‘Rich’ countries, poor people?

An empirical study on the resource curse hypothesis.

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics and Business Economics

Department of Economics

Master Thesis International Economics & Business Studies

Supervisor: dr. Vladimir A. Karamychev

Name: Malou Lilian Janet van den Berg

Exam number: 314829

E-mail address: loutje2001@hotmail.com / 314829mb@student.eur.nl

Date: December 2008

Table of contents

List of tables 3
List of appendices 4
Abstract 5
1. Introduction 6
2. Theory and literature review 10
  2.1 Theory on the resource curse 10
  2.2 A resource curse in the empirical literature 15
3. Model and data description 26
4. Empirical analysis 31
  4.1 Panel Least squares regressions 31
  4.2 Panel data model with interaction term 36
  4.3 Cross-section OLS regressions 39
  4.4 Cross-sectional model with interaction term 42
  4.5 2SLS regressions 44
  4.6 Comparison: Panel data or cross-section? 47
5. Conclusion 49
References 52
Appendices 55

List of tables

Table 4.1: Descriptive statistics 32
Table 4.2: GLS Regression output 33
Table 4.3: Regressions with interaction term 38
Table 4.4: Descriptive statistics 40
Table 4.5: OLS regression output 41
Table 4.6: Regression with interaction term 43
Table 4.7: 2SLS regression output 46

List of appendices

Appendix A: Overview of empirical literature 56

Appendix B: Country Sample 58

Appendix C: Variable description and sources 59

Appendix D: Testing both models 63

D.1: Testing for multicollinearity 63
D.2: Testing for robustness 67
D.3: Testing for heteroskedasticity 70
D.4: Testing for serial correlation 72
D.5: Unit root test 74
D.5: Unit root test 74
D.6: Testing for multicollinearity 77
D.7: Testing for robustness 77
D.8: Testing for heteroskedasticity 78
D.9: Testing for outliers 78
D.10: Testing joint significance 81
D.11: Sample size and variables 81

Appendix E: Regression Outputs 85

E.1: Correlation matrix for the measures of natural resources 85
E.2: Descriptive statistics of the variables in the model 85
E.3: Scatter plots of the growth variable (loggdp8906) versus natural resources 87
E.4: Fixed effects estimation (GLS SUR estimations) 89
E.5: Wald coefficient restrictions test 93
E.6: The Hausman test for endogeneity 94
E.7: Interaction term 95
E.8: Testing for joint significance 100
E.9: Correlation matrix for the measures of natural resources 101
E.10: Descriptive statistics of the variables in the model 101
E.11: Scatter plots of the growth variable (lgdpgrowth8906) versus natural resources 103
E.12: OLS regressions 105
E.13: The Hausman test for endogeneity 108
E.14: Correlation between instrumental variables 110
E.15: Interaction term 110
E.16: Testing for joint significance 114
E.17: 2SLS regressions 115

Abstract

Natural resource analysis has been a topic of discussion for several years. One of these issues is the natural resource curse hypothesis. This hypothesis states that countries having an abundance of natural resources have lower rates of economic growth relative to countries with no resources. This research investigates the existence of a resource curse empirically. Using a country dataset, different methods have been applied in order to find evidence of a resource curse. To obtain more reliable results, different measurements for resources are used including resources in per capita terms, and both cross-sectional and panel data methods are analyzed. Special attention is given to institutions and their indirect effect on resources and growth. No strong evidence is found to support the existence of a resource curse. In the panel data model, a significant positive relationship is found between resources and growth, whereas in the cross-sectional model, there is no clear relationship. When including an interaction term for institutions, the positive relationship remains in the panel data model. The cross-sectional model provides negative coefficients thus supporting the resource curse hypothesis: for subsoil assets, in particular, it is shown that institutions influence the negative effects on growth rates.

1. Introduction

"Our country is rich, but our people are poor."[1]

Vladimir Putin, former president of Russia.

Natural resources, energy resources and non-renewable energy resources, in particular, are currently very up-to-date topics. The issue of energy is becoming more and more important in public debate and international relations. Currently, global energy use consists mainly of oil (37%), coal (23%) and natural gas (21%). This means that more than 80 percent of the World energy consumption is produced through these fossil mineral resources that are exhaustible and yet energy has never been of more importance to the world economy. The challenges lie in the continued provision of energy into the future and the economic development of the poorer countries. Along with these issues, the sustainable development of energy sources and the changes in the global climate are the major topics of discussions today. Consequently, an extensive body of literature has been written on the subject.

The academic field of energy and resource economics has been transformed in the last twenty years. Market liberalization and globalization have increased in the past decades. In the field of energy economics, new economic issues have emerged, for example, in macro-economics, investment, economic policy, industrial organization and the economics of regulation. The econometric methods on the subject which can help analyze the new energy economic topics have considerably been developed in the past twenty years (Keppler, Bourbonnais & Girod. Ed., 2007, pp. xiii-xxiii).

Energy production is necessary to human needs, such as food, temperature and well-being, by using energy for heating, cooling, electricity, transportation, lightning and more. A significant proportion of the world population consumes energy for these purposes. However, there is still a large number of the world population that does not have access to modern energy sources. This means that they lag behind in economic development. Energy, therefore, plays an important role in economic growth and economic development (Keppler, Bourbonnais & Girod. Ed., 2007, p. xiv).

---

[1] Gylfason, (2007), p. 2.

However, energy resources do not always lead to increased economic growth and development *per se*. The possession of natural energy resource sometimes creates disadvantages for the country that possesses the resource, as was observed in the Netherlands in the 1960's after the discovery of large pockets of natural gas in Groningen. This was later referred to as the 'Dutch disease'. This term refers to the believe that, because of the natural resources, the equilibrium real exchange rate and other exporting sectors are affected thus lowering overall economic growth. This also occurred in Great Britain in the late 1970's (Herbertsson, Skuladottir & Zoega, 2000, p. 2). Theories on this phenomenon were developed after the natural gas discovery in the Netherlands.

Subsequently, the Dutch disease was considered to be one of the explanations for a more widely-spread observation, namely the relationship between countries with natural resources and their lower overall economic growth. This is called the resource curse hypothesis and has been topic of research ever since. Important contributions to this subject are those of Gelb (1988), Corden (1982, 1984), van Wijnbergen (1984a & b) and Auty (1993). These theories all concentrate on the issues involved with lower economic growth for countries with natural resources through, among other factors, the Dutch disease. In addition, empirical research on this issue has taken off from the 1990's on. In this research, there has been no agreement on the results. The resource curse is therefore a topic of an active discussion among researchers.

This research contributes to these discussions, by empirically studying the economic growth in a country due to the effects of natural resources in general, and subsoil resources, in particular. This is analyzed, taking into consideration multiple channels and reasons. These channels comprise the many factors that influence economic growth, such as human capital, investment, openness and trade. One important channel through which economic growth can be influenced is that of institutions (such as the quality of government in the country, the political freedom of the citizens and the quality of bureaucracy). Although the importance of institutions on economic growth is recognized, little attention has been paid to the role of institutions on economic growth and natural resources. Institutions however, may play a crucial role in determining economic growth through natural resources (Brunnschweiler, 2008, p. 400). Therefore, extra attention is paid to the role of institutions in this research.
The aim of this research is, therefore, to investigate if the theory of the resource curse can be substantiated, or to demonstrate if this theory is not supported by the data. The countries used

in this research are investigated to see if they are suffering from a resource curse or not. After a theoretical review, empirical analysis is applied. A large dataset of 101 countries is analyzed from which the conclusions are drawn. This is done by a panel data analysis and a cross-sectional analysis. In this analysis the emphasis is on the applicability of different variables for natural resources and on the specific influence of institutions.

The main research question that will be investigated is the following:

Is there evidence of a resource curse in countries with natural resource abundance?

Sub-questions to support this research question are:

- What exactly is the resource curse and through which channels can it influence economic growth?
- Is there empirical evidence of the existence of a resource curse?
- Is there a difference between the results of the cross-sectional analysis and the panel-date analysis?
- Is there a difference in the way natural resources are measured and thus the different variables for resource abundance included in the model?
- Do institutions influence growth through resources?

The research tries to contribute to the field of economic growth and natural resources, since there are still some interesting insights to be gained, that are of relevance for understanding this complex relationship better. Some issues can be mentioned here.

Firstly, empirical research on the resource curse, since it was first identified in 1995, is relatively new. This means there is room for new insights in this field of research.

Secondly, there has been lively discussion in the past decade on the reasons and existence of the resource curse. Not only do researchers differ in opinion on the channels and reasons for the resource curse, there is no unanimous concurrence as to the existence of a resource curse. It is the objective of this paper to contribute to this discussion since previous results have been ambiguous and the resource curse thesis is no longer an accepted fact.

Thirdly, some new insights and findings are incorporated into this research: both panel data and cross-sectional data are used for analysis, however, this research also uses other measures of resource wealth. Instead of only looking at the export of resources as a percentage of GDP, as is most common in resource curse empirics, a measurement that is used is per capita wealth in natural resources, and subsoil wealth. Although the different models and the different resource measures have already been used in previous studies, they have never been combined in any research. This research therefore incorporates the latest findings in recent empirical research, basing variables on the latest discussions and applying the methods used in recent empirical literature, and combines these to highlight the important and relevant aspects and methods.

In addition, newer data is used for the analysis. The period covered in this research is from 1989 to 2006 whereas previous research concentrated on the period 1970-1989.

The outline of this research is as follows.

Section 2 provides an overview of the theoretical and empirical literature available on the resource curse. It also includes the initial steps for the further analysis of the paper.

Section 3 introduces the empirical model and data description for the analysis on the resource curse.

Section 4 provides the empirical analysis. Subsections 4.1 and 4.2 concentrate on the panel-data analysis for a sample of 101 countries for the period from 1989 to 2006. A least squares model with and without an interaction term for institutions is used. Subsections 4.3 to 4.5 analyse the cross-sectional model. For this, an OLS, 2SLS and an interaction term model are applied. The last subsection 4.6 will compare the results for both models and its implications.

Finally, section 5, draws the overall conclusions, presents the main findings, and a summary of the research.

2. Theory and literature review

In this section, the theoretical and empirical work in the field of the resource curse is reviewed. These studies and methods of research form the guidelines of the empirical analysis. Subsection 2.1 reviews the theory on the resource curse. Subsection 2.2 reviews the empirical literature and divides the different literature into different categories. (A summarizing table of main results and variables from the empirical literature can be found in Appendix A.)

2.1 Theory on the resource curse

In one of the earlier works, Gelb (1988) identifies the resource curse. He distinguishes four theoretical approaches to the debate on the effect of resources on economic growth. These are the following: the linkage theory, the neoclassical and related growth theories, export instability theories and the booming sector and Dutch disease theory (Gelb, 1988, pp. 14-29).

The linkage theory uses interrelations and interactions. The effects of the interactions between the leading sector and other sectors in a country are divided into three linkages: production linkages, consumption linkages and fiscal linkages.

The first category, production linkages, deals with the input and output of a sector. Production linkages are of less importance in high-rent activities, since intermediate inputs account for a smaller part of the final value here.

The consumption linkages refers to the extent to which the surplus is accrued by private owners. These linkages are either favourable or adverse to development.

Fiscal linkages are the most important linkage since resource rents almost always go to governments. The resource rents can be consumed directly by public services and private sectors but not efficiently. On the other hand the supply-side can be stimulated with resource revenues. Also, the fiscal linkage, here, is not proven to stimulate long-term growth. Another problem with fiscal linkages is the fact that the surplus is in the hands of a relatively small group of decision makers which can lead to bad planning, dependence on only a few investments and rent-seeking. A fourth problem is the fluctuating fiscal revenues, where investments made during a boom cannot easily be reversed in periods of recession.

In the neoclassical growth model, growth is a process of expanding on a production possibility set, with a frontier set by quantity instead of quality of production factors and by the efficient allocation of these factors across activities. Growth depends on factors of production, labour and capital. If imports are an important factor in production, exports cannot easily be increased, meaning that foreign exchange can put a constraint on growth. Fiscal revenues can further constrain development. In this model, resources and resource gains have a very favourable effect on growth, since rent-intensive activities help to relax constraints on domestic saving, foreign exchange and fiscal revenues, when rents are invested rather than consumed (Gelb, 1988, pp. 17-18). Convergence happens between countries with low levels of GDP per capita and countries with a high GDP-level per capita, where the former will grow at a quicker pace than the latter (Farmer, 2002, p. 367).

The export instability theory weighs up the benefits of a temporary high income against the adverse effects from the variability of resource income. Developing countries have more concentrated exports which depend more on agriculture and minerals. Agricultural exports are relatively price-inelastic in demand and subject to supply shocks, whereas mineral exports are in-elastic on both sides, since demand and supply sensitive to economic activity. Terms of trade variations are larger for developing countries than for developed countries and mineral-rich countries are even more vulnerable to fluctuations in export prices and revenues. Buffering here is difficult, since this instability leads to uncertainty where adjustment is slow in response to fluctuations. When domestic demand increases, supply will be constrained at a certain point. To clear the market, inflation will increase, the exchange rate will appreciate and imports will rise. When demand decreases, unemployment will rise due to sticky prices or wage rigidity. Demand fluctuations thus raise average imports and lower capacity use, output and income. When mineral revenues are primarily put in investments, stability is important. If saving and investment fall with income, this will have an adverse impact on growth. This impact is even more significant when changes in resource revenue manifest themselves mainly through the rate of public investment. Some policies and government programs implemented during the boom years may not easily be reversed as resource income falls.

The fourth theory is the booming sector theory. Unlike the other three theories, this theory focuses on the sectoral reallocation of productive factors in response to a shock, resulting from a discovery of resources or an increase in a commodity price. If income is spent, rather than saved, abroad, then this results in a resource movement effect. This effect draws factors of production out of the other sectors into the booming sector. It also results in a spending effect, which draws factors of production out of sectors producing traded commodities, that are substituted by imports, and into non-traded sectors (Gelb, 1988, pp. 14-29).

An important observation by Neary and van Wijnbergen (1986) on Dutch disease theory is that the initial impact of a resource discovery is beneficial for the economy as a whole, only at the expense of reallocation of production, with a rise in the output of non-traded goods and a fall in manufacturing output. The spending effect thus results in deindustrialization and a real appreciation (pp. 13-40).

This contraction or stagnation of the traded sector is referred to as the Dutch disease. The economy consists of three sectors: A tradable natural resource sector, a tradable (often manufacturing) sector and a non-tradable sector.

The traded sector follows the law of one price where the domestic price follows the international price because of the flow of goods across international boundaries. The greater the natural resource endowment, the higher the demand for non-tradables will be. This drives up non-traded prices, including in particular, non-traded input costs and wages. This leads to a smaller allocation of capital and labour to the manufacturing sector and to lower profits since this sector uses those non-traded goods as inputs and yet sells its products on the international markets at relatively fixed prices. When natural resources are abundant, tradables production is concentrated on natural resources rather than in manufacturing, and capital and labour that would otherwise be employed in the manufacturing sector are pulled into the non-tradables sector.

When domestic spending power increases, an appreciation in the real exchange rate shifts production to the non-tradable sector and demand to the tradables. This leads to increased real income, in the form of higher absorption of non-tradables and higher net imports of tradables (manufactures). The real effective exchange rate is measured here by an index of domestic prices (for tradables and non-tradables) relative to the prices of major trading partners converted at market exchange rates.

One other consequence of an appreciating real effective exchange rate occurs, when capital is internationally mobile. In this case, with the inflation of domestic prices, the real interest rate on foreign funds (corrected for domestic prices) falls. This will stimulate foreign borrowing for consumption and investment and further boost domestic absorption. When resource prices and domestic spending fall and the real exchange rate depreciates on the other hand, capital movements will reverse. This means foreign assets are more profitable to hold than domestic assets, which means a falling demand. In addition, a real exchange rate depreciation means a nominal exchange rate devaluation and a loss in value of assets in domestic currency relative to assets in foreign currency. These capital outflows will result in economic contraction (Gelb, 1988, pp. 14-29).[2]

Auty (1993) develops the theory on the resource curse by using the linkage theory and the booming sector theory. Auty (1993) finds failures in resource-led development in his study. According to Auty, it was always assumed that resource richness was most harmful to the economy in the early and low-income stages of the development process of a country. However, evidence also points in the direction of the negative effects in the low- and mid-income stages of development. This new evidence suggests that not only do resource-rich countries fail to benefit from resource endowment, but they also perform worse than resource-poor countries. Resource-rich countries, or mineral economies, are countries that generate at least eight percent of their GDP and 40 percent of their export earnings from the mineral sector. The mineral sector is divided into two main categories: the ore (hard mineral exports) and the fuels (hydrocarbon) sectors.

The core of the low-growth problem lies in the mining sector's production function, the capital to labour ratio and the deployment of mineral rents. Mineral production is highly capital intensive, employing only a small fraction of the labour force. Since the capital inputs originate from foreign sources there are not many opportunities for local production linkages and consequently the creation of local factories for processing. This also leads to low revenue retention since most of the export earnings flow abroad As this is different to other primary sectors, fiscal linkages dominate the mining sector's contribution to the economy. When the rents on minerals are captured by the government through taxation, it can, in turn, destabilize the economy. The way this happens is through the domestic absorption of resource rents that leads to the decrease in international competitiveness of the agricultural and manufacturing

[2] For a more extensive description on booming sector and Dutch disease refer to Corden (1984), Neary & van Wijnbergen (1986) and Bruno & Sachs (1982).

sectors. The agricultural sector shrinks prematurely and the manufacturing sector becomes under-developed or extremely overprotected. This occurs in the Dutch disease. It results from an appreciation in the exchange rate as a consequence of the rapid inflow of resource rents into domestic economy. This loss in competitiveness is not easily restored (Auty, 1993, pp. 1-20).

Ellman (1981, pp. 163-165) concludes that some positive effects of oil and gas exports are a relaxation of the balance of payments constraint, additional imports of technically advanced investment goods and complementary goods, cheap energy and a positive influence on labour productivity through investment effects and real wages. There can also be negative effects of resource abundance: mainly on employment and industry. This is caused by increased competition in imports, higher domestic costs, diminishing profits in internationally competitive sectors due to an exchange rate appreciation, and the substitution of non-labour intensive export for labour-intensive exports. These refer to the booming sector theory.

A fifth theory on economic growth is the endogenous growth theory. This theory is an alternative growth theory, to explain some of the reasons for growth that the neoclassical growth theory did not explain. This theory states that long-run growth depends on investment decisions, rather than technological progress. This investment also includes factors such as research and development expenditures and human capital formation. The factor capital includes investment in knowledge as well as the physical capital in the model. Knowledge spillovers occur, where knowledge cannot be internalized entirely and spills over to the whole economy. Technology in the model is endogenous, and not, as the neoclassical growth model states, exogenous. The reason is that technology is not freely available without limitations in the whole economy. A technology gap between developed and developing countries is the result of this. Convergence between countries in this theory happens through following up of countries that do not have the means to develop technology by themselves. Imitation is cheaper than innovation, and this way follower economies will grow relatively faster than the leader economies and converge towards these leaders (Snowdon &Vane, 2005, pp. 625-632).

2.2 A resource curse in the empirical literature

The empirical finding that resource abundant countries have a lower economic growth rate was first identified by Sachs and Warner in 1995. This means that the field is relatively new and only in the past decade has it become of increasing interest to academics thus resulting in new studies.

The studies base their models on the different theories of the resource curse described in section 2.1, or combine several theories. Different variables are often included in the models to account for the different theories and channels trying to explain the resource curse. The emphasis then is on one of these channels, but the other channels are also incorporated into the regression as control variables.

The literature discussed here is divided into different categories by theory.

The main reasons for the resource curse that are identified are in linkage theory: forward and backward linkages, political factors (fiscal linkages), rent-seeking, corruption and institutions. In neoclassical growth theory, the main channels of influence are saving and investment. A third category of studies base their argument on the Dutch disease theory. The fourth category seeks an explanation in endogenous growth theory: in human capital and debt. A final category is that of studies combining channels and incorporating more than one theory into the model.

Linkage theory

Sachs and Warner (1995), in this pioneering work, identify two reasons for the resource curse phenomenon.

The first reason is related to linkage theory and states that political factors play a role in the lagging development of resource-rich countries. Resource abundance leads to more rent-seeking behaviour, since national politics are oriented on grabbing the rents from these resources; this is even more the case in countries with corruption and lax legal systems. In their model a windfall coming from a terms-of-trade improvement or a discovery of new resources leads to the inefficient exhaustion of a public good due to competing parties fighting for the resource rents.

A second line of reasoning in the paper by Sachs and Warner (1995) is strictly economic, drawing upon development economics and the Dutch disease theory. One argument is that demand for manufactures would grow faster than the demand for primary products and prices would increase more for manufactures than for primary exports. Rich countries would also be more protectionist against primary imports than manufacturing imports. Forward and backward linkages are identified as an argument, where benefits of forward and backward linkages from the primary exports to the rest of the economy are small. Manufacturing would lead to a higher standard of living because it has a more complex division of labour. Another argument is that the manufacturing sector is characterized by learning-by-doing. Sachs and Warner extend their model to include the Dutch disease case.

Sachs and Warner (1995) look at the natural resource exports as a percentage of GDP in relation to the annual growth rate per capita of a specific country. In resource exports they include agriculture, minerals and fuels. The model is specified below;

$$G = \delta_0 + \delta_1 \log(Y_0^i) + \delta_2 Z + \varepsilon$$

Economic growth G, which is calculated as ln (Y^i_t / Y^i_{t-1}), in an economy should be negatively related to initial income Y_0, supporting the conditional convergence, and Z, a coefficient for other structural characteristics of the economy. This research investigates if there is a measure of resource dependence in the Z.

A cross-country regression is performed, regressing the share of primary sector exports and initial income on the growth variable. After this, a number of other variables are included in the regression. These are *openness* (integration with the global economy), *investment*, *quality of the bureaucracy*, *terms-of-trade index*, a variable to measure inequality (incomes in highest 20% and lowest 20% of population) and the *investment deflator* (a significant determinant for investment rates).

The regression is optimised by omitting outliers and countries with little available data.

In addition to the variable *'share of primary sector exports as a percentage of GDP'*, four other variables are used to measure resource intensity. One is *mineral and fuel production in GDP* for one year (import prices / GDP). Another measure is the *fraction of primary exports in total exports* for a particular year. The fourth measure is the *logarithm of land area per person* in that year (since land abundance and resource abundance tend to be positively correlated). The first result is that high natural resource abundance leads to increased rent-seeking and corruption, directly or through lower investment. The hypotheses set in the paper are confirmed and evidence of the resource curse is found.

In a later paper, Sachs and Warner (2001), expand on the earlier research (1995). They investigate whether other factors, that have no relation to resources, influence the empirical results that prove the resource curse. A variable that is omitted in the empirics can influence the relationship between resource abundance and economic growth. This is verified by controlling for previous growth rates in the regression or, if the variable is known, by controlling the omitted variable in the regression. One of the variables that is omitted, according to Sachs and Warner, is geography. A cross-country regression is performed to test whether this affects results. The same variables as in the previous study (1995) are used and indicators for geography and climate are added to the model. The results, however, are not influenced by these variables.

The crowding-out hypothesis is tested to see if it is supported by the empirics. The systematic relationship between price levels and (non-natural resource) GDP is controlled for and tested is whether natural resource intensive economies had higher relative prices on top of this. The model is as follows:

$$Log(PLEVEL_t) = \beta_0 + \beta_1 \log(Y_{0t}) + \beta_2 R_t + \varepsilon$$

Where the price level in year *t*, ($PLEVEL_t$), is a function of the natural resource variable in year t (R_t) and of a control variable for the income effect of non-natural resource GDP, (Y_{0t}) in year *t*. This model proves to be significant. The resource rich countries also failed to achieve export-led growth, because of their high prices.

Rent-seeking, corruption and institutions

Good policy is the topic of focus in Larsen (2005), investigating whether rich countries are immune to the resource curse, using Norway as an example. As a benchmark, the other Scandinavian countries are used to compare to Norway. This is the first paper using structural break empirics to detect a possible resource curse. Two structural breaks are detected in the GDP per capita of Norway: in the mid-1970's and 1996. After the first break, there is no evidence of a deceleration in the Norwegian GDP, implying no resource curse, until 1990. In 1996, however, there is a breakpoint, where Norway's growth decreased. There seems to be a late appearance of a resource curse after all, however, there are no specific signs that point to the Dutch disease in Norway.

Leite and Weidmann (1999) find that there is an important interrelationship between resource abundance, lower economic growth and corruption (rent-seeking).
The main determinants of corruption are the absence of political stability, absence of transparency of rules and a bad institutional framework. A further determinant is the presence of trade restrictions, which leads to premium seeking, tariff seeking and revenue seeking. There is a direct effect of corruption on growth and an indirect effect of corruption on growth through resources. Rent-seeking encourages corruption and affects growth.
In the empirical part of this paper, the cross-country Sachs and Warner (1995) procedure is used. The extension to the model is the inclusion of the variable *corruption*. The resource variable is split up into four different groups: *fuels*, *ores, agriculture* and *food*. Regressions are run that include variables similar to those used by Sachs and Warner. The variable, corruption is endogenized and a regression is run on corruption, including the variable, *rule of law*, which they state is only of influence through the channel of corruption. This is different from the Sachs and Warner paper. The finding of a negative relationship by Sachs and Warner remains unchanged after correction for corruption, price variability and trade liberalization.

