Data Notes

Ivan Rodriguez

March 10, 2015

Abstract

This document is meant to be a short summary on the raw data that was compiled to create the analytical database for the CDS project. The sources and the dates of access will be documented along with the scripts that will merge and transform the data into the final product.

Status

- **3/10/2015** - Plan to add in ratings data
- **3/9/2015** - Creation of this document from prior scripts - does not include **ratings** or **spread** data

Summary of Project

The first step of our analysis will try and replicate Cantor and Packers *Determinants and Impact of Sovereign Credit Ratings* paper (Cantor Packer 1996). The main model the paper tests is a classical lineaer regression model of the form

$$Ratings = \mathbf{x}'\boldsymbol{\beta}$$

where $\mathbf{x}'$ is a vector of our regressors (with 1 as the first element). The regressand is

- **Ratings**: ratings as of t assigned by Moody's or SP (average, separately, difference)

and the regressors are:

- **Per Capita Income**: measured by GNP per capita as of $t - 1$
- **GDP Growth**: measured by an average annual GDP growth averaged for the past $t - 3$ years
- **Inflation**: measured by average CPI averaged for the past $t - 2$ years
- **Fiscal Balance**: measured by average annual government budget surplus relative to GDP averaged for the past $t - 2$ years
- **External Balance**: measured by average annual current accout surplus relative to GDP averaged for the past $t - 2$ years
- **External Debt**: measured by foreign currency debt relative to exports as of $t - 1$
- **Economic Development**: IMF classification (indicator variable) as of t
- **Default History**: indicator variable that is 1 if the country has defaulted since 1970.

The table summarizing the variables on page 40 of Cantor Packer 1996 has more detailed information on the variables. Once the summary statistics and original model parameters have been replicated, a CDS spread variable will be added.

General Information on the Raw Data

The raw data that will be used for the replication is based on the following three sources:

- The **macroeconomic variables** are from the World Bank's **World Development Indicators** (WDI) database
- The **default history data** is from 2014 update building on AEJ Macro 2013 paper ***Sovereign Defaults: The Price of Haircuts***, by Juan Cruces and Christoph Trebesch, which can be found at https://sites.google.com/site/christophtrebesch/data
- The **spread data** (both bond and CDS) is from **Bloomberg**.

Macroeconomic Variables

The major macroeconomic variables come from the WDI. It represents the most current and accurate global development data available, and includes national, regional and global estimates. The World Bank has created an R package to allow automated queries into their databases. I use this package in order to simplify and automate the data collection process. The indicators downloaded can be found using the URL *http* : *//data.worldbank.org/indicator/*[*indicator*] and replacing [*indicator*] with the variables indicator. For example, to view the original GDP per capita database from the world bank, the URL would be http://data.worldbank.org/indicator/NY.GDP.MKTP.CD

Indicators from WDI

- **GNI per Captia**: GNI per capita is the gross national income, converted to U.S. dollars using the World Bank Atlas method, divided by the midyear population. Corresponds to Cantor Packer 1996 variable **Per Capita Income**.
- **GDP Growth**: Annual percentage growth rate of GDP at market prices based on constant local currency. Corresponds to Cantor Packer 1996 variable **GDP Growth**.
- **Annual Inflation**: Inflation as measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly. Corresponds to Cantor Packer 1996 variable **Inflation**.
- **Port Infrastructure**: The Quality of Port Infrastructure measures business executives' perception of their country's port facilities. Scores range from 1 (port infrastructure considered extremely underdeveloped) to 7 (port infrastructure considered efficient by international standards). Will be converted to an indicator variable later. Corresponds to Cantor Packer 1996 variable **Economic Development**.
- **Current Account Surplus**: Current account balance is the sum of net exports of goods and services, net primary income, and net secondary income as a percentage of GDP. Corresponds to Cantor Packer 1996 variable **External Balance**.
- **Central Government Surplus**: Cash surplus or deficit is revenue (including grants) minus expense, minus net acquisition of nonfinancial assets as a percentage of GDP. Corresponds to Cantor Packer 1996 variable **Fiscal Balance**.
- **Exports**: Exports of goods and services represent the value of all goods and other market services provided to the rest of the world. Corresponds to denominator of the Cantor Packer 1996 variable **External Debt**.

External Debt

The external debt variable is in a seperate World Bank database. In October 2014, the World Bank launched the new Quarterly External Debt Statistics (QEDS) SDDS database. The original data can be found at

http://data.worldbank.org/data-catalog/quarterly-external-debt-statistics-ssds and the variable taken was **0059_T1_Gross External Debt Position**. This variable is used as the numerator of the Cantor Packer 1996 variable **External Debt**.

