CESifo *Working Paper Series*

NO NEED TO RUN MILLIONS OF REGRESSIONS

Jan-Egbert Sturm
Jakob de Haan*

Working Paper No. 288

April 2000

CESifo
Poschingerstr. 5
81679 Munich
Germany
Phone: +49 (89) 9224-1410/1425
Fax: +49 (89) 9224-1409
http://www.CESifo.de

* We would like to thank Jan Jacobs, Gerard Kuper and Tom Wansbeek for their comments on a previous version of this paper. The usual disclaimer applies.

CESifo *Working Paper No. 288*
April 2000

NO NEED TO RUN MILLIONS OF REGRESSIONS

Abstract

We argue that in modelling cross-country growth models one should first identify so-called outlying observations. For the data set of Sala-i-Martin, we use the least median of squares (LMS) estimator to identify outliers. As LMS is not suited for inference, we then use reweighted least squares (RLS) for our cross-country growth models. We identify 27 variables that are significantly related to economic growth. Subsequently, applying Sala-i-Martin's approach for the data set without outliers hardly reveals any additional information. Variables that are insignificant according to the RLS method are generally not significantly related to economic growth under the Sala-i-Martin approach.

Keywords: Sensitivity analysis, outliers, economic growth

JEL Classification: O4, C21, C52

Jan-Egbert Sturm
Department of Economics
University of Groningen
9700 AV Groningen
The Netherlands

Jakob de Haan
Department of Economics
University of Groningen and CESifo
9700 AV Groningen
The Netherlands
e-mail: j.de.haan@eco.rug.nl

Empirical research on economic growth is plagued by the fact that economic theory does not provide enough guidance for the proper specification of the empirical model. Sala-i-Martin (1997a,b) identifies, for instance, around 60 variables that have been suggested to be correlated with economic growth. The so-called extreme bound analysis of Leamer (1983) and Levine and Renelt (1992) is therefore often used to examine how 'robust' a certain variable of interest is related to economic growth (see e.g. De Haan and Sturm, 2000). In this approach equations of the following general form are estimated:

$$\Delta Y_i = \alpha M_i + \beta F_i + \gamma Z_i + u_i \qquad (1)$$

where the subscript refers to country i; ΔY_i is the average growth of per capita GDP of country i; M_i is a vector of 'standard' economic explanatory variables; F_i is the variable of interest; Z_i is a vector of up to three possible additional economic explanatory variables, which according to the literature may be related to economic growth; and u_i is an error term. The extreme bounds test for variable F says that if the lower extreme bound for b - i.e. the lowest value for b minus two standard deviations - is negative, while the upper extreme bound for b - i.e. the highest value for ß plus two standard deviations - is positive, the variable F is not robust.

Sala-i-Martin (1997a,b) argues that the test applied in the extreme bound analysis is too strong for any variable to really pass it. If the distribution of b has some positive and some negative support, then one is bound to find one regression for which the estimated coefficient changes sign if enough regressions are run. Instead of analyzing the extreme bounds of the estimates of the coefficient of a particular variable, Sala-i-Martin suggests to analyze the entire distribution. This implies, of course, that a large number of regressions have to be run. Sala-i-Martin (1997a,b) considers the distance of the point estimates for ß from zero, averaged over a large set of regression models. Broadly speaking, if the averaged 90 per cent confidence interval of a regression coefficient does not include zero, Sala-i-Martin classifies the corresponding regressor as a variable that is strongly correlated with economic growth. He concludes that a substantial number of variables are strongly related to growth.

