Collaborative Research Project - Assignment 3

N. Alvarado and G. Tarriba

13 November 2015

```
## [1] "/Users/Gabriel/Desktop/Third-Assignment"
```

```
## 'data.frame':    1172 obs. of  10 variables:
##  $ X        : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ iso2c    : Factor w/ 169 levels "AE","AF","AL",..: 1 1 1 1 1 1 1 2 2 2 ...
##  $ year     : int  2005 2006 2007 2008 2009 2010 2011 2005 2006 2007 ...
##  $ country  : Factor w/ 170 levels "Afghanistan",..: 161 161 161 161 161 161 161 1 1 1 ...
##  $ corrupest: num  1.069 0.947 1.077 1.131 0.955 ...
##  $ gasrents : num  6.18 5.01 4.42 5.29 2.97 ...
##  $ gdppc    : num  40299 38363 34060 30613 25977 ...
##  $ oilrents : num  24.4 24.9 23.3 27.1 17.7 ...
##  $ totrents : num  30.5 29.9 27.8 32.4 20.6 ...
##  $ unemp    : num  3.1 3.3 3.4 4 4.2 ...
```

Revisiting the natural resource curse: does the relationship between natural resource rents and corruption persist into the 21st Century?

1. Introduction

Our collaborative research project seeks to provide an answer to the following question: **How significant is the relationship between a country's degree of dependence on natural resource rents and its corruption level?** In the 1990s and early 2000s, several scholars (see among others Leite and Weidmann 1999, Sala-i-Martin and Subramanian (2003), Pendergast (2007) and Shaxson (2007)) identified this relationship and developed theories about the interplay of governance and the fiscal system to account for the "Natural Resource Curse" first identified by Sachs and Werner (1995). However, in light of the rapid economic development of many resource-dependent developing countries since the turn of the century (notably in sub-Saharan Africa, which is now the second fastest growing region in the world after Asia-Pacific - see Chuhan-Pule 2014), it is worth revisiting this relationship. Could it be that in the current century the relationship between reliance on natural resource rents and corruption has been broken?. This is, in a nutshell, our motivation for this project.

2. Data preparation

2.1 General plan

We seek to collect and analyze data on corruption and natural resource rents for all countries in the world for the last 20 years, because the existing empirical literature covers the previous years. As our research question suggests, we plan to use a *Corruption* indicator as our dependent variable and measures of dependence on natural resource rents as our explanatory variables of interest. As for our key explanatory variable, we want to have both *Total Natural Resource Rents* and *Oil Rents*. We want to examine oil separately because there is significant empirical evidence that oil rents have a particularly detrimental effect on governance

(see among others Ross 2012; Shaxson 2007). We also want to include control variables such as *GDP/GNI per capita* and *Unemployment*. These variables are included in similar empirical studies. Please see our Assignment 2 for more details on the variables to be included in our analysis, their conceptual definition and their operationalization.

2.1 Data sources

The data necessary for our empirical analysis is located in two sources, or rather two different research projects from the World Bank:

- The World Development Indicators (WDI) also from the World Bank. This is "The primary World Bank collection of development indicators, compiled from officially-recognized international sources. It presents the most current and accurate global development data available, and includes national, regional and global estimates" (The World Bank Group (2014)).
- The World Governance Indicators (WGI) of the World Bank, defined as "Aggregate and individual governance indicators for 215 countries and territories over the period 1996–2014, for six dimensions of governance" (The World Bank Group (2015)).

We need to collect data from these two sources and merge it.

In Assignment 2 we also considered including in our analysis an indicator of prevalence of conflict, using data from the UCDP/PRIO Armed Conflict Dataset. However, upon close inspection of the data we realized that it was not useful since it does not include information on the magnitude of conflicts -it is rather a list of all conflicts, the conflicting parties involved and the duration of the conflict.

2.2 Data gathering

2.2.1 Data from the World Development Indicators (WDI)

We gathered data directly from the **WDI** using the WDI Package for R. This package enables us to automatically download WDI indicators using the World Bank's APIs. The following five variables were downloaded from the **WDI**:

- **Total natural resources rents (% of GDP)** - This is "the sum of oil rents, natural gas rents, coal rents (hard and soft), mineral rents, and forest rents". World Bank code: NY.GDP.TOTR.RT.ZS
- **Oil rents (% of GDP)** - "Oil rents are the difference between the value of oil production at world prices and total costs of production". World Bank code: NY.GDP.PETR.RT.ZS
- **Natural gas rents (% of GDP)** - "Natural gas rents are the difference between the value of natural gas production at world prices and total costs of production". World Bank code: NY.GDP.NGAS.RT.ZS
- **GDP per capita (constant 2005 US$)** "GDP per capita is gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in constant 2005 U.S. dollars". World Bank code NY.GDP.PCAP.KD.
- **Unemployment, total (% of total labor force) (modeled ILO estimate)** "Unemployment refers to the share of the labor force that is without work but available for and seeking employment". World Bank code: SL.UEM.TOTL.ZS

