



DIVISION OF ECONOMICS AND BUSINESS
WORKING PAPER SERIES

A simple Monte Carlo approach to examine sample robustness in growth regressions

Daniel Kaffine
Graham A. Davis

Working Paper 2013-04
<http://econbus.mines.edu/working-papers/wp201304.pdf>

Colorado School of Mines
Division of Economics and Business
1500 Illinois Street
Golden, CO 80401

July 2013

Colorado School of Mines
Division of Economics and Business
Working Paper No. **2013-04**
July 2013

Title: A simple Monte Carlo approach to examine sample robustness in growth regressions*

Author(s):
Daniel Kaffine
Division of Economics and Business
Colorado School of Mines
Golden, CO 80401-1887
dkaffine@mines.edu

Graham A. Davis
Division of Economics and Business
Colorado School of Mines
Golden, CO 80401-1887
gdavis@mines.edu

ABSTRACT

Growth regressions are often influenced by extreme observations in the sample. We demonstrate the usefulness of a simple Monte Carlo method as a diagnostic for sample robustness. We apply the technique to a data set used by Mehlum et al. (2006), who show that institutional quality is decisive for growth in resource rich countries. Monte Carlo sampling reveals that this result hinges crucially on the inclusion of Malaysia in the sample of countries. Inclusion of Malaysia yields robust, significant estimates of the key interaction term between resource abundance and institutions, whereby strong institutions can turn resource abundance into a blessing. Exclusion of Malaysia yields robust, insignificant estimates of the interaction term, whereby institutions cannot overcome the resource curse. Further explorations find that the remaining results in Mehlum et al. (2006) are similarly sensitive to the sample of countries included. We argue that the Monte Carlo method utilised provides easily interpretable representations of the robustness of estimates to an arbitrary sample of countries and should become standard practice in growth regression diagnostics, similar to robustness testing of alternative specifications and additional covariates.

JEL classifications: **C21, O43, Q32, Q33**

Keywords: **DFBETAS, robust regression, sample robustness, Monte Carlo, growth regressions, institutions, resource curse.**

*We thank seminar participants at the University of Wyoming and the Front Range Energy Camp for useful comments. Shelley Norman and Ben Crost provided useful comments on an early draft.

1 Introduction

Growth regression methods have been criticised on many levels, one of which is the instability of ordinary least-squares (OLS) coefficient estimates as the sample is altered in small ways (Brock and Durlauf 2001; Durlauf et al. 2005). An example is the finding by Auerbach et al. (1994) that the external benefits of investment on growth found by DeLong and Summers (1991) hinge on the inclusion of Botswana in the sample. Another is the finding in Temple (1998) that the explanatory power of the augmented Solow model for growth in the OECD countries is almost zero once Portugal and Turkey are removed from the sample. Statistical significance in a model of OECD growth by Nonneman and Vanhoudt (1996) is shown to depend on the inclusion of Japan (Temple 1998). More recently, Herndon et al. (2013) examine the results in Reinhart and Rogoff (2010), and find that their average growth estimates for countries with high public debt/GDP ratios are sensitive to exclusion or inclusion of country-year observations for New Zealand. Easterly (2005) argues more generally that the independent data in growth regressions have long tails, and that extreme observations drive the results.

Growth regressions are typically subject to specification robustness (Levine and Renelt 1992; Sala-i Martin 1997a; Sala-i Martin 1997b; Sala-i Martin et al. 2004; Hauk and Wacziarg 2009), while sample robustness has received less systematic attention. One way to test for specification robustness is to run Monte Carlo simulations to sample over all possible regressors and test for those regressors that are most frequently and consistently influential (Sala-i Martin et al. 2004). In this paper we show that it is a fairly simple procedure to apply Monte Carlo techniques to examine sample robustness - in this case by randomly drawing

near-N subsamples of countries in a delete-k jackknifing exercise.

Multiple-row deletion tests have not been widely adopted by economists, perhaps because they are complex and do not provide a formal criterion by which to identify extreme observations.¹ In this paper we show that plotting the distribution of coefficient estimates and t-statistics from Monte Carlo analysis yields a simple and intuitive graphical approach to check for coefficient robustness to sample. We use the cross-section growth regressions of Mehlum et al. (2006) (hereafter MMT) as a vehicle to illustrate the important aspects of our technique. In this case, Monte Carlo analysis shows that a single country is necessary and sufficient for their key result to hold; inclusion of Malaysia yields robust, significant estimates of the key resource dependence-institutional quality interaction term, suggesting that strong institutions can turn resource abundance into a blessing. Exclusion of Malaysia yields robust, insignificant estimates of the interaction term, and the point estimates suggest that strong institutions cannot overcome the resource curse. Further jackknife sampling exercises find that the remaining results in MMT are also sensitive to the sample of countries included.

More generally, we stress the ease of implementation and interpretability of our Monte Carlo approach for exploring sample robustness in cross-sectional regressions. As demonstrated below, this approach can provide diagnostic information to the applied researcher about subsets of influential points, sample robustness, and covariate interpretation, and can provide a signal that additional analysis should be considered. Given increases in computing power, testing the stability of results to sample should be standard practice, similar to exist-

¹ This is true even if a Gaussian distribution is assumed. As such, Belsley et al. (1980) suggest that the researcher resort to graphical diagnostics.

ing practices for robustness testing with conditioning variables. For the example presented here, estimating each 100,000 random sample draw was completed in Matlab in less than a minute on a vanilla desktop setup.

2 Comparison with other regression diagnostic methods

As noted above, growth regressions have been criticised as being driven by extreme observations. It is useful to formalise the various types of extreme observations. Following the exposition in Birkes and Dodge (1993) let h_{ii} be the i th diagonal entry in the matrix $X(X'X)^{-1}X'$, where X is the $N \times (p + 1)$ matrix of explanatory variables x_{ij} augmented by a column of 1's. h_{ii} is the “potential” of the i th data point to influence the regression. The least squares estimate \hat{y}_i can be expressed as $\hat{y}_i = (1 - h_{ii})\hat{y}_i^* + h_{ii}y_i = \hat{y}_i^* + h_{ii}(y_i - \hat{y}_i^*)$, where \hat{y}_i^* is the estimate obtained when observation i is removed from the data.

An observation that has an extreme value for one of the independent variables is called a *leverage* point, corresponding to a large h_{ii} . An observation is an *outlier* if it is extreme in the dependent variable conditional on the independent variables, given by the residual $y_i - \hat{y}_i$. By contrast, an observation has an *outlier effect* if it is extreme in the dependent variable relative to its predicted value when removed from the regression, given by $y_i - \hat{y}_i^*$. Observations that have a convex combination of leverage and outlier effect $h_{ii}(y_i - \hat{y}_i^*)$ are *influential* points, such that their inclusion or exclusion has substantial impact on model estimates. In the economic growth literature, and in empirical testing in general, researchers

expose their empirics to a variety of diagnostic tests that variously identify outliers, leverage points, and influence points.