In line with the rent-seeking argument, the transmission channel for the resource curse given in the research by Atkinson and Hamilton (2003) is the saving rate of a country. Lower growth is attributed to the mismanagement of the portfolio of wealth of resource-abundant countries. Bad government and rent seeking cause persistently low genuine saving rates with adverse consequences for economic growth. The work of Sachs and Warner (1995) is extended, with cross-country data, to include sustainability. The definition of resources is different from the original Sachs and Warner model as this model does not include food and agriculture, but concentrates on resources such as *net timber accumulation* and the *depletion of energy and mineral resources*. Atkinson and Hamilton investigate three aspects:

- if the curse is explained by government policy that leads to a reduction in the rate of genuine saving;
- if countries that have avoided negative genuine saving rates have thus avoided the curse;

- to what extent resource abundant countries have financed additional saving and investment using the proceeds of this natural wealth.

A regression is run and a negative relationship is found between resources and growth. Government policy is also suspected to influence the resource curse. By examining the interaction of the resource variable with the government variables, it is found that countries that are resource rich and where government consumption is relatively high, have experienced, on average, lower economic growth. Countries with higher rates of genuine saving enjoyed, on average, higher growth rates and these countries did not suffer from a resource curse.

In a recent paper, Brunnschweiler (2008) re-examines the resource curse by using different measures of resource endowment, and secondly, investigating the resource curse in relation to institutional quality. Brunnschweiler (2008) claims that he natural resource proxy of Sachs and Warner (1995) is not a suitable measure. Therefore two new measures for resources are proposed: namely the *subsoil wealth* and *natural resource wealth per capita*. For the analysis, cross-country data are used and an OLS regression is performed to find a positive significant relationship between GDP per capita growth and natural resource abundance, measured as *natural wealth in USD per capita* and *subsoil wealth in USD per capita*. The focus is on institutions of a country in relation to economic growth and resources, which includes six variables for institutions in the regression. The OLS regression is run without, and with, an interaction term for institutions. The results remain the same, the relationship between resources and growth is positive. There exists a negative relationship between institutional quality and economic growth.

Secondly, a 2SLS regression is run, in which *geographical latitude* functions as a proxy for institutional quality. In this regression, the outcome is the same. To check for the robustness of these results, a variety of other variables are added, to see if there is omitted variable bias, but the results remain the same even after including the variables.

Neoclassical growth theory

One of the first papers to contest the view of Sachs and Warner (1995) is a paper by Davis (1995). He finds no evidence of the resource curse in his research.

Davis reviews the developing mineral economies in the world. He examines the mean and median economic performance of the mineral-based economies, over two years, for a group of mineral-based economies and a group of non mineral-based economies. The first group is shown to out-perform the second group. But, since these results can be biased, a second measure is used: the human development index and indicators such as education, mortality rates and sanitation in the two different groups of countries are assessed for this reason. There is again evidence that mineral economies in both years were at a higher average level of development than the non-mineral economies in all development categories.
Davis also looks at the possibility that non-mineral countries have converged towards the initial development level of the mineral countries. The gap between more developed mineral economies and the less-developed never-mineral economies is widening. This shows that there is no evidence of a resource curse in the group analysis, maybe only in a country-specific case.

Gylfason and Zoega (2001) investigate the link between the resource curse and lagging economic growth through investment and saving. Four other channels are included in the model: the Dutch disease, rent-seeking, human capital and overconfidence (social capital crowding-out).
Optimal saving, and hence also the rate of growth of output and capital, depends on the abundance of natural resources. The *genuine saving rate*, *gross saving* and *gross investment* are used in a cross-sectional analysis, which accounts for depreciation of natural capital, among others, which can be an indicator for sustainability of natural resource management. Other variables in the model on economic growth are *natural capital*, *initial income*, *financial depth* and the *enrolment rate*. Investment in physical capital is found to be inversely related to the share of natural capital in national wealth, and directly related to the development of the financial system. The development of the financial system is also inversely related to the share of natural capital in national wealth. When the share of output that accrues to the owners of natural resources rises, the demand for capital falls, leading to lower real interest rates, less saving and less rapid growth.

Dutch disease theory

Herbertsson, Skuladottir and Zoega (2000) look at the Dutch disease case in Iceland. This country has a natural resource in its fish, but unlike other resource abundant countries, has good data, being an OECD country.

Three symptoms of the Dutch disease are tested:

- the appreciation of the real exchange rate, leading to lower secondary-sector employment and output,
- the more volatile the primary sector, the more volatile is the real exchange rate and the higher the investment threshold,
- primary sector and secondary-sector wages move together. This affects employment and output, and affects investment in the medium-run as well.

In the paper, these three symptoms are tested empirically. The outcomes are as follows: primary exports have a positive short-run relationship with secondary-sector output, and a negative relationship in the long run. The real exchange rate depreciates with an increase in primary-sector output and wages increase with primary-sector wages. This means that the Dutch disease effects occur through the labour market instead of the exchange rate.

The paper of Sala-i-Martin and Subramanian (2003) also contains, after the general results, an investigation on the possibility of the Dutch disease in Nigeria. In examining the real exchange rates and prices of tradable to non-tradables, no evidence is found of the Dutch disease in their study.

Égert (2005) analyzes the exchange rates of a group of specific countries in Eastern Europe. The paper uses both time series and panel data to study deviations from the equilibrium exchange rate. The main finding is that the currencies of the countries investigated are undervalued in terms of absolute PPP. To check for the Dutch disease, the real exchange rate is compared, in relation to the Dollar and Euro, to the ratio of oil, gas and fuels in total exports. There is evidence of some Dutch disease symptoms in this data.

Endogenous growth theory

Human capital

In a recent World Bank study, Bravo-Ortega and De Gregorio (2005) show that having a high level of human capital can offset the negative relationship between natural resources and economic growth. In the paper growth is analyzed and the level of income, which is positively related to resources. The resource curse is attributed to rent-seeking and the Dutch disease. Here, immobile human capital instead of mobile physical capital is the factor that diminishes growth enhances activities.
Bravo-Ortega and De Gregorio (2005) use panel data for the analysis. The proxy for the resource variable included in the model is *the share of natural resources exports in GDP* and *in total exports.* Instrumental variables are applied to overcome bias in the human capital variable. Instruments are the five-year-lagged value of *human capital* and the five-year-lagged value of *government expenditure in education*. As control variables, *human capital*, *government expenditure*, *openness*, *terms-of-trade shocks*, *investment* and *initial income* are added.
The first regression has the growth rate as the dependant variable, where both the random effects and the fixed effects are considered: a negative relationship is found. Secondly, a regression is run on the level of income that results in a small positive relationship with resources. An interaction term between resources and human capital is also added in later regressions. The result is that resources could impede growth in countries with low levels of human capital, but in economies with an abundance of human capital, natural resources could boost growth. The resource curse can be fully explained by the level of human capital in a country.

Gylfason (2000) finds that neglect of the development of human resources and inadequate attention and expenditure on education is one of the main factors in explaining the resource curse. This means lower growth is not due to natural resource abundance *per se*, but to the bad policy making of governments in countries with these resources.
Three additional reasons for the resource curse are identified. One is the Dutch disease. The second reason is rent-seeking. Governments may offer tariff protection to domestic producers and provide other privileges; rent seeking may also breed corruption in business and

government, thereby distorting the allocation of resources and reducing both economic efficiency and social equity. The third argument is overconfidence.

The variables: *public expenditure on education relative to national income*, *expected years of schooling for girls* and *gross secondary-school enrolment*, are inversely related to natural resource abundance, measured by the *share of the labour force engaged in primary production*, across countries. For the empirics, OLS and SUR (seemingly unrelated regression) methods are used on 85 countries. The model consists of 2 equations. The first equation is economic growth regressed on the *natural capital share* variable, the *secondary school enrolment rate*, the *share of gross domestic investment in GDP* for the period covered, and the *logarithm of initial per capita income*. The second equation includes the *enrolment rate*, the *natural capital share* and *initial income*. Lagging growth through inadequate attention to education is confirmed empirically in the cross-country analysis.

Stijns (2006) studies the link between resource abundance and human capital accumulation. Resource rich countries tend to invest more in human capital than countries with no natural resources. This is in contradiction to the claim of Gylfason (2000) and supports the findings of Davis (1995).

Stijns (2006) first provides an overview of all the measures that are used in the literature as a proxy for natural resource abundance and human capital in a country. For all these variables, the correlation coefficients are calculated. The measure of Gylfason (2000) for resource abundance is biased according to Stijns (2006), thus biasing the results on human capital. It is better not to include agricultural resources and concentrate on mineral and subsoil resources, as this gives better results than total resources. There is a positive and significant correlation between the subsoil wealth and physical capital ratio and total years of education, these are the best proxies for both resource abundance and human capital.

Debt

Manzano and Rigobon (2001) show that a debt overhang causes the appearance of a resource curse in the data, where countries with a large amount of resources used their resources as collateral for debt. Countries did this when commodity prices were high, but in the 1980's prices decreased and this resulted in high debts for these countries.

Biased results are expected in the Sachs and Warner (1995) paper, since total GDP is used as a variable instead of non-resource GDP. However, the results are almost the same when correcting for this. Results are not the same, however, when panel data is used instead of cross-sectional data. When panel data is used, the negative relationship disappears.

Data for the non-resource sector, net of the resource-sector, is used to see whether the effect found by Sachs and Warner (1995) on total GDP is still present. For the cross-sectional data, the results give the same outcomes. The effect disappears, however, when a panel regression with fixed effects is performed. Two different explanatory variables are used: the *nominal share of primary exports to GDP* and the *real share*. To correct for endogeneity problems, an instrument is used for institutions: the variable *ethno-linguistic fractionalization*. Resources are divided into different groups: *agricultural*, *food*, *mineral* and *fuel exports*.

An explanation of why the curse only exists in the cross-section estimation and not in the panel estimation is that other variables, not included in the model, are correlated with resource abundance. Adding *development level* and *institutions* as variables does not change the situation. However, when a debt variable is introduced, this explains the resource curse. The resource abundance variable was picking up the fact that some countries were greatly in debt at the beginning of the decade, thereby explaining the apparent negative relationship in previous works.

Combining different theories

In a recent paper by Gylfason (2007) an empirical research on the resource curse is applied to Norway. Eleven channels of growth are identified: real capital, human capital, foreign capital, social capital, financial capital and natural capital, rent-seeking and overconfidence, human and social capital crowding out, less saving and investment, financial markets failure and finally, the Dutch disease and less openness.

Another variable is added to the existing model: that of political liberty. There is a cut-off point in the data, implying that above a certain income per capita, economic growth does increase with natural resource wealth. This means that while an increase in the natural capital share tends to reduce growth in developing countries, it may increase growth in industrial countries.

Sala-i-Martin and Subramanian (2003) test the resource curse in Nigeria. They hypothesize that the resource curse is due to the bad institutional quality of the country. Three channels of causation of the resource curse are used, based on the different theories (linkage, export

instability and Dutch disease). Firstly, it is caused by the institutional impact, which include rent-seeking and corruption. Secondly through a volatility channel, and thirdly, through the Dutch disease. A model is used which has these three causes as variables, namely *institutions, overvaluation of the exchange rate* and *price volatility* and a natural resource variable. Cross-sectional data and instrumental variable estimation (due to endogeneity and measurement error in the institutional variable) are used. As instruments, the *mortality rates of colonial settlers* and *fraction of the population speaking English and European languages* are applied. To avoid endogeneity in the other variables, variables are used that refer either to the beginning of the overall period or to average values for the entire period.
No direct, significant, impact of resources on growth is observed. They do, however, have an indirect impact through institutions on growth. Resources are also split up in the research: they are grouped into two categories - *fuels and ores*, and *food and agriculture*. Fuels and ores are found to have the biggest impact on institutions and thus on growth.

Papyrakis and Gerlagh (2004) investigate the effect of natural resources on *corruption*, *investment*, *terms of trade*, *openness*, *schooling* and *economic growth*. There are four negative transmission channels: bad institutional quality, a false sense of security, the Dutch disease and human capital, combining linkage theory, endogenous growth theory and the Dutch disease theory. A cross-country regression is used to show the effects of resources on the economy. When the variables mentioned above are incorporated in the model, the negative relationship between economic growth and resources disappears.
The model also includes *economic growth*, *initial income*, *natural resources* and a vector for all other explanatory variables, as is used in the papers by Sachs and Warner (1995) and Leite and Wiedmann (1999), but get different results. As the resource proxy, *the share of mineral production in GDP* is applied. To avoid endogeneity, variables are chosen that refer either to the initial or average values of a period.
First, a regression is run with only the resource and growth variable, and further variables are added with every subsequent run, until all variables are included. The more variables in the regression, the less negative the relationship between growth and resources becomes and eventually even a positive relationship is found, but with a lower level of significance. The negative relationship is only caused by the indirect effects. *Education* has a larger effect as a transmission channel than *corruption*, which is contradictory to the findings by Sachs and Warner (1995).

3. Model and data description

To construct an empirical model on the resource curse, the starting point is the basic empirical growth regression model used in Sachs and Warner (1995) and other literature[3]:

$$G^i = \alpha_0 + \alpha_1 \ln (Y_0^i) + \alpha_2 R^i + \alpha_3 Z^i + \varepsilon^i \quad (1)$$

In the model, different growth rates of countries are explained by several country characteristics. This empirical growth model is based on the conditional convergence hypothesis, which states that countries with high incomes, however, have a lower growth rate than low-income countries, ceteris paribus. This means that economic growth in a country between time t=0 and t=T, G^i, is a negative function of initial income, Y_0^i. It also depends on natural resources R^i and on other characteristics of the economy, Z^i.
$G^i = \ln (Y^i_t / Y^i_{t-1})$ and ε^i is the error term. Superscript i corresponds to a country in the sample. R^i is the variable that indicates the resource abundance in country i and Z^i is a term that captures all other control variables, per country, that are included in the model.

The model can be estimated on cross-sectional data, as in Sachs and Warner (1995) or on panel data, as in Bravo-Ortega & De Gregorio (2005).[4] The cross-sectional data model, and panel data model have considerable advantages and disadvantages. A cross-sectional analysis gets around problems such as the lack of yearly data, non-stationarity and autocorrelation. A big disadvantage of cross-sectional analysis is the number of observations that can be used in the analysis. In the cross-sectional model, omitted variable bias can influence the results.
Panel data has more observations and thus makes the outcomes more reliable than in the case of cross-sectional data, since the larger dataset has more variability and less collinearity among the variables (Mills & Patterson, 2006, pp. 633-660). It is also known that panel data can diminish problems that arise from omitted variable bias (Pindyck and Rubinfeld, 1998, pp. 250-251).

[3] Followed also by Papyrakis and Gerlagh, (2004).
[4] And others. Appendix A provides the model used in the different literature, cross-section or panel data.

A country sample of 101 countries is used in both cases, which includes European countries, Asian countries, African countries and American countries. (The list of countries used is provided in Appendix B.) The time period covered in the analysis is from 1989 to 2006. The use of the most available data leads to new insights. The initial year of the dataset is 1989 because the most data is available from this year on. In the panel data section, all years are used individually. In the cross-section analysis, an average of the same years is used. This applies to all variables described below.[5]

The dependent variable G^i in the regression, GDP growth, is calculated as the log of PPP adjusted real GDP growth per capita for the period 1989-2006. In the panel-data model this variable is named *loggdp8906* and in the cross-section model, it is *lgdpgrowth8906*. Data originates from the World Bank (WB), World Development Indicators (WDI). The Y^i, initial income, is the variable *inigdp* for the panel and *lgdp89* in the cross-section model and is measured as the log of previous year GDP or 1989 GDP. This also comes from the WDI.

R^i captures resource abundance in a country. The proxy for this variable is topic of discussion since a number of proxies used in research are not a suitable indicator for measuring resource abundance in a country.[6]

For the proxy of resource abundance it would be best to use natural wealth (*lognwpc*) and subsoil wealth per capita (*logswpc*) since they correct not only for the population, but also a distinction is made between the value of different types of resources and minerals. Natural wealth per capita is the wealth arising from subsoil assets and other resources, including timber resources, non-timber forest resources, protected areas, cropland and pastureland. Subsoil wealth per capita is measured from energy resources and other mineral resources. These variables are also best in the sense that income in the model is also measured per capita. However, the per capita wealth data are only available for two years: 1994 and 2002 from the World Bank (World Bank, 1997 & 2006), which means that the panel regression in this case is only of a limited size. For the cross-sectional case, the average of the variable from both years is taken.

Since there is only per capita data available for two years, and in order to be able to compare results, the proxy's mineral exports, as a share of GDP (*smex*), and total primary exports, as a

[5] For a more extensive data description and the methods and calculations using this data in both the panel and cross-section model, refer to Appendix C.

[6] Stijns, (2006) and Brunnschweiler, (2008) provide a more extensive discussion on this topic.

share of GDP (*spex*), are used as well. For the resource proxy it is best to use the subsoil and mineral wealth indicators, because, when using natural wealth or primary exports variables, agricultural products and timber resources are also included in the resources variable. This would not provide a good analysis, however, as it is doubtful that the agricultural and mining sectors play a similar role in the economic development. Not only do the sectors use different technologies, the rents that are generated in both sectors are also substantially different. It is therefore best to make a distinction between the two sectors.[7] In addition to the separation of all natural resources and subsoil (mineral) wealth, a further distinction is made. The fuel exports as a share of GDP are also included, which are a part of the more general variable mineral wealth exports in GDP. This variable is called *sfex* and is interesting to analyze, since oil and gas may have a different influence on growth in a country than more general measures. All different resource measures are positively correlated, as is shown in Appendix E. The main issue of this research depends on the sign and magnitude of the coefficient of this variable: the question is whether this variable is positive or negative related to economic growth and with what magnitude.

In the Z^i variable, all other channels are captured that are of influence and importance on the growth regression. The following variables are used in the regression: institutions, investment, openness, human capital and terms of trade.

For the institutional variable (*inst*), political and civil liberty is used. This variable is taken from Freedom House and rates the institutional freedom of a specific country, giving every country a grade varying between 1(very free) and 7(not free). This is the only institutional variable included in the model, since the methodology to construct variables such as corruption and rule of law is practically the same for all. For this reason it is not necessary to include all these variables in the regression.[8] Institutions are expected to have a positive relationship with economic growth and are expected to be negatively related to resources. In the models, this would mean that a negative sign appears in the regression. The reason for this is that the lower the value of the variable the better institutions are. This means that economic growth will be positively influenced with decreasing values for institutions.

---

[7] Stijns, (2006) provides a further and more detailed discussion.
[8] See for explanation Manzano & Rigobon, (2001), p. 19.

In addition, a variable for investment is included (*inv*).[9] This is the investment percentage share in total GDP in a given year or average of years for a given country. It is expected to have a positive relationship to GDP. The data for this have been taken from the Penn World tables.

Openness (*open*) is measured here as the value of imports and exports relative to GDP for every year (or average) and is calculated using WB, WDI data. The more open a country is, the better this is for trade and exports and the better for economic growth. This implies a positive sign in the regression outputs.

As a proxy for human capital, the gross enrolment ratio in primary education is used (*sch*). Primary enrolment is chosen because of the availability of more data and because it is a better proxy for the human capital of a country.[10] Data has been taken from UNESCO and the World Bank. It is expected that education is positively related to economic growth. The relationship between resources and education is theoretically expected to be negative.

Terms of trade growth (*gtt*) is a variable that is the log of the annual growth the terms of trade, which is given by the ratio of an export price index to an import price index and comes from the World Bank. It is assumed that a terms of trade improvement has a negative effect on GDP growth, and thus has a negative sign (Oppenheimer & Maslichenko, 2006, pp. 15-31).

In addition to these important channels of economic growth, a group of control variables is included to prevent omitted variable bias. The variables included are used frequently in the literature and have found to influence economic growth. These variables are: population, life expectancy at birth, government expenditure as a share of GDP, financial depth, and the debt to GDP ratio.

Population (*lpop* in the panel and *logpop89* in the cross-section) is measured in logs, either yearly or on average, and is sourced from the World Bank, WDI. Life expectancy at birth in years (*lib*) is also taken from World Bank data. The life expectancy of a person is supposed to increase when economic growth is higher.

Also, government expenditures as a share of GDP (*gov*) are taken from the WB. It is expected that government expenses are negatively related to growth, and thus have a negative sign.

---

[9] Gylfason, (2001) and Atkinson & Hamilton, (2003).

[10] See Sala-i-Martin & Subramanian, (2003), it is robust and exogenous in their model.

Financial depth is a measure of liquidity to GDP in a country. The variable name is *fd* and data is taken from the WDI. Financial depth is expected to be of positive influence on economic growth and is assumed to have a negative relationship with natural resources.[11] Finally, the debt to GNP ratio of all countries is included (*debt*). This is measured as the total external debt to GNP ratio, taken from the WDI and is a measure of credit rationing. Higher debt will lead to lower growth.

There are no variables or dummies included for geography since Sachs and Warner (2001) find no influence of these variables on the model.

Over-valuation of the exchange rate would be an important variable to include in the model as this would be an indicator of the Dutch disease in a country. However, it has been found not to be significant (Sala-i-Martin & Subramanian, 2003, p. 8).

[11] See for example to Gylfason & Zoega, (2001).

4. Empirical analysis

In this section, the resource curse is investigated empirically. The section is divided into five subsections. Section 4.1 and 4.2 use a panel data model, whereas sections 4.3, 4.4 and 4.5 use a cross-sectional model. In section 4.1, panel least squares analysis is performed. Section 4.2 uses an interaction term for institutions to see how this influences the results. Section 4.3 gives the cross-section OLS estimations, and in section 4.4, an interaction term is added to the cross-sectional model. Section 4.5 gives the 2SLS regressions and analysis. Section 4.6 evaluates both methods and compares results.

4.1 Panel Least squares regressions

In order to analyze the resource curse, the regression that is described in section 3 is used. However, some alterations are made to the model. The reason for this is the existence of multi-collinearity between some of the variables, including schooling, financial depth and debt. Also the problem of non-robustness makes it necessary to remove these three variables from the equation. (Details and tests on this can be found in Appendix D.) Once these adjustments have been made, the regression model becomes as follows:

$$Loggdp8906^{i} = \alpha_0 + \alpha_1\, inigdp^{i} + \alpha_2\, R^{i} + \alpha_3\, inst^{i} + \alpha_4\, inv^{i} + \alpha_5\, open^{i} + \alpha_6\, gtt^{i} + \alpha_7\, lpop^{i} + \alpha_8\, lib^{i} + \alpha_9\, gov^{i} + \varepsilon^{i} \qquad (2)$$

Where R^i will take on different variables for the different analyses using one of the five variables: *smex, spex, sfex, lognwpc or logswpc*.

In the regression, the fixed effects estimation is used. With fixed effects, the explanatory parameters are allowed to be correlated with the unobserved fixed effect. In the fixed effects method, the constant is treated as group-specific so that the model allows for different constants per group and countries in this case. With random effects, we would have to assume no correlation between the country specific constant and the explanatory variables.

Fixed effects should be the appropriate measure since the sample is made up of individual countries and not random observations from a large sample.[12]

After the model has been estimated with fixed effects, some tests are performed to determine whether the estimates for the variables are unbiased and consistent. If this is not the case, then several adjustments have to be made to the method used for estimating the regression. As can be seen in Appendix D, there is evidence of heteroskedasticity in the model. Also it can be expected that some of the variables have correlated error terms. To correct for this in the model and still have unbiased and consistent estimators, the Generalized Least Squares (GLS) method is used. The appropriate GLS estimation method, when both correlation in the errors and heteroskedasticity are present in the model, is the seemingly unrelated regression (period SUR) method.[13]

The descriptive statistics of the important variables are given in Table 4.1.[14]

Table 4.1: Descriptive statistics

	Loggdp8906	***inigdp***	***spex***	***smex***	***sfex***	***lognwpc***	***logswpc***
Mean	0.0152	7.538	24.881	9.85	5.478	8.322	5.804
Median	0.021	7.422	21.527	4.43	1.414	8.292	5.858
Maximum	0.319	10.612	95.486	87.337	86.975	10.912	10.817
Minimum	-0.637	4.61	0.351	0.0133	-0.0094	6.242	0.00
Std. Dev.	0.053	1.605	17.635	12.925	9.97	0.962	2.301
Observations	1715	1716	1420	1420	1426	180	137

In Table 4.2 the results of the different regressions are shown for the different measures of natural resources. The dependent variable is GDP growth in all five regressions (*loggdp8906*). The numbers one to five correspond to the five different resource variables used in the regression, *smex, spex, sfex, lognwpc* or *logswpc,* respectively. The regression is run using panel weighted least squares (SUR).

[12] See Wooldridge, (2003), p. 473. A Hausman test is not possible in Eviews 5 therefore it could not be performed to test for which estimation to use (fixed or random effects). In the absence of this testing tool, "fixed effects" is the best measure.
[13] For a more extensive discussion and the tests, see Appendix D.
[14] For a more detailed overview of all other variables see Appendix E.