Default History

The data is from 2014 update to the databse use in the AEJ Macro 2013 paper ***Sovereign Defaults: The Price of Haircuts***, by Juan Cruces and Christoph Trebesch, which can be found at https://sites.google.com/site/christophtrebesch/data. The sample covers the full universe of sovereign debt restructurings with foreign commercial creditors (banks and bondholders) from 1970 until 2010. Based on certain selection criteria, they identify 182 sovereign debt restructurings by 68 countries since 1978 (no restructurings occurred between 1970 and 1977).

Data Manipulation

Downloading the World Bank Indicators

```r
#This script downloads the raw data files for the variables
#from the World Bank and saves them as .csv
library('WDI') #load WDI function to download WB's World Development Indicators
```

```
## Warning: package 'WDI' was built under R version 3.1.2
```

```
## Loading required package: RJSONIO
```

```r
#Sets the working directory to save our input files
setwd("~/Documents/projects/research/201502_cds/cds/data/input")

#Per Capita Income ----
#Download the per Capita GDP
GDPpercap <- WDI(indicator = "NY.GDP.MKTP.CD", start=1960, end=2014)
names(GDPpercap)[3] <- "GDPpercap"
write.csv(GDPpercap, file="GDPpercap.csv", row.names=FALSE)

#Download the per Capita GNI
GNIpercap <- WDI(indicator = "NY.GNP.PCAP.CD", start=1960, end=2014)
names(GNIpercap)[3] <- "GNIpercap"
write.csv(GNIpercap, file="GNIpercap.csv", row.names=FALSE)

#GDP growth ----
#Download the annual percentage growth of GDP
GDPgrowth <- WDI(indicator = "NY.GDP.MKTP.KD.ZG", start=1960, end=2014)
names(GDPgrowth)[3] <- "GDPgrowth"
write.csv(GDPgrowth, file="GDPgrowth.csv", row.names=FALSE)

#Inflation ----
#Download the annual inflation
inflation <- WDI(indicator = "FP.CPI.TOTL.ZG", start=1960, end=2014)
names(inflation)[3] <- "inflation"
```

```r
write.csv(inflation, file="inflation.csv", row.names=FALSE)

#Economic Development ----
#Download port infrastructure data
development <- WDI(indicator = "IQ.WEF.PORT.XQ", start=1960, end=2014)
names(development)[3] <- "development"
write.csv(development, file="development.csv", row.names=FALSE)

#External Balance ----
#Download current account surplus as a percentage of GDP
currentaccount <- WDI(indicator = "BN.CAB.XOKA.GD.ZS", start=1960, end=2014)
names(currentaccount)[3] <- "currentaccount"
write.csv(currentaccount, file="current_account.csv", row.names=FALSE)

#Fiscal Balance ----
#Download central government budget as a percentage of GDP
budget <- WDI(indicator = "GC.BAL.CASH.GD.ZS", start=1960, end=2014)
names(budget)[3] <- "budget"
write.csv(budget, file="budget.csv", row.names=FALSE)

#External debt ----
#Download exports of goods and services in current dollars
export <- WDI(indicator = "NE.EXP.GNFS.CD", start=1960, end=2014)
names(export)[3] <- "export"
write.csv(export, file="export.csv", row.names=FALSE)
```

Data Cleaning and Merging

```r
#Load necessary packages for data manipulation
library("plyr")
library("reshape2")
library("dplyr")
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library("lubridate")
```

```
## Warning: package 'lubridate' was built under R version 3.1.2
```

```
##
## Attaching package: 'lubridate'
##
## The following object is masked from 'package:plyr':
##
##     here
```

All the raw files are in an input folder, this loop takes all the files in that input folder and imports them as data tables, which is the primary data structure in `R`

```r
setwd("~/Documents/projects/research/201502_cds/cds/data/input")
#Read in all the data files in working directory as data frames----
files <- dir()
for (file in files){
  perpos <- which(strsplit(file, "")[[1]]==".")
  assign(
    gsub(" ","",substr(file, 1, perpos-1)),
    read.csv(paste(file,sep=""), stringsAsFactors=FALSE))
}
```

The two files that were not downloaded directly from the WDI database do not have the correct 2 letter country codes. The external debt data file has the 3 letter country code while the default history file was converted from an excel file to a csv. It originally had a 3 letter country code which did not fully correspond with the iso3c standard. It was manually changed to the iso3c standard when the csv conversion took place.

```r
#Two data frames don't have iso2c codes, so merging them with code vector
code_country <- select(as.data.frame(WDI_data$country), iso3c, iso2c)

defaulthistory <- inner_join(defaulthistory, code_country, by = c("iso3c"))
externaldebt <- inner_join(externaldebt, code_country, by = c("iso3c"))
```

