasible values for r_{13} are given by $\underline{r_{13}}$ and by $\overline{r_{13}}$.

Table 2: Critical values of r_{13} (r_{13}^{C-}, r_{13}^{C+}) corresponding to $p = 0.05$ for the test of $H_0 : \beta_{12} = 0$ as a function of the number of observations N, of r_{12} and r_{23}

-	$r_{12} =$	0.00	0.00	0.00	0.00	0.00	0.00
-	$r_{23} =$	0.95	0.95	0.95	0.99	0.99	0.99
-	N	r_{13}	$\widehat{\beta}_{12}$	$\widehat{\beta}_{13}$	r_{13}	$\widehat{\beta}_{12}$	$\widehat{\beta}_{13}$
$\underline{r_{13}}$	-	-0.312	3.04	-3.2	-0.141	7.01	-7.08
r_{13}^{C-}	22	-0.127	1.23	-1.30	-0.057	2.83	-2.86
r_{13}^{C-}	102	-0.061	0.59	-0.62	-0.027	1.34	-1.36
r_{13}^{C-}	402	-0.031	0.30	-0.32	-0.014	0.69	-0.70
H_0: $r_{12}/r_{23} = 0$	-	0	0	0	0	0	0
r_{13}^{C+}	402	0.031	-0.30	0.32	0.014	-0.69	0.70
r_{13}^{C+}	102	0.061	-0.59	0.62	0.027	-1.34	1.36
r_{13}^{C+}	22	0.127	-1.23	1.30	0.057	-2.83	2.86
$\overline{r_{13}}$	-	0.312	-3.04	3.2	0.141	-7.01	7.08

For high correlation among regressors $r_{23} = 0.95$ and $r_{12} = 0$, for $\widehat{\beta}_{12}$ to be statistically significant ($p = 0.05$), it is sufficient that $|r_{13}| > 0.031$ for $N = 402$, $|r_{13}| > 0.061$ for $N = 102$, $|r_{13}| > 0.127$ for $N = 22$, knowing that $|r_{13}|$ cannot exceed 0.312.

For high correlation among regressors $r_{23} = 0.99$ and $r_{12} = 0$, for $\widehat{\beta}_{12}$ to be statistically significant ($p = 0.05$), it is sufficient and far less demanding that $|r_{13}| > 0.014$

for $N = 402$, $|r_{13}| > 0.027$ for $N = 102$, $|r_{13}| > 0.057$ for $N = 22$, knowing that $|r_{13}|$ cannot exceed 0.141.

In order to reach statistical significance of $\widehat{\beta}_{12}$ for a p-value such that: $p = 0.05$, a difference of a few percentage points between the correlation coefficient r_{12} and r_{13} is sufficient, and this required difference can to be very small $|r_{12} - r_{23}r_{13}| \geq 0.015$ as the number of observation reaches $N = 402$ and the correlation among regressors increases up to $r_{23} = 0.99$.

In table 3, we have five examples of near spurious regressions with both coefficients statistically significant ($p = 0.05$) for $N = 102$ observations: (orders of magnitude for statistical significance are $|r_{12} - r_{13}r_{23}| > 0.03$ when $r_{23} = 0.99$ and $|r_{12} - r_{13}r_{23}| > 0.10$ when $r_{23} = 0.95$).

Table 3: Statistically significant ($p = 0.05$) large parameters $\widehat{\beta}_1$ and $\widehat{\beta}_2$ for $N = 102$

r_{23}	r_{12}	r_{13}	$\widehat{\beta}_{12}$	$\widehat{\beta}_{13}$	PIF_{12}	PIF_{13}
0.99	0.00	-0.03	1.4925	-1.5075	$+\infty$	50.2
0.99	0.015	-0.015	1.5	-1.5	100	100
0.99	0.01	0.04	-1.4874	1.5126	-148.7	37.8
0.99	0.1	0.13	-1.4422	1.5578	-14.4	11.9
0.95	0.05	-0.05	1	-1	20	20

The parameter inflation factors PIF_{1j} are discussed in the following section. In the first example we have $r_{23} = 0.99$, $r_{12} = 0$ and $r_{13} = -0.03$, a value slightly below -0.027, the value required for statistical significance ($p \leq 0.05$), see table 1. In the five cases the standardized parameters $\widehat{\beta}_j$ of precise regressions are *very large* in absolute values: around 1.5 when $r_{23} = 0.99$ and 1 when $r_{23} = 0.95$. With opposite signs for r_{12} and r_{13}, for $r_{23} > 0$, the signs of the multiple correlations are identical to the signs of the simple correlation (regressions 1, 2 and 5). With $0 < r_{13} < r_{12}$, for $r_{23} > 0$, the sign of the multiple correlation for β_2 is the opposite of the sign of the simple correlation r_{23} (regressions 3 and 4). The fourth regression is a regression not with near zero effects but with small effects $r_{12} \geq 0.1$ and $r_{13} \geq 0.1$. The estimated parameters are very close to the ones of the regressions 1, 2 and 3. For near-spurious regressions, the conditions to reach statistical significance with $p \leq 0.05$ are **not at all constraining conditions** in empirical research. A graphical representation of r_{12} as a function of r_{13} for a given $r_{23} = 0.99$ is given in appendix 1.

An important issue is the sample distribution on (r_{12}, r_{13}) on the planes depicted in the figures 1 to 4 given r_{23}. This joined distribution may be 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 are easily obtained by Monte Carlo simulations. Samples have been drawn 1000 times from a multivariate normal distribution with the true correlation coefficients given as $r_{12}^T = 0$, $r_{13}^T = -0.03$, and $r_{23}^T = 0.99$ and a sample size of $N = 102$. This case corresponds to first line of table 3. The true correlation coefficients $r_{12}^T = 0$, $r_{13}^T = -0.03$ can be located on figure 4 for $r_{23}^T = 0.99$. The following table 4a reports 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.

Table 4a: 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%

The key result is that the proportion of spurious regressions (equal to $52, 1\%$) is not empty and not negligible for the finite sample distribution of coefficients. Table 4b reports results for a larger sample size $N = 1002$ (still with 1000 replications):

Table 4b: 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%

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.

A related issue is whether the probability weighted fraction of the area that is near the origin and yet in the region generating high t statistics (the type II discordances area for spurious regressions) is invariant or increases with respect to the degree of correlation between the two regressors. In tables 4c and table 4d, we replicate the two former simulations changing only the true correlation coefficient now equal to $r_{23}^T = 0.50$. The true correlation coefficients $r_{12}^T = 0$, $r_{13}^T = -0.03$ can be located on figure 2 for $r_{23}^T = 0.50$.

Table 4c: 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 4d: 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%

The proportion of type II discordance (spurious) remains very small for a correlation coefficient of 0.5 between the two regressors.

There is an important difference between near spurious regressions and the particular case of an exactly spurious and exactly collinear true model (we emphasize here the notation, with an exponent T for true model) ($r_{12}^T = 0$, $r_{23}^T = 1$ which imply $r_{13}^T = 0$). In this particular case, the sample correlations are such that researchers may reject the null hypothesis $\beta_{12}^T = 0$ (near spurious regression) only because the sample is too small, but when the sample tends to infinity, one has $|\widehat{r}_{12} - \widehat{r}_{13}\widehat{r}_{23}| \to \left|r_{12}^T - r_{13}^T r_{23}^T\right| = 0 = \left|\beta_{12}^T\right|$. The spurious regression is only a small sample problem. In the near-spurious regression, where the true parameter in the multiple regression is close to zero, but *not exactly equal* to zero $\left|\beta_{12}^T\right| \neq 0$ that is: $\left|r_{12}^T - r_{13}^T r_{23}^T\right| \neq 0$ with $\left|r_{12}^T - r_{13}^T r_{23}^T\right|$ very small, increasing the sample size increases the critical region of the t-test and increases the likelihood that the null hypothesis $\left|\beta_{12}^T\right| = 0$ will be rejected. In this much broader class of true models than the exact spurious and exact collinear true model, *increasing the sample size magnifies the possibilities for finding near-spurious multiple regressions* for applied researchers.