Table 4.2: GLS Regression output

Variable	**Regression**				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
inigdp	**-0.14***** (0.0169)	**-0.148***** (0.0172)	**-0.143***** (0.0168)	**-0.276***** (0.102)	**-0.142*** (0.0776)
inst	**0.0062**** (0.0027)	**0.0064**** (0.0028)	**0.0065**** (0.0027)	0.0060 (0.0145)	0.0042 (0.0116)
spex	**0.002***** (0.0003)				
smex		**0.0026***** (0.0006)			
sfex			**0.0033***** (0.00072)		
lognwpc				0.03195 (0.0278)	
logswpc					-0.0027 (0.0078)
Adj. R2	0.378	0.747	0.745	0.796	0.677
N	649	649	653	141	117

*, **, *** is the significance at the 10%, 5% and 1% level, respectively.

Notes:

1. The dependent variable is *loggdp8906*.
2. Standard errors are in parentheses.
3. Bold coefficients are significant.
4. For a more detailed regression output see Appendix E.

As can be seen in Table 4.2, there is a significant positive relationship between three of the resource variables and *GDP growth* up to the one percent level. The *natural wealth* and *subsoil wealth per capita* are not significant in the regression.
An increase of one percent in *resource exports in GDP* will, on average, increase *GDP growth* by 0.0025 percent. The greatest effect is that of *fuels as a percentage of GDP*, which has a coefficient of 0.0033. This suggests that fuels, such as oil and gas, would contribute the most to GDP growth. There is an overall positive relationship in the three significant variables, which means that no evidence is found for the existence of a resource curse.

Institutions have a positive effect in all five regressions. The variable is significant only in the first three regressions. The appearance of a negative sign is not as expected as this would mean the higher the value for institutions, and thus the worse they are, the higher the economic growth. This could be explained by the relationship of resources to institutions, as is explained in section 4.2.

The other variables in the model have the signs expected. (For the results with all the other variable coefficients see Appendix E.) *Initial GDP* has a negative relationship with *GDP growth*, which confirms conditional convergence. This variable is highly significant in the first four of the regressions ((1), (2,) (3), (4)) at the one percent significance level. Only in the fifth regression (5), does the level of significance reduce to ten percent.
Government expenditures are significant only in the first three regressions ((1), (2,) (3)). They are negatively related to the GDP growth, only in the fifth regression, with *subsoil wealth per capita* as resource variable, it has a positive sign, but is not significant.
The variable, *terms of trade changes*, is not significant in any of the regressions.
The *investment* variable is significant in all cases and has a positive coefficient. This variable proves to be a good estimate in the regression.
Life expectancy has a positive sign and is significant in all cases but one, the *subsoil wealth per capita* case. The higher *life expectancy*, the higher GDP growth, but the effect is relatively small.
Population is not significant for all but two regressions, the *spex* and *lognwpc*. There is a positive relationship between population and growth.
In the *natural wealth per capita*, the coefficient has a value of almost 0.5. This means that *GDP growth* would increase by almost half percent with a one percent increase in *population*.
Openness has a relatively large impact on growth, where it is highly significant at the one percent level in the three resource export regressions. There is a positive relationship between *openness* and growth, where a one percent increase in openness, increases growth by 0.06 percent.
The adjusted R-squared is highest in the last four regressions, but the proportion of the total sample variation that is explained by the independent variables is quite large, varying from 0.38 to 0.80. The number of observations is highest in the first three regressions, which can also explain some of the regression outcomes.

In the regressions in Table 4.2, it is clear that the regressions with the *exports in GDP* variables provide the most reliable and robust results. However, even when using these variables, a negative relationship between resources and GDP growth is not observed. This is another confirmation that, even when the variable for resource abundance is not the right proxy, no evidence can be found for the assumed negative relationship. Even when the more widely used variables, *resource exports as a percentage of GDP*, are used there is a positive and highly significant relationship between resources and economic growth. This relationship remains, even after including important variables such as *investment* and *terms of trade shocks*. This suggests that it is not the misspecification of the proxy for resources that affects the relationship, but the fact that a fixed panel is used in this case instead of cross-section data.

The higher number of observations means that the results are more reliable than in a cross-section analysis. It can be deduced that the empirical finding of a resource curse using cross-sectional data is possibly due to omitted variable bias and data restrictions due to the limitations in the number of observations. This topic will be investigated further in section 4.3.

When estimating this GLS model, there may be some problems. The variables in the OLS estimation can, due to omitted variable bias or measurement error, suffer from endogeneity. This means that one of the explanatory variables in the model might be correlated with the error term. This is not adequately allowed for in the normal regression and therefore some procedures may have to be performed, to correct for these problems.[15]

The problem of omitted variable bias is already controlled for in two ways: by using panel data instead of cross-sectional data, and by adding more variables to the model; however, it still cannot be ruled out. To improve the regression model, several actions can be performed to solve for problems of endogeneity of a variable. Instrumental variables can be used to correct for this. Also a two-stage least squares (2SLS) regression can be constructed and run.[16] The 2SLS procedure will be explained in more detail in section 4.5.

The variable that is most likely to be endogenous is *institutions*. Although the institutions variable is possibly less likely to suffer from endogeneity problems in the short run, since institutions evolve slowly, it might be endogenous in the long run (Papyrakis & Gerlagh, 2004, p.184). When a Hausman test for endogeniety is performed on the institutions variable,

[15] See Woolridge, (2003), chapter 15 and Pindyck and Rubinfeld, (1998), pp. 182-186.

[16] Woolridge, (2003), chapters 15 and 16: Besides measurement error and omitted variable bias, a third endogeneity problem may arise: this is simultaneity, which arises when one or more of the explanatory variables are jointly determined with the dependent variable. It can also be solved using instrumental variables.

there is no evidence of endogeneity at any significance level. This means a 2SLS regression in this case is not necessary. (A more detailed description of this is given in Appendix E.)

Main findings

When running the regressions for the different resource variables for the panel-data model, it shows that the *natural wealth per capita* and *subsoil wealth per capita* are not significant in the regression. The other three resource variables are however highly significant up to the one percent level. These variables show a positive relationship with GDP growth.

The variable with the largest impact on growth is that of *fuel exports as a percentage of GDP*. The other variables show the expected signs. *Initial GDP* has a negative relationship with *GDP growth*, which confirms conditional convergence. This variable is also significant in all five regression models. The institutions variable is significant in the first three models, but is not of the expected sign. The variables, *openness*, *investment*, *government expenditures*, and *life expectancy at birth* are all highly significant in the first three regressions.

4.2 Panel data model with interaction term

It is often claimed that natural resources do not have a direct negative relationship with economic growth, but that this occurs as a result of the influence of different channels. *Institutions*, in this respect, can be seen as the channel through which resource abundance has a negative impact on growth. Resources alone may have a positive effect on growth only in countries where the institutions are good. This means that the more resources a country has, and the worse the institutions in that country are, the lower economic growth will be. Even though resources may, themselves, have a positive effect on growth, their impact might be influenced by the quality of institutions in a country (Brunnschweiler, 2008, pp. 403-404).

To find out if the influence of resources on economic growth is, indeed, influenced by institutions, an interaction term is added to the regression. This interaction term allows resources to depend on institutions. The model will then look as follows:

$$Loggdp8906^{i} = \beta_0 + \beta_1\, lgdp89^{i} + \beta_2\, R^{i} + \beta_3\, inst^{i} + \beta_4\, R^{i} * inst^{i} + \beta_5\, Z^{i} + \delta^{i} \quad (3)$$

Where R^i is again one of the five measures of natural resources and Z^i captures all control variables in the regression (as in regression (2)).

If the interaction term, captured by β_5 R^i * $inst^i$ is positive, this means that more resources yield a higher growth rate for countries with worse institutions. If the interaction term is negative, more resources yield a lower growth rate for worse institutions. This last outcome is to be expected in the regression outputs, since the higher the *institutions* term, the lower the institutional quality; and the more resources a country with bad institutions has, the lower growth is expected to be.

When analyzing the results from this regression, the coefficient for R^i measures the effect on growth for a value for *inst* of zero, that is, very good institutions. When using an interaction term, the individual significance of the resource variables and the interaction term is no longer representative. To observe the significance of both variables, a test for joint significance has to be performed (Wooldrigde, 2003, pp. 194-196). The results of the regression with the interaction term are displayed in Table 4.3. (A more detailed output is given in Appendix E.)

As can be seen in Table 4.3, the variable institutions, *inst,* now carries a negative sign, as was discussed in section 4.1, in the *lognwpc* and *logswpc* regressions with the interaction term. When examining the significance of the interaction term, only *lognwpc* is significant. But this cannot be tested by a separate test on the significance of the variable. This has to be done by performing a test on the joint significance of the resource variable and the interaction term. On testing the joint significance of the resource variable and the interaction term the results show that the first three regressions are significant at the one percent level. The fourth regression gives only significant variables at the ten percent level, and the *subsoil wealth per capita* regression variables are not significant (See Appendix E). In regression (4), the resource variable *lognwpc* becomes significant, whereas it was insignificant in section 4.1. It shows a negative relationship between *resources per capita* and *growth*. This is the first instance of a resource variable showing a negative significant relationship with growth.

The findings provide evidence that resources do depend on institutions and jointly influence economic growth. As in section 4.1, the fuel export variable, *sfex,* still has the largest positive effect on *GDP growth*, now with a coefficient of 0.0028. This does not differ much from the earlier results.

Table 4.3: Regressions with interaction term

Variables	Regressions (1)	(2)	(3)	(4)	(5)
inigdp	**-0.140***** (0.0158)	**-0.149***** (0.0159)	**-0.141***** (0.016)	**-0.157***** (0.043)	**-0.1454***** (0.0467)
inst	0.0030 (0.0038)	**0.0067**** (0.0029)	**0.0057**** (0.0028)	**-0.0989**** (0.0425)	-0.0213 (0.0185)
spex	**0.0015***** (0.00053)				
smex		**0.0027***** (0.00085)			
sfex			**0.0028***** (0.0009)		
lognwpc				**-0.0452**** (0.0219)	
logswpc					-0.016 (0.0102)
Interaction	0.00017 (0.00014)	-0.00035 (0.00023)	0.00022 (0.00026)	**0.0124**** (0.0053)	0.004 (0.0027)
Adj. R2	0.759	0.748	0.746	0.716	0.688
N	649	649	653	140	117

*, **, *** is the significance at the 10%, 5% and 1% level, respectively.

Notes:

1. The dependent variable is *loggdp8906*.
2. Standard errors are in parentheses.
3. Bold coefficients are significant.
4. For a more detailed regression output see Appendix E.

Table 4.3 shows one other peculiar thing; the interaction term is positive, when the assumption was that it would be negative. This coefficient now states that countries with better institutions and higher economically developed countries demonstrate a declining growth due to the resources they have. A possible explanation of the positive interaction term in the regressions is the conditional convergence hypothesis. This means that countries that already have high quality institutions and possess resources do not grow much more, as compared to the lower institutions countries. The lower institutions countries demonstrate more growth due to resources, in this case. This means that income levels also react to growth rates in the same way, thus showing a negative relationship between institutions and initial income levels. The higher the initial income, the lower growth will be in this case.[17]

Main findings

After the inclusion of an interaction term between natural resources and institutions, the regressions for the resource variables as a percentage in GDP remain significant and have a positive influence on growth. The regression for *natural wealth per capita* becomes significant when the interaction term is included in the regression. This variable now shows a negative effect on growth. *Institutions* also remain significant in the first three regressions and also become significant in the natural wealth per capita case. Here the variable shows the expected sign. Initial GDP continues to demonstrate a negative relationship to growth. This variable also remains significant in all five regressions. The interaction term carries a positive sign, which is counter-intuitive. An explanation could be the conditional convergence hypothesis where, in this case, countries that already have good quality institutions grow less than institutionally under-developed countries.

4.3 Cross-section OLS regressions

In order to perform an OLS regression with the cross-sectional data, some tests were performed first. The model was tested for multicollinearity, robustness, heteroskedasticity and the sample size and variables were analyzed. (An extensive analysis and discussion is given in Appendix D.)

No signs of multicollinearity were found in the variables and they also seem to be robust. There was, however, a problem with the sample size. The variable terms of trade (*gtt*) was therefore removed from the model. When testing for heteroskedasticity, the null-hypothesis of no heteroskedasticity could not be rejected at all levels of significance, therefore, the White corrected standard errors are used in the regressions.

The model for the OLS regressions then looks as follows:

$$Lgdpgrowth8906^{i} = \gamma_0 + \gamma_1\, lgdp89^{i} + \gamma_2\, debt^{i} + \gamma_3\, fd^{i} + \gamma_4\, gov^{i} + \gamma_5\, inst^{i} + \gamma_6\, inv^{i} + \gamma_7\, lib^{i} + \gamma_8\, logpop89^{i} + \gamma_9\, open^{i} + \gamma_{10}\, sch^{i} + \gamma_{11}\, R^{i} + \upsilon^{i} \qquad (4)$$

[17] Brunnschweiler (2008) also suggested this after finding contradictory results in the interaction term.

Where R^i again takes on one of the five different resource measurement variables: *smex, spex, sfex, lognwpc* or *logswpc.* The descriptive statistics of the most important variables in the model are shown in Table 4.4.[18]

Table 4.4: Descriptive statistics

	lgdpgrowth8906	***lgdp89***	***lognwpc***	***logswpc***	***sfex***	***smex***	***spex***
Mean	0.0063	7.48	8.376	5.754	5.895	10.824	26.826
Median	0.007	7.393	8.366	5.792	1.573	4.495	23.763
Maximum	0.0345	10.378	10.761	10.462	77.443	77.878	85.482
Minimum	-0.0117	4.862	6.59	0.00	0.00078	0.0236	0.528
Std. Dev.	0.0072	1.55	0.891	2.368	11.249	14.0496	18.322
Observations	101	101	95	76	98	98	98

In Table 4.4, it can be seen that the number of observations is lower than in the panel-dataset. In Table 4.5 the results of the different OLS regressions are shown for all five measures of natural resources. The dependent variable is *GDP growth* in all five regressions (*lgdpgrowth8906*). Every number in the table is a regression with the resource variable, *smex, spex, sfex, lognwpc* and *logswpc,* respectively.

Table 4.5 shows that none of the resource variables are significant in the regression. The signs of two of the resource variables, *spex* and *lognwpc,* are negative. However, no conclusions can be drawn from this. Reason for the insignificance of the resource can be the small sample size.

The investment variable is significant in all cases and has a positive sign. This variable proves to be a good estimate in the regression.

Life expectancy, *lib,* has a positive sign and is significant at the ten percent level in all cases but one, the *subsoil wealth per capita* case. This was also the case in the panel-data regressions. There is a positive relationship between *growth* and *life expectancy.*

Population is only significant in the last two regressions.

Openness and *education* are not significant in any of the regressions.

There is not much difference between the five regressions with respect to the level of significance of the variables or the explanatory power of the variables. There is only a small

[18] For all other descriptive variables, see Appendix E.

sign that the last two regressions, with the *wealth per capita* as explanatory variable, are better estimates of the model as in this case, more variables are significant.

Table 4.5: OLS regression output

Variables	Regressions				
	(1)	(2)	(3)	(4)	(5)
lgdp89	**-0.0024**** (0.00095)	**-0.0024**** (0.001)	**-0.0025**** (0.001)	**-0.0024*** (0.0013)	**-0.0026**** (0.0011)
inst	-0.0004 (0.00067)	-0.0005 (0.00069)	-0.00058 (0.00073)	-0.00086 (0.0007)	-0.0013 (0.0008)
spex	-0.00005 (0.00007)				
smex		0.000007 (0.00006)			
sfex			0.00004 (0.00009)		
lognwpc				-0.00047 (0.0014)	
logswpc					0.00008 (0.0004)
Adj. R2	0.439	0.434	0.435	0.5	0.56
N	65	65	65	62	45

*, **, *** are the significances at the 10%, 5% and 1% level, respectively.

Notes:

1. The dependent variable is *lgdpgrowth8906.*
2. Standard errors are in parentheses.
3. Bold coefficients are significant.
4. For a more detailed regression output see Appendix E.

Main findings

In the OLS cross-sectional analysis, none of the resource variables are significant in the regressions. One reason for this can be the small sample size. *Initial GDP* remains highly significant in all five regressions, carrying a negative sign as expected. All other variables also show the expected signs, but not all are significant. The last two regressions, with the *wealth per capita* as explanatory variable, are somewhat better estimates of the model as more variables are significant in this case.

4.4 Cross-sectional model with interaction term

In order to, once again, ascertain if the influence of resources on economic growth is affected by institutions, this time in the cross-sectional model, an interaction term is added. The regression now looks as follows:

$$Lgdpgrowth8906^{i} = \gamma_0 + \gamma_1\, lgdp89^{i} + \gamma_2 R^{i} + \gamma_3\, inst^{i} + \gamma_4 R^{i} * inst^{i} + \gamma_5 Z^{i} + \upsilon^{i} \qquad (7)$$

Where R^i is, again, one of the five measures of natural resources and Z^i captures all the control variables in the regression. As before this regression calculates the effect on growth with a value of zero for the institutions variable, which means high quality institutions. In Table 4.6, the regression output is presented for the five regressions with the interaction term. (A more detailed output is presented in Appendix E.)

When the joint significance is tested for the different resource variables and the interaction term, only the *fuel exports in GDP*, *sfex,* and the *subsoil wealth per capita*, *logswpc,* are significant along with the interaction term at the ten percent level.
The interaction term again is expected to show a negative sign because of the higher institutional quality (the lower the variable *inst*) and more resources have positive growth effects. However, the interaction term is positive in three of the regressions. Only in the regression for *lognwpc* and *spex* it has a negative sign. These two regressions, however, show no significance in the interaction term, or in the resource variable. The explanation for the positive interaction term in the remaining regression suggests that higher institutional quality and more resources have a diminishing effect on growth. This can be due to the conditional convergence hypothesis, as was explained in section 4.2.

The effect of fuels and minerals from regressions (3) and (5) on growth are small but negative. This is an interesting result since it means that when fuels and subsoil wealth are taken separately, there is a significant negative relationship between resources and growth. This is not the case when all resources are considered as a whole. These findings support the resource curse hypothesis; also for the *subsoil wealth per capita* variable. It shows that a doubling of subsoil resources and thus an increase of a hundred percent, decreases growth by 0.16 percent.

Table 4.6: Regression with interaction term

Variables	**Regressions**				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
lgdp89	**-0.0022**** (0.0009)	**-0.0024**** (0.001)	**-0.0026**** (0.0011)	**-0.0026**** (0.0013)	**-0.0024**** (0.001)
inst	0.00014 (0.0012)	-0.00075 (0.0009)	-0.0012 (0.0009)	0.0057 (0.0067)	**-0.0038***** (0.0011)
spex	0.00003 (0.00014)				
smex		-0.0001 (0.0002)			
sfex			**-0.00043**** (0.0002)		
lognwpc				0.0024 (0.0029)	
logswpc					**-0.0016**** (0.0007)
Interaction	-0.00002 (0.00004)	0.00003 (0.00005)	**0.00012**** (0.00005)	-0.0008 (0.0008)	**0.0005**** (0.0002)
Adj. R2	0.432	0.425	0.453	0.5	0.59
N	65	65	65	62	45

*, **, *** are the significances at the 10%, 5% and 1% level respectively.

Notes:

1. The Dependent variable is *lgdpgrowth8906*.
2. Standard errors are in parentheses.
3. Bold coefficients are significant.
4. For a more detailed regression output see Appendix E.

Main findings

After including an interaction term between resources and institutions in the cross-sectional model, there are some interesting results. Two of the five resource variables, namely *fuel exports as a percentage of GDP* (*sfex*) and *subsoil wealth per capita* (*logswpc*), become significant. These two variables both measure the same thing, but in different ways. The sign of the coefficients is negative in both cases, whereas the interaction term is positive. This can be due to the conditional convergence hypothesis.

The resource variables *sfex* and *logswpc* both indicate that countries with oil, gas and minerals have relatively lower growth than countries that don't. Also the value of the change in the resource variables points to a negative effect on growth through institutions. This is measured for a value of zero for institutions, which means that even when countries have the highest quality of institutions, a significant negative effect is still experienced on growth attributable to the presence of resources.

Initial GDP remains highly significant and negative for all regressions. The explanatory power of the model is improved after including the interaction term.

4.5 2SLS regressions

As already discussed in section 4.1, the regression can, also for this case, suffer from endogeneity problems. Endogeneity problems are likely to arise sooner in the cross-sectional model than with the panel data. To ascertain if there is indeed endogeneity in the institutions variable, a Hausman test is performed on the consistency of the institutions variable. This test rejects the null-hypothesis at the ten percent level. In order to correct for this endogeneity problem, a 2SLS regression is performed, with institutions exogenized, by using instrumental variables for estimating the influence of institutions on the model. When there is one endogenous explanatory variable in the model, along with multiple exogenous instrumental variables (IV) for this one explanatory variable, the IV variable is called the two-stage least squares (2SLS) estimator and a 2SLS procedure is used. The conditions are that the two instruments do not appear in the original regression, that either one of the exogenous variables that is used as IV is different from zero and that the IV also have to be jointly significant at least at the 5% level (Wooldridge, 2003). (Appendix E provides the further details.)

The main instruments used for institutions in research on the resource curse are: data on *settler mortality* collected by Acemoglu et al. (2001); and the *fraction of the population speaking English and European languages* (due to Hall and Jones, 1998).[19] Also *geographical latitude* is used as an instrument (Brunnschweiler, 2006).

However, as this data is not available, the Polity IV variable (*reg*) is used as an instrument instead, as has also been considered in Brunnschweiler (2006). This variable is a measure of regime structure and political changes and indicates whether a country experienced a transition of regime or violent change during the period analysed.

Another instrument that is frequently used is that of ethno-linguistic fractionalization. The *ethno-linguistic fractionalization* variable (*eth*) is the probability of two random people in a country not speaking the same language and is taken from the IMF (2001). The higher this

[19] Discussed by Sala-i-Martine & Subramanian, (2003), p. 6.

index, the lower growth is assumed to be. (The correlation matrix of these three variables is given in Appendix E.)

Both *reg* and *eth* are used to instrument for *inst* in the regressions. The first stage regression, with institutions as a dependent variable and with the two instruments, is as follows:

$$inst^i = \delta_0 + \delta_1\, lgdp89^i + \delta_2\, debt^i + \delta_3 fd^i + \delta_4\, gov^i + \delta_5\, inv^i + \delta_6\, lib^i + \delta_7\, logpop8^i + \delta_8\, open^i + \delta_9\, sch^i + \delta_{10} R^i + \delta_{11}\, reg^i + \delta_{12}\, eth^i + \eta^i \quad (5)$$

The second stage is:

$$Lgdpgrowth8906^i = \gamma_0 + \gamma_1\, lgdp89^i + \gamma_2\, debt^i + \gamma_3 fd^i + \gamma_4\, gov^i + \gamma_5\, inst^i + \gamma_6\, inv^i + \gamma_7\, lib^i + \gamma_8\, logpop89^i + \gamma_9\, open^i + \gamma_{10}\, sch^i + \gamma_{11} R^i + \upsilon^i \quad (6)$$

Where R^i in both regressions, is again, one of the variables of natural resources, *smex, spex, sfex, lognwpc* and *logswpc.*

The results of the 2SLS regressions are shown in Table 4.7. Note that the numbers one to five correspond again to the five different resource variables used in the regression. The table shows that the number of observations has not changed when using the instruments. The adjusted R-squared of the different regressions has changed somewhat, explaining less in the variation of the variables, but this is expected when using instruments.
In Table 4.7 it is shown that after using the instruments for *inst*, the institutional variable is significant at the five percent level and it remains insignificant only in regression (1). The coefficient is again negative, as expected. *Life expectancy* is no longer significant in four of the five regressions. The resource variables remain insignificant in all five regressions. The signs and size of the resource variables remain close to the results in the normal OLS regressions in section 4.4.

Smex, sfex, and *subsoil wealth per capita* have a larger coefficient compared to the OLS model. *Spex* becomes smaller. *Financial depth*, *initial income* and *debt* remain significant in the majority of cases. *Schooling*, *openness* and *government expenditures* remain insignificant in all regressions. It is worth noting that *investment* loses significance in all regressions, whereas it was significant in the OLS model.

Table 4.7: 2SLS regression output

Variables	Regressions				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
lgdp89	**-0.0024**** (0.0011)	**-0.0027**** (0.0012)	**-0.0029**** (0.0012)	-0.0025 (0.0016)	**-0.0028**** (0.0012)
inst	-0.0019 (0.0012)	**-0.0025*** (0.0013)	**-0.0028**** (0.0013)	**-0.0027**** (0.0012)	**-0.0023**** (0.0011)
spex	-0.00002 (0.00007)				
smex		0.00007 (0.00007)			
sfex			0.00015 (0.00012)		
lognwpc				-0.00049 (0.0015)	
logswpc					0.00017 (0.00038)
Adj. R2	0.385	0.343	0.323	0.415	0.535
N	65	65	65	62	45

*, **, *** are the significances at the 10%, 5% and 1% level, respectively.

Notes:

1. The dependent variable is *lgdpgrowth8906*.
2. Standard errors are in parentheses.
3. Bold coefficients are significant.
4. For a more detailed regression output see Appendix E.

Main findings

On testing for endogeneity of the institutions variable, this appeared to reject the null-hypothesis of consistent estimators. To correct for this, a 2SLS regression was run, with institutions exogenized and the instruments *political regime shift* (*reg*) and *ethno-linguistic fractionalization* (*eth*) were used for institutions. On using the instruments for institutions in the regression, the institutional variable becomes significant at the five percent level in four out of the five regressions and the sign of institutions remains as expected. None of the resource variables, however, become significant after the 2SLS regressions.