Next, we will get the external debt data into the correct shape needed to be merged with the other files. It's original structure is the rows are countries while the columns are quarter/year. First, we will "melt" the table into having each country and quarter/year as rows with the external debt as the only column. Some cleanup will then take place. Finally, since the data is quarterly, the annual external debt amount will be the average over the four quarters.

```r
#Clean up external debt
externaldebt <- externaldebt %>%
  melt(value.name = "ext_debt") %>%
  mutate(year = as.integer(substr(variable, 2, 5))) %>%
  select(iso2c, year, ext_debt)
```

```
## Using iso3c, country, Indicator.Name, Indicator.Code, X2014Q4, iso2c as id variables
```

```
externaldebt$iso2c <- as.character(externaldebt$iso2c)
externaldebt <- aggregate(externaldebt, by=list(externaldebt$year, externaldebt$iso2c),
                          FUN=mean, na.rm=TRUE)
externaldebt <- select(externaldebt, year = Group.1, iso2c = Group.2, ext_debt)
```

Next, all the other data sets are cleaned. Mostly just selecting the variables we are interested in and renaming them.

```
#Cleanup everything else
budget <- select(budget, -country)
current_account <- select(current_account, -country)
defaulthistory <- select(defaulthistory, iso2c, year, default_date = date)
development <- select(development, -country)
export <- select(export, -country)
GDPgrowth <- select(GDPgrowth, -country)
GDPpercap <- select(GDPpercap, -country)
GNIpercap <- select(GNIpercap, -country)
inflation <- select(inflation, -country)
```

All of these data sets are then merged together.

```
#Merge data
data <- inner_join(GDPpercap, GDPgrowth, by = c("iso2c","year"))
data <- inner_join(data, inflation, by = c("iso2c","year"))
data <- inner_join(data, development, by = c("iso2c","year"))
data <- inner_join(data, GNIpercap, by = c("iso2c","year"))
data <- inner_join(data, export, by = c("iso2c","year"))
data <- inner_join(data, current_account, by = c("iso2c","year"))
data <- inner_join(data, budget, by = c("iso2c","year"))
data <- left_join(data, externaldebt, by = c("iso2c","year"))
data <- left_join(data, defaulthistory, by = c("iso2c","year"))

#Merge country info
country <- select(as.data.frame(WDI_data$country), iso2c, country)
data <- left_join(data, country, by = c("iso2c"))
```

Then the data set is exported into the output folder.

```
#Finalize data set ---
setwd("~/Documents/projects/research/201502_cds/cds/data/output")
write.csv(data, file="data.csv", row.names=FALSE)
```

Finalizing Data Prior to Ratings Data

Now, the two indicator variables, **Default History** and **Development Indicator**, are created. The **External Debt** variable is also finalized. Then we select the variables of interest to use and *voila*! We have the analytical database and can start our exploratory analysis.

```
#Create indicators
#Default - if country defaulted after 1990, output 1.
#Development - if port index > 3.5, output 1.
```

```
setwd("~/Documents/projects/research/201502_cds/cds/data/output")
data <- read.csv(file = "data.csv", head = TRUE, sep = ",",stringsAsFactors=FALSE)

data$default_date <- as.Date(data$default_date, "%m/%d/%y")
default_include <- data[year(data$default_date) > 1990,1]

data1 <- data %>%
  mutate(development_indicator = ifelse(development < 3.6 | is.na(development),0,1),
         default_indicator = ifelse(iso2c %in% default_include, 1, 0),
         external_debt = ext_debt/export * 100) %>%
  select(country, iso2c, year, GNIpercap, GDPgrowth, inflation, fiscal_balance = budget,
         external_balance = currentaccount, external_debt,
         development_indicator, default_indicator)

#Only complete cases
data1 <- tbl_df(data1)
data1 <- data1[complete.cases(data1),]

#Selecting variable of interest only
write.csv(data1, file="database_priorRatings.csv", row.names=FALSE)
```

Exploratory Analysis

How many complete cases do we have in each individual year?

```
dat <- data1 %>%
  group_by(year) %>%
  summarise(freq = n())
dat
```

```
## Source: local data frame [9 x 2]
##
##   year freq
## 1 2005   51
## 2 2006   55
## 3 2007   56
## 4 2008   61
## 5 2009   63
## 6 2010   63
## 7 2011   70
## 8 2012   67
## 9 2013    3
```