2.4. The distributions of true correlation coefficients and variables selection with publication bias.

The frequency of spurious regressions depends also upon the distribution on true correlation coefficients (r_{12}, r_{13}) on the planes depicted in the figures 1 to 4 given r_{23} for any type of empirical work. 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 usually advanced appeals to the fact 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) in 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 also select variables while doing exploratory regressions and data mining, in order to reach statistical significance (reject $\beta_{12} = 0$) 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. Let us give four ways to construct a pair of highly correlated regressors. First, the other regressor may measure the same phenomenon or may be driven by the same missing causal variable that the first regressor. Second, the other regressor may be the square or other powers of the first regressor. Third, the other regressor may be an interaction term of the first regressor with another variable (see the case study). Fourth, the other regressor may be a lagged value of the first regressor. Hence, publication bias when facing a classical suppressor may lead to an increased concentration of the distribution on the true correlation coefficients in the area of spurious regressions depicted in the figures 1 to 4. Doucouliagos and Paldam (2009) meta analysis of conditional foreign aid effectiveness on growth reveals that this field of research has a biased distribution of true correlation coefficients in the spurious regression area.

3. How to Avoid Spurious Regressions with High Correlation among Regressors

3.1. Detecting spurious regressions with Parameter Inflation Factor

With high correlation among regressors, precise regressions with large t- statistics are easily found. In these cases, a measure able to evaluate whether coefficients are oversized or not is needed for the applied researcher as well as 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 $\beta^{NS}_{1.2}$.

$$PIF_{12} = \frac{\beta_{1.2/.3...k}}{r_{12}} = \frac{\beta^{NS}_{1.2/3...k}}{\beta^{NS}_{1.2}}. \text{ For } k = 2: \; PIF_{12} = \frac{1 - \frac{r_{13}}{r_{12}} r_{23}}{1 - r^2_{23}}.$$

The third equality is the PIF-formula for the trivariate case. The PIF is a measure of the relative omitted variable bias in the simple regression in proportional terms with respect to the multiple regression. As compared to the variance inflation factor (VIF) which depends only on elements of det $(R_{2...k+1})$, the PIF takes also into account the vector r_{1j}. 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. For the particular case of near-spurious regressions where r_{12} is close to zero, the PIF takes into account the magnifying effect of a small value of r_{12} on the size of the parameter $\beta_{1.2/.3...k}$. As seen on figures 5 to 8, the absolute value of the slope of parameter as a function of r_{13} $(\partial\beta_{1.2/.3...k}/\partial r_{13})$ is much higher for small absolute values of r_{12} than for large values. In other words, near-spurious precise regressions are likely to provide the largest values of standardized coefficients when facing high correlation among regressors.

In the extreme examples of table 3, PIF exceeds 10, so that the simple regression parameter is multiplied by more than 10 in the multiple regression. In practice, even a PIF above 2 may be a signal of a potential spurious regression. Even if this parameter is statistically different from zero with $p < 0.05$, doubts may be raised with respect to the substantive adequacy of the size of the parameter with a "ceteris paribus" partial effect interpretation, despite its statistical adequacy. Once the problem is highlighted, we suggest that applied researchers perform preliminary tests on simple correlation coefficients.

When the correlation matrix is not reported in the paper, one may use Ioannidis' [2008] "vibration ratio" as the ratio of the largest vs. smallest effect on the same association approached with different analytic choices. 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.*"

3.2. Avoiding spurious regressions with preliminary tests on simple correlation coefficients with the dependent variable

We define a near-spurious regression as a regression in which 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 ($H_0 : r^T_{1j} < 0.1$ when $r^T_{1j} > 0$, or $H_0 : r^T_{1j} > -0.1$ when $r^T_{1j} < 0$) 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 ($H_0 : \beta^T_{1j'} = 0$). The letter T refers to the "true" correlation coefficient and not the sample correlation coefficient. The test of a null effect ($H_0 : \beta^T_{1j'} = 0$) instead of a negligible effect ($H_0 : \beta^T_{1j'} < 0.1$ when $\beta^T_{1j} > 0$, or $H_0 : \beta^T_{1j'} > -0.1$ when $\beta^T_{1j'} < 0$) in multiple regression corresponds to the current practice of applied research, although this practice may change in the future. 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: $r_{1j} = 0.1$ or $r_{1j}^2 = 1\%$: small effect, $r_{1j} = 0.3$ or $r_{1j}^2 = 9\%$ medium size effect, $r_{1j} = 0.5$ or $r_{1j}^2 = 25\%$ large effect. The definition of near-spurious regressions extends Spanos' [2006] (pp.203-205) definition of an exact spurious regression ($H_0 : r_{1j}^T = 0$) when revisiting the omitting variable bias.

We suggest to begin with these tests before starting multiple regression analysis. These tests are based on Fisher's [1921] Z transformation and are available in many statistical softwares, such as SAS 9.1 (the instruction is: *proc corr data=database fisher (rho0=0.1 lower);*). Spanos [2006] recommends to supplement these tests "*with a post data evaluation of inference based on the notion of severe testing, see Mayo [1996]*" and Mayo and Spanos [2006] in order to get a more robust inference.

3.3. Conditioning the "Ceteris Paribus" interpretation of coefficients on tests on simple correlation coefficients between regressors

Once the spurious regressions have been ruled out, we suggest to condition the ceteris paribus interpretation of coefficients based on prior tests of small correlation between regressors. If the above test led to reject the null hypothesis of a less than small effect, we suggest a second step where the researcher performs the tests of null hypothesis of a negligible effect of the first regressor on the second one ($H_0 : r_{23}^T < 0.1$ when $r_{23}^T > 0$, or $H_0 : r_{23}^T > 0.1$ when $r_{23}^T < 0$, where the letter T refers to the "true" correlation coefficient and not the sample correlation coefficient). This is a condition for applying the "ceteris paribus" interpretation of a parameter as a partial effect on a dependent variable. With high correlation among regressors, this test is likely to be heavily rejected. For example, for a very small sample $N = 11$ and a sample correlation coefficient $r_{23} = 0.83$, the null hypothesis ($H_0 : r_{23}^T < 0.5$) for an effect where each of the regressors explain less than 25% of the variance of the other regressor is already rejected.