2.2.1 Data from the World Governance Indicators (WGI)

The dataset of the **WGI** consists of six aggregate indicators of governance: - Voice and Accountability - Political Stability and Absence of Violence/Terrorism - Government Effectiveness - Regulatory Quality - Rule of Law - Control of Corruption

In contrast with the **WDI**, for the **WGI** there was no API. Hence, the data was automatically downloaded as a .XLSX file and then imported into R. We are only interested in the Control of Corruption estimate, because we want it to be our dependent variable. We don't need the other variables so we only keep Control of Corruption. The estimates begin in 1996 and end in 2014. The entire WGI dataset) is available as a CSV file. Please refer to our Analysis file for more details on how the data was downloaded and cleaned up.

The variable **Control of Corruption** "reflects perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as"capture" of the state by elites and private interests" (The World Bank Group 2015). It is an estimate of governance, ranging from -2.5 (weak) to 2.5 (strong) governance performance.

2.2.1 Cleaning and merging the data

Having downloaded the data from the **WGI** and **WDI**, we then proceeded to clean it by removing missing observations, standardizing country codes (the two datasets use different coding systems), transforming string variables into numeric variables, and other necessary transformations. Once we are done with the data preparation, we have a dataset with 1155 observations, corresponding to 165 countries and 7 years (2005 to 2011).

We saved the resulting dataset as a CSV file under the name MergedData.csv.

With this dataset, we can explore the relationship between **Control of Corruption** and three possible dependent variables (**Total natural resource rents**, **Oil rents** and **Natural gas rents**).

3. Descriptive and inferential statistics

Before starting with our analysis, we want to examine and describe the key variables.

We want to be create indicators to run the following model:

$$Corruption^*_{it} = \beta_1 NaturalResourceRents^*_{it} + \beta_2 GDPpercapita^*_{it} + \beta_3 Unemployment^*_{it} + \epsilon$$

4. Limitations of our data

Before conducting our analysis we want to characterize our data, in particular our dependent variable *Corruption* and our interest explanatory variables *Total Natural Resource Rents* and *Oil Rents*.

5. Next steps

Corruption over time

Regressions

Following the model

$$Corruption_i = \beta_0 + \beta_1 NaturalResourceRents_i + \beta_2 GDPpercapita_i + \beta_3 Unemployment_i + \epsilon_i$$

The model one is a regression with the control variables specified.

```
##                      2.5 %        97.5 %
## (Intercept) -0.5588637652 -4.295447e-01
## totrents    -0.0152637697 -1.183046e-02
## gdppc        0.0000492183  5.310915e-05
## unemp        0.0027087866  1.272930e-02
```

```
##
## Call:
## lm(formula = corrupest ~ totrents + gdppc + unemp, data = merged)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.95468 -0.32190 -0.02724  0.28425  1.76540
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.942e-01  3.296e-02 -14.996  < 2e-16 ***
## totrents    -1.355e-02  8.750e-04 -15.483  < 2e-16 ***
## gdppc        5.116e-05  9.916e-07  51.600  < 2e-16 ***
## unemp        7.719e-03  2.554e-03   3.023  0.00256 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5165 on 1168 degrees of freedom
## Multiple R-squared:  0.7367, Adjusted R-squared:  0.736
## F-statistic:  1089 on 3 and 1168 DF,  p-value: < 2.2e-16
```

Heteroskedasticity

When testing for heterosckedasticity, ...

`{}r plot(model1, which = 1, main = "Residual vs.  Fitted Plot")` The graphical inspection suggests there is heterosckedastocity. For this reason, it is necessary

Pearson's correlation

$$Corruption_i = \beta_0 + \beta_1 NaturalResourceRents_i + \beta_2 GDPpercapita_i + \beta_3 ArmedConflict_i + \beta_4 Unemployment_i + \epsilon_i$$

The next step will be to run a Fixed Effects regression, which will allow us to get rid of the country-specific fixed effects (such that we will only be analyzing the variation in variables within one country across time). To run a Fixed Effects regression, we will subtract from each term its mean. Thus, the regression equation will be:

$$Corruption^*_{it} = \beta_1 NaturalResourceRents^*_{it} + \beta_2 GDPpercapita^*_{it} + \beta_3 ArmedConflict^*_{it} + \beta_4 Unemployment^*_{it} + \epsilon$$