4.6 Comparison: Panel data or cross-section?

There are some important considerations in choosing between the cross-sectional analysis and the panel data analysis. The panel data model provides more observations which can provide more reliable results. The panel data model is, most likely, to correct for omitted variable bias. The cross-sectional model, on the other hand, avoids problems such as shortcomings in the amount of data available, non-stationarity, and autocorrelation. The biggest disadvantage of this model is the small number of observations.

Main differences and similarities

When examining the results obtained using the two models, the following points are observed.

1. With the least squares regressions, three of the five resource variables in the panel are significant (all *exports as a share of GDP*). In the cross-sectional model, this is zero.
2. On including the interaction term in both models, in the panel model, the same three resource variables as in 1 continue to be significant and a further variable, *lognwpc*, becomes significant as well. The results are that three of the resource variables, *spex, smex* and *sfex* show a positive relationship to resources in the panel model. *Lognwpc* shows a negative relation to resources.
3. In the cross-sectional model, two of the five variables become significant after adding an interaction term. These are *subsoil wealth per capita* (*logswpc*) and *fuel exports* (*sfex*). Both variables show a negative relationship to growth.
4. The findings indicate that the inclusion of the interaction term is the best option for both models as it results in the resource variables in the cross-sectional model becoming significant in two cases, and in the panel data model four variables become significant.
5. If the findings are compared, only the variable *fuel exports* (*sfex*) is significant in both models. In the panel data model it shows a positive relationship and in the cross-sectional, it shows a negative relationship.
6. The one thing both models support, for every model and every regression, is the conditional convergence hypothesis. The initial income variable is significant and negative in all cases.
7. What is interesting is that only the case of minerals and fuels show a negative significant sign in the cross-sectional model. This implies that these types of resources, specifically, have an effect on growth. In the panel data model, this

assumption is confirmed, since the coefficient of fuels (*sfex*) is the largest of the three resource exports variables, implying that, of all resources, fuels have the largest impact on growth. The sign of the variables differs in both models however.

Type of measurement of resources

In the panel data model, it is proven that the type of measurement for resources is, in fact, of influence on the outcome. When it is assumed that per capita measures are better measurements than the exports as a percentage of GDP, or the more 'old', measurements, this shows that there is a negative relationship of resources to growth. The cross-sectional model provides confirmation of this statement since the subsoil wealth per capita for this model has a negative coefficient. This means that the resource curse hypothesis is confirmed.
A disadvantage of measuring resources per capita is the availability of data, with only two data points per country. This limits the available results. The 'old' measurements give ambiguous results since in the panel data model they show a positive relationship, whereas in the cross-sectional model there is evidence of a negative relationship.

Number of observations

If the number of observations in both models is considered, it is most likely, that the panel-data model provides the most robust results. In addition, if omitted variables are taken into account this model appears to be the most reliable model of the two.

5. Conclusion

In a world where energy and natural resources are becoming scarce, there is room for many studies in the academic field on these topics. Energy production is necessary for human needs and economic development. However, natural resources in general, and energy resources in particular, do not always lead to more economic growth and development of a country *per se*. There is some evidence of the existence of a curse on resources. This curse means that having more resources as a country leads to lower economic growth.

Theories on the existence of a resource curse have abounded for decades. The main transmission channels of the resource curse that are identified in the literature are: the Dutch disease, human capital, corruption, rent seeking, political liberty, investment and savings, debt overhang, openness to trade and terms of trade.

The link between resources, government quality and economic growth was already identified early on. Rent-seeking is seen to be an important problem in the relationship of resources to growth. Fiscal linkages influence economic growth through resources. Starting in 1995 with the research of Sachs and Warner, different studies have empirically tested the issue, with no clear consensus. The discussion is still prevalent.

In this research, the resource curse hypothesis was tested empirically. What is different from previous research is that this study reviews both the cross-sectional and the panel data model, to compare results. In addition, the measure for natural resource abundance is different from most of previous research in that per capita measures are used instead of only percentages of resources in exports. Also more recent data is utilized.

Using the panel data model, the regressions shows that resource variables as a percentage of GDP are significant in all cases and show a positive relationship to growth. The per capita measures are not significant. The variable that has the largest impact on growth is fuel exports. After including the interaction term between natural resources and institutions, the regressions remain significant and continue to have a positive influence on growth. The regression for the natural wealth per capita becomes significant after including the interaction term in the regression, carrying a negative sign. The interaction term seems to be the right approach for this case.

For the cross-sectional model, the results differ. In the OLS regressions, none of the resource variables are significant. The same applies for the 2SLS regressions. After including the interaction term, two resource variables become significant, namely *fuel exports as a percentage of GDP*, and *subsoil wealth per capita*. The sign of the coefficients is negative in both cases, whereas the interaction term is positive. This can be due to the conditional convergence hypothesis.

There are some important differences in the use of one or the other model. The panel data model provides more observations and is more likely to correct for omitted variable bias, which can be a problem when using the cross-sectional model. The panel data model has other disadvantages.
What is interesting is that only the minerals and fuels variables show a negative significant sign in the cross-sectional model, which implies that especially these types of resources have an effect on growth. In the panel data model, this assumption is confirmed, since the coefficient of fuels is the largest, implying that fuels have the largest impact on growth out of all the types of resources.
Whether this relationship is positive or negative remains ambiguous however. Only the fuel exports variable is significant in both of the models. In the panel data model it shows a positive relationship and in the cross-sectional it shows a negative relationship.
When assumed that per capita measures are the best measures for resources, this implies that there a negative relationship of recourses to growth does exist. Both models confirm this and thus support the resource curse hypothesis.
A disadvantage of measuring resources per capita is the data availability, with only two data points per country which limits the results available. Using the 'old' measures for resources, there is no evidence found of the resource curse hypothesis in the panel data model, however. The cross-sectional model shows a negative sign.

The influence of the resource curse can only be accurately measured with an adequate number of data points, which is difficult to accomplish because some of the variables used in the model. For the cross-sectional dataset, the danger of bias is also present. Omitted variables can influence the outcome of the regressions even though many variables are already included in the model. The possibility of other variables affecting growth in the resource regression should be analyzed.

I recommend that the focus of future research should be the role of subsoil resources, in particular.

The results of this research have shown that, in the cross-sectional model, there is evidence of a negative relationship between growth and subsoil resources in combination with institutions. This means that recent data shows that oil, gas, fuels and minerals are the main resources influencing growth though institutions.

What can be useful are the policy implications that arise from this research making the need for further research in this direction very valuable.

References

Annet, A., (2001). *Social Fractionalization, Political Instability, and the Size of Government.* IMF Staff Papers, Vol. 48, No. 3.

Asteriou, Dimitros & Hall, Stephen G., (2006). *Applied econometrics. A modern approach using Eviews and Microsiftl.* PalgraveMacMillian.

Atkinson, Giles and Hamilton, Kirk, (2003). Savings, growth and the resource curse hypothesis. *World Development*, Vol. 31, No. 11, pp. 1793–1807.

Auty, R.M., (1993). *Sustaining development in mineral economies : the resource curse thesis.* London: Routledge.

Bravo-Ortega, C., & De Gregorio, J. (2005). *The relative richness of the poor? Natural resources, human capital and economic growth.* World Bank, working paper, No. 3484.

Bruno, M. and Sachs, J.D., (1982). Energy and resource allocation: A dynamic model of the "Dutch disease". *The Review of Economic Studies,* Vol. 49, No. 5, pp. 845-859.

Brunnschweiler, Christa N., (2008). Cursing the blessings? Natural resource abundance, institutions, and economic growth. *World Development,* Vol. 36, No. 3, pp. 399–419.

Corden, W.M., (1984). Booming sector and Dutch disease economics: Survey and consolidation. *Oxford Economic Papers*, Vol. 36, No. 3, pp. 359-380.

Davis, Graham A., (1995). Learning to love the Dutch disease. Evidence from the mineral economies. *World Development,* Vol. 23, No. 10, pp. 1765-1779.

Égert, Balázs, (2005). Equilibrium exchange rates in South Eastern Europe, Russia, Ukraine and Turkey: Healthy or (Dutch) diseased? *Economic Systems*, Vol. 29, pp. 205–241.

Ellman, Michael, (1981). Natural gas, restructuring and re-industrialization: The Dutch experience of industrial policy, in Barker, T. and Brailovsky, V. (eds), *Oil or industry? : energy, industrialisation and economic policy in Canada, Mexico, the Netherlands, Norway and the United Kingdom.* London: Academic press.

EViews 5 User's Guide, (2004). Quantitative Micro Software, LLC: USA.

Farmer, Roger E.A., (2002). *Macroeconomics.* Mason: South Western Thomson Learning.

Gelb, Alan H. & Associates, (1988). *Oil windfalls: Blessing or curse?* New York: Oxford University press.

Gylfason, T., (2007). *The international economics of natural resources and growth.* CESIFO, working paper, No. 1994.

Gylfason, T., (2000). *Natural resources, education and economic development.* CEPR, discussion paper, No. 2594.

Gylfason, T. & Zoega, G.,(2001). *Natural resources and economic growth: The role of investment.* CEPR, discussion paper, No. 2743.

Herbertsson, T.T., Skuladottir, M. and Zoega G., (2000). *Three symptoms and a cure: a contribution to the economics of the Dutch disease.* CEPR, discussion paper, No. 2364.

Heston, A., Summers, R. & Aten, B., (2004). *Penn World Tables.* Version 6.2. Through website: http://pwt.econ.upenn.edu/

International Monetary Fund (IMF), (2008). *World economic outlook database.* Through website: http://www.imf.org/external/pubs/ft/weo/2008/01/weodata/index.aspx

Keppler, Jan Horst, Bourbonnais, Régis & Girod, Jaques. Ed., (2007). *The econometrics of energy systems.* New York: Palgrave MacMillan.

Larsen, E.R., (2005). Are rich countries immune to the resource curse? Evidence from Norway's management of its oil riches. *Resources Policy*, Vol. 30, pp. 75–86.

Leite, C., & Weidmann, J. (1999). *Does mother nature corrupt? Natural resources, corruption, and economic growth.* IMF, working paper, No. 99/85.

Manzano, Osmel and Rigobon, Roberto, (2001). *Resource Curse or Debt Overhang?* NBER, Working Paper, No. 8390.

Marshall, M.G. & Jaggers, K., (2006). *Polity IV Project dataset: Political Regime Characteristics and Transitions, 1800-2006*. Center for Systemic Peace. Via website: http://www.systemicpeace.org/polity/polity4.htm

Mills, Terence C. & Patterson, Kerry. Ed., (2006). *Palgrave handbook of econometrics. Volume I: Econometric theory*. New York: Palgrave Macmillan.

Neary, Peter, & van Wijnbergen, Sweder, (eds.), (1986). *Natural resources and the macroeconomy*. Oxford: Basil Blackwell.

Oppenheimer, Peter & Maslichenko, Sergiy, (2006). Energy and the economy: an introduction, in Ellman, Michael (eds.), *Russia's oil and natural gas. Bonanza or curse?* London: Anthem press.

Papyrakis, Elissaios and Gerlagh, Reyer, (2004). The resource curse hypothesis and its transmission channels. *Journal of Comparative Economics,* Vol. 32, pp. 181–193.

Pindyck, R.S. & Rubinfeld, D.L., (1998). *Econometric models and economic forecasts.* Boston: McGraw-Hill.

Sachs, J. D. and Warner, A.M., (2001). Natural resources and economic development. The curse of natural resources. *European Economic Review*, Vol. 45, pp. 827-838.

Sachs, J. D. and Warner, A.M., (1995). *Natural resource abundance and economic growth*. NBER, working paper, No. 5398.

Sala-i-Martin, X., & Subramanian, A. (2003). Addressing the natural resource curse: An illustration from Nigeria. NBER, working paper, No. 9804.

Snowdon, Brian & Vane, Howard R., (2005). *Modern Macroeconomics. Its origins, development and current state.* Cheltenham: Edward Elgar Publishing Limited.

Stijns, J-P., (2006). Natural resource abundance and human capital accumulation. *World Development,* Vol. 34, No. 6, pp. 1060–1083.

Transparency International, (2007). *Corruption Perceptions Index 2007.* Through website: http://www.transparency.org/policy_research/surveys_indices/cpi

UNESCO, (2007). *Data Centre*. UNESCO, Institute for statistics. Through website: http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=136&IF_Language=eng&BR_Topic=0

Wijnbergen, S. van, (1984a). The `Dutch disease': A disease After All? *The Economic Journal*, Vol. 94, No. 373, pp. 41-55.

Wijnbergen, S. van, (1984b). Inflation, employment, and the Dutch disease in oil-exporting countries: A short-run disequilibrium analysis. *The quarterly journal of economics*, Vol. 99, No. 2, pp. 233-250.

Woolridge, J.M., (2003). *Introductory econometrics. A modern approach*. Mason: South Western Thomson Learning.

World Bank (2007). *World development indicators (WDI*). The World Bank Group.

World Bank, (2006). *Where is the wealth of nations? Measuring capital for the XXI century.* Washington, DC: The International Bank for Reconstruction and Development/The World Bank.

World Bank, (1997). *Expanding the measure of wealth: Indicators of environmentally sustainable development.* Environmentally sustainable development studies and monographs series, no. 17.

Appendices

Appendix A: Overview of empirical literature

author	**year**	**resource curse?**	**data-set**	**analysis type**	**endogeneity correction**
Sachs & Warner	1995	yes	cross-section	OLS	
Sachs & Warner	2001	yes	cross-section	OLS	
Gylfason	2000	yes	cross-section	OLS	
Gylfason & Zoega	2001	yes	cross-section	OLS	
Gylfason	2007	yes	cross-section	OLS	
Leite & Weidmann	1999	yes	cross-section	OLS	IV (instrumental variable)
Atkinson & Hamilton	2003	yes	cross-section	OLS	
Sala-i-Martin & Subramanian	2003	yes	cross-section	OLS & TSLS	IV (instrumental variable)
Papyrakis & Gerlagh	2004	yes	cross-section	OLS	
Larsen	2005	yes	time series	structural break	
Bravo-Ortega & De Gregorio	2005	yes	panel data	OLS	IV (instrumental variable)
Davis	1995	no	/	/	
Manzano & Rigobon	2001	no	panel data	OLS	extra exogenous variable for institutions
Stijns	2006	no	cross-section & panel data	correlations	
Brunnschweiler	2008	no	cross-section	OLS & TSLS	IV (instrumental variable)
		Dutch disease?			
Herbertsson, Skuladottir & Zoega	2000	yes	panel data	error correction regression	
Sala-i-Martin & Subramanian	2003	no	/	/	
Égert	2005	yes	time series & panel data	DOLS & ARDL	

Appendix A continued

extra/other data	**no. of countries**	**other findings / main focus**
	100	
geography	100	crowding out is also present
	85	education is reason of resource curse
	85	investment and saving are reason of resource curse
political liberty	85	cut-off point in income
corruption	72	corruption is reason of resource curse
natural resources	99	saving is reason of resource curse
volatility of prices, overvaluation of the exchange rate	71	institutions are reason of resource curse
	47	when controlled for other variables, negative relation disappears
	3	
	40	high level of human capital can offset negative growth
	43	
	79	debt is reason of resource curse
	90	finds best resource and human capital variable
natural resources	84	with other variable for resources no curse
	1	
	71	
	6	exchange rates

Appendix B: Country Sample

A list of all the countries used in the analysis. There are 101 countries in total. Included are the most widely used countries in empirical works and countries with sufficient data.

Argentina	Ghana	Niger
Australia	Greece	Norway
Austria	Guatemala	Pakistan
Bangladesh	Guinea-Bissau	Panama
Belarus	Haiti	Papua New Guinea
Belgium	Honduras	Paraguay
Benin	Hungary	Peru
Bolivia	India	Philippines
Botswana	Indonesia	Poland
Brazil	Ireland	Portugal
Bulgaria	Italy	Romania
Burkina Faso	Jamaica	Russia
Burundi	Japan	Rwanda
Cameroon	Jordan	Senegal
Canada	Kazakhstan	Sierra Leone
Central African Rep.	Kenya	South Africa
Chad	Latvia	South Korea
Chile	Lesotho	Spain
China	Lithuania	Sri Lanka
Colombia	Madagascar	Sweden
Congo	Malawi	Switzerland
Costa Rica	Malaysia	Thailand
Côte D'Ivoire	Mali	Togo
Denmark	Mauritania	Trinidad and Tobago
Dominican Republic	Mauritius	Tunisia
Ecuador	Mexico	Turkey
Egypt	Mongolia	Ukraine
El Salvador	Morocco	United Kingdom
Estonia	Mozambique	United States
Finland	Namibia	Uruguay
France	Nepal	Venezuela
Gambia, The	Netherlands	Zambia
Georgia	New Zealand	Zimbabwe
Germany	Nicaragua	

Appendix C: Variable description and sources

Below are the variable descriptions of all variables included in the model. The variables are both used in the panel and the cross-section model. Sources are also provided.

DEBT Debt/GNP ratio. Total External debt divided by the GNP for the countries for which this ratio is available in World Bank, World Development Indicators (WDI). This is a measure of credit rationing. In the cross-section dataset this variable is the average of the years 1989-2006.

ETH Ethno-linguistic fractionalization index. Taken from Annet, IMF staff paper calculations (2001). It is the probability of two random people in a country not speaking the same language. The variable takes on a value between 1 (highly fractionalized) and 0 (no fractionalization) and is calculated for the year 1982. No other years are available, but this variable is expected to change slowly over time.

FD Financial depth. Measured as liquidity (money and quasi-money: M2) as a percentage share of GDP. For the panel this is a yearly variable, for the cross-section this is the average of the years 1989-2006. Source: World Bank, WDI.

GOV General government final consumption expenditure as a % of GDP. Yearly in the panel, averaged in the cross-section. Source: World Bank, WDI.

INIGDP / LGDP89

ln (Y_{t-1}^{i}), initial per period income. Calculated as the difference in the log of real GDP *per capita* in the previous year (PPP adjusted). In the cross-section this is the log of GDP per capita in 1989. Data are from the World Bank, WDI and are in constant 2000 US $.

INST Data from Freedom House for 1989-2006. The variable is an average of the two measures used in the dataset. Those are political and civil liberty. The variable takes on a value between 1 (very free) to 7 (not free). It is an overall measure of policy and institutions in a country. Yearly in the panel, averaged in the cross-section.

INV Investment. Measured as real gross investment as a percentage share of GDP (in constant 2000 prices) for the period 1989-2004. Yearly in the panel, averaged in the cross-section. Source: Penn World Tables, version 6.2.

LOGDP8906 / LGDPGROWTH8906

Calculated as the log of PPP adjusted annual real GDP growth *per capita,* measured as the difference between GDP per capita in year t and year t-1, for the period 1989-2006, $\ln(Y^i_t) - \ln(Y^i_{t-1})$. For the cross-section the growth from 1989 to 2006 is taken and calculated as $(\ln(Y^i_t) - \ln(Y^i_0)) / N$: LGDPGROWTH8906= (logGDP06 – logGDP89) / 18. Data are from the World Bank, WDI, in constant 2000 US $.

LIB Total life expectancy at birth in years. Source: World Bank, WDI.

OPEN Measure of trade openness of a country. This is calculated as the value of imports and exports relative to GDP for every year or the average of years: (Exports + Imports)/GDP.

Openness is given by the ratio of (nominal) imports plus exports to GDP (in nominal (current) US dollars). Data for imports, exports and GDP come from the World Bank, WDI.

LOGNWPC Log of natural wealth per capita, in US dollars. This is the wealth coming from subsoil assets and other resources, including timber resources, non-timber forest resources, protected areas, cropland and pastureland. The data are available for two years, 1994 and 2000. For the cross-section the average value of these two years is taken. Source: World Bank (1997 & 2006).

LOGSWPC Log of subsoil wealth per capita, in US dollars. This includes energy resources (oil, natural gas, hard coal, and lignite) and other mineral resources (bauxite, copper, gold, iron, lead, nickel, phosphate, silver, tin, and zinc). Subsoil wealth is the present value of a constant stream of economic profits on ''resource rents'' on various fuels and minerals, that is, gross profit on extraction less depreciation of capital and normal return on capital. Data are for 1994 and 2000. For the cross-section the average value of these two years is taken. Source: World Bank (1997 & 2006).

LPOP / LOGPOP89

Population. The annual population, in logs for 1989-2006 for the panel data dataset. For the cross-section, the log of the beginning year 1989 population is taken. Data come from WB, WDI.

REG Regime structure. A variable derived from the Polity IV database indicating whether a country experienced a regime transition or violent change during the period 1989-2006. It codes annual democratic and autocratic "patterns of authority" and regime changes in all countries. This variable takes a value between -10, for very autocratic regimes and 10 for highly democratic regimes. A value of + 6 and higher indicates democracy and- 6 and lower indicates autocracy. Cases of interregnum or anarchy are converted to a "neutral" polity score of 0.

SCH The primary school gross enrolment ratio. Total enrolment in a primary level of education, regardless of age, expressed as a percentage of the eligible official school-age population corresponding to the same level of education in a given school-year. All types of schools are included (both private and public). This is used as a proxy for human capital. Data come from the World Bank, WDI and from UNESCO, institute for statistics.

SFEX Fuel exports as a share of GDP. These include all fuel exports (oil and gas). Data on exports and GDP from World Bank, WDI. Own calculations.

SMEX Mineral exports as a share of GDP. These include fuel and non-fuel mineral exports. Data on exports and GDP come from World Bank, WDI. Own calculations.

SPEX The share of total primary exports in GDP (in nominal dollars). This includes subsoil resources, agricultural and other land resources (food). Data on exports and GDP from World Bank, WDI. Own calculations.

GTT Terms of trade growth: The annual growth in the log (change in log) of external terms of trade between 1989 and 2006, where the terms of trade is given by the ratio of an export price index to an import price index. Calculated as $GTT = \ln(TT_t) - \ln(TT_{t-1})$ in the panel dataset. In the cross-section dataset this is measured as $GTT = \ln(TT_{06}) - \ln(TT_{89})$.

Appendix D: Testing both models

Below, both the panel data model and the cross-sectional model are tested for the most common and important problems that can arise when empirically analyzing the data. Those are multicollinearity, robustness and heteroskedasticity. All test are performed using Eviews and by the instructions and theory in Asteriou and Hall (2006) and Wooldrige (2003).

Section 1: the panel data model

The equation that is worked with initially is the following:

$Loggdp8906^{i} = \alpha_0 + \alpha_1\, inigdp^{i} + \alpha_2\, R^{i} + \alpha_3\, inst^{i} + \alpha_4\, inv^{i} + \alpha_5\, open^{i} + \alpha_6\, gtt^{i} + \alpha_7\, lpop^{i} + \alpha_8\, lib^{i} + \alpha_9\, gov^{i} + \alpha_{10}\, sch^{i} + \alpha_{11}\, fd^{i} + \alpha_{12}\, debt^{i} + \varepsilon^{i}$

In the regression equations, R takes on the value of one of the five different resource measures.

D.1: Testing for multicollinearity

To check for multicollinearity of the model, there are several methods to check for this.

First, the correlation matrix can show this. When there are variables that have correlation coefficients higher than 0.9, this can point to multicollinearity. When looking at the correlation matrix in table D.1, it can be seen that the highest correlation is 0.765, between initial income and natural wealth per capita. This means there are no excessive high or suspect correlations.

Second way to check for multicollinearity between one or more variables, is to make auxiliary regressions; regressing one explanatory variable on all the other individual explanatory variables. When performing this tolerance test, the value of the R^2 is of interest. When the value of R^2 is very high, 0.9 or higher, between two variables, this means the variables together are not good estimates in the model and are multicollineair. Other indicators for this are a small standard errors and significant, very high, t-value in the auxiliary regression. Also signs of the variables can change, so they are not stable.

To check this, the explanatory variables that are expected to have problems are regressed against all combinations of all individual explanatory variables. This means a variable is regressed against one other variable, after which the R^2 is evaluated. For three variables, the regressions turned out to show signs of multicollinearity. The results are displayed below:

R^2 higher than 0.9 for *sch* regressed on:

Fd:

0.950300

Lib:

0.991655

Lpop:

0.997755

R^2 higher than 0.9 for *fd* regressed on:

Lib:

0.963592

Lpop:

0.995441

R^2 higher than 0.9 for *debt* regressed on:

Lib:

0.952244

Lpop:

0.994602

Inigdp:

0.985092

Here, the R^2 is displayed below the variable. After all iterations, it shows that the variables schooling, financial depth and debt show high R^2 values when regressed on the other explanatory variables. Schooling shows a high R^2 value with financial depth, population and life expectancy, but not vice versa. This means schooling is the term that should be considered to be dropped. Debt shows high R^2 values with population, life expectancy and initial GDP, the reason why debt also has to be considered to be left out. The third variable that shows signs of multicollinearity is financial depth, which interacts with life expectancy and population.