It is useful to distinguish between the activity of filtering the sample for erroneous data points and that of testing the regression output for sample robustness. The former involves the search for outliers without regard for regression robustness. The latter involves the search for influential points that affect regression robustness - those that individually or as a group can change coefficient values and t-statistics. Not all sample robustness tests are equally effective at identifying influence points. DFFITS and DFBETAS, the most commonly used diagnostics for sample robustness, produce summary statistics based on a test for individual influential points in a dataset by removing one observation at a time (Belsley et al. 1980). DFFITS identifies the effect of the observation on the forecast, while DFBETAS identifies the effect of the observation on the coefficient estimate.² The delete-one jackknifing undertaken by some researchers to explore sample stability is a diagnostic that mimics the DFBETAS test (e.g., Egorov et al. (2009), Egert (2012)), though the results are examined without recourse to statistical theory. DFBETAS, DFFITS, and delete-one jackknifing have the problem that there may be groups of data points that are together influential but that are not influence points individually. Belsley et al. (1980) and Breiman (1996) provide some ideas for multiple-row deletion tests that can overcome this problem, but as noted in the

² We observe that researchers who are mainly interested in the statistical significance of a particular regression coefficient may mistakenly use the DFFITS test rather than the DFBETAS test in this regard. Researchers also often exclude all data points flagged as influential through either test, even though this is technically not how these statistics should be interpreted and used. Instead, one should perform an iterated test, where the statistics are recomputed for all combinations of omitted observations to see if additional influence points exist conditional on exclusion of the first influential point. Each and every influential observation would then be observed, and “is legitimately deleted altogether only if, once identified, it can be shown to be uncorrectibly in error” (Belsley et al. 1980, p. 16).

introduction, such tests have not been adopted by economists.

Another robustness diagnostic, robust regression, downweights or drops one or more outliers in a regression analysis (Temple 1998; Temple 2000; Clyde and Lee 2001). For example, the least-trimmed squares (LTS) method considered in Temple (1998) minimises the sum of squared errors over the half of the total observations that has the smallest residual sum of squares. Because not all influential points are outliers, and not all outliers are influential points, robust regression will identify influential points only to the extent that they also happen to be outliers.

Our proposal of delete-k jackknifing undertakes an N choose K Monte Carlo sampling exercise, whereby the regression coefficients and t-statistics of interest are collected for each trial and ultimately plotted as a distribution. The presence of individual influence points or groups of influential points is visually detected, and their exact identification made via a max or min search of the stored sampling results.

[Figure 1 about here]

To illustrate the different effectiveness of the diagnostic techniques (DFBETAS, robust regression, and Monte Carlo delete-k jackknifing) in terms of identifying outliers and influence points, Figure 1 below depicts stylised data in a univariate OLS regression. Given the data in the figure, the OLS regression is nominally upward sloping, despite the neutral orientation of the “non-extreme” cloud of points defined as Zone I. Outliers are defined as points above line aa and below line bb, Zone III. Leverage points are defined as points to the left of line cc and to the right of line dd, Zone II. Influence points (Zone IV) can be high residual points that are outliers (Zone IVa), or low residual points that are not outliers (Zone IVb). Per the definition of influence as $h_{ii}(y_i - \hat{y}_i^*)$, Zone IV is strictly convex due to substitutability

between leverage and outlier effect in determining influence.

Points in Zone II (leverage points that are neither outliers nor influential) generally receive little or no attention in diagnostic testing, even though they may well be related to data errors in the dependent variable and thus should be scrutinised for elimination in the same way that outliers are. Outliers and influence points garner more attention. Per our discussion above, DFBETAS and the Monte Carlo method will identify only those outliers that also happen to be influence points (Zone IVa), but will not identify outliers with insufficient leverage to be considered influential (Zone III). By contrast, robust regression will identify all outliers regardless of influence. DFBETAS and the Monte Carlo method will identify all influential points (Zones IVa and IVb), while robust regression will only identify influential points if they also happen to be high residual outliers (Zone IVa). This point is important when considering sample stability of estimates, as robust regression estimates may still be unstable with respect to inclusion or exclusion of influential observations in Zone IVb, as we show in the following section.

While the preceding discussion outlines the broad differences between diagnostic techniques, there is an additional nuance to consider when there are multiple extreme observations. Consider the four extreme points in Zones III, IVa and IVb in Figure 1. The two adjacent observations clustered together in Zone IVa are influence points that may not be identified by single-row deletion methods such as DFBETAS (or delete-one jackknifing) due to masking. By contrast, robust regression and Monte Carlo delete-k jackknifing can accommodate multiple-row deletion (or downweighting) and can thus identify influential groups of observations.

[Table 1 about here]

Two additional considerations are: Does the diagnostic method provide information regarding the economic importance of an identified influence point or influence group? Does the method utilise a formal statistical criterion? DFBETAS only provides a simple influence statistic for each observation, with further analysis required to determine if that observation “matters” in terms of the economic conclusions of the study. By contrast, robust regression provides coefficient estimates and Monte Carlo sampling provides a distribution of coefficient estimates and t-statistics, which can then be compared with OLS estimates to provide inference regarding any economic conclusions. Both DFBETAS and robust regression rely on formal statistical tests in identifying extreme values, while Monte Carlo sampling relies on graphical interpretation.³

Table 1 provides a summary of the above discussion for the three diagnostic methods considered. Returning to the question of sample robustness in growth regressions, we note that our proposed Monte Carlo delete-k jackknife method has some advantages over DFBETAS in that it can identify groups of influence points and provides readily interpretable information. It also has some advantages over robust regression in that it will identify all individual influential points and groups of influential points in the sample. It has the disadvantage with respect to DFBETAS and robust regression as it does not provide a formal criterion for the identification of influential points. As such, the three methods may be seen to be complementary in a sample robustness analysis.

³ Andrews and Pregibon (1978) provide a more sophisticated graphical analysis technique associated with delete-k jackknife tests for outliers. Though their work has been widely cited by statisticians and in econometric methodology texts, we are unable to find a single use of the technique in the growth literature.

3 Identifying influence points that “matter”

3.1 Institutions and the Resource Curse - MMT

No one who studies economic development will be surprised to learn that the quality of government institutions matters for growth - Rodrik (1997)

To illustrate the effectiveness of the different sample robustness methods introduced in the previous section, we test for sample robustness of MMT’s cross-sectional growth regression coefficients. A substantial body of literature has developed to extend and explore the empirical studies of Sachs and Warner (1995) and Sachs and Warner (1997a), who find that resource-abundant countries grow more slowly than similar but resource-poor countries. In a recent review of this ‘resource curse’ literature, Van der Ploeg (2011) casually observes that tests of the impact of resource abundance on growth are “sensitive to changing the sample period, the sample of countries, or the definition of various explanatory variables” [p. 381]. While many authors have explored sample period and alternative explanatory variables, less attention has been paid to the importance of the sample of countries. Exceptions include Knabb (2005), who addresses the issue of country sample by extending DFBETA tests to a 2SLS framework, and Norman (2009), who uses imputation to overcome the listwise deletion problem as additional covariates are introduced. As Norman (2009) notes, “If inclusion in the sample is driven by the availability of conditioning variables, choice of variables amounts to choice of sample, and different results may confuse the different effects of the controlling variables and the changing of the sample” [p. 193].