In this note we argue that a more useful (and, for sure, less time-consuming) estimation strategy is to identify so-called outlying observations first.[1] It is our contention that after a careful analysis of outliers, the approach suggested by Sala-i-Martin (1997a;b) does not yield much more information. Employing the data set of Sala-i-Martin, we use the least median of squares (LMS) estimator of Rousseeuw (1984, 1985) to identify outlying observations. This technique can cope extremely well with data sets containing outliers. The basic principle of LMS is to fit the majority of the data, after which outliers may be identified as those points that lie far away from the robust fit, i.e. the cases with large positive or negative residuals.[2] LMS by itself is not suited for inference. As proposed by Rousseeuw (1984), this can be resolved by using reweighted least squares (RLS). Subsequently, applying Sala-i-Martin's (1997a,b) approach for the data set without outliers hardly reveals any additional information. Variables that are insignificant according to the LMS/RLS method are generally not significantly related (in the sense of Sala-i-Martin) to economic growth.

The remainder of this note is organized as follows. Section I discusses the concept of outlying observations. Section II briefly explains the estimation technique that we apply, while Section III presents the empirical results. Finally, Section IV offers some concluding remarks.

[1] As will be explained in more detail in Section I, simply looking at the OLS residuals cannot discover the worst type of outliers.

[2] Note that we employ robust in here in a different sense than in the literature referred to above. So far, an explanatory variable has been defined to be 'robust' in case changes in the conditioning information set, i.e. the list of explanatory variables, do not alter its estimated coefficient too much. We define robustness in terms of the observations included in the regression.

I. Outliers

Following Barnett and Lewis (1994, p. 316) we define an outlier as an observation 'lying outside' the typical relationship between the dependent and explanatory variables revealed by the remaining data. For instance, point A in figure 1(a) is clearly an outlier. Outliers in the dependent variable - i.e. in the *y*-direction - often possess large positive or large negative residuals, which are easy to detect by plotting the residuals.[3] Observations may be outlying for several reasons. The most obvious ones involve problems with the quality of the data, and non-linearities in the data that - by definition - cannot be captured by a linear regression model. Outliers in the explanatory variables are even more likely as the number of explanatory variables (k) is usually greater than 1, and hence there are more opportunities for something to go wrong. As figure 1(b) shows an unusual observation in the *x*-direction (B) can actually tilt the OLS regression line. In such a case we call the outlier a (bad) leverage point. Note that looking at the OLS residuals cannot discover bad leverage points. If a leverage point tilts the regression line, deleting the points with the largest OLS residuals implies that some 'good' points would be deleted in stead of the 'bad' leverage point.

[FIGURE 1 ABOUT HERE]

Basically, there are two ways to deal with outliers: regressions diagnostics and robust estimation. Diagnostics are certain statistics mostly computed from the OLS regression estimates with the purpose of pinpointing outliers and leverage points.[4] Often the unusual observations are then removed or corrected after which an OLS analysis on the remaining observations follows. When there is only one unusual observation, some of these methods work quite well. It is, however, much more difficult to diagnose outliers and leverage points when there are several of them.

[3] Note, however, that if x_i is near the center of the set of explanatory observations, as is the case in figure 1(a), it will mainly affect the constant and hardly alter the slope.
[4] See, for instance Belsey et al. (1980) and Chatterjee and Hadi (1988).

Take for instance figure 1c. Deleting either of the two outliers will have little effect on the regression outcome and will therefore not be spotted by the single-case diagnostics. The potential effect of one outlying observation is clearly masked by the presence of the other. Testing for groups of observations to be influential might solve this masking effect problem. However, a serious problem in the multiple observation case is how to determine the size of the subset of jointly influential observations. Suppose we are interested in detecting all subsets of size m=2,3,..., of observations that are considered to be jointly outliers and/or high-leverage. A sequential method might be useful, but where to stop? In the multiple observation case the number of possible subsets for which each diagnostic measure of interest can be computed is: $\frac{n!}{m!(n-m)!}$ where n is number of observations. For m=5 and n=50 this results in over 2 million diagnostics.

Therefore we prefer so-called robust regression techniques that employ estimators that are not strongly affected by outliers. Diagnostics and robust regression have the same goals, only in the opposite order: When using diagnostic tools, one first tries to delete the outliers and then to fit the 'good' data by OLS, whereas a robust analysis first wants to fit a regression to the majority of the data and then to discover the outliers as those points which possess large residuals from that robust solution. One robust estimator is the Least Median of Squares estimator.