Table D.1: Correlation matrix between all variables

	LOGGDP8906	INIGDP	DEBT	FD	GOV	GTT	INST	INV	LIB
LOGGDP8906	1.00	0.377	-0.207	0.353	0.0459	0.0973	-0.277	0.510	0.449
INIGDP	0.377	1.00	-0.518	0.249	0.363	0.153	-0.402	0.539	0.596
DEBT	-0.207	-0.518	1.00	-0.145	-0.274	0.0548	0.175	-0.144	-0.285
FD	0.353	0.249	-0.145	1.00	0.244	-0.214	-0.0427	0.33	0.344
GOV	0.0459	0.363	-0.274	0.244	1.00	-0.0738	-0.174	0.0621	0.0613
GTT	0.0973	0.153	0.0548	-0.214	-0.0738	1.00	0.118	-0.020	0.0115
INST	-0.277	-0.402	0.175	-0.0427	-0.174	0.118	1.00	-0.326	-0.284
INV	0.510	0.539	-0.144	0.33	0.0621	-0.020	-0.326	1.00	0.451
LIB	0.449	0.596	-0.285	0.344	0.0613	0.0115	-0.284	0.451	1.00
SMEX	-0.0353	-0.115	0.428	-0.132	-0.184	0.55	0.138	-0.0741	-0.144
LPOP	0.174	-0.0546	-0.334	0.142	-0.226	0.108	0.127	0.064	0.111
OPEN	0.248	0.130	0.192	0.528	-0.0321	-0.190	-0.099	0.360	0.116
SCH	0.189	0.538	-0.351	0.116	0.333	0.216	-0.348	0.415	0.442
LOGNWPC	0.266	0.765	-0.273	-0.0191	0.0257	0.492	-0.314	0.374	0.523
LOGSWPC	0.403	0.545	-0.352	0.110	-0.0026	0.542	-0.0864	0.417	0.339
SFEX	0.052	0.125	0.0525	-0.0834	-0.198	0.659	0.19	-0.143	0.151
SPEX	-0.331	-0.397	0.714	-0.190	-0.26	0.0039	0.0218	-0.198	-0.171

	SMEX	LPOP	OPEN	SCH	LOGNWPC	LOGSWPC	SFEX	SPEX
LOGGDP8906	-0.0353	0.174	0.248	0.189	0.266	0.403	0.052	-0.331
INIGDP	-0.115	-0.0546	0.130	0.538	0.765	0.545	0.125	-0.397
DEBT	0.428	-0.334	0.192	-0.351	-0.273	-0.352	0.0525	0.714
FD	-0.132	0.142	0.528	0.116	-0.0191	0.11	-0.0834	-0.190
GOV	-0.184	-0.226	-0.0321	0.333	0.0257	-0.0026	-0.198	-0.26
GTT	0.55	0.108	-0.190	0.216	0.492	0.542	0.659	0.0039
INST	0.138	0.127	-0.0991	-0.348	-0.314	-0.0864	0.19	0.0218
INV	-0.0741	0.064	0.360	0.415	0.374	0.417	-0.143	-0.198
LIB	-0.144	0.111	0.116	0.442	0.523	0.339	0.151	-0.171
SMEX	1.00	-0.196	0.173	-0.172	0.30	0.323	0.643	0.537
LPOP	-0.196	1.00	-0.383	0.0895	0.0163	0.332	0.0448	-0.484
OPEN	0.173	-0.383	1.00	-0.146	0.047	-0.066	0.116926	0.362
SCH	-0.172	0.0895	-0.146	1.00	0.459	0.439	-0.0328	-0.313
LOGNWPC	0.30	0.0163	0.047	0.459	1.00	0.678	0.575	-0.0643
LOGSWPC	0.323	0.332	-0.066	0.439	0.678	1.00	0.452	-0.311
SFEX	0.643	0.0448	0.117	-0.0328	0.575	0.452	1.00	0.260
SPEX	0.537	-0.484	0.362	-0.313	-0.0643	-0.311	0.260	1.00

D.2: Testing for robustness

To check for the robustness of variables, it is important to include several variables and to see whether the relationship between natural resources and GDP growth does not change. It is case to find how sensitive the results are to variables that explain growth. In particular, resource abundance could be negatively correlated with many different variables that are used to explain growth. To test this, regressions are run with *loggdp8906* as dependent variable and initial GDP (*inigdp*) and one of the five measures for resources as the explanatory variables. All control variables are then added to see whether the relation changes by inclusion of the specific variable. This means there are 4*11 regressions run.

Some variables in the model appear not to be robust after performing the required regressions. First variable is education. Running the regression of the resource variables and the schooling variable leads to mixed results, where with *lognwpc, logswpc, smex* and *spex* there is a positive relation between schooling and economic growth, where with *sfex* there is a negative relation. Whereas it should be negatively correlated with resource abundance in all cases.
When including life expectancy (*lib*), schooling changes sign, which means schooling is not robust.

Schooling before the addition of *lib*

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INIGDP	-0.089604	0.015788	-5.675489	0.0000
GOV	-0.005250	0.000900	-5.833218	0.0000
GTT	0.008350	0.014937	0.559035	0.5764
INST	0.000861	0.002832	0.304059	0.7612
LPOP	0.159907	0.025025	6.390024	0.0000
SCH	-0.000595	0.000227	-2.620626	0.0091
OPEN	0.054270	0.013179	4.117991	0.0000
SPEX	0.000716	0.000213	3.361680	0.0008
C	-1.869234	0.376439	-4.965574	0.0000

After the addition of *lib*

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INIGDP	-0.170754	0.029392	-5.809613	0.0000
GOV	-0.013685	0.001548	-8.840576	0.0000
GTT	-0.012729	0.023337	-0.545456	0.5859

INST	-0.010785	0.004783	-2.254692	0.0249
LPOP	0.206842	0.059095	3.500138	0.0005
SCH	0.000345	0.000376	0.916838	0.3600
OPEN	0.073899	0.021002	3.518643	0.0005
LIB	0.004677	0.002815	1.661353	0.0978
SPEX	0.000975	0.000393	2.479274	0.0138
C	-2.222837	0.872456	-2.547792	0.0114

There are also robustness problems with financial depth (*fd*). When *sch* or *inv* is included in the model, the sign of financial depth changes into a negative sign, which means that the variable *fd* is not robust.

Fd changes signs after adding *sch* to the regression:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DEBT	-2.44E-05	2.82E-05	-0.865834	0.3869
GOV	-0.001958	0.000643	-3.044243	0.0024
INIGDP	-0.102226	0.016414	-6.227794	0.0000
INV	0.004058	0.000514	7.893562	0.0000
INST	-0.004876	0.002133	-2.286303	0.0225
POP	4.11E-10	1.26E-10	3.256753	0.0012
FD	0.000107	0.000222	0.481500	0.6303
SMEX	0.001757	0.000370	4.750753	0.0000
C	0.685428	0.108283	6.329958	0.0000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DEBT	-3.97E-05	4.73E-05	-0.839112	0.4021
GOV	-0.005437	0.001288	-4.220830	0.0000
INIGDP	-0.048040	0.027695	-1.734597	0.0839
INV	0.003292	0.000943	3.489839	0.0006
SCH	-0.000174	0.000310	-0.561217	0.5751
INST	-0.004798	0.004192	-1.144569	0.2533
POP	4.50E-10	1.91E-10	2.352510	0.0193
FD	-0.000202	0.000350	-0.575248	0.5656
SMEX	0.001830	0.000608	3.008520	0.0029
C	0.392961	0.190017	2.068033	0.0395

Change in *fd* after adding *inv* to the regression

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DEBT	-5.06E-05	2.54E-05	-1.991582	0.0467
GOV	-0.002388	0.000594	-4.018455	0.0001

INIGDP	-0.028515	0.013217	-2.157404	0.0312
INST	-0.004676	0.001937	-2.413599	0.0160
POP	2.54E-10	1.05E-10	2.418866	0.0158
FD	-0.000170	0.000198	-0.861034	0.3895
SMEX	0.001538	0.000258	5.967282	0.0000
C	0.249816	0.088877	2.810794	0.0051

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DEBT	-2.44E-05	2.82E-05	-0.865834	0.3869
GOV	-0.001958	0.000643	-3.044243	0.0024
INIGDP	-0.102226	0.016414	-6.227794	0.0000
INV	0.004058	0.000514	7.893562	0.0000
INST	-0.004876	0.002133	-2.286303	0.0225
POP	4.11E-10	1.26E-10	3.256753	0.0012
FD	0.000107	0.000222	0.481500	0.6303
SMEX	0.001757	0.000370	4.750753	0.0000
C	0.685428	0.108283	6.329958	0.0000

Variables that prove to be robust after the regressions are initial GDP, all four resource variables, population, terms of trade, investment, institutions, openness, life expectancy at birth, debt and government expenditures.

Sch and *fd* are the two variables found not to be robust.
Besides the fact that schooling is not robust to additions of variables to the model and it shows high R-squared values with life expectancy and debt, also the number of observations for this variable is limited. The data for schooling are limited to 5 years from the 18 years covered. Excluding this variable from the model will mean more observations.

After running the regressions to check for multicollinearity and after the robustness checks, it shows that the variables that show problems are schooling, financial depth and debt.
These variables are thus removed from the initial model.

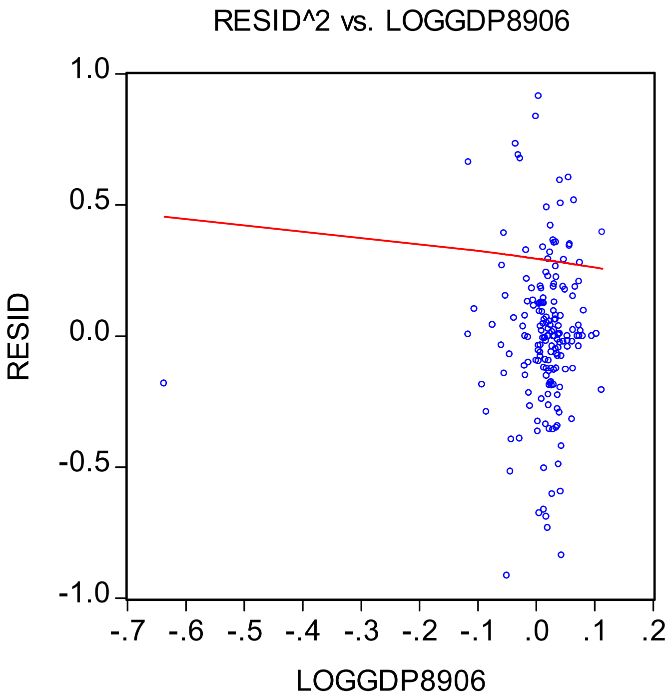
D.3: Testing for heteroskedasticity

When homoskedasticity is present, this means that the variance of the error term is different for values of the explanatory variables. The variables are still unbiased and consistent, but the standard errors, F and t statistics are no longer valid.

To see whether the model suffers from heteroskedasticity, the model should be tested.

An informal way to check this, is to look at the scatter plot of the dependent variable and the square of the residuals. A clear pattern would indicate heteroskedasticity, while no systematic pattern would indicate a healthy model. It can be seen in graph D.2 that there is evidence of a pattern in the residuals, indicating heteroskedasticity.

Graph D.2: scatter plot



To test in a more formal way, the Breusch-Pagan test is performed by hand, since there are no tests available in Eviews for this when using panel data.

1) estimate the equation
2) obtain the residuals of this regression: resid02
3) obtain the square of the residuals: resid02sq
4) then regress all the explanatory variables from the first regression on resid02sq
5) the outcome is shown below:

Dependent Variable: RESID02SQ
Method: Panel Least Squares
Sample (adjusted): 1991 2004
Cross-sections included: 53
Total panel (unbalanced) observations: 137

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INIGDP	-0.007209	0.002513	-2.868906	0.0054
DEBT	1.01E-06	1.22E-05	0.083242	0.9339
FD	6.45E-05	4.50E-05	1.433430	0.1560
GOV	-0.000277	0.000163	-1.706489	0.0922
GTT	-0.001640	0.001927	-0.851192	0.3974
INST	-0.000217	0.000382	-0.569296	0.5709
INV	-0.000156	8.74E-05	-1.783766	0.0786
LIB	1.70E-05	0.000241	0.070460	0.9440
LPOP	0.006771	0.005343	1.267390	0.2090
OPEN	0.000275	0.002486	0.110742	0.9121
SPEX	5.14E-05	3.80E-05	1.354692	0.1797
C	-0.058484	0.083850	-0.697481	0.4877

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.660424	Mean dependent var	0.000733
Adjusted R-squared	0.367366	S.D. dependent var	0.001759
S.E. of regression	0.001399	Akaike info criterion	-10.00113
Sum squared resid	0.000143	Schwarz criterion	-8.637053
Log likelihood	749.0775	F-statistic	2.253559
Durbin-Watson stat	3.094113	Prob(F-statistic)	0.000446

The number of degrees of freedom is 11 (p-1) and this test has a chi-squared distribution. So for a significance of 5 percent, the chi-square critical value is 19.68.

When the LM statistic is larger than the critical value, the null hypothesis is rejected. H0= no heteroskedasticity, against H1: heteroskedasticity, so there is heteroskedasticity.

The LM statistic is calculated as:

$LM = obs * R^2$

and is thus equal to: 137*0.660424= 90.4781

This means 90.4781 > 19.68 and we should reject H0, which means there is evidence of heteroskedasticity in the model. Also the p-value is less than the level of significance 0.05, which leads to the same conclusion.

Since there is evidence of heteroskedasticity, the model must be estimated using generalized (weighted) least squares (GLS), which recognizes the presence of heteroskedasticity. This results in more efficient estimates than OLS.

D.4: Testing for serial correlation

When serial correlation (or autocorrelation) is present in the model, this means there is correlation between the error terms in different time periods in the panel data model.

The variables are still unbiased, but are no longer BLUE.

For testing the presence of serial correlation, the hypotheses are the following:

H0: The errors are serially uncorrelated

H1: The errors are serially correlated

The test is performed as follows:

1) make a regression as the one above
2) regress the residuals from this regression on the explanatory variables and the lagged value of the residuals for p periods (the number of lags)
3) calculate the LM statistic as (n-p)*R2 = 78-4=74*0.489766= 36.24268

Dependent Variable: RESID02
Method: Panel Least Squares
Sample (adjusted): 1994 2004
Cross-sections included: 8
Total panel (unbalanced) observations: 78

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INIGDP	-0.018536	0.118206	-0.156812	0.8760
DEBT	-5.52E-05	0.000615	-0.089672	0.9289
FD	-0.003774	0.001765	-2.138263	0.0370
GOV	-0.006561	0.004291	-1.529157	0.1320
GTT	-0.108502	0.054926	-1.975406	0.0532
INST	0.014744	0.010112	1.458034	0.1505
INV	0.008471	0.002119	3.997394	0.0002
LIB	0.033383	0.019335	1.726559	0.0899
LPOP	0.033540	0.242377	0.138378	0.8904
OPEN	-0.058023	0.061608	-0.941806	0.3504
SPEX	0.002098	0.001770	1.185301	0.2410
RESID02(-1)	-0.235843	0.194849	-1.210386	0.2313
RESID02(-2)	-0.260033	0.200049	-1.299847	0.1991
RESID02(-3)	-0.178711	0.179472	-0.995762	0.3237
RESID02(-4)	-0.188265	0.162647	-1.157505	0.2521

C	-2.739059	3.579334	-0.765243	0.4474
	Effects Specification			
Cross-section fixed (dummy variables)				

R-squared	0.489766	Mean dependent var	6.94E-05
Adjusted R-squared	0.285672	S.D. dependent var	0.037425
S.E. of regression	0.031631	Akaike info criterion	-3.829004
Sum squared resid	0.055028	Schwarz criterion	-3.134077
Log likelihood	172.3311	F-statistic	2.399713
Durbin-Watson stat	1.459091	Prob(F-statistic)	0.004518

The chi-distribution critical value is 25, which is smaller than the LM statistic of 36.24. This means we reject H0 that there is no serial correlation. This means that serial correlation is present.

Another way to test for serial correlation is by the Durbin-Watson statistic.
This means that for 0 serial correlation, the DW statistic should be around 2. Is it lower than 2, than the null-hypothesis should be rejected. So DW < 2.
When running the regression that is used in all other tests as well, the DW statistic is
Durbin-Watson stat 1.734558
This means 1.74 < 2, so reject H0 of no serial correlation.
When there are a lot of observations (>50), a value of 1.5 or lower points to serious serial correlation. This means in this case it is present, but not in the worst way.

As could be seen in both the serial correlation and the heteroskedasticity tests, both are present in the model. Therefore we have to correct for this. This can be done in Eviews. To correct for both, the method of Generalized (weighted) least squares is used. The appropriate GLS estimation method when both are present in the model is the seemingly unrelated regression (period SUR) method, also known as the multivariate regression, since this accounts for heteroskedasticity and correlation in the error terms.[20]

[20] See Eviews user's guide (2004) for further information on this topic.

D.5: Unit root test

Because of the time periods in the dataset, all variables in the model must be stationary and therefore do not have unit root. To make sure all variables are specified as stationary processes, tests on the variables are performed. Variables that are suspected to follow a unit root process are population, human capital, debt, government expenditures and life expectancy. To test for unit root the following steps can be made:

1) Run a regression with the suspect variable as explanatory variable and the difference of this variable as the dependent variable. Some other set of variables can be included (*Z*) to check the results. The regression will look as follows:

$$\Delta y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_3 Z + \varepsilon$$

2) Test the null hypothesis H0: unit root, against H1: stationarity. If the p-value of the α_1 coefficient is not significant (thus high), H0 is not rejected and there is unit root in the variable.

3) If unit root is found, the difference in the variable has to be used as a variable, since this variable is stationary in the model.

The variables have to be tested per country, since only than the unit-root process can be detected. The test is available in Eviews.

When unit-root is present in a variable, the difference should be taken. However, this means less data-points, since the first observations is always dropped. Then there are also some variables with some missing data in between the years. These also cause extra observation points to be lost in the differences. The problem with this is, that among the 101 countries, not all might have unit-root in the specific variables. This gives us another problem in the panel-data model, since the choice between differencing in the case of non-stationarity is not that clear in this case.

Below is one of the unit root tests of all tests that were performed for all countries in the panel. Variable used here is population, since this is expected to have unit-root.

In a lot of the cases, population does have unit-root. However, this is not always the case.

Below are the unit-root tests for two random countries: France and China. As can be seen, according to the test, the population in France has unit-root, as was expected. However, the population variable for China is stationary and the unit-root hypothesis is rejected.

Unit-root test for France, population

Null Hypothesis: POPFRA has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic based on SIC, MAXLAG=3)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		1.677722	0.9989
Test critical values:	1% level	-3.920350	
	5% level	-3.065585	
	10% level	-2.673459	

*MacKinnon (1996) one-sided p-values.
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 16

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(POPFRA)
Method: Least Squares
Sample (adjusted): 3 18
Included observations: 16 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
POPFRA(-1)	0.030964	0.018456	1.677722	0.1173
D(POPFRA(-1))	0.524447	0.227044	2.309889	0.0380
C	-1675070.	1040361.	-1.610086	0.1314

R-squared	0.638556	Mean dependent var	282600.0
Adjusted R-squared	0.582949	S.D. dependent var	105787.4
S.E. of regression	68316.94	Akaike info criterion	25.26906
Sum squared resid	6.07E+10	Schwarz criterion	25.41392
Log likelihood	-199.1525	F-statistic	11.48343
Durbin-Watson stat	1.799553	Prob(F-statistic)	0.001340

Unit-root test for China, population

Null Hypothesis: POPCHIN has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic based on SIC, MAXLAG=3)

		t-Statistic	Prob.*

Augmented Dickey-Fuller test statistic		-17.21908	0.0000
Test critical values:	1% level	-3.886751	
	5% level	-3.052169	
	10% level	-2.666593	

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(POPCHIN)
Method: Least Squares
Sample (adjusted): 2 18
Included observations: 17 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
POPCHIN(-1)	-0.048362	0.002809	-17.21908	0.0000
C	70511572	3438878.	20.50424	0.0000

R-squared	0.951845	Mean dependent var	11361629
Adjusted R-squared	0.948635	S.D. dependent var	2916275.
S.E. of regression	660939.7	Akaike info criterion	29.75084
Sum squared resid	6.55E+12	Schwarz criterion	29.84887
Log likelihood	-250.8822	F-statistic	296.4966
Durbin-Watson stat	1.332501	Prob(F-statistic)	0.000000

It can be expected that other countries and other variables also have this problem, of not all being non-stationary. This makes the choice of the variables (differences or not) in the model even more difficult. It can even be expected that the dependent variable GDP growth is also non-stationary. This gives even more problems in analyzing this type of model.
This is a large shortcoming of the panel-dataset.

Section 2: Cross-sectional data

In this section the same analysis as in the previous section is performed, only this time for the cross-section dataset and model. The equation that is worked with here initially is also the following:

$$Lgdpgrowth8906^{i} = \gamma_0 + \gamma_1\, lgdp89^{i} + \gamma_2\, debt^{i} + \gamma_3 fd^{i} + \gamma_4\, gov^{i} + \gamma_5\, inst^{i} + \gamma_6\, inv^{i} + \gamma_7\, lib^{i} + \gamma_8\, logpop89^{i} + \gamma_9\, open^{i} + \gamma_{10}\, sch^{i} + \gamma_{11}\, R^{i} + \gamma_{12}\, gtt^{i} + \upsilon^{i}$$

R^{i} takes on the value of one of the five different resource measures.
Again is tested for the most common problems in the regression model and outcomes.

D.6: Testing for multicollinearity

We proceed in the same way as described in E.1.
In table D.3, the correlation matrix is presented. It can be seen that there are no high correlations between the different variables. The highest correlation exists between initial GDP and life expectancy, with a coefficient of 0.658. This suggests no multicollinearity in the model.

To be sure that no multicollinearity exists, again auxiliary regressions are run. Here regressions are run of one explanatory variable on another, checking for high R^2, small standard errors and high t-values.
After running all the auxiliary regressions, none of them show signs of multicollinearity. No R^2 is extremely high.

D.7: Testing for robustness

Just as in the panel model, we also test for robustness of the cross-sectional model. For this, a regression is run with GDP growth as dependent variable and initial GDP and a resource variable as explanatory variables. After this, step by step the other explanatory variables are included in the regression. If the relationship does not change after including the variables, the model is robust. After running the regressions, the relationship stays robust after including the other variables.

After testing for multicollinearity and robustness, it can be concluded that the cross-sectional model is much less problematic than the panel-data model in both respects. There are no signs of multicollinearity and the model is robust this time.

D.8: Testing for heteroskedasticity

When heteroskedasticity is present consistent covariance estimator should be used in the form of White estimators. In this case, the standard errors are again valid. To test for heteroskedasticity, a White test is performed. White's test is a test of the null hypothesis of no heteroskedasticity against the alternative hypothesis that heteroskedasticity of some unknown general form is present. In table D.4 the results of the test are presented.

Table D.4: Results from White heteroskedasticity test

White Heteroskedasticity Test:

F-statistic	1.750105	Probability	0.077779
Obs*R-squared	31.80076	Probability	0.132011

The test shows a p-value of 0.078. This means the null-hypothesis is not rejected at the 5% and 1% level, but is rejected at the 10% level. Since there is still some small evidence of heteroskedasticity, this has to be corrected for.

Two things can be done to correct for heteroskedasticity. First is OLS, but with adjusted standard errors to get heteroskedastic-consitent standard errors. The most common way to do this is by using the White standard errors. This is only justified for large samples.

D.9: Testing for outliers

The existence of outliers in the model can be of influence on the regression results. In order to see if this is the case, the outliers were detected in the dataset by constructing scatter plots of the different variables and normal line graphs to see if there were abnormal values in the dataset. Some outliers were detected, but only one outlier changes regression results significantly. This was financial debt for Japan, which was on average four times higher than the other values in the dataset. This one data point was therefore removed from the dataset.