MMT’s influential paper provides evidence that good institutions can overturn the re-

source curse.⁴ MMT propose that resource-rich countries can obtain either a production equilibrium or a grabber equilibrium. The former is more likely when institutional quality is high, the latter when institutional quality is low. Production equilibria make good use of natural resource rents, as exhibited by high economic growth. Grabber equilibria result in unproductive rent seeking that slows growth. MMT test their proposition using the popular growth data from Sachs and Warner (1997b). However, they do not subject their results to extensive sample robustness tests, providing an opportunity for Monte Carlo sampling.⁵

The main specification in MMT, which builds on Sachs and Warner (1997a), is a simple cross-sectional regression on the 87 countries that have complete data across all of the regressors:

$$Growth_i = \alpha + \beta_1 Res_i + \beta_2 Inst_i + \beta_3 Res_i * Inst_i + \gamma X_i + \epsilon_i \quad (1)$$

where $Growth_i$ is average growth rate of real GDP per capita between 1965 and 1990, Res_i is the measure of resource abundance defined as primary export share of GNP in 1970, $Inst_i$ is an index of institutional quality ranging from zero to one (Knack and Keefer 1995), and additional controls in X_i include initial income level (log of GDP per economically active

⁴ As of February 24h, 2013, the paper has been cited 776 times per Google Scholar since its publication in *The Economic Journal* in 2006. According to Harzig's Publish or Perish shareware, it has the 9th highest average annual citation rate of all papers published in that journal. Alternative papers to consider would be Boschini et al. (2007) or Boschini et al. (2012), who use an approach similar to MMT to examine the extent to which the types of resources and institutions matter for resource curse outcomes. The simple Monte Carlo method we describe below could be applied to their study to test the robustness of their conclusions regarding the importance of alternative measures of mineral intensity and institutions to small changes in sample.

⁵ MMT do exhibit some sensitivity to the concern of sample robustness, as they exclude the African countries from the 87 country sample to see if their results still hold for the 59 non-African countries. In this case their proposition that resource-abundant countries either land in grabbing or producing equilibrium depending on the quality of their institutions remains intact.

population in 1965), openness (index of a country’s openness from 1970 to 1990) and investments (average ratio of real gross domestic investments over GDP from 1970 to 1989).⁶ The coefficients of interest are β_1 , which captures the direct effect of resource abundance on growth, β_2 , which captures the direct effect of institutions on growth, and β_3 , which captures the interaction between institutions and resource abundance. Importantly, under this specification the marginal effect of an increase in resource abundance is given by $\frac{dGrowth_i}{dRes_i} = \beta_1 + \beta_3 * Inst_i$. Thus, even if $\beta_1 < 0$, as long as β_3 is positive and sufficiently larger than $|\beta_1|$, there exists a level of institutional quality whereby the total effect of resource abundance on growth is positive. MMT’s key and somewhat surprising result is that, contrary to received wisdom, institutional quality per se does not impact growth ($\beta_2 = 0$), but the interaction effect β_3 is positive and statistically significant. In a result that has been widely heralded as verifying that institutions nevertheless matter, the point estimates of the coefficients show that resource-abundant countries with “good enough” institutional quality can overcome the resource curse.

[Table 2 about here]

Columns I and III in our Table 2 replicate Regressions 3 and 4 in Table 1 of MMT that produce this result. In column I, the resource-institution interaction term is not included, and as expected resource abundance has a negative impact on growth - evidence of the resource curse. Institutional quality has no bearing on this outcome, which is at first puzzling given that this is the main case-study explanation of why countries like Norway and Botswana escaped the resource curse. However, when the interaction term is included in column

⁶ Data exists for 211 countries, of which 113 have growth data, and the addition of controls results in listwise deletion down to 87 countries.

III, there is a positive and significant interaction effect between resource abundance and institutional quality, and the magnitude on the interaction effect is 7% larger in absolute magnitude than the coefficient on resource abundance. Thus, for countries with institutional quality in excess of $0.93 = 14.34/15.37$ on a scale of 0 to 1 (there are 15 of these in the sample, all of which are developed countries), resource abundance was actually a blessing due to the rents being put to good use in producer equilibria. The policy implication is that institutions are indeed decisive for growth, but only in resource rich countries.

Curiously, based on these results not a single resource-abundant developing country benefited from its resource rents, because each has an institutional quality index below 0.93. Robinson et al. (2002) (later published as Robinson et al. (2006)), which is cited by MMT, suggest that Botswana, Chile, Malaysia, Oman and Thailand all escaped the resource curse and are suggested to have done so because they had good institutions. MMT make particular note of Botswana as a resource growth winner, and yet their results suggest that Botswana, Chile, Malaysia, and Thailand were all harmed by their resources.⁷ In these countries, grabber institutions combined with resource abundance are estimated to have slowed growth from -0.2%/year (Botswana) to -1.4%/year (Malaysia) over the 25 year sample period.

3.2 Monte Carlo analysis of sample robustness

The 87 countries in this analysis are a subset of the 211 countries listed in the Sachs and Warner data set. Countries for which a single data entry was missing were listwise deleted from the regression and excluded from the analysis.⁸ We can never know how different

⁷ There is no institutional quality data for Oman.

⁸ One may rightly question whether the excluded countries are missing completely at random, as this could potentially bias results. However, we are not interested in bias per se,

the results would be if the regression was run over a sample of 88 countries, for example, or whether the results can be generalised beyond the sample of countries used to generate them. But we can get an inference as to regression stability, and thus the need for further analysis, by running regressions over a sample with an arbitrary and small subset of countries removed.

Consider the following thought experiment: Imagine alien researchers investigating the resource curse on several thousand near-identical planets. The only difference between these near-identical planets was that instead of a dataset with 87 countries, one data point for two different countries was arbitrarily missing, such that each researcher was examining a particular 85 country subsample after listwise deletion. How different would the conclusions reached by these researchers be, depending on the particular countries dropped from the sample?

[Figure 2 about here]

We can formally examine the outcome of our thought experiment with an $N = 87$ choose $K = 85$ Monte Carlo simulation whereby 85 countries are randomly drawn and estimated via Equation (1). Figure 2 displays the Monte Carlo distribution of the estimated interaction coefficient and its corresponding t-statistic from 100,000 simulations. While the bulk of coefficient estimates are in line with the results in Table 2, column III, the eye is immediately drawn to a cluster of subsamples with coefficient estimates centered at roughly 12. Similarly, there is a cluster of subsamples with t-statistics centered at roughly 1.65. These points are worthy of further investigation as they represent samples where the coefficient estimate is such that the resource curse is not overturned or where the interaction term is not significant but rather the stability of regression results to small changes in sample.

at the 5% level.⁹ Calculating the probability that $t < 1.96$, we see that in 2.3% of the estimations the interaction term is not significant at the 95% level, which precisely equals the probability of excluding a particular country given $N = 87$ choose $K = 85$.

Examining these 2.3% of draws, we determined that Malaysia was the pivotal country for both the coefficient estimates and statistical significance - any subsample of 85 countries that includes Malaysia yields a significant interaction term and a resource curse that can be overcome with sufficient institutional quality. Any subsample that excludes Malaysia yields an insignificant interaction term and a point estimate suggesting that institutional quality cannot overcome the resource curse. Thus Malaysia is both necessary and sufficient for the results in MMT to hold. Returning to our alien researchers on near-identical planets, their conclusions would be substantially different depending on whether or not they received a dataset that included Malaysia.

Our Table 2 columns II and IV confirm the sensitivity of the “institutions are decisive” result to the inclusion of Malaysia. Column II suggests that removing Malaysia from the dataset has no obvious impact on the resource abundance coefficient (or any other coefficient). However, when Malaysia is excluded from the interaction regression in Column IV, the point estimate on the interaction term falls to 11.95, is insignificant at 5% level, and is insufficient in magnitude to overcome the resource abundance coefficient of -13.14.¹⁰ Thus, had Malaysia been missing any one of the independent variables and therefore been listwise deleted, one

⁹ Of course, there is no a priori guarantee that the subsamples with a coefficient estimate of roughly 12 correspond to those subsamples with t-statistics near 1.65, and so at this point these are two separate concerns.