II. Least Median of Squares

The effectiveness of estimators in dealing with 'contaminated' observations can be determined using the so-called breakdown point (see e.g. Hampel, 1971). The breakdown point is the smallest fraction of contamination that can cause the estimator used to take on values for $\hat{b}$ arbitrarily far from b.[5] In this section, we will discuss one very robust method, which can be used to analyze extremely messy data sets as well as clean ones: Least Median of Squares.

[5] As shown in figure 1(b), one observation - like point B - is sufficient to break down the OLS estimator. This is independent of the total number of observations n available. Hence, its breakdown point equals $1/n$ that tends to zero for increasing sample size, which reflects the extreme sensitivity of the OLS method to outliers.

The least median of squares (LMS) estimator introduced by Rousseeuw (1984) can be written as:

$$\min_{\hat{b}} \operatorname*{median}_{i=1,\dots,n} e_i^2 \tag{2}$$

where e_i is the residual of case i with respect to the LMS fit. The LMS line has an intuitive geometric interpretation, because it lies at the center of the narrowest strip covering half of the observations. The LMS solution is found by fitting regression surfaces to subsets of k points, where k equals the number of explanatory variables. LMS regression is typically computed by approximate algorithms based on Rousseeuw and Leroy (1987), as e.g. available in S-Plus, SAS/IML 7 and TSP 4.5.

The LMS regression method attains the highest possible breakdown value, namely $([n-k)/2]+1)/n$, which asymptotically equals 50 per cent. This means that the LMS fit stays in a bounded region whenever $(n\text{-}k)/2$ or fewer observations are replaced by arbitrary points and hence is very robust with respect to outliers in the dependent as well as the explanatory variables. The standard error is estimated as:

$$\hat{s} = 1.483(1+\frac{5}{n-k})\sqrt{\operatorname*{median}_{i=1,\dots\dots,n} e_i^2} \tag{3}$$

The constant term is merely a factor used to achieve consistency at Gaussian error distributions (Rousseeuw and Leroy, 1987). The factor $(1+\frac{5}{n-k})$ is a finite-sample correction factor. We define observations with outlying residuals as those observations whose residual is greater than 2.5 times $\hat{s}$, i.e. $\frac{e_i}{\hat{s}} > 2.5$. We use this (rough) yardstick for the standardized residuals since it would determine a (roughly) 99 per cent tolerance interval if they had a Gaussian distribution.

As shown by Rousseeuw (1984), the LMS unfortunately has an abnormally slow convergence rate and hence performs poorly from the point of view of asymptotic efficiency. Because of its low finite-sample efficiency, LMS is not suited for inference. As proposed by Rousseeuw (1984), this can be resolved by using reweighted least squares (RLS). A simple, but effective, way is to put weight zero if the observation is an outlier and weight one otherwise.[6] The resulting estimator is more efficient and yields all the usual inferential output such as *t*-statistics and R^2.

III. Empirical Results

In the empirical analysis we stick as closely as possible to Sala-i-Martin (1997a,b), both in terms of the variables taken into account and in using his data set.[7] As in Sala-i-Martin (1997a,b) the 'standard' variables in our regression are the level of income in 1960, life expectancy in 1960, and the primary-school enrollment rate in 1960. Each of the other 59 variables used by Sala-i-Martin (1997a,b) is added as additional explanatory variable to this base equation. For comparison purposes we first apply OLS. Then the LMS regression technique is used to detect outlying countries, i.e. countries that do not follow the general pattern of the data. After having detected the outliers we apply reweighted least squares (RLS). The results are shown in Table 1. To save space we only present *t*-statistics of the tested variables. The variables are ordered by their absolute *t*-values in the RLS regressions.