Table D.3: correlation matrix of the cross-sectional variables

	LGDP89	DEBT	FD	GOV	GTT	INST	INV	LIB
LGDP89	1.00	-0.254	0.0577	0.0624	0.159	-0.429	0.381	0.658
DEBT	-0.254	1.00	-0.159	0.167	-0.0306	-0.0821	-0.221	-0.202
FD	0.0577	-0.159	1.00	0.195	-0.239	0.238	0.56	0.338
GOV	0.0624	0.167	0.195	1.00	0.0114	0.071	0.0753	-0.164
GTT	0.159	-0.0306	-0.239	0.0114	1.00	-0.0865	-0.267	0.06
INST	-0.429	-0.0821	0.238	0.0711	-0.0865	1.00	-0.151	-0.406
INV	0.381	-0.221	0.56	0.0753	-0.267	-0.151	1.00	0.472
LIB	0.658	-0.202	0.338	-0.164	0.06	-0.406	0.472	1.00
LOGPOP89	-0.212	-0.265	0.287	-0.222	-0.0683	0.345	0.221	0.0574
OPEN	0.045	0.159	0.39	0.164	-0.382	0.0428	0.294	0.106
SCH	0.47	-0.201	0.143	0.0513	-0.191	-0.155	0.465	0.502
LOGNWPC	0.676	-0.157	-0.155	-0.190	0.243	-0.299	0.321	0.474
LOGSWPC	0.507	-0.288	0.107	-0.0484	0.248	-0.104	0.387	0.372
SPEX	-0.285	0.603	-0.209	0.105	-0.0748	0.0984	-0.264	-0.291
SMEX	0.00199	0.218	-0.081	0.141	0.171	0.199	-0.0964	-0.192
SFEX	0.231	-0.0795	-0.0294	-0.205	0.316	0.173	-0.0768	0.138

	LOGPOP89	OPEN	SCH	LOGNWPC	LOGSWPC	SPEX	SMEX	SFEX
LGDP89	-0.212	0.045	0.47	0.676	0.507	-0.285	0.00199	0.231
DEBT	-0.265	0.159	-0.201	-0.157	-0.288	0.603	0.218	-0.0795
FD	0.287	0.39	0.143	-0.155	0.107	-0.209	-0.081	-0.0294
GOV	-0.222	0.164	0.0513	-0.190	-0.048	0.105	0.141	-0.205
GTT	-0.0683	-0.382	-0.191	0.243	0.248	-0.0748	0.171	0.316
INST	0.345	0.0428	-0.155	-0.299	-0.104	0.0984	0.199	0.173
INV	0.221	0.294	0.465	0.321	0.387	-0.264	-0.0964	-0.0768
LIB	0.0574	0.106	0.502	0.474	0.372	-0.291	-0.192	0.138
LOGPOP89	1.00	-0.306	0.250	0.0338	0.128	-0.373	-0.162	0.0656
OPEN	-0.306	1.00	-0.187	0.0186	-0.00867	0.428	0.197	0.0367
SCH	0.250	-0.187	1.00	0.381	0.361	-0.373	-0.142	0.00065
LOGNWPC	0.0338	0.0186	0.381	1.00	0.647	-0.00595	0.336	0.584
LOGSWPC	0.128	-0.0087	0.361	0.647	1.00	-0.185	0.444	0.472
SPEX	-0.373	0.428	-0.373	-0.00595	-0.185	1.00	0.54	0.163
SMEX	-0.162	0.197	-0.142	0.336	0.444	0.54	1.00	0.506
SFEX	0.0656	0.0367	0.00065	0.584	0.472	0.163	0.506	1.00

D.10: Testing joint significance

To make sure the variables are jointly significant, a Wald test is performed. In this the hypothesis is tested that the variables are jointly = 0, against H1: ≠ 0. We reject the null-hypothesis, since the p-value is 0.0001. Results are displayed in table D.5.

Table D.5: Wald test on the joint significance

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	4.844673	(12, 40)	0.0001
Chi-square	58.13607	12	0.0000

D.11: Sample size and variables

There is one more problem that arises when using the cross-section data. Although there are 101 countries in the sample, when all variables are used, only 53 observations remain. This is a relatively small sample and decisions have to be made in order to improve this.

If a variable with few observations is deleted from the model, this means a variable that also has influence on GDP growth is removed, but more observations remain. More observations are better for the explanatory power of the model.

Variables that have this problem are debt (with 76 obs.), financial depth (with 89 obs.) and terms of trade (with 86 obs.).

To see whether there are major changes when one of these variables is removed, we look at the different regressions. Below is the original model with all variables included. The N is 53 here. The R^2 is 0.59.

Dependent Variable: LGDPGROWTH8906
Method: Least Squares
Sample: 1 101
Included observations: 53

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002357	0.001468	-1.605249	0.1163
DEBT	-1.30E-05	1.31E-05	-0.993471	0.3265
FD	0.000103	4.50E-05	2.287216	0.0276
GOV	-6.61E-05	0.000208	-0.317709	0.7524
GTT	0.001549	0.004536	0.341401	0.7346

INST	-0.000682	0.000689	-0.990753	0.3278
INV	0.000253	0.000171	1.477798	0.1473
LIB	0.000343	0.000156	2.196307	0.0339
LOGNWPC	-0.000732	0.001393	-0.525712	0.6020
LOGPOP89	0.001167	0.000676	1.725651	0.0921
OPEN	-0.001216	0.002598	-0.468139	0.6422
SCH	-3.13E-05	5.60E-05	-0.558895	0.5793
C	-0.010471	0.014889	-0.703272	0.4860

R-squared	0.592403	Mean dependent var	0.007333
Adjusted R-squared	0.470123	S.D. dependent var	0.007179
S.E. of regression	0.005226	Akaike info criterion	-7.461294
Sum squared resid	0.001092	Schwarz criterion	-6.978015
Log likelihood	210.7243	F-statistic	4.844673
Durbin-Watson stat	2.819710	Prob(F-statistic)	0.000074

Regression without debt variable

Dependent Variable: LGDPGROWTH8906
Method: Least Squares
Sample: 1 101
Included observations: 63

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002337	0.001256	-1.860581	0.0686
FD	7.05E-05	4.01E-05	1.758124	0.0847
GOV	-6.55E-05	0.000169	-0.387112	0.7003
GTT	0.002497	0.004535	0.550646	0.5843
INST	-1.90E-05	0.000609	-0.031124	0.9753
INV	0.000328	0.000144	2.277929	0.0270
LIB	0.000370	0.000154	2.400769	0.0200
LOGNWPC	-0.000614	0.001151	-0.533153	0.5962
LOGPOP89	0.000703	0.000625	1.123419	0.2665
OPEN	-0.000668	0.002543	-0.262654	0.7939
SCH	-1.14E-05	5.36E-05	-0.211668	0.8332
C	-0.011001	0.013111	-0.839057	0.4054

R-squared	0.488401	Mean dependent var	0.007719
Adjusted R-squared	0.378056	S.D. dependent var	0.006906
S.E. of regression	0.005447	Akaike info criterion	-7.417975
Sum squared resid	0.001513	Schwarz criterion	-7.009759
Log likelihood	245.6662	F-statistic	4.426124
Durbin-Watson stat	2.552870	Prob(F-statistic)	0.000117

This table shows that the R^2 decreases, but the significance of the separate variables does not improve much. Therefore dropping debt is not a good option.

Regression without financial depth

Dependent Variable: LGDPGROWTH8906
Method: Least Squares
Sample: 1 101
Included observations: 53

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002284	0.001542	-1.481290	0.1462
DEBT	-1.44E-05	1.38E-05	-1.044696	0.3023
GOV	7.60E-06	0.000216	0.035199	0.9721
GTT	0.001346	0.004763	0.282511	0.7790
INST	-0.000149	0.000681	-0.218699	0.8280
INV	0.000347	0.000174	1.994737	0.0527
LIB	0.000485	0.000151	3.221567	0.0025
LOGNWPC	-0.001737	0.001389	-1.250664	0.2181
LOGPOP89	0.001446	0.000698	2.070369	0.0448
OPEN	0.000308	0.002637	0.116659	0.9077
SCH	-3.97E-05	5.87E-05	-0.675403	0.5032
C	-0.016638	0.015380	-1.081840	0.2856

R-squared	0.539095	Mean dependent var	0.007333
Adjusted R-squared	0.415438	S.D. dependent var	0.007179
S.E. of regression	0.005489	Akaike info criterion	-7.376119
Sum squared resid	0.001235	Schwarz criterion	-6.930015
Log likelihood	207.4671	F-statistic	4.359593
Durbin-Watson stat	2.809903	Prob(F-statistic)	0.000254

The R^2 also decreases in this regression and the number of observations stays the same. The variables do not change much in significance either. Dropping financial depth is also not an option.

Without terms of trade change

Dependent Variable: LGDPGROWTH8906
Method: Least Squares
Sample: 1 101
Included observations: 62

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002423	0.001311	-1.848570	0.0704
DEBT	-1.86E-05	1.14E-05	-1.630169	0.1094
FD	0.000128	4.10E-05	3.118101	0.0030
GOV	-6.84E-05	0.000193	-0.354919	0.7241
INST	-0.000862	0.000627	-1.375580	0.1751
INV	0.000284	0.000133	2.139275	0.0373

LIB	0.000252	0.000124	2.024833	0.0482
LOGNWPC	-0.000465	0.001221	-0.380791	0.7050
LOGPOP89	0.001158	0.000596	1.941748	0.0578
OPEN	-0.001274	0.002351	-0.541824	0.5903
SCH	-2.28E-05	5.21E-05	-0.437563	0.6636
C	-0.007644	0.013032	-0.586569	0.5601

R-squared	0.590056	Mean dependent var	0.006314
Adjusted R-squared	0.499869	S.D. dependent var	0.007490
S.E. of regression	0.005297	Akaike info criterion	-7.471263
Sum squared resid	0.001403	Schwarz criterion	-7.059559
Log likelihood	243.6091	F-statistic	6.542546
Durbin-Watson stat	2.107660	Prob(F-statistic)	0.000001

Here, the R^2 remains almost the same as in the original model. Also the significance of 7 variables in the model increases. Therefore the variable *gtt* is candidate for removing from the model. It also leaves 62 observations, 9 more than in the old model.

Removing more than one variable from the model also does not improve its explanatory power. Dropping two variables together also does not necessarily mean more observations (such as with removing terms of trade and financial depth). This would also compromise the explanatory power of the model.

Appendix E: Regression Outputs

In this appendix all regressions and other calculations are presented.

Section I: panel-data analysis

E.1: Correlation matrix for the measures of natural resources

As can be seen from the matrix in E.1, all resources are positively correlated with each other, as is expected. *Smex* has the highest correlation with *sfex* and *spex*, as is also expected. Also *logswpc* and *lognwpc* are correlated the highest among each other.

Table E.1: Correlation matrix

	SPEX	SMEX	SFEX	LOGNWPC	LOGSWPC
SPEX	1.00	0.715	0.555	0.0397	0.026
SMEX	0.715	1.00	0.876	0.234	0.375
SFEX	0.555	0.876	1.00	0.362	0.444
LOGNWPC	0.0397	0.234	0.362	1.00	0.702
LOGSWPC	0.026	0.375	0.444	0.702	1.00

E.2: Descriptive statistics of the variables in the model

Below in table E.2, the descriptive statistics of the variables are shown. Included are mean, median, standard deviation, the minimum and maximum value of the variables and the number of observations per variable.

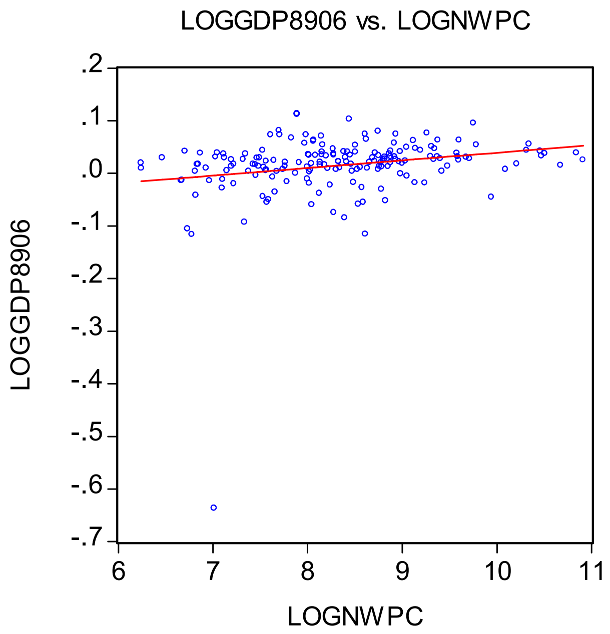
As can be seen, schooling (*sch*) has the least amount of observations (apart from the resource measure variables). This is unbeneficial to the model.

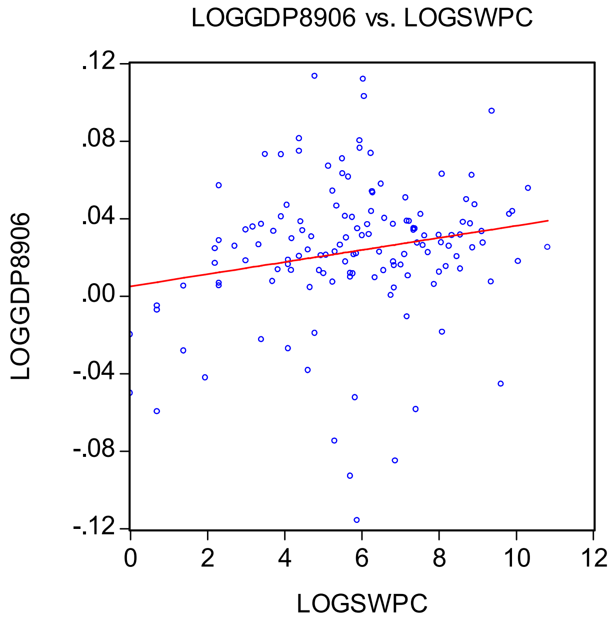
Table E.2: Descriptive statistics for all variables

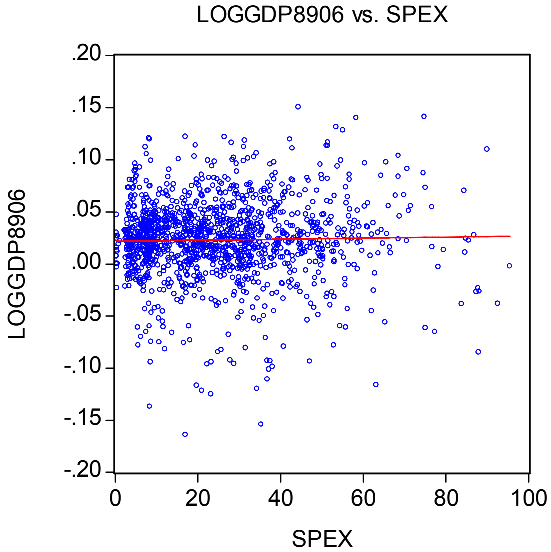
	DEBT	ETH	FD	GOV	GTT	INIGDP	INST	INV	LIB
Mean	78.814	0.502	41.149	14.97	0.00092	7.538	3.206	14.7	68.164
Median	58.669	0.56	31.165	13.897	-0.0021	7.422	3.00	13.224	70.86
Maximum	1209.303	0.91	242.239	43.479	0.701	10.612	7.00	44.135	82.322
Minimum	0.144	0.00	5.844	2.90	-0.741	4.610	1.00	0.15	23.641
Std. Dev.	85.283	0.273	32.328	5.456	0.095	1.605	1.760	7.808	10.383
Observations	1333	1818	1529	1789	1372	1716	1817	1536	1229

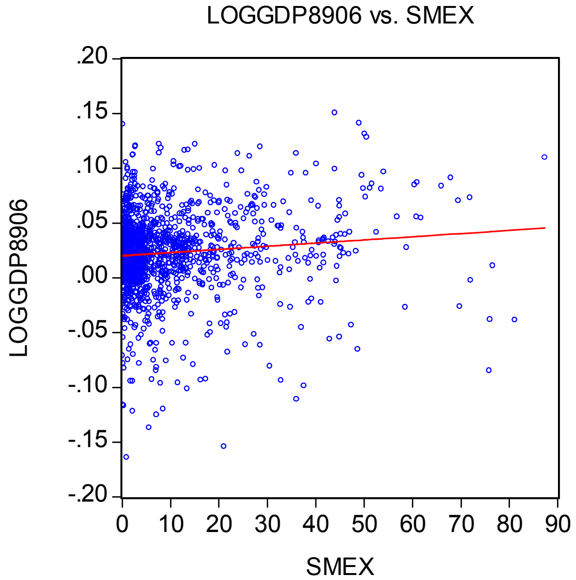
LOGGDP8906	LOGNWPC	LOGSWPC	LPOP	OPEN	SCH	SFEX	SMEX	SPEX	
0.0152	8.322	5.804	16.391	0.735	99.401	5.478	9.85	24.881	Mean
0.0211	8.292	5.858	16.143	0.657	101.864	1.414	4.43	21.527	Median
0.319	10.912	10.817	20.995	3.505	154.441	86.975	87.337	95.486	Maximum
-0.637	6.242	0.00	13.739	0.107	27.814	-0.0094	0.0133	0.351	Minimum
0.0532	0.962	2.301	1.418	0.395	17.0938	9.97	12.925	17.635	Std. Dev.
1715	180	137	1818	1610	840	1426	1420	1420	Observations

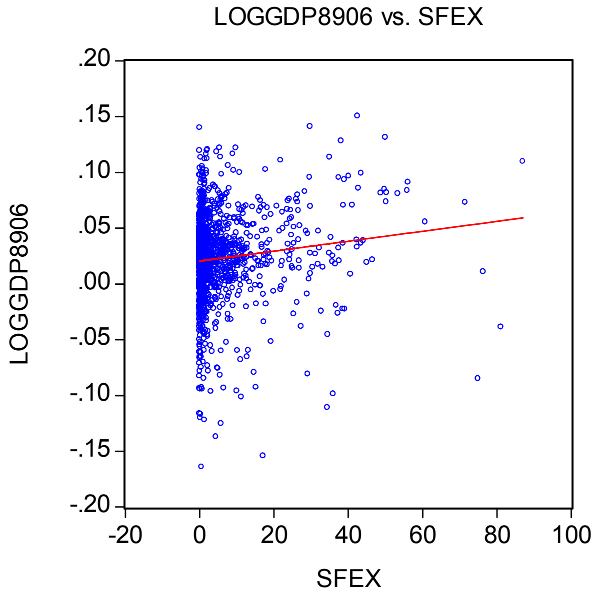
E.3: Scatter plots of the growth variable (loggdp8906) versus natural resources






LOGGDP8906 vs. SPEX
LOGGDP8906
SPEX
.20
.15
.10
.05
.00
-.05
-.10
-.15
-.20
0
20
40
60
80
100


LOGGDP8906 vs. SMEX
LOGGDP8906
SMEX
.20
.15
.10
.05
.00
-.05
-.10
-.15
-.20
0
10
20
30
40
50
60
70
80
90


LOGGDP8906 vs. SFEX
LOGGDP8906
SFEX
.20
.15
.10
.05
.00
-.05
-.10
-.15
-.20
-20
0
20
40
60
80
100

E.4: Fixed effects estimation (GLS SUR estimations)

The regression in the following outputs is the following:

$Loggdp8906^{i} = \alpha_0 + \alpha_1\ inigdp^{i} + \alpha_2\ R^{i} + \alpha_3\ inst^{i} + \alpha_4\ inv^{i} + \alpha_5\ open^{i} + \alpha_6\ gtt^{i} + \alpha_7\ lpop^{i} + \alpha_8\ lib^{i} + \alpha_9\ gov^{i} + \varepsilon^{i}$

Regression output with GDP growth and *spex*

Dependent Variable: LOGGDP8906
Method: Panel EGLS (Period SUR)
Sample (adjusted): 1990 2004
Cross-sections included: 77
Total panel (unbalanced) observations: 649
Linear estimation after one-step weighting matrix
Period SUR (PCSE) standard errors & covariance (d.f. corrected)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INIGDP	-0.139735	0.016906	-8.265294	0.0000
GOV	-0.002676	0.000671	-3.990110	0.0001
GTT	0.003488	0.013418	0.259944	0.7950
INST	0.006210	0.002718	2.284921	0.0227
INV	0.004622	0.000512	9.025846	0.0000
LIB	0.004712	0.001148	4.106004	0.0000
LPOP	0.048223	0.028106	1.715783	0.0868
OPEN	0.067326	0.012866	5.232680	0.0000
SPEX	0.001958	0.000328	5.965339	0.0000
C	-0.100507	0.432512	-0.232380	0.8163

Effects Specification

Cross-section fixed (dummy variables)

Weighted Statistics

R-squared	0.788804	Mean dependent var	1.215102
Adjusted R-squared	0.756919	S.D. dependent var	1.919143
S.E. of regression	0.946200	Sum squared resid	504.0512
F-statistic	24.73851	Durbin-Watson stat	1.666690
Prob(F-statistic)	0.000000		

Unweighted Statistics

R-squared	0.378027	Mean dependent var	0.019742
Sum squared resid	0.487011	Durbin-Watson stat	1.368339

Regression output with GDP growth and *smex*

Dependent Variable: LOGGDP8906
Method: Panel EGLS (Period SUR)
Sample (adjusted): 1990 2004
Cross-sections included: 77
Total panel (unbalanced) observations: 649
Linear estimation after one-step weighting matrix
Period SUR (PCSE) standard errors & covariance (d.f. corrected)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INIGDP	-0.148021	0.017149	-8.631414	0.0000
GOV	-0.001715	0.000671	-2.554851	0.0109
GTT	0.005816	0.013474	0.431697	0.6661
INST	0.006366	0.002810	2.265689	0.0238
INV	0.004747	0.000520	9.131929	0.0000
LIB	0.004643	0.001134	4.093367	0.0000
LPOP	0.039679	0.028423	1.396038	0.1633
OPEN	0.061954	0.012812	4.835697	0.0000
SMEX	0.002571	0.000565	4.549308	0.0000
C	0.125230	0.435193	0.287757	0.7736

Effects Specification

Cross-section fixed (dummy variables)

Weighted Statistics

R-squared	0.779763	Mean dependent var	1.215021
Adjusted R-squared	0.746512	S.D. dependent var	1.909110
S.E. of regression	0.961191	Sum squared resid	520.1489
F-statistic	23.45099	Durbin-Watson stat	1.621859
Prob(F-statistic)	0.000000		

Unweighted Statistics

R-squared	0.376157	Mean dependent var	0.019742
Sum squared resid	0.488475	Durbin-Watson stat	1.325647

Regression output with GDP growth and *sfex*

Dependent Variable: LOGGDP8906
Method: Panel EGLS (Period SUR)
Sample (adjusted): 1990 2004
Cross-sections included: 77
Total panel (unbalanced) observations: 653
Linear estimation after one-step weighting matrix
Period SUR (PCSE) standard errors & covariance (d.f. corrected)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INIGDP	-0.142614	0.016810	-8.483998	0.0000
GOV	-0.001472	0.000660	-2.229293	0.0262
GTT	0.003735	0.013544	0.275747	0.7828
INST	0.006470	0.002735	2.365502	0.0183
INV	0.004814	0.000522	9.218134	0.0000
LIB	0.004172	0.001101	3.789706	0.0002
LPOP	0.031399	0.027558	1.139380	0.2550
OPEN	0.061618	0.012680	4.859490	0.0000
SFEX	0.003256	0.000717	4.539215	0.0000
C	0.249926	0.422489	0.591555	0.5544

Effects Specification

Cross-section fixed (dummy variables)

Weighted Statistics

R-squared	0.777937	Mean dependent var	1.204312
Adjusted R-squared	0.744647	S.D. dependent var	1.884078
S.E. of regression	0.952070	Sum squared resid	513.9501
F-statistic	23.36861	Durbin-Watson stat	1.630020
Prob(F-statistic)	0.000000		

Unweighted Statistics

R-squared	0.383185	Mean dependent var	0.019656
Sum squared resid	0.487610	Durbin-Watson stat	1.342138

Regression output with GDP growth and *lognwpc*

Dependent Variable: LOGGDP8906
Method: Panel EGLS (Period SUR)
Sample (adjusted): 1994 2000
Cross-sections included: 74
Total panel (unbalanced) observations: 141
Linear estimation after one-step weighting matrix
Period SUR (PCSE) standard errors & covariance (d.f. corrected)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INIGDP	-0.275832	0.101612	-2.714561	0.0087
GOV	-0.004770	0.003520	-1.355306	0.1806
GTT	0.046459	0.076758	0.605267	0.5474
INST	0.006037	0.014519	0.415798	0.6791
INV	0.004766	0.002609	1.827078	0.0728
LIB	0.011240	0.004615	2.435523	0.0180
LPOP	0.486884	0.130050	3.743832	0.0004

OPEN	0.011142	0.077937	0.142957	0.8868
LOGNWPC	0.031948	0.027752	1.151172	0.2544
C	-7.026091	2.199807	-3.193958	0.0023

Effects Specification

Cross-section fixed (dummy variables)

Weighted Statistics

R-squared	0.993339	Mean dependent var	1.945093
Adjusted R-squared	0.983921	S.D. dependent var	8.847587
S.E. of regression	1.121883	Sum squared resid	73.00007
F-statistic	105.4790	Durbin-Watson stat	0.142023
Prob(F-statistic)	0.000000		

Unweighted Statistics

R-squared	0.796129	Mean dependent var	0.014747
Sum squared resid	0.125173	Durbin-Watson stat	1.036765

Regression output with GDP growth and *logswpc*

Dependent Variable: LOGGDP8906
Method: Panel EGLS (Period SUR)
Sample (adjusted): 1994 2000
Cross-sections included: 61
Total panel (unbalanced) observations: 117
Linear estimation after one-step weighting matrix
Period SUR (PCSE) standard errors & covariance (d.f. corrected)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INIGDP	-0.141781	0.077641	-1.826102	0.0742
GOV	0.001288	0.003285	0.392253	0.6966
GTT	0.029193	0.062970	0.463600	0.6451
INST	0.004150	0.011550	0.359355	0.7209
INV	0.003468	0.001925	1.801149	0.0781
LIB	0.003298	0.003717	0.887361	0.3794
LPOP	0.112862	0.110093	1.025151	0.3105
OPEN	0.071767	0.064995	1.104182	0.2751
LOGSWPC	-0.002700	0.007830	-0.344772	0.7318
C	-1.103247	1.705823	-0.646753	0.5209

Effects Specification

Cross-section fixed (dummy variables)

Weighted Statistics

R-squared	0.989532	Mean dependent var	4.394623
Adjusted R-squared	0.974164	S.D. dependent var	7.087682
S.E. of regression	1.139242	Sum squared resid	61.00001
F-statistic	64.38952	Durbin-Watson stat	0.168248
Prob(F-statistic)	0.000000		

	Unweighted Statistics		
R-squared	0.676844	Mean dependent var	0.023185
Sum squared resid	0.047994	Durbin-Watson stat	1.026316

E.5: Wald coefficient restrictions test

Also the joint significance of the variables in the model is tested. Therefore a Wald test is used.