¹⁰ While one could argue that the interaction term is still marginally significant at the 10% level, the crucial test of whether or not the point estimate on the interaction term exceeds the point estimate on resource abundance fails.

might conclude that institutional quality may be irrelevant in explaining the divergent growth outcomes of the resource rich countries.

The results of our Monte Carlo analysis are also useful in interpreting the stability of regression results to additional conditioning variables. In Regressions 3 through 6 of their Table 2, MMT perform standard robustness tests over additional conditioning variables - secondary school enrollment rate, ethnic fractionalization, and language fractionalization.¹¹ In each case the interaction term remains positive and larger than the direct resource abundance effect. Importantly, these four tests of specification robustness drop the country sample from 87 to as low as 74 due to listwise deletion, but all include Malaysia.

In regressions I through IV of our Table 3, we present the results for the conditioning variable regressions after removing Malaysia. In all cases, the interaction effect remains statistically insignificant and the point estimate remains smaller in magnitude than the resource abundance effect.¹² Had Malaysia been dropped from the sample in these supplemental regressions either due to lack of data or accidental omission, any lack of significance on the interaction effect would have been attributed to the added conditioning variable. As a counterexample, had the Monte Carlo distribution of the interaction term t-statistics in a $N = 87$ choose $K = 74$ experiment always exceeded 1.96 in the base regression specification,

¹¹ Secondary school enrollment rate is taken from from Sachs and Warner (1995) and ethnic and language fractionalization are from Alesina et al. (2003).

¹² Note that MMT appear to have lost one country in these regressions, as we have one more country in each regression than they do. We have determined that they omitted Guyana from regressions 3 and 6 in their Table 2, and omitted Ivory Coast in regressions 4 and 5. We dropped Malaysia from this subsample and obtain the same qualitative results as in our Table 3; the interaction coefficient is not significant and the point estimate is too small to overcome the resource curse. We also note that their secondary education data, which is available to four decimal points in the Sachs and Warner data file, is truncated to two decimal places in their regressions.

one would be reasonably confident that changes in the significance of the coefficient of interest during robustness checks were truly driven by the conditioning variables and not the change of sample.¹³ This highlights the utility of our method for assessing sample stability when interpreting the addition of conditioning variables across supplemental regressions.

3.3 Comparison with other regression diagnostics

Monte Carlo analysis has determined that whether or not institutional quality matters for growth in resource-rich countries depends on whether Malaysia is included in the analysis. As Belsley et al. (1980, p. 16) note, there is always the philosophical question as to whether high-influence data points should be included in a regression. They suggest that if an influence point is due to data error, it should be excluded. Our inspection of the data series for Malaysia reveals no obvious problems.

Temple (1998) suggests that robust regression, specifically least-trimmed squares (LTS), can be helpful in objectively adjudicating choice-of-sample questions as it downweights outliers, a suggestion echoed by other authors (Brock and Durlauf 2001; Durlauf et al. 2005). How would robust regression deal with an influential observation such as Malaysia? Estimating Equation (1) for the full sample of 87 countries using LTS (as well as other robust regression techniques with alternative downweighting methods) yields qualitatively identical estimates to those in MMT (our Table 2, Column III). However, LTS estimates of Equation (1) for the sample with Malaysia excluded yields qualitatively identical estimates to

¹³ For example, the coefficient on resource abundance (β_1) is always strongly statistically significant in the $N = 87$ choose $K = 85$ experiment. In fact, the interaction term stands out as the one coefficient that is not robust to sample - initial income level, openness, resource abundance, and investments remain statistically significant regardless of sample.

those in Table 2, Column IV, whereby the resource curse is not overturned. The intuition is straightforward: Malaysia's strong leverage leads to small residuals associated with that data point, and thus it is not downweighted in robust regression. Because Malaysia is not downweighted, robust regression would not have identified Malaysia as an influence point. As a result, the LTS estimates are just as unstable as the OLS estimates with respect to inclusion or exclusion of Malaysia. Returning to Figure 1, the data point in Zone IVb is indicative of Malaysia's influence on the regression. As we note in Table I, robust regression does not identify influence points in this region.

This limitation of LTS is not unique to this dataset. We also examined the growth data used in Auerbach et al. (1994), who found removing Botswana significantly affects the results of DeLong and Summers (1991). Applying our Monte Carlo analysis, we find a bimodal distribution of coefficients nearly identical in shape to Figure 2, where inclusion or exclusion of Botswana has a similar effect as Malaysia in the MMT dataset. Applying robust regression techniques per Temple (1998) and Durlauf et al. (2005) yields a similar story as well: LTS estimates for the full sample are consistent with those in DeLong and Summers (1991), while LTS estimates when Botswana is excluded are consistent with those in Auerbach et al. (1994). As in the case of Malaysia in the MMT dataset, Botswana in this dataset appears to be an influence point, but not an outlier due to its strong leverage (Zone IVb in Figure 1), and thus is not downweighted in the robust estimation.

Finally, Malaysia's influence in MMT's results could have been detected using the DFBE-TAS test. The DFBE-TAS statistic for Malaysia is 0.53, which does not exceed the absolute cutoff of 2.0 but does exceed the sample-size adjusted cutoff of $2/\sqrt{n} = 0.21$. The same statistic also flags Guyana and Switzerland as single influence points. The Monte Carlo

approach shows that the MMT results are in fact not dependent on the inclusion of these last two countries in the sample, and that removing them, as is common practice following a DFBETAS test, unnecessarily reduces the sample size. Monte Carlo sampling has the additional advantage of illustrating the qualitative impact of Malaysia graphically, as in Figure 2, where it can immediately be seen that the range of interaction term coefficient estimates and t-statistics across subsamples is enough to bring the main results of the analysis into doubt. In a DFBETAS test, this impact may not be evident, as if Malaysia, Guyana, and Switzerland are removed prior to analyzing the regression output, one would not see that the results for or against institutions mattering for growth are conditional on Malaysia's inclusion in the sample.

Monte Carlo analysis can also identify whether or not there are groups of countries that are acting together as influence points. The next section exemplifies the use of the Monte Carlo technique in identifying the influence of clusters of countries, something that the DFBETAS test cannot do.

4 Identifying groups of influential observations

When testing different regression specifications or different measures of specific independent variables, researchers often use the largest subsample for which all of the data is available across all tests. For example, Sachs and Warner (1997a) use a subsample of 87 countries in the first three regressions of their paper, even though data is also available for Cape Verde Islands, Iceland, Fiji, and Panama in the first regression. In cases like this, one can use Monte Carlo sampling to check for subsample robustness in the usual way. Here, though,

one can also test for robustness by *adding* countries or groups of countries to the sample. We provide such an example, again in reference to the impact of institutions on the resource curse as investigated by MMT.

Aware that their theory of institutions applies best to lootable resources like minerals and energy, MMT estimate a version of Equation (1) in which share of domestic mineral and energy production in GNP is used instead of primary resource exports. The mineral production data is available in the Sachs and Warner (1997a) data set. There are 91 countries for which all of the data for Equation (1) are available, and yet in keeping with standard practice, MMT restrict their analysis to the 87 country sample used in their base case regression in Table 2. Thus, in this exercise, the 87 countries represent a subsample of the 91 available countries. In Column V of our Table 2 we present the results of this estimation.¹⁴ The coefficients on both the direct resource curse effect and the resource-interaction effect grow. MMT argue these results suggest that their model is more appropriate for explaining differences in growth outcomes for mineral- and energy-abundant countries than for primary resource-abundant countries. We would add that the institutional quality threshold here is $17.71/29.43 = 0.60$. This makes more sense than the results based on resource exports, as Malaysia, along with Botswana, Chile and Thailand, would now be seen as having benefited from resources due to producer-friendly institutions. Furthermore, we have confirmed that the inclusion of Malaysia is not driving this result.