As seen in figures 5-8, in multiple regressions with high correlation among regressors, the standardized parameters can easily be much larger than one while being precisely estimated, if the sum of squares of residuals is very small. The size of a given coefficient is often interpreted as a ceteris paribus effect ("all other things being equal", or "the other regressors being unchanged"). For example, the "ceteris paribus" interpretation of the partial correlation coefficient is proposed by Moore's [1917] estimation of the demand for cotton (see also Morgan [1992]).[6] But the "ceteris paribus" conditions are hard to maintain: the β_{12} coefficient may not be very well "identified" from a given sample due to the high correlation among regressors. The "ceteris paribus" interpretation when there is high correlation among regressors leads to misleading forecasts of extreme values of the dependent variable. Let us consider

[6]Marshall was skeptical on Moore's application of the "ceteris paribus" interpretation to observational data: "*No important chain of events seems likely to be associated with any one cause so predominantly that a study of the concomitant variations of the two can be made as well by Mathematics, as by comparison of a curve representing those two elements with a large number of other curves representing other operative causes: the ceteris paribus clause though formally adequate seems to me impracticable.*" (Marshall to Moore, 5/6/1912 [1013], Marshall and Whitaker [1996])".

the parameters of the standardized regression:

$$\frac{x_1 - \overline{x}_1}{\sigma_1} = \beta_1 \frac{x_2 - \overline{x}_2}{\sigma_2} + \beta_2 \frac{x_3 - \overline{x}_3}{\sigma_3} + \varepsilon_{1.23}.$$

The interpretation of a standardized parameter is as follows: a deviation from the mean of the regressors $\overline{x}_2$ by one standard error σ_2 of this regressor (that is: $\frac{x_2 - \overline{x}_2}{\sigma_2} = 1$) implies a prediction $\widehat{x}_1$ which deviates from the mean of the dependent variable $\overline{x}_1$ by β_1^S times the standard error of the dependent variable σ_1. In the case of the simple regression, the standardized parameter is exactly equal to the correlation coefficient r_{12}, which is such that $|r_{12}| \leq 1$. For example, if x_1 and x_2 are normally distributed, and for a standardized parameter equal to $\beta_{12} = 2$:

$$\underbrace{]-\infty, \overline{x}_2 - \sigma_2[\bigcup]\overline{x}_2 + \sigma_2, +\infty[}_{\text{33\% observations of } x_2} \rightarrow \underbrace{]-\infty, \overline{x}_1 - 2\sigma_1[\bigcup]\overline{x}_1 + 2\sigma_1, +\infty[}_{\text{Predictions of } x_1 \text{ in 5\% extreme tails}} .$$

One third of the observations of x_2 imply extreme predictions of $\widehat{x}_1$ in the 5% tails of the distribution of x_1, according to the "ceteris paribus" interpretation. The "ceteris paribus" interpretation predicts one time over three extreme events that are expected to occur for only one observation over twenty. These are unreliable "extreme" forecasts due to oversized standardized parameters. In fact, when the first regressor moves from one standard error from the mean $\overline{x}_2 + \sigma_{x_2}$, the second regressor deviates from its mean with nearly one standard error $\overline{x}_3 + r_{23}\sigma_{x_3}$ with high correlation among regressors. Then, the explained variable deviation due to a shock of $\overline{x}_2 + \sigma_{x_2}$ boils down to the simple regression effect r_{12} with a more reliable predicted value of $\widehat{x}_1$:

$$\widehat{x}_1 = \overline{x}_1 + (\beta_{12} - \beta_{13} \cdot r_{23})\,\sigma_{x_1} = \overline{x}_1 + r_{12}\sigma_{x_1}.$$

The correction is the subtraction of the omitted variable bias on the estimated parameter β_{12} in the simple regression, which could be very large for precise regression with high correlation among regressors:

$$\beta_{12} - r_{12} = -\beta_{13} r_{32}.$$

The estimated parameters when there is high correlation among regressors are the best linear unbiased estimates, but they are in practice *useless*, even when the regression is not spurious (r_{12} sufficiently large). They are useless for practitioners, because they require a sophisticated interpretation of the interaction among regressors (typically described in an auxiliary regression with the explanatory variable explained by other explanatory variables highly correlated), instead of the usual "ceteris paribus" interpretation. In fact, the parameters which are easy to interpret as partial effects, "ceteris paribus" are given by two separate regressions in our example:

$$x_1 = r_{12}x_2 + \varepsilon_{12} \text{ and } x_2 = r_{23}x_3 + \varepsilon_{23},$$

where r_{12} is sufficiently large (no spurious regressions) and x_2 and x_3 are highly correlated. Informative policy simulations (if x_3 or x_2 is related to policy or treatment) should reflect the underlying correlation structure among such variables.

4. Case study: Aid and Growth

4.1. Burnside and Dollar (2000), "Aid, Policies and Growth"

In this example, we show that it would not have taken a lengthy refereeing process for the editors of *American Economic Review* to ask for dramatic changes in Burnside and Dollar's (2000) paper, including the PIF and tests on correlation coefficients. Then, they might have decided to publish the paper for its interesting negative results and the novelty of its database, like in the *Journal of Interesting Negative Results* that do exist now in many fields, such as BioMedecine.

Burnside and Dollar state 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). The paper, however, faces two main problems: It presents a spurious regression the results are sensitive to including or excluding certain 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 the expected change of signs). With respect to the control variables, there are spurious regression effects, with statistical significance at the 10-percent level, for Assassinations and 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 other correlation coefficients among regressors are at most around 0.5.

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: $H_0 : r^T_{1j} < 0.1$ when $r^T_{1j} > 0$ 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: $H_0 : r^T_{1j} > -0.1$ when $r^T_{1j} < 0$ 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. 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 $(\text{Aid/GDP})^2$×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).[7]

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)×Policy and have a large negative or positive influence on the slope of (Aid/GDP)×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 (Bostwana 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 Bostwana 1978-1981, the coefficient of (Aid/GDP)×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).[8] The 0.19^* statistically significant parameter of regression (5) is driven by Botswana data. Bostwana, a high growth country in Africa, is an outlier because its value is one for the Sub-Saharan dummy

[7]It is no surprise that by adding new data to the Burnside and Dollar sample of 275 observations, Easterly, Levine and Roodman [2004] found weaker correlations between Aid/GDP, (Aid/GDP)×Policy $(\text{Aid/GDP})^2$×Policy and Growth.

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

regressor, which has a negative effect on growth. The choice of 5 outliers out of 9 in regression (5) corresponds to the largest value of the parameter of Aid/GDP×Policy.

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

Finally, the GDP growth increases with the macroeconomic policy index, the institutional quality, and for East Asian countries and decreases for Sub-Saharan countries. As the correlation coefficient between institutional quality and policy is equal to 0.247, which reject the null hypothesis $H_0 : r_{ij} < 0.1$, a ceteris paribus interpretation of a change of the macroeconomic policy index on growth is not correct, because it is linked to a positive change of institutional quality.

4.2. Endemic spurious regressions with quadratic and interaction terms

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 adding 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 (table 4, equation 5B). If one eliminates the high correlation among regressors by dropping (aid/GDP)$^2\times$policy, the large t-statistic disappears throughout (table 4, equation (4B)). 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. As described in this paper, this creates at least another classical suppressor which is highly correlated with aid, itself a "classical suppressor" which has a zero correlation with growth. 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.* Much ado about nothing, due to highly correlated classical suppressors.

The Burnside and Dollar [2000] study is a typical "winner's curse" article as recently described in epidemiology (Ioannidis [2005]). Burnside and Dollar's study emphasizes a fragile and inflated effect on a hot political topic. Their sample is relatively small with respect to the full population and it faces an endogenous selection. Its key statistically significant effect (at the 5-percent level) comes from spurious regression effects, as described in this paper. It has been published in a top journal. Firstly, it led to an empirical controversy. Secondly, it paved the way for an industry of regressions with highly correlated pairs of regressors. As a consequence, it has been highly cited. It even turned out to be the most cited paper published in *American Economic Review* during the year 2000 (more than 2000 citations in google scholar in 2010). This may end in a meta-analysis concluding that there is no such effect, fifteen years after the release of the working paper. A similar path for winner's curse papers is found in epidemiology. For example, Ioannidis [2008] and Ioannidis and Trikalinos [2006] find support that finding unexpected and large effects leads to higher probability to be published in top journals in many fields of clinical research and epidemiology. These effects are then contradicted in lower ranked journals accepting replications of initial ideas.