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	26.70631	(8, 780)	0.0000
Chi-square	213.6505	8	0.0000

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(1)	-0.118877	0.016491
C(2)	-0.002719	0.000653
C(3)	0.002868	0.002566
C(4)	0.004861	0.000517
C(5)	0.005473	0.001200
C(6)	0.024100	0.028915
C(7)	0.066385	0.012905
C(8)	0.002179	0.000297

Restrictions are linear in coefficients.

This means that we reject H0: β1= β2= β3= β4= β5= β6= β7= β8= 0 against H1: β1= β2= β3= β4= β5= β6= β7= β8≠ 0 which means that the variables are jointly significant at the 10%, 5% and 1% level (as can be seen in the table).

E.6: The Hausman test for endogeneity

To check for endogeneity in the model, the Hausman test can be used. This test looks at the difference between two different estimates, by using instrumental variables.

To perform this test the following steps are followed:

1) estimate the normal regression:

$$Loggdp8906^{i} = \alpha_0 + \alpha_1\, inigdp^{i} + \alpha_2\, spex^{i} + \alpha_3\, inst^{i} + \alpha_4\, inv^{i} + \alpha_5\, open^{i} + \alpha_6\, gtt^{i} + \alpha_7\, lpop^{i} + \alpha_8\, lib^{i} + \alpha_9\, gov^{i} + \varepsilon^{i}$$

It is expected that institutions (*inst*) are endogenously determined with GDP growth. If endogeneity is found to be present, then the OLS estimates will be biased and inconsistent. To test this hypothesis, there are instrumental variables needed that are correlated with *inst*, but not with the error term. This instrument will be *reg*.

2) In the first regression, we regress the suspect variable *inst* on all exogenous variables from the normal model in (1) and the new instrumental variable *reg* and retrieve the residuals:

Dependent Variable: INST
Method: Panel EGLS (Period SUR)
Sample (adjusted): 1990 2004
Cross-sections included: 90
Total panel (unbalanced) observations: 789
Linear estimation after one-step weighting matrix
Period SUR (PCSE) standard errors & covariance (d.f. corrected)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INIGDP	-0.489931	0.215618	-2.272223	0.0234
GOV	0.011440	0.010394	1.100580	0.2715
INV	0.006426	0.006889	0.932879	0.3512
LIB	0.010705	0.019796	0.540765	0.5888
LPOP	-0.482582	0.502903	-0.959592	0.3376
OPEN	0.026990	0.212452	0.127039	0.8989
SPEX	0.004702	0.004287	1.096794	0.2731
REG	-0.117576	0.017249	-6.816438	0.0000
C	14.29888	8.019593	1.782993	0.0750

Where resid_inst is retrieved after running this regression.

3) Than, the original regression is run, including as a variable the residuals (*resid_inst*) from the previous regression.

Dependent Variable: LOGGDP8906
Method: Panel EGLS (Period SUR)
Sample (adjusted): 1990 2004
Cross-sections included: 90
Total panel (unbalanced) observations: 789
Linear estimation after one-step weighting matrix
Period SUR (PCSE) standard errors & covariance (d.f. corrected)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INST	0.003461	0.007967	0.434491	0.6641
INIGDP	-0.118584	0.016973	-6.986665	0.0000
GOV	-0.002721	0.000660	-4.123792	0.0000
INV	0.004852	0.000528	9.183276	0.0000
LIB	0.005457	0.001210	4.508586	0.0000
LPOP	0.024759	0.029867	0.828990	0.4074
OPEN	0.066446	0.012946	5.132692	0.0000
SPEX	0.002177	0.000298	7.308956	0.0000
RESID_INST	-0.000702	0.008631	-0.081309	**0.9352**
C	0.071636	0.477331	0.150077	0.8807

If the previous estimates are consistent, then the coefficient on the first stage residuals (*resid_inst*) should not be significantly different from zero.
The H0 in the test is consistent estimates for the variable tested (the covariance of the variable and the error term = 0). To see whether we reject H0 or not, the t-statistic and the p-value of the resid_inst variable are considered. As can be seen, the p-value is 0.935. This means we do not reject H0 of consistent estimates.[21]

E.7: Interaction term

Panel data results with interaction term for institutions and natural resources. Every resource variable has another regression again, with a different interaction term.

Regression for *spex*

Dependent Variable: LOGGDP8906
Method: Panel EGLS (Period SUR)
Sample (adjusted): 1990 2004

[21] Eviews user's guide (2004), p. 578.

Cross-sections included: 77
Total panel (unbalanced) observations: 649
Linear estimation after one-step weighting matrix

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INIGDP	-0.140360	0.015798	-8.884779	0.0000
GOV	-0.003016	0.000722	-4.176165	0.0000
GTT	0.001831	0.014543	0.125925	0.8998
INST	0.003011	0.003762	0.800402	0.4238
INV	0.004561	0.000490	9.311808	0.0000
LIB	0.004874	0.001119	4.355496	0.0000
LPOP	0.054638	0.026957	2.026885	0.0431
OPEN	0.063844	0.012880	4.956977	0.0000
SPEX	0.001462	0.000527	2.773241	0.0057
INTSPEX	0.000172	0.000142	1.206535	0.2281
C	-0.197669	0.414640	-0.476724	0.6337

	Effects Specification		
Cross-section fixed (dummy variables)			

	Weighted Statistics		
R-squared	0.791163	Mean dependent var	1.216301
Adjusted R-squared	0.759206	S.D. dependent var	1.922704
S.E. of regression	0.943486	Sum squared resid	500.2731
F-statistic	24.75693	Durbin-Watson stat	1.667117
Prob(F-statistic)	0.000000		

	Unweighted Statistics		
R-squared	0.367478	Mean dependent var	0.019742
Sum squared resid	0.495271	Durbin-Watson stat	1.382820

Regression with *smex*

Dependent Variable: LOGGDP8906
Method: Panel EGLS (Period SUR)
Sample (adjusted): 1990 2004
Cross-sections included: 77
Total panel (unbalanced) observations: 649
Linear estimation after one-step weighting matrix

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INIGDP	-0.148668	0.015876	-9.364266	0.0000
GOV	-0.001690	0.000665	-2.540031	0.0114
GTT	0.005053	0.014549	0.347302	0.7285
INST	0.006650	0.002905	2.289190	0.0224

INV	0.004770	0.000492	9.697758	0.0000
LIB	0.004641	0.001122	4.134998	0.0000
LPOP	0.041097	0.026705	1.538889	0.1244
OPEN	0.061963	0.012650	4.898166	0.0000
SMEX	0.002706	0.000846	3.196716	0.0015
INTSMEX	-3.50E-05	0.000225	-0.155581	0.8764
C	0.105144	0.409302	0.256886	0.7974

Effects Specification

Cross-section fixed (dummy variables)

Weighted Statistics

R-squared	0.781355	Mean dependent var	1.217794
Adjusted R-squared	0.747897	S.D. dependent var	1.914699
S.E. of regression	0.961368	Sum squared resid	519.4163
F-statistic	23.35322	Durbin-Watson stat	1.618040
Prob(F-statistic)	0.000000		

Unweighted Statistics

R-squared	0.375739	Mean dependent var	0.019742
Sum squared resid	0.488802	Durbin-Watson stat	1.326301

Regression with *sfex*

Dependent Variable: LOGGDP8906
Method: Panel EGLS (Period SUR)
Sample (adjusted): 1990 2004
Cross-sections included: 77
Total panel (unbalanced) observations: 653
Linear estimation after one-step weighting matrix

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INIGDP	-0.140967	0.015964	-8.830183	0.0000
GOV	-0.001468	0.000670	-2.192494	0.0288
GTT	0.004821	0.014597	0.330291	0.7413
INST	0.005689	0.002748	2.070065	0.0389
INV	0.004818	0.000494	9.753050	0.0000
LIB	0.004142	0.001109	3.733871	0.0002
LPOP	0.030138	0.026583	1.133751	0.2574
OPEN	0.061008	0.012625	4.832431	0.0000
SFEX	0.002766	0.000907	3.047958	0.0024
INTSFEX	0.000215	0.000259	0.829437	0.4072
C	0.261519	0.405178	0.645443	0.5189

Effects Specification

Cross-section fixed (dummy variables)

	Weighted Statistics		
R-squared	0.779138	Mean dependent var	1.206103
Adjusted R-squared	0.745579	S.D. dependent var	1.887620
S.E. of regression	0.952118	Sum squared resid	513.0953
F-statistic	23.21727	Durbin-Watson stat	1.618683
Prob(F-statistic)	0.000000		

	Unweighted Statistics		
R-squared	0.380200	Mean dependent var	0.019656
Sum squared resid	0.489970	Durbin-Watson stat	1.337647

Regression using *lognwpc*

Dependent Variable: LOGGDP8906
Method: Panel EGLS (Period SUR)
Sample (adjusted): 1994 2000
Cross-sections included: 74
Total panel (unbalanced) observations: 140
Linear estimation after one-step weighting matrix

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INIGDP	-0.157048	0.042958	-3.655828	0.0006
GOV	-0.001543	0.001470	-1.049732	0.2984
GTT	0.070423	0.031306	2.249524	0.0284
INST	-0.098928	0.042499	-2.327791	0.0236
INV	0.003240	0.001085	2.985996	0.0042
LIB	0.003693	0.002016	1.832388	0.0722
LPOP	0.132638	0.063796	2.079093	0.0422
OPEN	0.062644	0.032141	1.949017	0.0563
LOGNWPC	-0.045163	0.021845	-2.067491	0.0433
INTNWPC	0.012347	0.005260	2.347463	0.0225
C	-0.926507	1.086221	-0.852963	0.3973

Effects Specification

Cross-section fixed (dummy variables)

	Weighted Statistics		
R-squared	0.990397	Mean dependent var	3.758136
Adjusted R-squared	0.976165	S.D. dependent var	7.395398
S.E. of regression	1.141741	Sum squared resid	73.00000
F-statistic	69.58803	Durbin-Watson stat	0.171744
Prob(F-statistic)	0.000000		

	Unweighted Statistics		
R-squared	0.716276	Mean dependent var	0.019399
Sum squared resid	0.053014	Durbin-Watson stat	1.044776

Regression using *logswpc*

Dependent Variable: LOGGDP8906
Method: Panel EGLS (Period SUR)
Sample (adjusted): 1994 2000
Cross-sections included: 61
Total panel (unbalanced) observations: 117
Linear estimation after one-step weighting matrix

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INIGDP	-0.145387	0.046702	-3.113119	0.0032
GOV	0.000745	0.002007	0.371248	0.7122
GTT	0.022680	0.038071	0.595721	0.5543
INST	-0.021265	0.018511	-1.148761	0.2566
INV	0.003690	0.001166	3.164482	0.0028
LIB	0.003466	0.002235	1.550593	0.1279
LPOP	0.115830	0.066265	1.747985	0.0871
OPEN	0.065901	0.039238	1.679490	0.0998
LOGSWPC	-0.016036	0.010160	-1.578413	0.1213
INTSWPC	0.004029	0.002721	1.480858	0.1455
C	-1.040795	1.028236	-1.012214	0.3167

Effects Specification

Cross-section fixed (dummy variables)

	Weighted Statistics		
R-squared	0.989895	Mean dependent var	4.472903
Adjusted R-squared	0.974518	S.D. dependent var	7.213932
S.E. of regression	1.151559	Sum squared resid	61.00002
F-statistic	64.37558	Durbin-Watson stat	0.168248
Prob(F-statistic)	0.000000		

	Unweighted Statistics		
R-squared	0.688056	Mean dependent var	0.023185
Sum squared resid	0.046329	Durbin-Watson stat	1.026316

E.8: Testing for joint significance

To test the significance of the resource variable and the interaction term in these regressions, they have to be tested jointly for their significance.

Below are the Wald coefficient tests for the five regressions, testing the joint significance of the resource variable and the interaction term: H0: $\beta_2 R^i = \beta_5 R^i * inst^i = 0$

For *spex:*

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	19.85363	(2, 638)	0.0000
Chi-square	39.70725	2	0.0000

For *smex*

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	13.23953	(2, 638)	0.0000
Chi-square	26.47907	2	0.0000

For *sfex*

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	13.71960	(2, 642)	0.0000
Chi-square	27.43919	2	0.0000

For *lognwpc*

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	2.769097	(2, 129)	0.0665
Chi-square	5.538194	2	0.0627

For *logswpc*

Wald Test:
Equation: EQINTWSPC

Test Statistic	Value	df	Probability
F-statistic	1.261269	(2, 106)	0.2875
Chi-square	2.522537	2	0.2833

These tests show that the variables in the first three cases are significant up to the 1% level. The fourth regression variables are only significant at the 10% level and the last regression variables are not significant.

Section II: Cross-section analysis

E.9: Correlation matrix for the measures of natural resources

	LOGNWPC	LOGSWPC	SFEX	SMEX	SPEX
LOGNWPC	1.00	0.678	0.373	0.278	0.0546
LOGSWPC	0.678	1.00	0.454	0.417	0.0692
SFEX	0.373	0.454	1.00	0.799	0.504
SMEX	0.278	0.417	0.799	1.00	0.734
SPEX	0.0546	0.0692	0.504	0.734	1.00

The correlation matrix shows that all resources are positively correlated with each other, as is expected. *Smex* has the highest correlation with *sfex* and *spex,* as is also expected. Also *logswpc* and *lognwpc* are correlated the highest among each other.

E.10: Descriptive statistics of the variables in the model

Below in table E.3., the descriptive statistics of the cross-section variables are shown. Included are mean, median, standard deviation, the minimum and maximum value of the variables and the number of observations per variable.

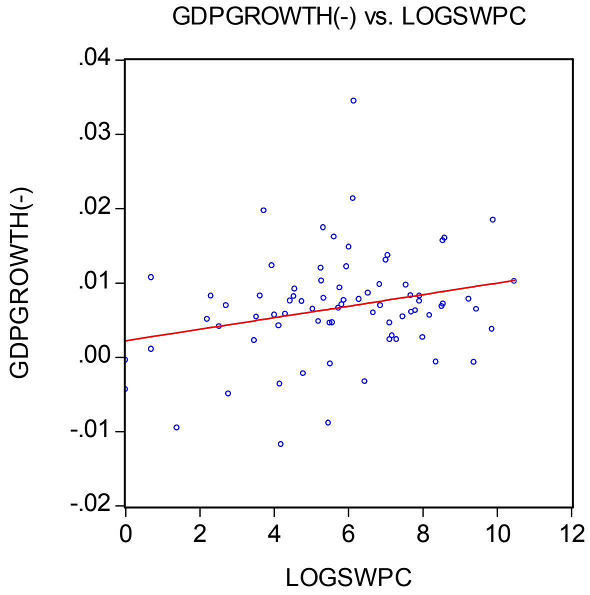
Table E.3: descriptive statistics for the cross-sectional variables

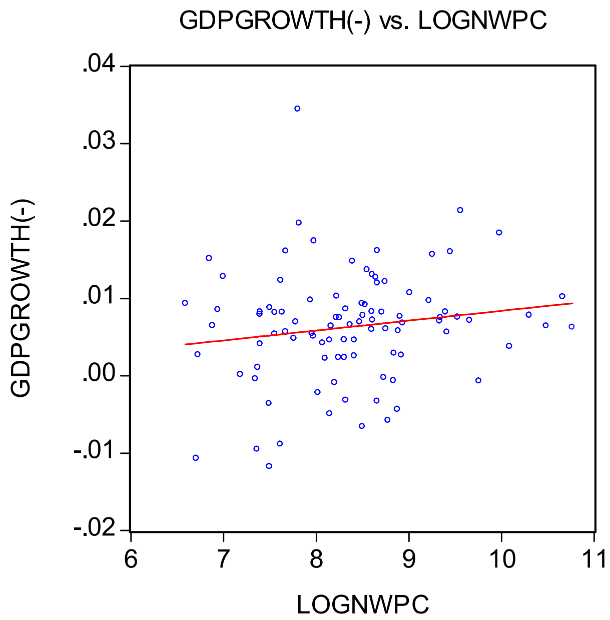
	LGDPGROWTH8906	LGDP89	DEBT	FD	GOV	GTT	INST	INV
Mean	0.00625	7.48	77.967	40.529	14.994	-0.0048	3.206	14.788
Median	0.00695	7.393	62.223	31.434	14.0998	-0.0119	2.972	12.967
Maximum	0.0345	10.378	422.55	208.536	28.967	0.499	6.722	37.629
Minimum	-0.0117	4.863	11.168	10.331	4.76	-0.516	1.00	3.131
Std. Dev.	0.00719	1.55	68.468	30.749	4.917	0.1596	1.586	7.536
Observations	101	101	76	89	101	86	101	101

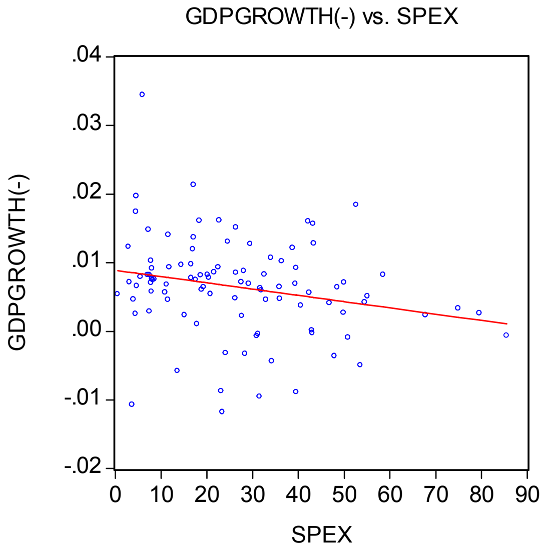
	LIB	LOGNWPC	LOGPOP89	LOGSWPC	OPEN	SCH	SFEX	SMEX	SPEX
Mean	65.816	8.376	16.265	5.754	0.736	98.217	5.895	10.824	26.826
Median	69.28	8.366	16.0235	5.792	0.651	101.378	1.573	4.495	23.763
Maximum	80.465	10.761	20.835	10.462	1.978	139.85	77.443	77.878	85.482
Minimum	36.444	6.59	13.738	0.00	0.186	38.494	0.00078	0.0236	0.528
Std. Dev.	10.832	0.891	1.435	2.368	0.363	16.99	11.249	14.0496	18.322
Observations	101	95	101	76	92	101	98	98	98

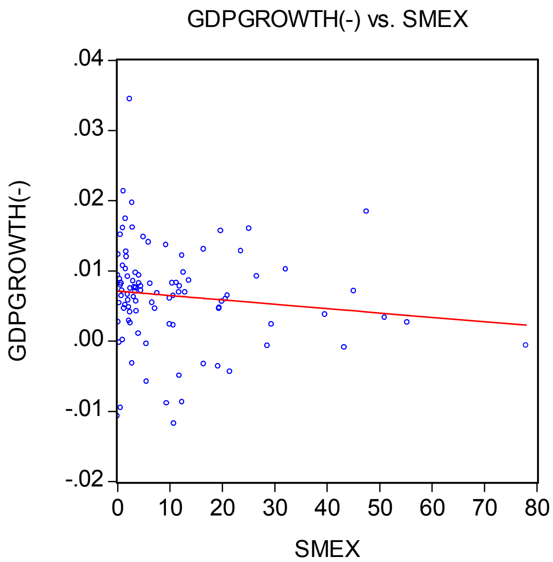
E.11: Scatter plots of the growth variable (lgdpgrowth8906) versus natural resources

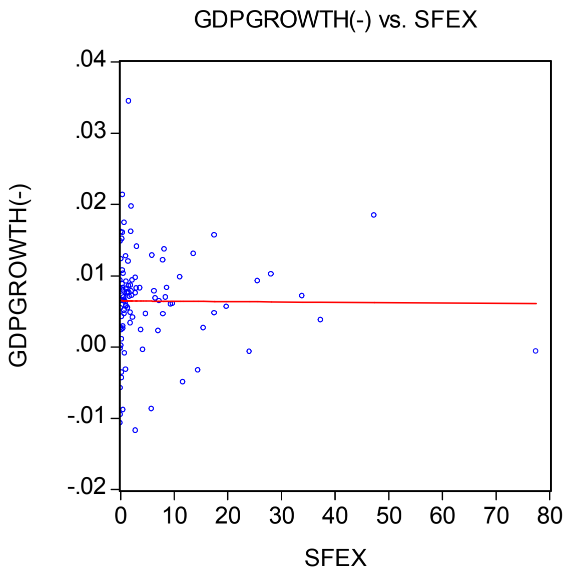
As can be seen in the scatter plots, the two per capita resource measures show a positive relationship with growth. The three variables of exports to GDP show a negative relationship with GDP growth.






GDPGROWTH(-) vs. SPEX
GDPGROWTH(-)
SPEX
.04
.03
.02
.01
.00
-.01
-.02
0
10
20
30
40
50
60
70
80
90


GDPGROWTH(-) vs. SMEX
GDPGROWTH(-)
SMEX
.04
.03
.02
.01
.00
-.01
-.02
0
10
20
30
40
50
60
70
80


GDPGROWTH(-) vs. SFEX
GDPGROWTH(-)
SFEX
.04
.03
.02
.01
.00
-.01
-.02
0
10
20
30
40
50
60
70
80

E.12: OLS regressions

Here the White coefficient covariances are used to correct for heteroskedasticity.

Regression output with GDP growth and *spex*

Dependent Variable: LGDPGROWTH8906
Method: Least Squares
Sample: 1 101
Included observations: 65
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002383	0.000953	-2.501274	0.0155
DEBT	-6.57E-06	1.16E-05	-0.566894	0.5732
FD	0.000114	4.37E-05	2.622383	0.0114
GOV	-8.69E-05	0.000223	-0.389167	0.6987
INST	-0.000402	0.000669	-0.600811	0.5505
INV	0.000327	0.000143	2.281155	0.0266
LIB	0.000261	0.000142	1.833281	0.0724
LOGPOP89	0.000799	0.000592	1.351462	0.1823
OPEN	-0.000481	0.002465	-0.195120	0.8460
SCH	-2.55E-05	4.49E-05	-0.567459	0.5728
SPEX	-4.97E-05	6.61E-05	-0.752593	0.4550
C	-0.007553	0.012607	-0.599133	0.5516

R-squared	0.535486	Mean dependent var	0.006415
Adjusted R-squared	0.439077	S.D. dependent var	0.007448
S.E. of regression	0.005578	Akaike info criterion	-7.374705
Sum squared resid	0.001649	Schwarz criterion	-6.973280
Log likelihood	251.6779	F-statistic	5.554338
Durbin-Watson stat	2.148160	Prob(F-statistic)	0.000008

Regression output with GDP growth and *smex*

Dependent Variable: LGDPGROWTH8906
Method: Least Squares
Sample: 1 101
Included observations: 65
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002370	0.001008	-2.350832	0.0225
DEBT	-1.30E-05	1.13E-05	-1.143274	0.2581
FD	0.000124	4.23E-05	2.944706	0.0048
GOV	-8.34E-05	0.000231	-0.360601	0.7198
INST	-0.000512	0.000691	-0.741866	0.4614
INV	0.000335	0.000142	2.353819	0.0223

LIB	0.000255	0.000147	1.735082	0.0885
LOGPOP89	0.000841	0.000583	1.442835	0.1550
OPEN	-0.001408	0.001988	-0.708483	0.4818
SCH	-2.23E-05	4.40E-05	-0.506119	0.6149
SMEX	6.94E-06	5.91E-05	0.117393	0.9070
C	-0.008736	0.012117	-0.720936	0.4741

R-squared	0.530907	Mean dependent var	0.006415
Adjusted R-squared	0.433548	S.D. dependent var	0.007448
S.E. of regression	0.005606	Akaike info criterion	-7.364897
Sum squared resid	0.001666	Schwarz criterion	-6.963471
Log likelihood	251.3591	F-statistic	5.453094
Durbin-Watson stat	2.184314	Prob(F-statistic)	0.000010

Regression output with GDP growth and *sfex*

Dependent Variable: LGDPGROWTH8906
Method: Least Squares
Sample: 1 101
Included observations: 65
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002490	0.001011	-2.462107	0.0171
DEBT	-1.27E-05	1.19E-05	-1.068645	0.2901
FD	0.000129	4.39E-05	2.933240	0.0049
GOV	-7.75E-05	0.000228	-0.339907	0.7353
INST	-0.000584	0.000734	-0.794904	0.4302
INV	0.000340	0.000143	2.380005	0.0209
LIB	0.000250	0.000149	1.672660	0.1003
LOGPOP89	0.000802	0.000578	1.386756	0.1713
OPEN	-0.001592	0.001983	-0.803005	0.4256
SCH	-2.11E-05	4.38E-05	-0.481321	0.6323
SFEX	4.34E-05	9.32E-05	0.465656	0.6434
C	-0.007148	0.012622	-0.566302	0.5736

R-squared	0.532379	Mean dependent var	0.006415
Adjusted R-squared	0.435326	S.D. dependent var	0.007448
S.E. of regression	0.005597	Akaike info criterion	-7.368040
Sum squared resid	0.001660	Schwarz criterion	-6.966614
Log likelihood	251.4613	F-statistic	5.485427
Durbin-Watson stat	2.206603	Prob(F-statistic)	0.000009

Regression output with GDP growth and *lognwpc*

Dependent Variable: LGDPGROWTH8906
Method: Least Squares
Sample: 1 101

Included observations: 62
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002423	0.001341	-1.806660	0.0768
DEBT	-1.86E-05	1.23E-05	-1.516073	0.1358
FD	0.000128	4.54E-05	2.816837	0.0069
GOV	-6.84E-05	0.000228	-0.300304	0.7652
INST	-0.000862	0.000687	-1.254099	0.2156
INV	0.000284	0.000120	2.355325	0.0225
LIB	0.000252	0.000144	1.754008	0.0856
LOGPOP89	0.001158	0.000522	2.216141	0.0313
OPEN	-0.001274	0.001592	-0.800105	0.4274
SCH	-2.28E-05	4.60E-05	-0.495982	0.6221
LOGNWPC	-0.000465	0.001432	-0.324699	0.7468
C	-0.007644	0.009694	-0.788566	0.4341

R-squared	0.590056	Mean dependent var	0.006314
Adjusted R-squared	0.499869	S.D. dependent var	0.007490
S.E. of regression	0.005297	Akaike info criterion	-7.471263
Sum squared resid	0.001403	Schwarz criterion	-7.059559
Log likelihood	243.6091	F-statistic	6.542546
Durbin-Watson stat	2.107660	Prob(F-statistic)	0.000001

Regression output with GDP growth and *logswpc*

Dependent Variable: LGDPGROWTH8906
Method: Least Squares
Sample: 1 101
Included observations: 45
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002581	0.001082	-2.385181	0.0230
DEBT	-1.72E-05	1.26E-05	-1.364135	0.1818
FD	0.000115	4.54E-05	2.532661	0.0163
GOV	-1.16E-05	0.000275	-0.042062	0.9667
INST	-0.001321	0.000803	-1.644241	0.1096
INV	0.000446	0.000220	2.025275	0.0510
LIB	0.000201	0.000157	1.277991	0.2102
LOGPOP89	0.001016	0.000556	1.825548	0.0770
OPEN	-0.002167	0.002053	-1.055397	0.2989
SCH	-1.19E-05	6.50E-05	-0.182925	0.8560
LOGSWPC	7.81E-05	0.000397	0.196773	0.8452
C	-0.006872	0.015556	-0.441776	0.6615

R-squared	0.669882	Mean dependent var	0.006477

Adjusted R-squared	0.559842	S.D. dependent var	0.007748
S.E. of regression	0.005140	Akaike info criterion	-7.480298
Sum squared resid	0.000872	Schwarz criterion	-6.998521
Log likelihood	180.3067	F-statistic	6.087655
Durbin-Watson stat	2.538792	Prob(F-statistic)	0.000024

E.13: The Hausman test for endogeneity

To check for endogeneity in the model, the Hausman test can be used for the cross-section model as well.