Column VI of Table 2 presents the regression results for the full 91 country sample. The

¹⁴ Our results differ from MMT's in the second decimal point in several cases. We can exactly replicate the results in MMT, save for the coefficient on institutional quality (-0.25 vs -0.20), if we round the SNR variable (available to 10 decimal places in the Sachs and Warner data set) to two decimal places.

point estimate of the interaction effect is now insignificant at the 5% level, and the mineral abundance effect has shrunk to roughly the same magnitude as the resource abundance effect. The four countries dropped in Column V due to lack of resource export data are the seemingly random grouping of Guinea-Bissau, Luxembourg, Mozambique, and Papua New Guinea. Yet whether this group of four countries are included yields substantially different interpretations of the effect of resources and institutions on economic growth. Before proceeding to a more detailed investigation of sample effects, regressions V through VIII of our Table 3 perform the MMT specification checks that add secondary education, ethnic fractionalization and language fractionalization, for the full 91 country sample. In each case, the coefficient on the interaction term is statistically insignificant at the 5% level (even though these regressions all include Malaysia).

Now, a skeptic might argue that the point estimate in the full 91 country sample is close to significant at the 5% level. Based on the point estimate, the producer equilibrium-grabber equilibrium cutoff is 0.64, which is also reasonable. Maybe one should conclude that the $N = 91$ and $N = 87$ regression results are not that far apart, and that institutions can be deemed generally important. But one would only want to make such a conclusion if one knew that the institution interaction term was significant or close to significant for *any* subsample of 90, 89, 88, or 87 countries, and not just the 87 country subsample that happened to exclude Guinea-Bissau, Luxembourg, Mozambique, and Papua New Guinea.

This potential sensitivity to groups of countries within the sample provides an excellent test case for Monte Carlo analysis. In the context of our alien researchers, suppose that each of 2,672,670 researchers is handed a unique dataset where four different countries are dropped from the initial 91 countries with mineral production data. How often would they

conclude that institutions matter for growth in mineral-producing countries, and how often would they not? We mimic this arbitrary process of country exclusion with an $N = 91$ choose $K = 87$ Monte Carlo simulation.

[Figure 3 about here]

Figure 3 shows the Monte Carlo distribution of the estimated interaction coefficient and corresponding t-statistic from 100,000 simulations. Strikingly, in roughly 70% of the cases examined, dropping four arbitrary countries from the sample results in a statistically insignificant interaction effect. In contrast to the bimodal distribution of the interaction coefficient in Figure 2, here the interaction coefficient exhibits a large range of values, from a low of -1.47 (good institutions are bad for growth) to 37.19.¹⁵ Again, depending on the dataset our alien researchers were handed, extremely different conclusions would have been reached regarding the impact of institutions on mineral-producing economies. Furthermore, the reported estimate in MMT of $\beta_3 = 29.43$ and $t = 2.66$ is in the 99.9% tail of each distribution. While it is standard practice to hold country sample fixed when conducting specification checks, it turns out that the exclusion of Guinea-Bissau, Luxembourg, Mozambique, and Papua New Guinea could hardly have been more influential in revealing that institutions are decisive for growth in mineral-producing countries.

While the Monte Carlo exercise in Section 3 revealed a lack of sample robustness to the inclusion of Malaysia, Monte Carlo sampling in this case has revealed a general lack of sample robustness, with substantial variation in the interaction coefficient and t-statistic estimates depending on small changes in country sample. Uncovering this general instabil-

¹⁵ These represent lower bounds on the range, as not every 87 country subsample is drawn from the 2,672,670 possible combinations given that we only run 100,000 trials.

ity to sample would have been difficult under DFBETAS, which flags Guyana, Papua New Guinea, Zaire, Australia, Guinea-Bissau and Trinidad and Tobago as individual influential points. Ignoring the issue that DFBETAS identifies influential observations conditional on the inclusion of other potentially influential points, simply excluding these six countries results in an interaction effect large enough to overturn the resource curse, but is statistically insignificant at the 10% level ($p = 0.13$). At best, this hints at the lack of sample robustness revealed in Figure 3. Similar to the results in Section 3, robust regression via LTS is just as sensitive as OLS to the sample of countries included. For the 91 country sample, LTS substantially downweights Hong Kong, Egypt, Congo, Tunisia, Guinea-Bissau, South Korea, Nicaragua, and Syria as outliers, with Guinea-Bissau the only country flagged by DFBETAS as a potentially influential point.

5 Conclusions

Growth regressions have been noted to be sensitive to extreme observations. In this paper we differentiate between extreme observations that are leverage points, outliers, and influence points. We show that two common sample robustness diagnostics, DFBETAS and robust regression, differ in their ability to identify outliers and influence points. We propose a third and simple complementary diagnostic, Monte Carlo delete-k jackknife sampling, that is specifically designed to identify influence points and that has certain advantages over the other two methods.

Through the use of an illustrative dataset used by MMT to test the importance of institutions for growth in resource-intensive economies, Monte Carlo sampling finds that the

statistical support for the proposition that good institutions can overcome the resource curse is not robust to sample selection. Specifically, had Malaysia been missing any of the independent variables, and thus been listwise deleted in the same way as 124 other countries in the population, the interaction effect between resource abundance and institutions would have been found to be statistically insignificant and too small in magnitude to overturn the resource curse. Alternatively, had Malaysia been missing data for one of the additional conditioning variables chosen for the specification checks, or had it been accidentally omitted in these robustness checks as was the case for Guyana and Ivory Coast, the conditioning variable would have been erroneously highlighted as nullifying the noted effect of institutions. While the DFBETAS statistic also flags Malaysia as being influential to the result, Monte Carlo analysis reveals that two other countries identified as influential by DFBETAS are nonetheless inconsequential in terms of the economic conclusions of the MMT paper. Robust regression analysis on the other hand, does not downweight Malaysia because it is not an outlier, but rather an influence point. As such, robust regression is just as sensitive as OLS to the inclusion of Malaysia in this dataset.

Monte Carlo analysis can also be used to examine whether groups of observations can influence regression estimates. In the MMT paper, when mineral production is instead used as the resource intensity measure, the result that institutions matter holds for the same 87 countries as in the baseline analysis. We confirm that in this case the result is not conditional on the inclusion of Malaysia. On the other hand, Monte Carlo sampling reveals that 70% of arbitrary sets of 87 out of the 91 countries with complete mineral production data yield a statistically insignificant interaction effect, with substantial variation in the coefficient estimates depending on the set of observations included. The particular 87 country

subsample that MMT report results for is in the extreme tail of the 30% of sets of 87 countries that do yield significant results.

MMT's finding that institutions affect growth in resource-rich countries is intuitively satisfying and robust to the standard investigative methods they employ. However, our Monte Carlo approach has shown that this is a data set that yields unstable OLS estimates and warrants further investigation. Problems found in the small number of other sample robustness investigations in the growth literature causes us to suspect that the sample instability found here is not unique to this data set, but rather is common to growth data in general. We recommend that the relatively simple exercise of Monte Carlo sampling over country observations become standard diagnostic practice in testing for influence points and groups of influence points in growth regressions.