Why spurious regressions using highly correlated classical suppressors, which is a "curiosa" or an odd artefact of an otherwise efficient statistical method, are likely to be endemic in top journals? They allow to find novel and unexpected correlations in the pool of the (correct) priors of no correlation among variables. The effects are found to be large and statistically significant. The tools to find at least a highly correlated twin to a classical suppressor corresponds to interesting models (dynamic models with lags of the suppressor variable, non linear model, quadratic, polynomial, interaction terms, and so on). They foster easily controversy, which in turn increases citations and the impact factor of top journals. This recent history of the field of Aid and Growth suggests that spurious regressions with highly correlated classical suppressors almost certainly exceeds 5% of published results.

A final remark is related to the current debate on the debate on statistical significance versus substantive significance raised by McCloskey and Ziliak [2008]. They emphasize that, on the one hand, with large samples, some effects are considered statistically significant whereas the size of the effect (for example, measured by the standardized estimated parameter) is minuscule and related to a negligible change of a loss function. On the other hand, when the size of the full population or of available samples is too small, they emphasize that effects are not considered statistically significant although their size and effect leads to a large charge of a loss function (they do have substantive significance). By contrast, "type II" spurious regressions described in this paper lead to **statistically significant and large** effects which are nevertheless meaningless.

5. Conclusion

A high correlation among regressors along with small, but slightly different effects on the dependent variable of each of these regressors (e.g. differing by a wedge of 0.01 to 0.1 units of correlation coefficients depending on the number of observations) leads to large parameters and high values of their t-statistic. High correlation among regressors may foster a publication bias for data sets where a couple of indicators are poorly correlated with the dependent variable (close to be classical suppressor variables), and highly correlated together. To help mitigate the problems due to high correlation among regressors, we suggest to use standardized variables and to proceed in the following five steps: (1) present the correlation matrix including the dependent variable, (2) test the null hypothesis $H_0 : r_{1j} < 0.1$ (for positive sample correlation coefficients) or $H_0 : r_{1j} > -0.1$ (for negative sample correlation coefficients) and *exclude* regressors who *do not* reject this null hypothesis, (3) test the null hypothesis for the *remaining* regressors ($i > 1$, $j > 1$, $i \neq j$) $H_0 : r_{ij} < 0.1$ (for positive sample correlation coefficients) or $H_0 : r_{ij} > -0.1$ (for negative sample correlation coefficients) and use the ceteris paribus interpretation of their estimated parameters only for regressors who *do* reject this null hypothesis, (4) for the *remaining* regressors for which one restrains from the ceteris paribus interpretation of their estimated parameters, first discuss the plausibility of the size of the (standardized) parameters using the parameter inflation factor PIF, then (5) explain the data generating process of the high correlation among regressors, and, for example, present an auxiliary regression where this regressor is the dependent variable and the other regressors are its explanatory variables and the partial R2 of each of its regressors in descending order.

Let us conclude with Yule [1897]: "*But, in every case, the means, standard deviations, and (gross [simple]) correlation coefficients should be tabulated in published work, as they enable readers of the paper to evaluate any of the omitted functions for themselves*".

References

[1] Aldrich [1995]. Correlations Genuine and Spurious in Pearson and Yule, *Statistical Science.* 10(4), 364-376

[2] Belsley [1991]. *Conditioning Diagnostics: Collinearity and Weak Data in Regression.* Wiley, New York.6.

[3] Bryant H.L., Bessler D.A. and Haigh M.S. [2009]. Disproving Causal Relationships Using Observational Data. *Oxford Bulletin of Economics and Statistics.* 71(3), 357-374.

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

[5] Cohen J. [1988]. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. Lawrence Erlbaum Associates, Routledge publisher.

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

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

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

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

[10] McGuirk, A. and Driscoll P. [1995]. The Hot Air in R2 and Consistent Measures of Explained Variation. *American Journal of Agricultural Economics*, 77 (2), 319-328.

[11] 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(3), 774-780.

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

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

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

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

[16] Fisher R. [1925]. Applications of "Student's" distribution. *Metron.* 5(3). 90-104.

[17] Frisch R. [1934]. *Statistical Confluence Analysis by Means of Complete Regression Systems.* University Institute of Economics, Oslo.

[18] Greene W.H. [2008]. *Econometric analysis.* Sixth edition. Prentice Hall. New York.

[19] Isotalo J., Puntanen S. and Styan G.H. [2006]. On the role of the constant term in linear regression. *Festschrift for Tarmo Pukkila on his 60th Birthday.* Eds. E. P. Liski, J. Isotalo, J. Niemelä, S. Puntanen, and G. P. H. Styan, Univ. of Tampere, 243-259.

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

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

[22] 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.

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

[24] Ioannidis J.P.A., Trikalinos .TA. [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(6) 543-549.

[25] Ioannidis J.P.A. [2008]. Why Most Discovered True Associations Are Inflated. *Epidemiology.* 19(5) 640-648..

[26] McCloskey D. and Ziliak S.T. [2008]. *The Cult of Statistical Significance: how the standard error cost us jobs, justice and lives.* University of Michigan Press. Ann Arbor.

[27] Marshall A., Whitaker J.K. [1996]. The correspondence of Alfred Marshall, Economist, volume 3, Towards the close, 1903-1924. *Cambridge University Press.*

[28] Mason C.H. and Perreault W.D. [1991]. Collinearity, Power and Interpretation of Multiple Regression Analysis, *Journal of Marketing Research,* 28(3), 268-80.

[29] Mayo D.G. [1996]. *Methodology and the Growth of Empirical Knowledge*, Chicago. The University of Chicago Press.

[30] Mayo D.G. and Spanos A. [2006]. Severe testing as a basic concept in a Neyman-Pearson philosophy of induction. *British Journal for the Philosophy of Science.* 57(2), 3253-357.

[31] Morgan M.S. [1992]. *The history of econometric ideas*, Cambridge University Press, Cambridge.

[32] Moore H.L. [1917]. *Forecasting the Yield and the Price of Cotton.* The Macmillan company.

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

[34] Puntanen S. and Styan G.P.H. [2005]. Schur complements in statistics and probability. In *The Schur complement and its applications,* F. Zhang editor, Springer Verlag, 163-226.

[35] 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.

[36] Shieh G. [2001]. The inequality between the Coefficient of determination and the sum of squared simple correlation coefficients. *The American Statistician,* 55(2), 121-124.

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

[38] Spanos [1995]. On Normality and the Linear Regression Model. *Econometric Reviews,* 14(2), 195-203.

[39] Spanos [2006]. Revisiting the omitted variables argument: Substantive vs. statistical adequacy. *Journal of Economic Methodology*, Volume 13, Issue 2, 179 - 218.