The terms in the equation above are demeaned to eliminate the fixed effects, such that:

$$NaturalResourceRents^*_{it} = NaturalResourceRents_i t - \overline{NaturalResourceRents_i}$$

$$GDPpercapita^*_{it} = GDPpercapita_{it} - \overline{GDPpercapita_i}$$
$$ArmedConflict^*_{it} = ArmedConflict_{it} - \overline{ArmedConflict_i}$$
$$Unemployment^*_{it} = Unemployment_{it} - \overline{Unemployment_i}$$

```
##       exp              wks        bluecol         ind         south
##  Min.   : 1.00   Min.   : 5.00   no :2036   Min.   :0.0000   no :2956
##  1st Qu.:11.00   1st Qu.:46.00   yes:2129   1st Qu.:0.0000   yes:1209
##  Median :18.00   Median :48.00              Median :0.0000
##  Mean   :19.85   Mean   :46.81              Mean   :0.3954
##  3rd Qu.:29.00   3rd Qu.:50.00              3rd Qu.:1.0000
##  Max.   :51.00   Max.   :52.00              Max.   :1.0000
##   smsa      married        sex       union            ed        black
##  no :1442   no : 773   female: 469   no :2649   Min.   : 4.00   no :3864
##  yes:2723   yes:3392   male  :3696   yes:1516   1st Qu.:12.00   yes: 301
##                                                 Median :12.00
##                                                 Mean   :12.85
##                                                 3rd Qu.:16.00
##                                                 Max.   :17.00
##      lwage
##  Min.   :4.605
##  1st Qu.:6.395
##  Median :6.685
##  Mean   :6.676
##  3rd Qu.:6.953
##  Max.   :8.537
```

```
##
## Call:
## lm(formula = lwage ~ ed, data = Wages)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.92996 -0.26863  0.00931  0.28453  1.83076
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.838779   0.030997  188.37   <2e-16 ***
## ed          0.065204   0.002358   27.65   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4243 on 4163 degrees of freedom
## Multiple R-squared:  0.1552, Adjusted R-squared:  0.155
## F-statistic: 764.5 on 1 and 4163 DF,  p-value: < 2.2e-16
```

```
##                 2.5 %     97.5 %
## (Intercept) 5.7780088 5.89954938
## ed          0.0605805 0.06982708
```

```
##
## <table style="text-align:center"><caption><strong>OLS regression of the Percentage of Wages Variatio
## <tr><td colspan="3" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"
```

```
## <tr><td></td><td colspan="2" style="border-bottom: 1px solid black"></td></tr>
## <tr><td style="text-align:left"></td><td colspan="2">lwage</td></tr>
## <tr><td style="text-align:left"></td><td>(1)</td><td>(2)</td></tr>
## <tr><td colspan="3" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"
## <tr><td style="text-align:left"></td><td>(0.002)</td><td>(0.002)</td></tr>
## <tr><td style="text-align:left"></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">exp</td><td></td><td>0.01<sup>***</sup></td></tr>
## <tr><td style="text-align:left"></td><td></td><td>(0.001)</td></tr>
## <tr><td style="text-align:left"></td><td></td><td></td><td></td></tr>
## <tr><td style="text-align:left">Constant</td><td>5.84<sup>***</sup></td><td>5.44<sup>***</sup></td><
## <tr><td style="text-align:left"></td><td>(0.03)</td><td>(0.03)</td></tr>
## <tr><td style="text-align:left"></td><td></td><td></td><td></td></tr>
## <tr><td colspan="3" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"
## <tr><td style="text-align:left">R<sup>2</sup></td><td>0.16</td><td>0.25</td></tr>
## <tr><td style="text-align:left">Adjusted R<sup>2</sup></td><td>0.15</td><td>0.25</td></tr>
## <tr><td style="text-align:left">Residual Std. Error</td><td>0.42 (df = 4163)</td><td>0.40 (df = 4162
## <tr><td style="text-align:left">F Statistic</td><td>764.53<sup>***</sup> (df = 1; 4163)</td><td>681.
## <tr><td colspan="3" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"
## </table>
```

6. References

Chuhan-Pule, Punam. 2014. *Africa's Pulse*. 9. World Bank.

Leite, Carlos A, and Jens Weidmann. 1999. "Does Mother Nature Corrupt? Natural Resources, Corruption, and Economic Growth." *Natural Resources, Corruption, and Economic Growth (June 1999). IMF Working Paper*, no. 99/85.

Pendergast, Shannon Marie. 2007. "Corruption and the Curse of Natural Resources." PhD thesis, University of Victoria.

Ross, Michael. 2012. *The Oil Curse: How Petroleum Wealth Shapes the Development of Nations*. Princeton University Press.

Sachs, Jeffrey D, and Andrew M Warner. 1995. *Natural Resource Abundance and Economic Growth*. National Bureau of Economic Research.

Sala-i-Martin, Xavier, and Arvind Subramanian. 2003. *Addressing the Natural Resource Curse: An Illustration from Nigeria*. National Bureau of Economic Research.

Shaxson, Nicholas. 2007. "Oil, Corruption and the Resource Curse." *International Affairs* 83 (6). Wiley Online Library: 1123–40.

The World Bank Group. 2014. "World Development Indicators (WGI) Project." http://wdi.worldbank.org/table/3.15.

———. 2015. "World Governance Indicators (WGI) Project." http://info.worldbank.org/governance/wgi/index.aspx#home.