References

- Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003). Fractionalization. *Journal of Economic Growth* 8(2), 155–194.
- Andrews, D. F. and D. Pregibon (1978). Finding the outliers that matter. *Journal of the Royal Statistical Society. Series B (Methodological)* 40(1), 85–93.
- Auerbach, A., K. Hassett, and S. Oliner (1994). Reassessing the social returns to equipment investment. *Quarterly Journal of Economics* 109(3), 789–802.
- Belsley, D., E. Kuh, and R. Welsch (1980). *Regression diagnostics*. J Wiley, New York.
- Birkes, D. and Y. Dodge (1993). *Alternative Methods of Regression*, Volume 282. Wiley-Interscience.
- Boschini, A. D., J. Pettersson, and J. Roine (2007). Resource curse or not: A question of appropriability. *Scandinavian Journal of Economics* 109(3), 593–617.
- Boschini, A. D., J. Pettersson, and J. Roine (2012). The resource curse and its potential reversal. *World Development In Press*.
- Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *Annals of Statistics* 24(6), 2350–2383.
- Brock, W. and S. Durlauf (2001). Growth empirics and reality. *World Bank Economic Review* 15(2), 229–272.

- Clyde, M. and H. Lee (2001). Bagging and the Bayesian bootstrap. In T. Richardson and T. Jaakkola (Eds.), *Artificial Intelligence and Statistics 2001*, pp. 169–174.
- DeLong, J. and L. Summers (1991). Equipment investment and economic growth. *Quarterly Journal of Economics* 106(2), 445–502.
- Durlauf, S., P. Johnson, and J. Temple (2005). Growth econometrics. In P. Aghion and S. N. Durlauf (Eds.), *Handbook of economic growth*, Volume 1A, pp. 555–677. North-Holland: Amsterdam.
- Easterly, W. (2005). National policies and economic growth: a reappraisal. In P. Aghion and S. N. Durlauf (Eds.), *Handbook of economic growth*, Volume 1A, pp. 1015–1059. North-Holland: Amsterdam.
- Egert, B. (2012). Fiscal policy reaction to the cycle in the oecd: Pro- or counter-cyclical? *CESifo Working Paper No. 3777*.
- Egorov, G., S. Guriev, and K. Sonin (2009). Why resource-poor dictators allow freer media: A theory and evidence from panel data. *American Political Science Review* 103(4), 645–668.
- Hauk, W. and R. Wacziarg (2009). A Monte Carlo study of growth regressions. *Journal of Economic Growth* 14(2), 103–147.
- Herndon, T., M. Ash, and R. Pollin (2013). Does high public debt consistently stifle economic growth? a critique of reinhart and rogoﬀ. *Political Economy Research Institute Working Paper Series* (322).
- Knabb, S. (2005). The contribution of institutions, trade, and geography to the development process: How robust is the empirical evidence to variations in the sample? *Empirical Economics* 30(2), 393–409.
- Knack, S. and P. Keefer (1995). Institutions and economic performance: Cross-country tests using alternative institutional measures. *Economics & Politics* 7(3), 207–227.
- Levine, R. and D. Renelt (1992). A sensitivity analysis of cross-country growth regressions. *American Economic Review* 82(4), 942–963.
- Mehlum, H., K. Moene, and R. Torvik (2006). Institutions and the resource curse. *Economic Journal* 116(508), 1–20.
- Nonneman, W. and P. Vanhoudt (1996). A further augmentation of the Solow model and the empirics of economic growth for OECD countries. *Quarterly Journal of Economics* 111(3), 943–953.
- Norman, C. (2009). Rule of law and the resource curse: Abundance versus intensity. *Environmental and Resource Economics* 43(2), 183–207.
- Reinhart, C. M. and K. S. Rogoﬀ (2010). Growth in a time of debt. *The American Economic Review* 100(2), 573–78.
- Robinson, J., R. Torvik, and T. Verdier (2002). Political foundations of the resource curse. *CEPR Discussion Paper 3422*.
- Robinson, J., R. Torvik, and T. Verdier (2006). Political foundations of the resource curse. *Journal of Development Economics* 79(2), 447–468.

- Rodrik, D. (1997). TFPG controversies, institutions and economic performance in East Asia. *NBER Working Paper No. 5914*.
- Sachs, J. D. and A. M. Warner (1995). Natural resource abundance and economic growth. *NBER Working Paper No. 5398*.
- Sachs, J. D. and A. M. Warner (1997a). Natural resource abundance and economic growth. *Working Paper, Harvard University*.
- Sachs, J. D. and A. M. Warner (1997b). Sources of slow growth in African economies. *Journal of African Economies* 6(3), 335–376.
- Sala-i Martin, X. (1997a). I just ran four million regressions. *NBER Working Paper 6252*.
- Sala-i Martin, X. (1997b). I just ran two million regressions. *American Economic Review* 87(2), 178–183.
- Sala-i Martin, X., G. Doppelhofer, and R. Miller (2004). Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. *American Economic Review*, 813–835.
- Temple, J. (1998). Robustness tests of the augmented Solow model. *Journal of Applied Econometrics* 13(4), 361–375.
- Temple, J. (2000). Growth regressions and what the textbooks don't tell you. *Bulletin of Economic research* 52(3), 181–205.
- Van der Ploeg, F. (2011). Natural resources: Curse or blessing? *Journal of Economic Literature* 49(2), 366–420.

Table 1: Comparison between alternative regression diagnostics

Criterion	DFBETAS	Robust regression	Monte Carlo
Identifies all outliers	No ^a	Yes	No ^a
Identifies all individual influence points	Yes	No ^b	Yes
Identifies groups of influence points	No	No ^b	Yes
Economic inference	No	Yes	Yes
Formal criterion	Yes	Yes	No

Notes: a - Outliers are identified only if they also happen to be influence points. b - Influence points are identified only if they also happen to be outliers.

Table 2: Comparison with Mehlum et al. (2006)

Dependent variable: GDP growth						
	I	II	III	IV	V	VI
Initial income level	-1.28*	-1.24*	-1.26*	-1.24*	-1.33*	-1.24*
	(-6.65)	(-6.52)	(-6.71)	(-6.59)	(-6.25)	(-5.68)
Openness	1.45*	1.27*	1.66*	1.50*	1.86*	1.92*
	(3.36)	(2.95)	(3.87)	(3.35)	(3.76)	(3.75)
Resource abundance	-6.69*	-7.53*	-14.34*	-13.14*		
	(-5.43)	(-5.89)	(-4.21)	(-3.68)		
Mineral abundance					-17.70*	-13.60*
					(-3.17)	(-2.43)
Institutional quality	0.58	0.59	-1.35	-0.91	-0.26	-0.26
	(0.64)	(0.66)	(-1.13)	(-0.73)	(-0.24)	(-0.23)
Investments	0.15*	0.15*	0.16*	0.15*	0.15*	0.15*
	(6.73)	(6.77)	(7.15)	(7.02)	(6.27)	(6.22)
Interaction term			15.37*	11.95	29.43*	21.23
			(2.40)	(1.68)	(2.67)	(1.94)
Observations	87	86	87	86	87	91
Adjusted R ²	0.69	0.70	0.71	0.71	0.63	0.61

Notes: Dependent variable is real per capita GDP growth. *t*-values in brackets. Columns I and III are replications of Regressions 3 and 4 in Table 1 of Mehlum et al. (2006). Column V is a replication of Regression 1 in Table 2 of Mehlum et al. (2006). Columns II and IV drop Malaysia. Column VI uses the 91 countries for which mineral abundance data is available. * indicates 5 percent significance.