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

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

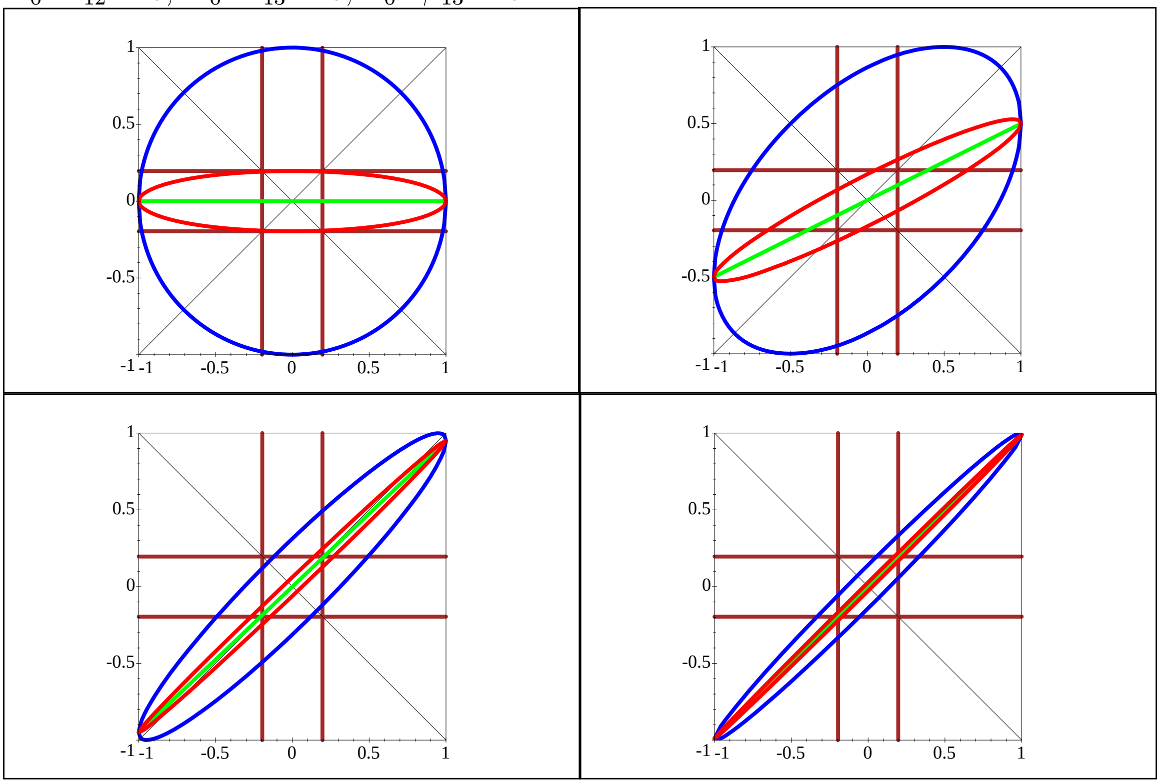
[42] Stanley, T.D., [2005]. Beyond publication bias. *Journal of Economic Surveys.* 19, 309-45.

[43] Weber J.E. and Monarchi D.E. [1977]. Graphical Representation of the Effect of multicollinearity, *Decision Sciences*, 8, 534-547.

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

Figure 1, 2, 3 and 4: $r_{12} = f(r_{13})$, for $N = 102$ and $r_{23} = 0$ (graph 1) then 0.5 (graph 2), 0.95 (graph 3) and 0.99 (graphe 4). The criticial regions for the tests: $N = 102$, $t = 2$.

$H_0 : r_{12} = 0$, $H_0 : r_{13} = 0$, $H_0 : \beta_{13} = 0$.



Feasible trivariate regressions are inside the blue ellipse, which corresponds to exact regressions. 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 brown line around $r_{12} = r_{13} = 0$ is the location of near-spurious trivariate regressions, such that a relatively large area lies inside the critical region of the t-test to H_0: $\beta_{12} = 0$

when $r_{23} \geq 0.95$.

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 × 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)×Policy	-	0.013 (0.049)	0.20* (0.09)	0.19* (0.07)	-0.13 (0.15)	-0.025 (0.09)
$(\text{Aid/GDP})^2\times$Policy	-	-	-0.019* (0.0084)	-	0.065* (0.028)	-

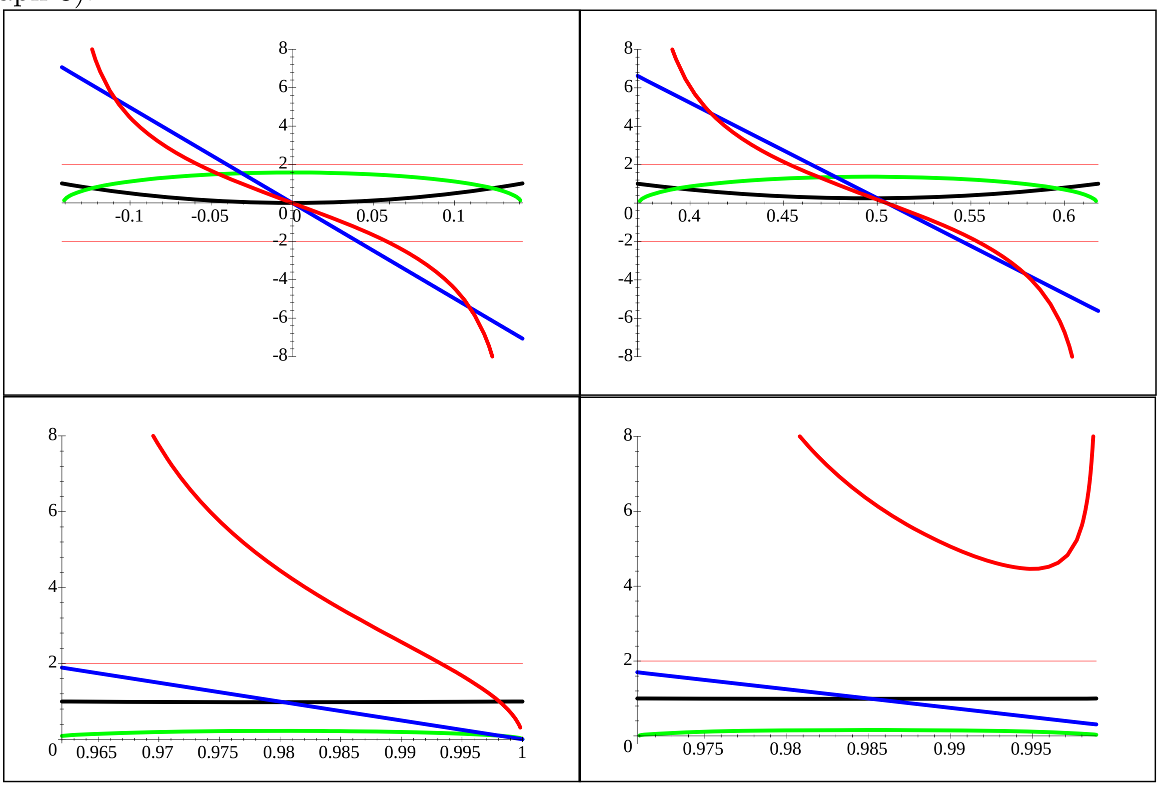
Notes: The dependent variable is real per capita GDP growth. White heteroskedasticity consistent standard errors are in parentheses. Regressions (3), (4) and (5) are in Burnside and Dollar [2000] article (there is a typo in BD's article for regressions 4: the parameter is 0.049 for the variable Aid/GDP).

* statistical significance at the 5-percent level (but not necessarily substantive significance).

Appendix 1: Regression statistics with high correlation among regressors

Appendix 3 presents the details of the computation of the regression statistics while figures 5, 6, 7, 8 present original graphical views of these statistics for a high correlation between regressors ($r_{23} = 0.99$), as a function of r_{13}. The range of feasible values for r_{13} corresponds to four horizontal cross sections of figure 4, with first $r_{12} = 0$ (graph 5), then $r_{12} = 0.5$ (graph 6), $r_{12} = 0.99$ (graph 7), $r_{12} = 0.995$ (graph 8). On these graphs, the standardized parameter $\widehat{\beta}_{12}$ (blue line) exceeds one for a wedge $|r_{12} - r_{13}r_{23}|$ bounded away from zero. The maximal value of the standardized parameter is reached when $r_{12} = 0$ (the case for spurious regressions, graph5). The R^2 (black line) is equal to one on the boundaries of the wedge $r_{13} - r_{12}$ corresponding to an exact regression ($\det(\mathbf{R}_3) = 0$). It reaches a minimum when $\widehat{\beta}_{12} = 0$. It remains always very close to one in the limit case $r_{13} \geq r_{23} = 0.99$ (graph 7 and 8). Because a larger R^2 decreases the root mean square error, the estimated standard error $\widehat{\sigma}_{\widehat{\beta}_{12}}$ is equal to zero on the on the boundaries of the wedge $r_{13} - r_{12}$ corresponding to an exact regression. Finally, the $t_{\widehat{\beta}_1}$- statistic (red line) exceeds 1.96 for gaps $|r_{13} - r_{12}|$ sufficiently large (cf. table 1), except for very high correlation of regressor x_2: $r_{12} \geq r_{23} = 0.99$ with the dependent variable, where a small wedge $|r_{13} - r_{12}|$ is no longer a condition for statistical significance.