What is the composition of these completed cases?

```
dat <- data1 %>%
  group_by(country) %>%
  summarise(freq = n(),
            min_year = min(year),
```

```
            max_year = max(year))
dat
```

```
## Source: local data frame [80 x 4]
##
##                  country freq min_year max_year
## 1               Armenia    8     2005     2012
## 2             Australia    8     2005     2012
## 3               Austria    8     2005     2012
## 4           Bahamas, The    2     2011     2012
## 5            Bangladesh    1     2011     2011
## 6               Belarus    8     2005     2012
## 7               Belgium    8     2005     2012
## 8               Bolivia    2     2006     2007
## 9                Brazil    8     2005     2012
## 10             Bulgaria    8     2005     2012
## 11          Burkina Faso    3     2008     2010
## 12               Canada    9     2005     2013
## 13                Chile    3     2010     2012
## 14             Colombia    5     2008     2012
## 15           Costa Rica    5     2008     2012
## 16              Croatia    8     2005     2012
## 17               Cyprus    3     2008     2010
## 18       Czech Republic    8     2005     2012
## 19              Denmark    8     2005     2012
## 20             Dominica    2     2011     2012
## 21   Dominican Republic    1     2011     2011
## 22     Egypt, Arab Rep.    8     2005     2012
## 23          El Salvador    8     2005     2012
## 24              Estonia    8     2005     2012
## 25              Finland    8     2005     2012
## 26               France    8     2005     2012
## 27              Georgia    6     2007     2012
## 28              Germany    8     2005     2012
## 29               Greece    8     2005     2012
## 30            Guatemala    1     2012     2012
## 31 Hong Kong SAR, China    5     2005     2009
## 32              Hungary    8     2005     2012
## 33              Iceland    8     2005     2012
## 34                India    7     2006     2012
## 35            Indonesia    5     2005     2009
## 36              Ireland    8     2005     2012
## 37               Israel    8     2005     2012
## 38                Italy    8     2005     2012
## 39              Jamaica    1     2012     2012
## 40                Japan    8     2005     2012
## 41               Jordan    7     2006     2012
## 42          Korea, Rep.    7     2005     2011
## 43      Kyrgyz Republic    6     2006     2011
## 44               Latvia    7     2005     2011
## 45            Lithuania    7     2005     2011
## 46           Luxembourg    8     2005     2012
## 47        Macedonia, FYR    8     2005     2012
```

```
## 48             Malaysia   1    2012    2012
## 49               Malta    4    2008    2011
## 50            Mauritius    2    2011    2012
## 51              Moldova    8    2005    2012
## 52              Morocco    4    2009    2012
## 53                Nepal    2    2012    2013
## 54          Netherlands    8    2005    2012
## 55          New Zealand    2    2011    2012
## 56            Nicaragua    2    2011    2012
## 57              Nigeria    2    2011    2012
## 58               Norway    8    2005    2012
## 59             Paraguay    8    2005    2012
## 60                 Peru    8    2005    2012
## 61          Philippines    3    2010    2012
## 62               Poland    8    2005    2012
## 63             Portugal    8    2005    2012
## 64   Russian Federation    8    2005    2012
## 65            Singapore    4    2009    2012
## 66       Slovak Republic    8    2005    2012
## 67             Slovenia    8    2005    2012
## 68         South Africa    8    2005    2012
## 69                Spain    8    2005    2012
## 70            Sri Lanka    1    2012    2012
## 71               Sweden    8    2005    2012
## 72          Switzerland    7    2005    2011
## 73             Thailand    8    2005    2012
## 74              Tunisia    8    2005    2012
## 75               Turkey    5    2008    2012
## 76               Uganda    2    2011    2012
## 77              Ukraine    8    2005    2012
## 78       United Kingdom    8    2005    2012
## 79        United States    9    2005    2013
## 80              Uruguay    8    2005    2012
```

Incorporating Ratings Data

Firstly, in order to accomodate the two ratings agencies, we will extend our prior database.

```
data2 <- data1

data1$agency <- "S&P"
data2$agency <- "Moody's"

data1 <- rbind(data1,data2)
head(data1)
```

```
## Source: local data frame [6 x 12]
##
##    country iso2c year GNIpercap GDPgrowth inflation fiscal_balance
## 1 Armenia    AM 2012      3700   7.20000  2.555471     -1.3910222
## 2 Armenia    AM 2011      3430   4.70000  7.651234     -2.7644068
## 3 Armenia    AM 2010      3370   2.20000  8.181576     -4.9676618
## 4 Armenia    AM 2009      3240 -14.14999  3.406767     -7.6008154
```

```
## 5 Armenia     AM 2008     3510   6.90000  8.949953     -0.5427496
## 6 Armenia     AM 2007     2710  13.74920  4.407361     -0.5668837
## Variables not shown: external_balance (dbl), external_debt (dbl),
##   development_indicator (dbl), default_indicator (dbl), agency (chr)
```