[INSERT TABLE 1 ABOUT HERE]

In the full sample we have 103 countries. As not all variables are available for all countries, the sample is sometimes reduced to even only 65. As follows from the third column of Table 1 the number of outlying observations as indicated by the LMS technique varies between 2 and 17. The final column of Table 1 reports the *t*-statistics in case the outlying countries get weight zero in the RLS regression. There are some noteworthy differences between the OLS and RLS estimates. For instance, the variable terms of trade growth is definitely not significant in the standard OLS equation. However, after reweighing 16 outlying countries it becomes highly significant. More or less the same holds for the following variables: revolutions and coups, democratic freedom, ratio workers to population, public consumption share and the fraction of the population speaking a foreign language. Sometimes the

[6] Alternatively, we have used a weight of 0.50 for the outliers.

opposite is also true. For instance, the war dummy becomes insignificant once the outliers are effectively removed. As discussed in the previous section we reweigh those countries whose robust residual is greater than 2.5 times the robust standard error. To check the robustness of our findings to this somewhat arbitrary yardstick we have also used others. The main results are not very sensitive to this and, hence, the qualitative conclusions do not change.[8]

It is interesting to see which countries behave poorly in the sense that they do not follow the general pattern in the data. Table 2 shows those countries outlying in at least 10 per cent of the regressions. Often they are countries with extremely high or low growth rates. As a matter of fact, the 7 fastest growing countries are all included in Table 2. The same is true for the two countries with the lowest growth rates (Madagascar and Iraq). However, some countries that have less extreme growth rates are also identified as outlying. Guyana, for instance, is not a country with an extremely high or low growth rate but in most models this country is considered to be an outlier. Note there is no country outlying in all specifications.

[INSERT TABLE 2 ABOUT HERE]

Finally, we have applied a similar procedure as Sala-i-Martin (1997a,b) but starting from an outlier-free sample.[9] In order to determine the entire distribution of the estimated coefficient of a specific variable of interest, Sala-i-Martin combines the remaining 58 variables in sets of three and adds all possible sets of three variables to the equation. Hence, for each variable he estimates 30,856 $(\frac{58!}{3!55!})$ models. Sala-i-Martin (1997a) finds only three variables (fraction Confucian, equipment investment, and number of years open economy) having a significant *t*-statistic more than 95 per cent of the time. The first column of Table 3 reports the fraction of the 30,856 regressions in which the tested variable is significantly different from zero (defined as

[7] The data set is available at http://www.columbia.edu/~xs23/data/millions.htm.
[8] In case we use a weight of 0.50 for the outliers, the main conclusions do also not change. All results are available on request.

a *t*-statistic with an absolute value larger than two) if outliers are removed from the sample. Using a robust sample the number of variables having a significant *t*-statistic more than 95 per cent of the time increases to 13.

[INSERT TABLE 3 ABOUT HERE]

Like Sala-i-Martin (1997a,b), we have also calculated the fraction of the cumulative distribution function of the estimated coefficient for the tested variable lying on each side of zero (CDF(0)). To be on the safe side, we do not want to impose normality on the density function of the estimates. Therefore, we first compute the area under the density function to the right of zero for each of the 30,856 regressions. We then compute the aggregate CDF(0) of the tested variable as the average of all these individual cumulative distribution functions.[10] The final column in Table 3 shows the results. In case the largest part of the aggregate CDF(0) lies to the left of zero, we report 1 - CDF(0). Hence, the numbers presented in the last column of Table 3 will always be between 0.5 and 1. We only report those 28 variables with an aggregate CDF(0) of more than 0.9. Comparing the results of Table 3 with those of Table 1 reveals that all 27 variables with an absolute RLS *t*-value above 2.0 are also included in Table 3. The only additional variable is exchange rate distortions, which has an absolute *t*-value of 1.74 (and rank 29) in Table 1.[11] Hence, we gain hardly any additional insights from applying the Sala-i-Martin (1997a,b) approach.

Finally, let's compare the variables that we find to be related to economic growth to those identified by Sala-i-Martin (1997a,b). Of the 22 variables selected by Sala-i-Martin only the fraction of GDP in mining and the war dummy are not on our

---

[9] A more profound strategy would of course be to determine for each of the nearly two million (30,856 times 59) regressions which countries are outlying. Due to the search algorithm behind LMS, this would have taken nearly 2 years using S-Plus on a Pentium 200 Mhz.