Figures 5, 6, 7 and 8: R^2 (black), Parameter $\widehat{\beta}_1$ (blue), its estimated standard error $\widehat{\sigma}_{\widehat{\beta}_1}$ (green), and $t_{\widehat{\beta}_1}$ statistic (red), all functions of r_{13} for $r_{23} = 0.99$ and $N = 22$, with first $r_{12} = 0$ (graph 5), then $r_{12} = 0.5$ (graph 6), $r_{12} = 0.99$ (graph 7), $r_{12} = 0.995$ (graph 8).



The feasible values of r_{13} are related to horizontal cross sections of diagram 4, bounded by the blue curve of exact regressions.

Appendix 2: high correlation among many regressors and precise regression (NOT FOR PUBLICATION)

In what follows, we consider a regression on standardized variables. Then, there is no constant in the model and all variables have mean zero and a variance of one. This is no restriction since it is always possible to transform units accordingly. Bold letters correspond to matrices and vectors

$$\mathbf{x}_1 = \mathbf{X}_k\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{x}_1$ is the vector of N observations of the dependent variable, $\mathbf{X}_k$ is the matrix where column i corresponds to the N observations of the regressor $\mathbf{x}_i$ for $2 \leq i \leq k+1$, $\boldsymbol{\beta}$ is a vector of k parameters to be estimated, and $\boldsymbol{\varepsilon}$ is a vector of random disturbances that follow a normal distribution with mean zero and variance σ.

In a linear regression model with standard assumptions on the error term ($E(\varepsilon_t|\mathbf{X_{kt}}) = \mathbf{0}$ and $E(\varepsilon_t^2|\mathbf{X}_{kt}) = \boldsymbol{\sigma^2}$), Spanos and McGuirk [2002] derive in their theorem 1 a relation between the model parameters $(\boldsymbol{\beta}, \boldsymbol{\sigma^2})$ 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. This defines the set of possible values for the primary parameters (see also Spanos [1995]). In what follows we take up the same approach, considering correlations coefficients for N observations.

Let $\mathbf{R}_{k+1}$ be a block 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 $\mathbf{R}_k$ corresponds to the correlation matrix of the regressors. One has $r_{ij}^2 \leq 1$ for $1 \leq i \leq k+1$ and $1 \leq j \leq k+1$.

$$\mathbf{R}_{k+1} = \begin{bmatrix} 1 & \mathbf{r}_1^{'} \\ \\ \mathbf{r}_1 & \mathbf{R}_k \end{bmatrix} \text{ with } \mathbf{R}_k = \frac{1}{N}\mathbf{X}_k^{\prime}\mathbf{X}_k = [r_{ij}]_{\substack{2\leq i\leq k+1 \\ 2\leq j\leq k+1}} \text{ and } \mathbf{r}_1 = \frac{1}{N}\mathbf{X}_k^{\prime}\mathbf{x}_1 = [r_{1i}]_{2\leq i\leq k+1} \ .$$

A correlation matrix is non-negative definite. Its determinant is such that $0 \leq \det(\mathbf{R}_{k+1}) \leq 1$, for all values of $k \geq 1$. The Schur complement $\mathbf{R}_{k+1}/\mathbf{R}_k$ of the matrix $\mathbf{R}_{k+1}$ defined by:

$$\mathbf{R}_{k+1}/\mathbf{R}_k = 1 - \mathbf{r}_1^{'}\mathbf{R}_k^{-1}\mathbf{r}_1 = 1 - R_{1.23\ldots k+1}^2 \geq 0.$$

Because $\mathbf{r}_1$ is a column, the Schur complement $\mathbf{R}_{k+1}/\mathbf{R}_k$ is a scalar. It is equal to one minus the coefficient of determination of the multiple correlation $R_{1.23\ldots k+1}^2$, so that: $0 \leq \mathbf{R}_{k+1}/\mathbf{R}_k \leq 1$. From the Schur lemma, we know: $\det(\mathbf{R}_{k+1}/\mathbf{R}_k) = \det \mathbf{R}_{k+1}/\det \mathbf{R}_k$. Then we get the Schur inequality (Puntanen and Styan [2005]):

$$0 \leq \det(\mathbf{R}_{k+1}) = \det(\mathbf{R}_{k+1}/\mathbf{R}_k)\cdot\det(\mathbf{R}_k) = \left(1 - R_{1.23\ldots k+1}^2\right)\det(\mathbf{R}_k) \leq \det(\mathbf{R}_k) \leq 1.$$

¿From now on, we assume that there is no perfect collinearity between the explanatory variables: $\det(\mathbf{R}_k) \neq 0$. Then, we can express the estimated parameters of the regression and their estimated variances and t-values in terms of the correlation matrix. For the ordinary least square estimate of the vectors of the standardized parameters $\widehat{\boldsymbol{\beta}}$ we get:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}_k'\mathbf{X}_k\right)^{-1}\mathbf{X}_k'\mathbf{x}_1 = \mathbf{R}_k^{-1}\mathbf{r}_1 = \frac{\mathbf{C}_k\mathbf{r}_1}{\det(\mathbf{R}_k)} = \frac{\mathbf{C}_k\mathbf{r}_1}{\prod_{i=1}^{i=k}\lambda_{k,i}},$$

where $\mathbf{C}_k = [c_{ij}]_{\substack{2\leq i\leq k+1\\ 2\leq j\leq k+1}}$ is the cofactor matrix of the correlation matrix of explanatory variables of order k: $\mathbf{R}_k$, and $\mathbf{c}_{ii}$ is the vector of its diagonal elements. $\lambda_{k,i}$ are the real, positive eigenvalues of $\mathbf{R}_k$. The variance of the error term can be estimated in the following way:

$$\widehat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{N-k} = \frac{N}{N-k}\left(1-\mathbf{r}_1'\mathbf{R}_k^{-1}\mathbf{r}_1\right) = \frac{N}{N-k}\mathbf{R}_{k+1}/\mathbf{R}_k = \frac{N}{N-k}\left(1-R^2_{1.23\ldots k+1}\right).$$

Then we get for the estimated variances of the parameters $\widehat{\boldsymbol{\sigma}}^2_{\widehat{\beta}_i}$ and the t-statistic (we define the ratio of a vector as the ratio of each of its elements):

$$\begin{aligned}
\widehat{\boldsymbol{\sigma}}^2_{\widehat{\beta}} &= \widehat{\sigma}^2\cdot diag\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \frac{1}{N-k}\frac{\det(\mathbf{R}_{k+1})}{\det(\mathbf{R}_k)}\frac{\mathbf{c}_{ii}}{\det(\mathbf{R}_k)} \\
\mathbf{t} &= \frac{\widehat{\boldsymbol{\beta}}}{\widehat{\boldsymbol{\sigma}}_{\widehat{\beta}}} = \frac{\frac{1}{\det(\mathbf{R}_k)}\mathbf{C}_k\mathbf{r}_1}{\sqrt{\frac{1}{N-k}\frac{\det(\mathbf{R}_{k+1})}{\det(\mathbf{R}_k)}\frac{1}{\det(\mathbf{R}_k)}\mathbf{c}_{ii}}} = \frac{\sqrt{N-k}}{\sqrt{\det(\mathbf{R}_{k+1})}}\frac{\mathbf{C}_k\mathbf{r}_1}{\sqrt{\mathbf{c}_{ii}}}
\end{aligned}$$