Table 3: Covariate sensitivity in Mehlum et al. (2006)

Dependent variable: GDP growth								
	I	II	III	IV	V	VI	VII	VIII
Initial income level	-1.32*	-1.32*	-1.33*	-1.46*	-1.24*	-1.38*	-1.44*	-1.45*
	(-5.91)	(-6.89)	(-6.03)	(-5.55)	(-4.68)	(-6.28)	(-5.57)	(-4.50)
Openness	1.40*	1.45*	1.50*	1.34*	1.73*	1.88*	1.92*	1.71*
	(2.93)	(3.26)	(3.32)	(2.80)	(3.15)	(3.77)	(3.72)	(3.11)
Resource abundance	-12.90*	-12.36*	-13.48*	-12.58*				
	(-3.52)	(-3.48)	(-3.74)	(-3.39)				
Mineral abundance					-13.74*	-11.77*	-12.87*	-12.53*
					(-2.36)	(-2.15)	(-2.28)	(-2.13)
Institutional quality	-0.22	-0.73	-0.80	0.01	-0.05	0.00	0.28	0.24
	(-0.16)	(-0.59)	(-0.62)	(0.01)	(-0.04)	(0.00)	(0.23)	(0.18)
Investments	0.15*	0.14*	0.15*	0.13*	0.16*	0.14*	0.15*	0.14*
	(5.62)	(6.49)	(6.76)	(4.87)	(5.50)	(5.56)	(5.75)	(4.55)
Interaction term	11.01	12.00	13.02	11.36	18.33	19.64	21.09	18.09
	(1.46)	(1.71)	(1.81)	(1.50)	(1.51)	(1.85)	(1.92)	(1.49)
Secondary	-0.29			-0.24	0.07			0.28
	(-0.21)			(-0.17)	(0.04)			(0.18)
Ethnic Frac.		-0.89		-0.78		-1.40*		-1.09
		(-1.73)		(-1.15)		(-2.45)		(-1.39)
Language Frac.			-0.35	-0.21			-0.80	-0.30
			(-0.73)	(-0.35)			(-1.50)	(-0.42)
Observations	76	86	84	74	80	91	89	78
Adjusted R ²	0.71	0.71	0.70	0.71	0.62	0.64	0.61	0.62

Notes: Dependent variable is real per capita GDP growth. *t*-values in brackets. Columns I-IV replicate regressions 3-6 in Table 2 of Mehlum et al. (2006) but exclude Malaysia. Columns V-VIII consider all 91 countries for which mineral abundance data is available. * indicates 5 percent significance.