[10] Sala-i-Martin (1997a,b) also uses a weighted version of this, where the weights are the integrated likelihood function. As noted by Sala-i-Martin (1997a,b), a potential problem with that is that the integrated likelihood might not be a good indicator of the probability that a model is the true model. We therefore prefer the unweighted version.

[11] The Spearman rank correlations between the 59 ordered variables in Tables 1 and 3 are 0.92 in case the variables in Table 3 are ordered by the fraction of significant coefficients, and 0.96 in case they are ordered by the aggregate CDF(0).

list (the first 28 entries of Table 1).[12] According to the RLS results these variables are not robustly related to growth. Hence, for these two variables the results of Sala-i-Martin (1997a,b) are highly determined by a small subset of countries which do not follow the general pattern in the data.

IV.Concluding Remarks

In this note we argue that in empirical research on economic growth one should first identify outlying observations. Employing the data set of Sala-i-Martin, we first use the least median of squares (LMS) estimator to identify outliers and then employ reweighted least squares (RLS) for inference. Subsequently, applying Sala-i-Martin's (1997a,b) approach for the data set without outliers hardly reveals any additional information. Variables that are insignificant according to the LMS/RLS method are generally not significantly related (in the sense of Sala-i-Martin) to economic growth. Hence, detecting outliers seems to be a short-cut for solving the specification uncertainty.

[12] Sala-i-Martin (1997a,b) orders his variable according to the weighted CDF(0) results. In case his results are ordered by the unweighted CDF(0), the only variable in his list of 21 significant variables which is not on our list is the fraction of the population speaking English.

References

Barnett, Vic and Lewis, Toby *Outliers in Statistical Data*, New York, John Wiley and Sons, 1994.

Belsley, David A., Kuh, Edwin and Welsch, Roy E. *Regression Diagnostics*, New York, John Wiley and Sons, 1980.

Chatterjee, Samprit and Hadi, Ali S. *Sensitivity Analysis in Linear Regression*, New York, John Wiley and Sons, 1988.

De Haan, Jakob and Sturm, Jan-Egbert. "How Robust Is the Relationship Between Economic Freedom and Economic Growth?", forthcoming in *European Journal of Political Economy*, 2000.

Hampel, Frank R. "A General Qualitative Definition of Robustness", *The Annals of Mathematical Statistics*, 42, pp. 1887-1896.

Leamer, Edward E. "Let's take the Con Out of Econometrics." *American Economic Review*, June 1983, 73(3), pp.31-43.

Levine, Ross and Renelt, David. "A Sensitivity Analysis of Cross-Country Growth Regressions." *American Economic Review*, September 1992, 82(4), pp.942-963.

Rousseeuw, Peter J.. "Least Median of Squares Regression." *Journal of the American Statistical Association*, 1984, 79, pp. 871-880.

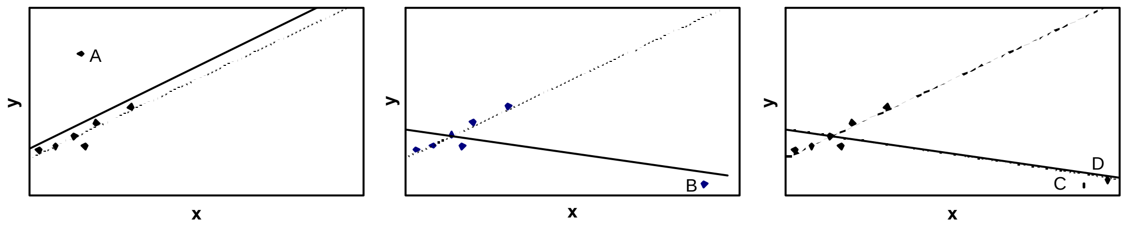
Rousseeuw, Peter J.. "Multivariate Estimation with High Breakdown Point." in Grossmann, W., Pflug, G. Vincze, I. and Wertz, W. (eds.), *Mathematical Statistics and Applications*, Reidel, Dordrecht, 1985, pp. 283-297.