The Student's t-statistic is the same for the standardized parameter (related to standardized variables) as for non-standardized variables since the ratio of standard errors of the dependent variable with respect to the regressor i multiplies both the parameter estimate and its standard error estimate:

$$\widehat{\beta}_i = \widehat{\beta}_{iS}\frac{\widehat{\sigma}_1}{\widehat{\sigma}_i} \text{ and } \widehat{\sigma}_{\widehat{\beta}_i} = \widehat{\sigma}_{\widehat{\beta_{iS}}}\frac{\widehat{\sigma}_1}{\widehat{\sigma}_i}.$$

Additionally we can compute the partial correlation coefficient of the kth regressor $1 - r^2_{1(k+1).2...k}$:

$$1 - r^2_{1(k+1).2...k} = \frac{1 - R^2_{1.2...k+1}}{1 - R^2_{1.2...k}} = \frac{\det(\mathbf{R}_{k+1})}{\det(\mathbf{R}_k)}\frac{\det(\mathbf{R}_{k-1})}{\det(\mathbf{R}_k)} = 1 - \frac{t^2_{1(k+1)}}{t^2_{1(k+1)} + N - k}.$$

In what follows we distinguish four different cases:

1. Exact collinearity between regressors occurs if $\det(\mathbf{R}_k) = 0$.

2. High correlation among regressors occurs if $0 < \det(\mathbf{R}_k) \leq \delta < 1$, with δ relatively small, defined by a rule of thumb such as $\delta = 0.1$.

3. An exact multiple regression (exact collinearity between the dependent variable and its regressors) occurs if $\det(\mathbf{R}_{k+1}) = 0$ and $\det(\mathbf{R}_k) \neq 0$.

4. A precise regression is a regression where at least one parameter is statistically significant: the p-value of type I error is below 5% ($p < 0.05$) for the null hypothesis: $H_0 : \beta_j = 0$.

Taking the above classification into account we have a first result:

Theorem 1. *In the case of an "***exact multiple regression***" (*$\det(\mathbf{R}_k) \neq 0$ *and* $\det(\mathbf{R}_{k+1}) = 0$*), one has:*

$$R^2_{1.23...k+1} = 1, \tag{5.1}$$

$$\widehat{\boldsymbol{\sigma}}_{\widehat{\beta}} = \mathbf{0} \tag{5.2}$$

$$\lim_{\substack{\det(\mathbf{R}_{k+1}) \to 0 \\ \det(\mathbf{R}_k) \neq 0}} t_{\widehat{\beta}_i} = \begin{cases} \pm\infty, \text{ if } \mathbf{C}_k \mathbf{r}_1 \neq 0. \\ 0, \text{ if } \mathbf{C}_k \mathbf{r}_1 = 0. \end{cases} \tag{5.3}$$

In an exact multiple regression, the variable x_1 is an **exact** linear combination of the k explanatory variables. The coefficient of determination $R^2_{1.23...k+1}$ is equal to 1 and the root mean square error of the regression is equal to zero. The standard error of the parameters is zero. The t-statistic tends to infinity, except in the case where the numerator of the parameter estimate is *exactly* equal to zero. This means that the null hypothesis that the explanatory variable has nothing to do with the dependent variable will be rejected. A second result can be stated for high correlation among regressors:

Theorem 2. *Precise regressions are compatible with high correlation among regressors.*

As $\widehat{\boldsymbol{\beta}}$, $R^2_{1.23...k}$, $\widehat{\boldsymbol{\sigma}}^2_{\widehat{\beta}_i}$ and $\det(\mathbf{R}_{k+1})$ are *continuous* functions of the correlation coefficients when $\det(\mathbf{R}_k) > 0$, there exist neighborhoods of the exact multiple regression case (defined by: $0 \leq \det(\mathbf{R}_{k+1}) < \varepsilon < 1$) where the standard error of the parameters is very close to zero and the t-statistic is very large when $\mathbf{C}_k\mathbf{r}_1 \neq 0$. Because of the Schur inequality, high correlation among regressors ($0 < \det(\mathbf{R}_k) \leq \delta < 1$) is compatible with the existence of precise regressions for ε sufficiently small, such that:

$$0 \leq \det(\mathbf{R}_{k+1}) < \varepsilon < \det(\mathbf{R}_k) \leq \delta \leq 1.$$

The Schur inequality states that when $\det(\mathbf{R}_k)$ is close to zero (and different from zero, precise regressions exists such that $\det(\mathbf{R}_{k+1})$ is relatively much closer to zero than $\det(\mathbf{R}_k)$. In other words, the residual sum of squares may be much smaller than the determinant of the correlation matrix of the regressors.

Appendix 3: The regression statistics as functions of r_{13} (NOT FOR PUBLICATION).

In order to get some insights how the estimated parameter values and their standard deviation and t-values evolve, we now rewrite the above relations as depending on r_{13}. This means that we assume the correlation coefficients r_{12} and r_{23} as given, $0 \leq r_{23} < 1$ and $-1 \leq r_{12} \leq 1$. A graphical representation is depicted in figures 5 to 8, for a small number of observations $N = 22$, so that it is easier to see the standard error curve on the graphs.

The correlation coefficient of the second explanatory variable with respect to the dependent variable has to remain in the interval:

$$r_{13} \in [\underline{r}_{13}, \overline{r}_{13}] = \left[r_{12}r_{23} - \sqrt{\left(1-r_{12}^2\right)\left(1-r_{23}^2\right)}, r_{12}r_{23} + \sqrt{\left(1-r_{12}^2\right)\left(1-r_{23}^2\right)} \right].$$

The limits of the interval are found by a horizontal slice of the ellipse that determines the possible values of the correlation coefficient r_{13}, with the limits corresponding to the ellipse determined by $\det(\mathbf{R}_3) = 0$, for a given $r_{23} < 1$:

$$\begin{aligned} \det(\mathbf{R}_3) &= \left(1-r_{12}^2\right)\left(1-r_{23}^2\right) - \left(r_{13} - r_{12}r_{23}\right)^2 = \left(\overline{r_{13}} - r_{13}\right)\left(r_{13} - \underline{r_{13}}\right) \geq 0 \Longleftrightarrow \\ -1 &\leq r_{12}r_{23} - \sqrt{\left(1-r_{12}^2\right)\left(1-r_{23}^2\right)} = \underline{r_{13}} \\ \leq r_{13} &\leq \overline{r_{13}} = r_{12}r_{23} + \sqrt{\left(1-r_{12}^2\right)\left(1-r_{23}^2\right)} \leq 1. \end{aligned} \tag{5.4}$$

The estimated standardized parameter $\widehat{\beta}_1$ is a linear and decreasing function of r_{13}. The slope of the line increases non linearly with r_{23}. (This can be interpreted as a multiplier effect.)

$$\frac{\partial \widehat{\beta}_1}{\partial r_{13}} = \frac{-r_{23}}{1 - r_{23}^2} < 0.$$

The sensitivity of the coefficient of determination with respect to r_{13} is:

$$\frac{\partial R_{1.23}^2}{\partial r_{13}} = \frac{2r_{13} - 2r_{12}r_{23}}{1 - r_{23}^2} < 0 \Rightarrow r_{13} < r_{12}r_{23}.$$

The coefficient of determination starts from the value 1 for $\underline{r}_{13}$, then decreases till it reaches its minimal value $r_{13} = r_{12}r_{23}$ (which is the condition for the other parameter to be equal to zero: $\beta_2 = 0$), then increases and reaches again the value 1 for $\overline{r}_{13}$. The coefficient of determination is depicted as the black line in figures 5 to 8.