2) estimate the normal regression:

$Lgdpgrowth8906^{i} = \gamma_0 + \gamma_1\, lgdp89^{i} + \gamma_2\, debt^{i} + \gamma_3 fd^{i} + \gamma_4\, gov^{i} + \gamma_5\, inst^{i} + \gamma_6\, inv^{i} + \gamma_7\, lib^{i} + \gamma_8\, logpop89^{i} + \gamma_9\, open^{i} + \gamma_{10}\, sch^{i} + \gamma_{11}\, R^{i} + \upsilon^{i}$

Since *inst* is expected to be endogenously determined with GDP growth, the instruments *eth* and *reg* are used to test if this is the case.

2) In the first regression, we regress the suspect variable *inst* on all exogenous variables from the normal model in (1) and the new instrumental variable *eth* and *reg* and retrieve the residuals:

Dependent Variable: INST
Method: Least Squares
Sample: 1 101
Included observations: 62
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	0.427178	0.245624	1.739156	0.0883
DEBT	-0.002157	0.001800	-1.198411	0.2365
FD	0.012719	0.007872	1.615609	0.1126
GOV	-0.049222	0.029420	-1.673066	0.1007
INV	-0.044191	0.020858	-2.118674	0.0392
LIB	-0.040434	0.023391	-1.728594	0.0902
LOGNWPC	-0.200588	0.246780	-0.812822	0.4203
LOGPOP89	0.236964	0.117806	2.011487	0.0498
OPEN	0.713030	0.514955	1.384645	0.1724
SCH	-0.007879	0.010422	-0.755941	0.4533
ETH	1.629358	0.570496	2.856035	0.0063
REG	-0.142982	0.038740	-3.690850	0.0006

C	1.754016	2.950974	0.594386	0.5550

R-squared	0.593612	Mean dependent var	3.614247
Adjusted R-squared	0.494088	S.D. dependent var	1.416788
S.E. of regression	1.007726	Akaike info criterion	3.037310
Sum squared resid	49.76004	Schwarz criterion	3.483322
Log likelihood	-81.15660	F-statistic	5.964531
Durbin-Watson stat	2.391863	Prob(F-statistic)	0.000003

Where *residinst* is retrieved after running this regression.

3) Than, the original regression is run, including as a variable the residuals (*residinst*) from the previous regression.

Dependent Variable: LGDPGROWTH8906
Method: Least Squares
Sample: 1 101
Included observations: 62
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002458	0.001264	-1.943761	0.0577
DEBT	-2.29E-05	1.25E-05	-1.834552	0.0726
FD	0.000169	5.09E-05	3.322588	0.0017
GOV	-7.78E-05	0.000219	-0.354656	0.7244
INV	0.000183	0.000126	1.447363	0.1542
LIB	0.000145	0.000163	0.891204	0.3772
LOGNWPC	-0.000487	0.001320	-0.368610	0.7140
LOGPOP89	0.001615	0.000522	3.093665	0.0033
OPEN	-0.000369	0.001848	-0.199636	0.8426
SCH	-5.15E-05	4.36E-05	-1.179851	0.2438
INST	-0.002690	0.001158	-2.322670	0.0244
RESIDINST	0.002624	0.001170	2.241908	**0.0295**
C	0.001084	0.010980	0.098710	0.9218

R-squared	0.620424	Mean dependent var	0.006314
Adjusted R-squared	0.527466	S.D. dependent var	0.007490
S.E. of regression	0.005149	Akaike info criterion	-7.515969
Sum squared resid	0.001299	Schwarz criterion	-7.069957
Log likelihood	245.9950	F-statistic	6.674278
Durbin-Watson stat	2.070619	Prob(F-statistic)	0.000001

If the previous estimates are consistent, then *residinst* should again not be significantly different from zero.

The H0 in the test is consistent estimates for the variable tested. As can be seen, the p-value is 0.0295. This means we reject H0 up to the 10% level, but not at the 5% and 1% level.

E.14: Correlation between instrumental variables

Below, in table E.4 is the correlation matrix of the three variables used for institutions in the model. It can be seen that *reg* is (negatively) correlated with *inst*, with a coefficient of 0.71. The negative sign again is explained by the fact that *inst* is lower the better institutions and *reg* is higher the better the regime is. *Eth* is positively correlated with *inst*, since a higher value of *inst* and bad institutions also means higher fractionalization. *reg* and *eth* are negatively correlated, which is explained by the higher the value of the regime variable (the better the regime), the lower ethno-linguistic fractionalization.

Table E.4: Correlation matrix

	ETH	INST	REG
ETH	1.00	0.404	-0.378
INST	0.404	1.00	-0.711
REG	-0.378	-0.711	1.00

E.15: Interaction term

Below are the regression outputs for the cross-section model with the interaction term between resources and institutions included in the regressions. Every resource variable has another regression again, with a different interaction term.

Regression for *spex*

Dependent Variable: LGDPGROWTH8906
Method: Least Squares
Sample: 1 101
Included observations: 65
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002213	0.000899	-2.462873	0.0171
DEBT	-9.29E-06	1.33E-05	-0.697102	0.4888
FD	0.000112	4.46E-05	2.504072	0.0155
GOV	-0.000104	0.000226	-0.459919	0.6475
INST	0.000139	0.001204	0.115305	0.9086
INV	0.000318	0.000143	2.221036	0.0307
LIB	0.000237	0.000134	1.766691	0.0831

LOGPOP89	0.000798	0.000597	1.336275	0.1873
OPEN	-6.45E-05	0.002576	-0.025052	0.9801
SCH	-3.35E-05	5.06E-05	-0.661771	0.5110
SPEX	2.65E-05	0.000140	0.189054	0.8508
INTSPEX	-2.24E-05	4.12E-05	-0.542888	0.5895
C	-0.007825	0.012813	-0.610687	0.5441

R-squared	0.538149	Mean dependent var	0.006415
Adjusted R-squared	0.431568	S.D. dependent var	0.007448
S.E. of regression	0.005616	Akaike info criterion	-7.349686
Sum squared resid	0.001640	Schwarz criterion	-6.914809
Log likelihood	251.8648	F-statistic	5.049205
Durbin-Watson stat	2.144050	Prob(F-statistic)	0.000017

Regression for *smex*

Dependent Variable: LGDPGROWTH8906
Method: Least Squares
Sample: 1 101
Included observations: 65
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002400	0.001040	-2.307929	0.0250
DEBT	-1.35E-05	1.11E-05	-1.216905	0.2291
FD	0.000127	4.32E-05	2.936108	0.0049
GOV	-7.69E-05	0.000234	-0.329077	0.7434
INST	-0.000748	0.000881	-0.848618	0.4000
INV	0.000338	0.000147	2.304145	0.0252
LIB	0.000260	0.000147	1.763867	0.0836
LOGPOP89	0.000851	0.000595	1.431793	0.1582
OPEN	-0.001576	0.002007	-0.785250	0.4359
SCH	-2.45E-05	4.53E-05	-0.539743	0.5917
SMEX	-9.80E-05	0.000163	-0.600294	0.5509
INTSMEX	2.76E-05	4.74E-05	0.583612	0.5620
C	-0.008044	0.012398	-0.648851	0.5193

R-squared	0.532904	Mean dependent var	0.006415
Adjusted R-squared	0.425113	S.D. dependent var	0.007448
S.E. of regression	0.005647	Akaike info criterion	-7.338393
Sum squared resid	0.001658	Schwarz criterion	-6.903516
Log likelihood	251.4978	F-statistic	4.943845
Durbin-Watson stat	2.122228	Prob(F-statistic)	0.000022

Regression for *sfex*

Dependent Variable: LGDPGROWTH8906

Method: Least Squares
Sample: 1 101
Included observations: 65
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002548	0.001047	-2.434100	0.0184
DEBT	-1.30E-05	1.14E-05	-1.142518	0.2585
FD	0.000137	4.42E-05	3.105355	0.0031
GOV	-7.20E-05	0.000224	-0.321456	0.7492
INST	-0.001154	0.000894	-1.291140	0.2024
INV	0.000355	0.000148	2.399325	0.0200
LIB	0.000248	0.000147	1.683493	0.0983
LOGPOP89	0.000757	0.000581	1.301058	0.1990
OPEN	-0.002376	0.002069	-1.148378	0.2561
SCH	-2.07E-05	4.58E-05	-0.452564	0.6527
SFEX	-0.000428	0.000187	-2.291371	0.0260
INTSFEX	0.000120	5.14E-05	2.344467	0.0229
C	-0.003954	0.013293	-0.297413	0.7673

R-squared	0.555129	Mean dependent var	0.006415
Adjusted R-squared	0.452467	S.D. dependent var	0.007448
S.E. of regression	0.005511	Akaike info criterion	-7.387144
Sum squared resid	0.001580	Schwarz criterion	-6.952267
Log likelihood	253.0822	F-statistic	5.407324
Durbin-Watson stat	2.133936	Prob(F-statistic)	0.000007

When using the interaction term in the regression with *sfex*, both the interaction term and the *sfex* variable become significant at the 5% level.

Regression for *lognwpc*

Dependent Variable: LGDPGROWTH8906
Method: Least Squares
Sample: 1 101
Included observations: 62
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002625	0.001285	-2.041945	0.0466
DEBT	-1.87E-05	1.24E-05	-1.509022	0.1377
FD	0.000128	4.58E-05	2.788092	0.0075
GOV	-7.48E-05	0.000228	-0.328831	0.7437
INST	0.005674	0.006742	0.841590	0.4041
INV	0.000286	0.000123	2.318713	0.0246
LIB	0.000239	0.000150	1.590342	0.1182

LOGPOP89	0.001210	0.000533	2.268805	0.0277
OPEN	-0.000887	0.001589	-0.558111	0.5793
SCH	-2.33E-05	4.64E-05	-0.502491	0.6176
LOGNWPC	0.002432	0.002845	0.854924	0.3968
INTNWPC	-0.000813	0.000833	-0.977067	0.3333
C	-0.029971	0.021729	-1.379306	0.1741

R-squared	0.598177	Mean dependent var	0.006314
Adjusted R-squared	0.499772	S.D. dependent var	0.007490
S.E. of regression	0.005298	Akaike info criterion	-7.459014
Sum squared resid	0.001375	Schwarz criterion	-7.013001
Log likelihood	244.2294	F-statistic	6.078695
Durbin-Watson stat	2.038082	Prob(F-statistic)	0.000002

Regression for *logswpc*

Dependent Variable: LGDPGROWTH8906
Method: Least Squares
Sample: 1 101
Included observations: 45
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002382	0.000977	-2.438932	0.0205
DEBT	-2.38E-05	1.21E-05	-1.966440	0.0580
FD	9.16E-05	4.73E-05	1.937652	0.0615
GOV	9.49E-07	0.000263	0.003607	0.9971
INST	-0.003828	0.001108	-3.455033	0.0016
INV	0.000526	0.000232	2.266987	0.0303
LIB	0.000170	0.000157	1.083710	0.2866
LOGPOP89	0.001080	0.000535	2.018721	0.0520
OPEN	-0.002393	0.002075	-1.153192	0.2574
SCH	-5.03E-05	6.75E-05	-0.745669	0.4613
LOGSWPC	-0.001578	0.000708	-2.227346	0.0331
INTSWPC	0.000484	0.000191	2.532195	0.0164
C	0.005676	0.017920	0.316771	0.7535

R-squared	0.701506	Mean dependent var	0.006477
Adjusted R-squared	0.589571	S.D. dependent var	0.007748
S.E. of regression	0.004964	Akaike info criterion	-7.536554
Sum squared resid	0.000788	Schwarz criterion	-7.014629
Log likelihood	182.5725	F-statistic	6.267068
Durbin-Watson stat	2.399693	Prob(F-statistic)	0.000016

E.16: Testing for joint significance

To test the significance of the resource variable and the interaction term in these regressions, they have to be tested jointly for their significance.

Below are the Wald coefficient tests for the five regressions, testing the joint significance of the resource variable and the interaction term: H0: $\beta_2 R^{i} = \beta R^{i} * inst^{i} = 0$

For *spex*

Wald Test:
Equation: EQSPEX

Test Statistic	Value	df	Probability
F-statistic	0.344241	(2, 52)	0.7104
Chi-square	0.688482	2	0.7088

For *smex*

Wald Test:
Equation: EQSPEX

Test Statistic	Value	df	Probability
F-statistic	0.181253	(2, 52)	0.8347
Chi-square	0.362506	2	0.8342

For *sfex*

Wald Test:
Equation: EQSPEX

Test Statistic	Value	df	Probability
F-statistic	2.820834	(2, 52)	0.0687
Chi-square	5.641669	2	0.0596

For *lognwpc*

Wald Test:
Equation: EQSPEX

Test Statistic	Value	df	Probability
F-statistic	0.477615	(2, 49)	0.6231
Chi-square	0.955230	2	0.6203

For *logswpc*

Wald Test:
Equation: EQSPEX

Test Statistic	Value	df	Probability
F-statistic	3.219531	(2, 32)	0.0532
Chi-square	6.439062	2	0.0400

The tests show that only the *sfex* and *swpc* variables are significant together with the interaction term. Both are significant at the 10% level.

E.17: 2SLS regressions

Since there is evidence of endogeneity of the institutions variable (*inst*), the 2SLS method is used in this part. As can be read in Wooldridge (2003), when there is one endogenous explanatory variable in the model, along with multiple exogenous instrumental variables (IV) for this one explanatory variable, the IV variable is called the two stage least squares (2SLS) estimator and a 2SLS procedure is used. Condition is that the two instruments do not appear in the original regression, they are called exclusion restrictions. When there are multiple instruments, the linear combination of them is a valid IV. Condition is that either one of the exogenous variables used as IV is different from zero. The IV also have to be jointly significant at least at the 5% level. As instruments for *inst* the variables *reg* and *eth* are used. Another condition in the 2SLS model is that there are at least as many excluded exogenous variables as there are included endogenous explanatory variables in the structural equation (Eviews also requires this).[22] The 2SLS regression is run for all five natural resource variables, just as in F.8. As an example, again *spex* is used.

Stage one:

$$inst = \beta_0 + \beta_1 \text{ lgdp89} + \beta_2 \text{ gov} + \beta_3 \text{ inv} + \beta_4 \text{ lib} + \beta_5 \text{ logpop89} + \beta_6 \text{ open} + \beta_7 \text{ spex} + \beta_8 \text{ reg} + \beta_9 \text{ eth} + \varepsilon$$

$$inst^i = \delta_0 + \delta_1\, lgdp89^i + \delta_2\, debt^i + \delta_3\, fd^i + \delta_4\, gov^i + \delta_5\, inv^i + \delta_6\, lib^i + \delta_7\, logpop89^i + \delta_8\, open^i + \delta_9\, sch^i + \delta_{10}\, R^i + \delta_{11}\, reg^i + \delta_{12}\, eth^i + \eta^i$$

[22] Wooldridge (2003), pp. 499-503.

As can be seen below, both *eth* and *reg* are significantly different from zero, as is a condition for 2SLS.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ETH	1.412883	0.494426	2.857626	0.0061
REG	-0.135478	0.034953	-3.875956	0.0003

The other condition is also valid in the model, where the IV variables have to be jointly significant at the 5% level. The Wald test below shows this.

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	13.07305	(2, 52)	0.0000
Chi-square	26.14609	2	0.0000

Stage 2:

$$Lgdpgrowth8906^{i} = \gamma_0 + \gamma_1\, lgdp89^{i} + \gamma_2\, debt^{i} + \gamma_3\, fd^{i} + \gamma_4\, gov^{i} + \gamma_5\, inst^{i} + \gamma_6\, inv^{i} + \gamma_7\, lib^{i} + \gamma_8\, logpop89^{i} + \gamma_9\, open^{i} + \gamma_{10}\, sch^{i} + \gamma_{11}\, spex^{i} + \upsilon^{i}$$

Dependent Variable: LGDPGROWTH8906
Method: Two-Stage Least Squares
Sample: 1 101
Included observations: 65
White Heteroskedasticity-Consistent Standard Errors & Covariance
Instrument list: LGDP89 DEBT FD GOV INV LIB LOGPOP89 OPEN
SCH SPEX ETH REG

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002363	0.001107	-2.134194	0.0375
DEBT	-1.66E-05	1.21E-05	-1.371514	0.1760
FD	0.000144	5.65E-05	2.548606	0.0137
GOV	-4.30E-05	0.000227	-0.189606	0.8503
INST	-0.001898	0.001180	-1.608447	0.1137
INV	0.000215	0.000181	1.191065	0.2389
LIB	0.000189	0.000165	1.148145	0.2561
LOGPOP89	0.001282	0.000617	2.075615	0.0428
OPEN	0.000515	0.002913	0.176827	0.8603
SCH	-4.02E-05	4.76E-05	-0.844466	0.4022
SPEX	-2.01E-05	6.67E-05	-0.300634	0.7649

C	-0.005209	0.013807	-0.377313	0.7074

R-squared	0.490711	Mean dependent var	0.006415
Adjusted R-squared	0.385010	S.D. dependent var	0.007448
S.E. of regression	0.005841	Sum squared resid	0.001808
F-statistic	5.247857	Durbin-Watson stat	2.335192
Prob(F-statistic)	0.000016		

Regression for *smex*

Dependent Variable: LGDPGROWTH8906
Method: Two-Stage Least Squares
Sample: 1 101
Included observations: 65
White Heteroskedasticity-Consistent Standard Errors & Covariance
Instrument list: LGDP89 DEBT FD GOV INV LIB LOGPOP89 OPEN
SCH SMEX ETH REG

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002656	0.001156	-2.297783	0.0256
DEBT	-2.36E-05	1.17E-05	-2.015330	0.0490
FD	0.000162	5.76E-05	2.808570	0.0070
GOV	-3.90E-05	0.000236	-0.165131	0.8695
INST	-0.002491	0.001252	-1.989471	0.0518
INV	0.000197	0.000191	1.033865	0.3059
LIB	0.000173	0.000168	1.029877	0.3077
LOGPOP89	0.001371	0.000628	2.183844	0.0334
OPEN	0.000180	0.002798	0.064453	0.9489
SCH	-4.21E-05	4.97E-05	-0.845537	0.4016
SMEX	6.61E-05	6.75E-05	0.979915	0.3316
C	-0.002246	0.014440	-0.155561	0.8770

R-squared	0.455653	Mean dependent var	0.006415
Adjusted R-squared	0.342676	S.D. dependent var	0.007448
S.E. of regression	0.006039	Sum squared resid	0.001933
F-statistic	4.952272	Durbin-Watson stat	2.395906
Prob(F-statistic)	0.000031		

Regression for *sfex*

Dependent Variable: LGDPGROWTH8906
Method: Two-Stage Least Squares
Sample: 1 101
Included observations: 65
White Heteroskedasticity-Consistent Standard Errors & Covariance
Instrument list: LGDP89 DEBT FD GOV INV LIB LOGPOP89 OPEN

SCH SFEX ETH REG

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002866	0.001223	-2.344312	0.0228
DEBT	-2.18E-05	1.19E-05	-1.836164	0.0719
FD	0.000174	5.93E-05	2.940038	0.0049
GOV	-8.00E-06	0.000231	-0.034587	0.9725
INST	-0.002813	0.001326	-2.121038	0.0386
INV	0.000194	0.000194	0.999240	0.3222
LIB	0.000143	0.000175	0.816913	0.4176
LOGPOP89	0.001332	0.000635	2.096812	0.0408
OPEN	1.35E-06	0.002902	0.000464	0.9996
SCH	-4.11E-05	5.07E-05	-0.809543	0.4218
SFEX	0.000148	0.000118	1.254404	0.2152
C	0.001846	0.015743	0.117251	0.9071

R-squared	0.439043	Mean dependent var	0.006415
Adjusted R-squared	0.322618	S.D. dependent var	0.007448
S.E. of regression	0.006130	Sum squared resid	0.001992
F-statistic	4.858196	Durbin-Watson stat	2.404460
Prob(F-statistic)	0.000038		

Regression for *lognwpc*

Dependent Variable: LGDPGROWTH8906
Method: Two-Stage Least Squares
Sample: 1 101
Included observations: 62
White Heteroskedasticity-Consistent Standard Errors & Covariance
Instrument list: LGDP89 DEBT FD GOV INV LIB LOGNWPC
LOGPOP89 OPEN SCH ETH REG

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002458	0.001565	-1.570193	0.1227
DEBT	-2.29E-05	1.12E-05	-2.044210	0.0462
FD	0.000169	5.44E-05	3.111936	0.0031
GOV	-7.78E-05	0.000230	-0.338417	0.7365
INST	-0.002690	0.001165	-2.309713	0.0251
INV	0.000183	0.000159	1.145284	0.2575
LIB	0.000145	0.000170	0.852849	0.3978
LOGNWPC	-0.000487	0.001493	-0.325932	0.7458
LOGPOP89	0.001615	0.000581	2.778682	0.0077
OPEN	-0.000369	0.002340	-0.157664	0.8754
SCH	-5.15E-05	5.33E-05	-0.965245	0.3391
C	0.001084	0.014196	0.076349	0.9394

R-squared	0.520333	Mean dependent var	0.006314

Adjusted R-squared	0.414806	S.D. dependent var	0.007490
S.E. of regression	0.005730	Sum squared resid	0.001642
F-statistic	5.878699	Durbin-Watson stat	2.180818
Prob(F-statistic)	0.000005		

Regression for *logswpc*

Dependent Variable: LGDPGROWTH8906
Method: Two-Stage Least Squares
Sample: 1 101
Included observations: 45
White Heteroskedasticity-Consistent Standard Errors & Covariance
Instrument list: LGDP89 DEBT FD GOV INV LIB LOGPOP89 OPEN SCH LOGSWPC ETH REG

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LGDP89	-0.002819	0.001172	-2.405487	0.0219
DEBT	-2.00E-05	1.20E-05	-1.667560	0.1049
FD	0.000135	4.52E-05	2.992709	0.0052
GOV	-8.56E-06	0.000271	-0.031646	0.9749
INST	-0.002268	0.001067	-2.124559	0.0412
INV	0.000372	0.000234	1.591190	0.1211
LIB	0.000142	0.000170	0.838046	0.4080
LOGPOP89	0.001212	0.000551	2.200786	0.0349
OPEN	-0.001727	0.002471	-0.698893	0.4895
SCH	-5.28E-06	7.80E-05	-0.067742	0.9464
LOGSWPC	0.000165	0.000379	0.436867	0.6651
C	-0.002446	0.018766	-0.130344	0.8971

R-squared	0.651278	Mean dependent var	0.006477
Adjusted R-squared	0.535037	S.D. dependent var	0.007748
S.E. of regression	0.005283	Sum squared resid	0.000921
F-statistic	5.815115	Durbin-Watson stat	2.492698
Prob(F-statistic)	0.000038		