Rousseeuw, Peter J. and Leroy, Annick M. *Robust Regression & Outlier Detection,* New York, John Wiley and Sons, 1987.

Sala-i-Martin, Xavier. "I Just Ran Four Millions Regressions." Mimeo, Columbia University, 1997a.

Sala-i-Martin, Xavier. "I Just Ran Two Millions Regressions." *American Economic Review*, May 1997, 87(2), pp.178-183.

Figure 1: Outlying observations and bad leverage points



The solid lines represent the OLS estimates including the unusual observation(s). The dotted lines represent the OLS estimates without the unusual observations A, B, or C. The dashed line represents the OLS estimate without observations C and D.

Table 1: Estimation results

	Variable	OLS Obs	OLS t-stat.	$\left\lvert \frac{e_i}{\hat{\sigma}} \right\rvert > 2.5$; RLS Outl.	$\left\lvert \frac{e_i}{\hat{\sigma}} \right\rvert > 2.5$; RLS t-stat.
1.	**Fraction Buddhist**	103	3.00	8	7.98
2.	**Fraction Confucian**	103	6.26	5	6.57
3.	**Fraction Muslim**	103	3.39	11	6.27
4.	**Sub-Saharan dummy**	103	−2.41	11	−6.15
5.	**Number of years open economy**	103	6.22	5	6.04
6.	**Non-Equipment investment**	82	2.15	11	5.94
7.	Terms of trade growth	89	−0.35	16	−5.20
8.	**Latin American dummy**	103	−4.04	14	−5.16
9.	**Revolutions and coups**	103	−1.48	12	−4.81
10.	**S.D. of black-market premium**	95	−2.94	17	−4.54
11.	**Political rights**	103	−2.28	10	−4.50
12.	**Equipment investment**	82	6.97	9	4.21
13.	Democratic freedom	93	−1.44	12	−4.06
14.	**Rule of law**	92	3.64	5	4.06
15.	**Absolute lattitude**	103	2.53	11	3.77
16.	Liquid liabilities	65	3.25	12	3.15
17.	**Spanish colony dummy**	103	−2.91	6	−3.13
18.	**Fraction Catholic**	103	−3.66	8	−2.83
19.	Public defence share	96	1.89	7	2.67
20.	**Fraction Protestant**	103	−2.25	4	−2.39
21.	**Primary exports**	100	−3.75	8	−2.33
22.	Ratio workers to population	100	−0.47	9	−2.32
23.	**Civil liberties**	103	−1.86	9	−2.24
24.	Public consumption share	96	−1.35	5	−2.18
25.	Fraction speaking foreign language	103	1.42	7	2.12
26.	**Degree of capitalism**	103	2.61	10	2.08
27.	Age	103	−2.46	8	−2.04
28.	Secondary school enrollment	100	1.07	7	1.74
29.	**Exchange rate distortions**	102	−2.18	7	−1.74

(to be continued)

Table 1: Estimation results (continued)