The estimated standard error sensitivity to r_{13} is given by

$$\frac{\partial \sigma_{\widehat{\beta}_{12}}}{\partial r_{13}} = \frac{1}{\sqrt{N-2}} \frac{r_{12}r_{23} - r_{13}}{\sqrt{\det(\mathbf{R}_3)}\,(1 - r_{23}^2)} < 0 \Rightarrow r_{13} > r_{12}r_{23},$$

when the regression is not exact: $\det(\mathbf{R}_3) \neq 0$).

The estimated standard error of the estimated parameter $\widehat{\beta}_1$ starts from the value 0 for $\underline{r}_{13}$ with a vertical tangent (tangent with infinite slope), then increases till its maximum $r_{13} = r_{12}r_{23}$ (which is the condition for the other parameter to be equal to zero: $\beta_2 = 0$), then decreases until it reaches again the value 0 for $\overline{r}_{13}$ with a vertical tangent. The corresponding curve is the green curve in figures 5 to 8.

The $t_{\widehat{\beta}_1}$ statistics sensitivity to r_{13} is given by ($\det(\mathbf{R}_3) \neq 0$)

$$\begin{aligned}
\frac{\partial t_{\widehat{\beta}_1}}{\partial r_{23}} &= \frac{\sqrt{N-2}}{[\det(\mathbf{R}_3)]^{\frac{3}{2}}} \left(-r_{23}\det(\mathbf{R}_3) - (r_{12} - r_{13}r_{23}) \frac{1}{2} \frac{\partial \det(\mathbf{R}_3)}{\partial r_{13}} \right) \\
&= \frac{\sqrt{N-2}}{[\det(\mathbf{R}_3)]^{\frac{3}{2}}} (r_{12}r_{13} - r_{23}) \left(1 - r_{23}^2\right).
\end{aligned}$$

Three scenarios can be distinguished: The absolute value of r_{12} is smaller than r_{23} (the latter is necessarily larger than zero), the absolute value of r_{12} is equal to the value of r_{23}, and the absolute value of r_{12} is larger than r_{23}.

Theorem 3. *Let* $|r_{12}| < r_{23}$. *Then* $\widehat{\beta}_1(\underline{r}_{13}) > 0$ *and* $\widehat{\beta}_1(\overline{r}_{13}) < 0$ *and there exists a*

$\tilde{r}_{13} = r_{12}/r_{23} \in (\underline{r}_{13}, \overline{r}_{13})$ *such that* $\widehat{\beta_1}(\tilde{r}_{13}) = 0$. *Furthermore* $t_{\widehat{\beta}_1}(\tilde{r}_{13}) = 0$ *and*

$$\lim_{r_{13}\to\underline{r}_{13}} t_{\widehat{\beta}_1}(r_{13}) = +\infty \text{ and } \lim_{r_{13}\to\overline{r}_{13}} t_{\widehat{\beta}_1}(r_{13}) = -\infty.$$

The proof follows from straightforward calculations, replacing the upper and lower bounds for r_{13} in the equations for $\widehat{\beta}_1$ and $t_{\widehat{\beta}_1}$. The parameter starts from a maximum for $r_{13} = \underline{r}_{13}$, is equal to zero when $r_{13} = r_{12}/r_{23}$, and reaches a minimum for $r_{13} = \overline{r}_{13}$. Because of the multiplier effect of the correlation coefficient r_{23}, when there is near collinearity (for example, when $r_{23} > 0.95$), the size of the estimated standardized parameter can be very easily too large (for example, exceeding 1 or 1.5) outside a neighborhood of $r_{13} = r_{12}/r_{23}$. This case is depicted in figures 5 and 6. The blue line is the curve for $\widehat{\beta}_1$. The parameters are large, and estimated with a standard error equal to zero: the t-statistic tends to infinity and leads to asymptotes. The estimated $t_{\widehat{\beta}_1}$-statistic of the parameter β_1 starts from the value $+\infty$ for $\underline{r}_{13}$, then decreases via the value $\beta_1 = t_{\widehat{\beta}_1} = 0$ to infinity with a vertical asymptote for $\overline{r}_{13}$. The t-statistic is depicted as the red curve in figures 5 and 6. The t statistic exceeds 2 ($p < 0.05$) for $|r_{12} - r_{13}r_{23}| > S(N)$, where $S(N)$ is a given threshold.

Theorem 4. *Let* $|r_{12}| = r_{23}$. *Then*

(i) If $r_{12} = -a$ *and* $r_{23} = a \neq 1$ *we have* $\underline{r}_{13} = -1$ *and* $\overline{r}_{13} = 1 - 2a^2$. *Furthermore,* $\widehat{\beta_1}(\underline{r}_{13}) = 0$ *and* $\widehat{\beta_1}(\overline{r}_{13}) - 2a < 0$. *In this case* $\tilde{r}_{13} = \underline{r}_{13}$ *with* $\widehat{\beta_1}(\tilde{r}_{13}) = 0$. *For the*

t-statistic we get:

$$\lim_{r_{13}\to\underline{r}_{13}} t_{\widehat{\beta}_1}(r_{13}) = 0 \text{ and } \lim_{r_{13}\to\overline{r}_{13}} t_{\widehat{\beta}_1}(r_{13}) = -\infty.$$

(ii) If $r_{12} = a$ *and* $r_{23} = a \neq 1$ *we have* $\underline{r}_{13} = 2a^2 - 1$ *and* $\overline{r}_{13} = 1$*. Furthermore,* $\widehat{\beta_1}(\underline{r}_{13}) = 0$ *(if* $a \neq 0$*) and* $\widehat{\beta_1}(\overline{r}_{13}) - 2a < 0$*. In this case* $\tilde{r}_{13} = \overline{r}_{13}$ *with* $\widehat{\beta_1}(\tilde{r}_{13}) = 0$*. For the t-statistic we get:*

$$\lim_{r_{13}\to\underline{r}_{13}} t_{\widehat{\beta}_1}(r_{13}) = +\infty \text{ and } \lim_{r_{13}\to\overline{r}_{13}} t_{\widehat{\beta}_1}(r_{13}) = 0.$$

The second part of the theorem corresponds to the scenario shown in figure 7.

Theorem 5. *Let* $|r_{12}| > r_{23}$*. Then:*

(i) If $r_{12} > 0$ *then* $\widehat{\beta}_1$ *is always positive;* $\widehat{\beta_1}(\underline{r}_{13}) > 0$ *and* $\widehat{\beta_1}(\overline{r}_{13}) > 0$*. In this case we get for the t-statistic:*

$$\lim_{r_{13}\to\underline{r}_{13}} t_{\widehat{\beta}_1}(r_{13}) = +\infty \text{ and } \lim_{r_{13}\to\overline{r}_{13}} t_{\widehat{\beta}_1}(r_{13}) = +\infty.$$

Furthermore there exists a minimum of $t_{\widehat{\beta}_1}$*,* $r^*_{13} = r_{23}/r_{12}$*, in the admissible interval.*

(ii) If $r_{12} < 0$ *then* $\widehat{\beta}_1$ *is always negative;* $\widehat{\beta_1}(\underline{r}_{13}) < 0$ *and* $\widehat{\beta_1}(\overline{r}_{13}) < 0$*. In this case we get for the t-statistic:*

$$\lim_{r_{13}\to\underline{r}_{13}} t_{\widehat{\beta}_1}(r_{13}) = -\infty \text{ and } \lim_{r_{13}\to\overline{r}_{13}} t_{\widehat{\beta}_1}(r_{13}) = -\infty.$$

Furthermore there exists a maximum of $t_{\widehat{\beta}_1}$, $r_{13}^* = r_{23}/r_{12}$, *in the admissible interval.*

An example corresponding to the first part of the theorem is shown in figure 8.