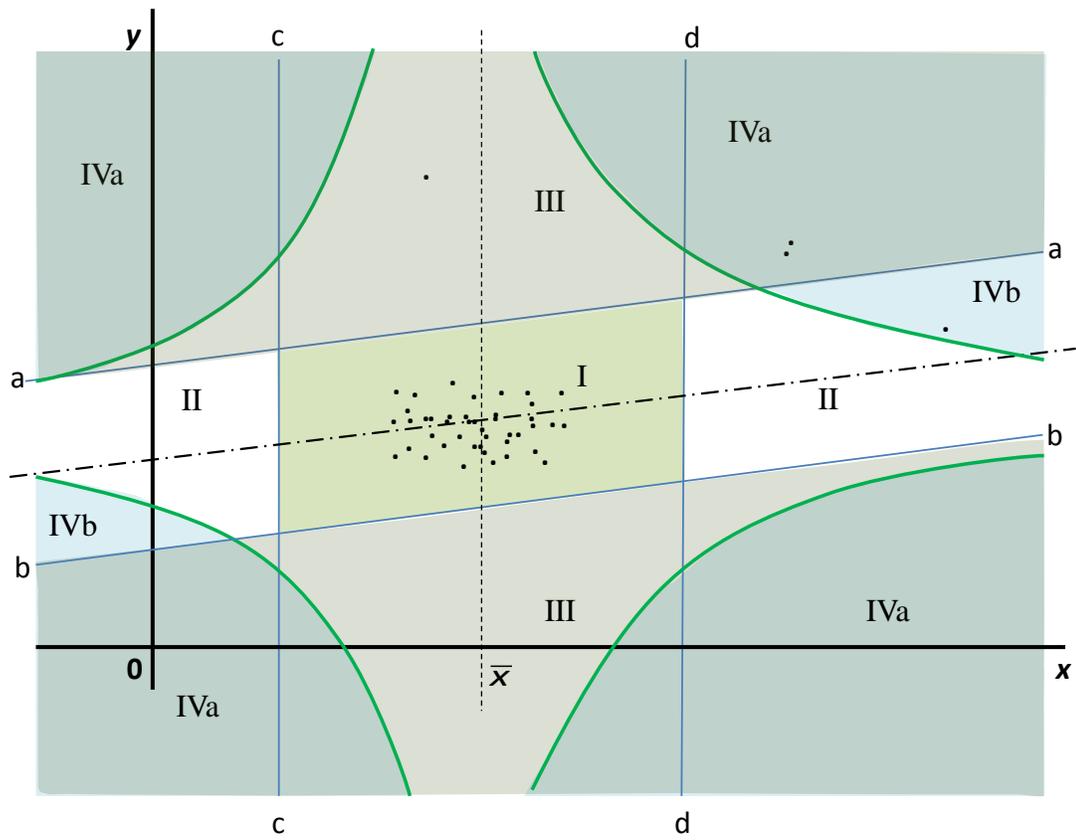


Fig 1. Illustration of outliers, leverage and influence points

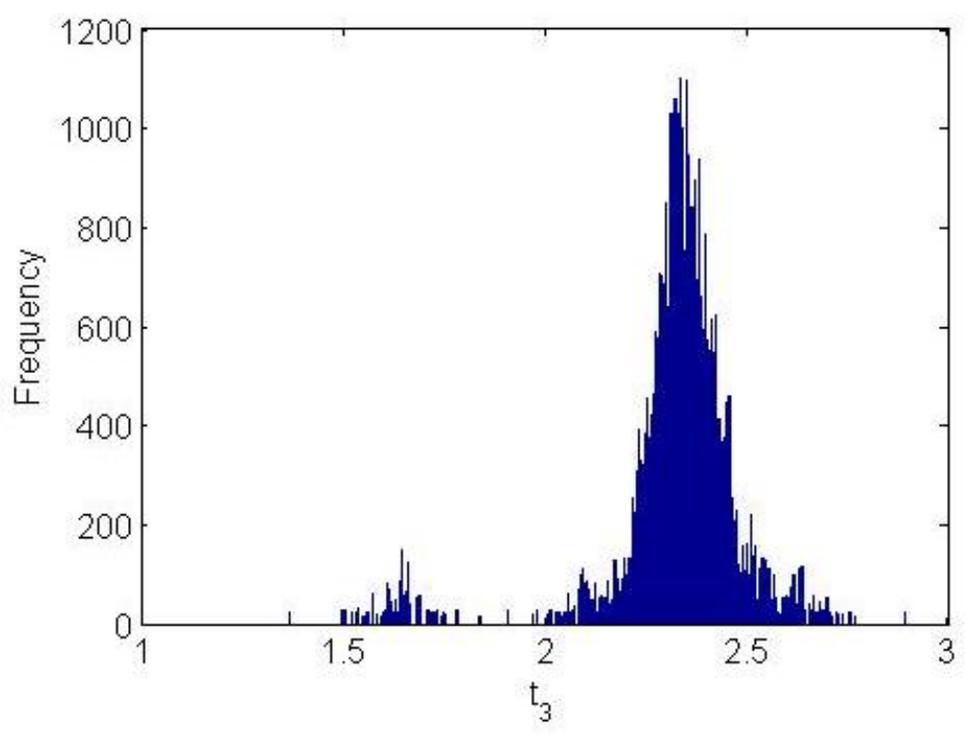
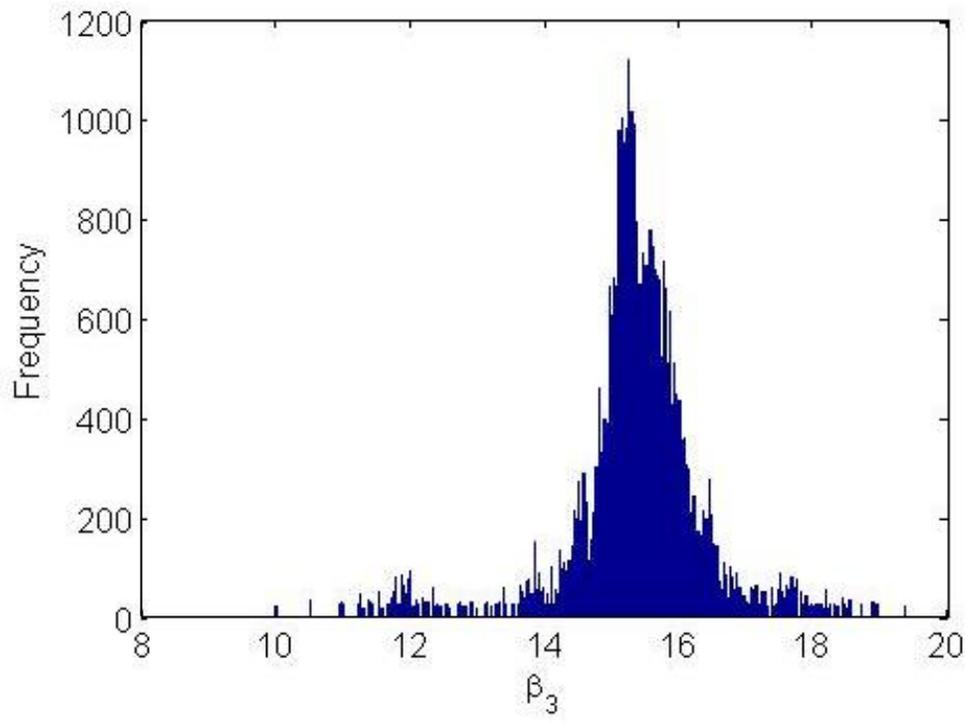


Fig 2. Monte Carlo distribution of the resource exports-institution interaction coefficient and corresponding t-statistic from 100,000 draws of $N = 87$ countries choose $K = 85$.

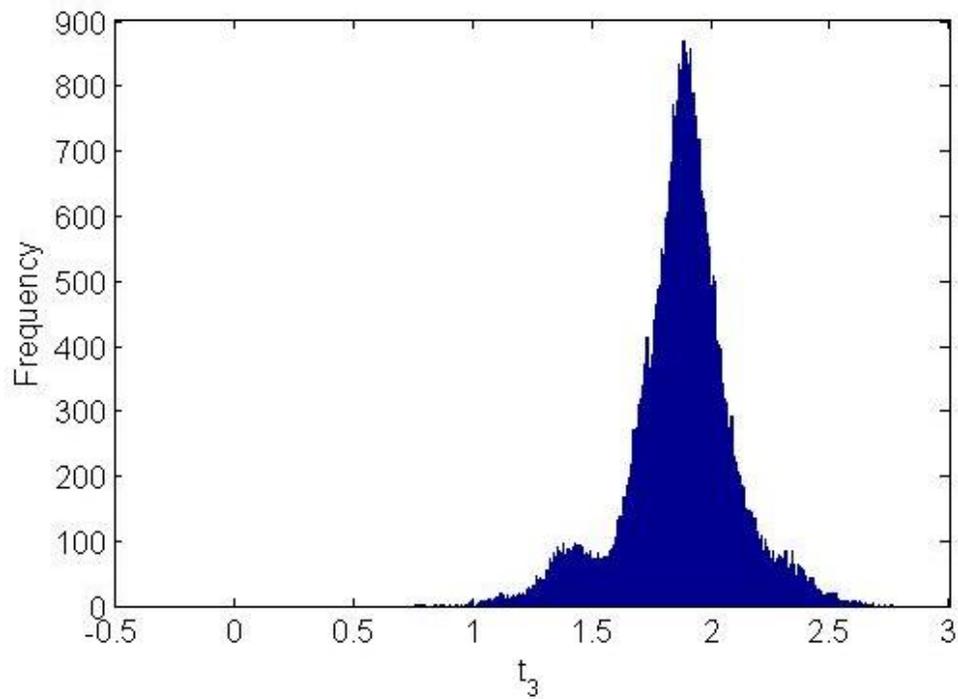
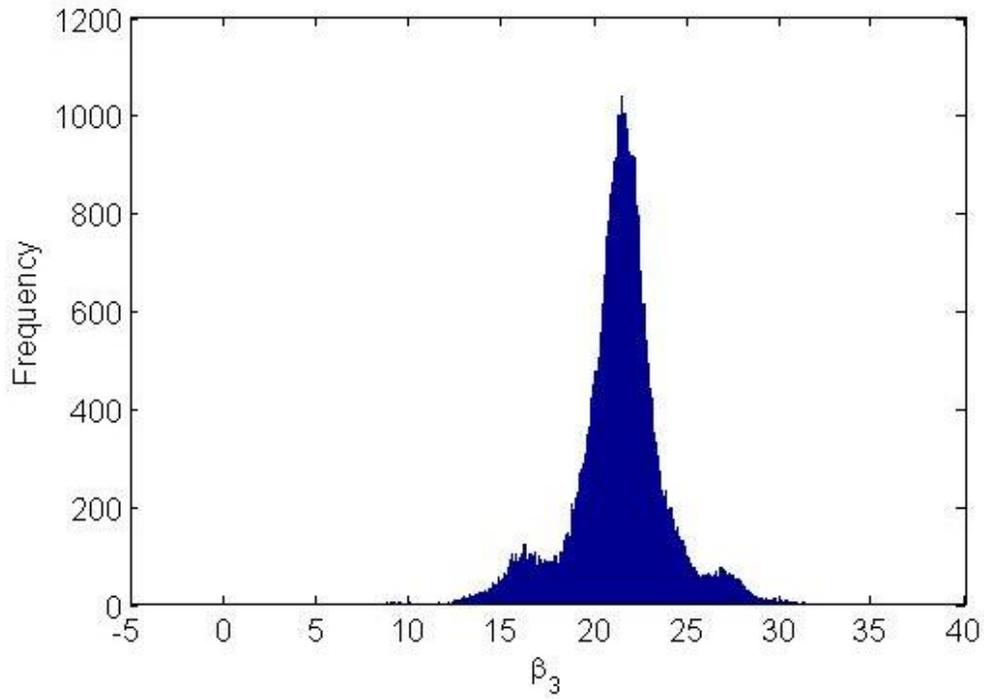


Fig 3. Monte Carlo distribution of the mineral production-institution interaction coefficient and corresponding t-statistic from 100,000 draws of $N = 91$ countries choose $K = 87$.