	Variable	OLS Obs	OLS t-stat.	$\left\lvert \frac{e_i}{\hat{\sigma}} \right\rvert > 2.5$; RLS Outl.	$\left\lvert \frac{e_i}{\hat{\sigma}} \right\rvert > 2.5$; RLS t-stat.
30.	French colony dummy	103	0.24	9	−1.73
31.	Tariff restrictions	83	−0.74	8	−1.66
32.	Fraction of pop. speaking English	103	−2.21	7	−1.52
33.	Size labor force	100	1.34	6	1.44
34.	**Fraction GDP in mining**	103	1.05	3	1.34
35.	Free trade	83	1.69	2	1.33
36.	**War dummy**	102	−1.83	8	−1.30
37.	Population growth	103	0.00	7	−1.12
38.	S.D. of domestic credit	88	−0.33	9	−1.12
39.	Urbanization rate	101	0.68	8	−1.11
40.	Average inflation	98	−1.05	8	−1.10
41.	Human capital * GDP per capita	85	−0.86	8	−0.90
42.	Average years of primary schooling	85	−1.40	11	−0.89
43.	Outward orientation	102	1.26	7	0.89
44.	Ethnolinguistic fractionalization	98	−0.27	5	0.87
45.	Average years of education	85	−0.70	8	−0.81
46.	Fraction Jewish	103	−1.60	11	0.81
47.	Political assassinations	95	−1.21	9	−0.75
48.	Public investment share	95	0.64	6	0.69
49.	S.D. of inflation rate	98	−0.60	8	−0.68
50.	Political instability	95	−0.34	9	0.65
51.	Black Market Premium	93	−0.69	5	0.64
52.	Public education share	98	0.83	9	0.58
53.	British colony dummy	103	0.18	5	0.55
54.	Higher education enrollment	102	0.38	7	−0.47
55.	Area	101	−0.09	7	0.43
56.	Average years of secondary schooling	93	1.48	10	0.35
57.	Fraction Hindu	103	−0.01	9	0.33
58.	Average years of higher education	94	0.77	4	0.30
59.	Domestic credit growth	85	0.34	7	0.24

Bold variables are found to be related to economic growth by Sala-i-Martin (1997a,b).

Table 2: Outlying countries

	Country	# Regressions	Outlying
1.	Korea, South	59	96.6%
2.	Botswana	50	90.0%
3.	Congo (former Zaire)	56	78.6%
4.	Guyana	53	75.5%
5.	Singapore	57	75.4%
6.	Hong-Kong	54	72.2%
7.	Indonesia	55	47.3%
8.	Taiwan	54	40.7%
9.	Madagascar	51	39.2%
10.	Gabon	49	34.7%
11.	Thailand	59	32.2%
12.	Ethiopia	53	26.4%
13.	Japan	59	25.4%
14.	Iraq	53	22.6%
15.	Swaziland	42	14.3%

Table 3: Applying Sala-i-Martin's method to an outlier free sample

		Variable	% Sign.	CDF(0)
	1.	Fraction Confucian	100.00%	1.000
→	2.	**Terms of trade growth**	99.99%	1.000
	3.	Fraction Muslim	99.98%	1.000
	4.	Non-Equipment investment	99.98%	1.000
	5.	Number of years open economy	99.96%	1.000
	6.	S.D. of black-market premium	99.90%	1.000
	7.	Equipment investment	99.94%	1.000
	8.	Latin American dummy	99.72%	1.000
	9.	Sub-Saharan dummy	99.58%	1.000
→	10.	**Democratic freedom**	98.82%	0.999
	11.	Rule of law	97.27%	0.998
	12.	Fraction Buddhist	95.96%	0.997
	13.	Political rights	96.40%	0.997
	14.	**Liquid liabilities**	89.16%	0.992
	15.	Absolute lattitude	88.30%	0.988
→	16.	Revolutions and coups	70.65%	0.970
	17.	**Public consumption share**	78.18%	0.968
	18.	Fraction Protestant	63.03%	0.965
	19.	Fraction Catholic	71.09%	0.964
→	20.	**Public defence share**	65.93%	0.955
→	21.	**Ratio workers to population**	63.35%	0.941
	22.	Primary exports	35.99%	0.940
	23.	Degree of capitalism	46.24%	0.931
→	24.	Spanish colony dummy	67.71%	0.931
	25.	Exchange rate distortions	38.18%	0.928
→	26.	**Fraction speaking foreign language**	48.57%	0.926
	27.	**Age**	10.06%	0.915
	28.	Civil liberties	32.99%	0.907

Bold variables are found *not* to be related to economic growth by Sala-i-Martin (1997a,b) when looking at his *weighted* aggregate CDF(0) results.
Variables with an arrow (→) in front are found *not* to be related to economic growth by Sala-i-Martin (1997a) when looking at his *unweighted* aggregate CDF(0) results.