

# DATA WRANGLING I

MPA 630: Data Science for Public Management

September 27, 2018

*Fill out your reading report  
on Learning Suite*

# PLAN FOR TODAY

---

iExam!

---

Reproducibility

Technical stuff

Answering your own  
questions with data

# EXAM

**B R E A K**

# REPRODUCIBILITY

# AUSTERITY AND EXCEL

---

Growth in a Time of Debt

Carmen M. Reinhart and Kenneth S. Rogoff

NBER Working Paper No. 15639

January 2010, Revised January 2010

JEL No. E2,E3,E6,F3,F4,N10

## ABSTRACT

We study economic growth and inflation at different levels of government and external debt. Our analysis is based on new data on forty-four countries spanning about two hundred years. The dataset incorporates over 3,700 annual observations covering a wide range of political systems, institutions, exchange rate arrangements, and historic circumstances. Our main findings are: First, the relationship between government debt and real GDP growth is weak for debt/GDP ratios below a threshold of 90 percent of GDP. Above 90 percent, median growth rates fall by one percent, and average growth falls considerably more. We find that the threshold for public debt is similar in advanced and emerging economies. Second, emerging markets face lower thresholds for external debt (public and private)—which is usually denominated in a foreign currency. When external debt reaches 60 percent of GDP, annual growth declines by about two percent; for higher levels, growth rates are roughly cut in half. Third, there is no apparent contemporaneous link between inflation and public debt levels for the advanced countries as a group (some countries, such as the United States, have experienced higher inflation when debt/GDP is high). The story is entirely different for emerging markets, where inflation rises sharply as debt increases.

**Debt:GDP = 90%+ → -0.1% growth**

## THE PATH TO PROSPERITY

RESTORING AMERICA'S PROMISE



FISCAL YEAR 2012 BUDGET RESOLUTION

House Committee on the Budget  
Chairman Paul Ryan of Wisconsin  
[budget.GOP.gov](http://budget.GOP.gov)

# AUSTERITY AND EXCEL

---



Finally, Ms. Reinhart and Mr. Rogoff allowed [researchers at the University of Massachusetts](#) to look at their original spreadsheet — and [the mystery of the irreproducible results was solved](#). First, they omitted some data; second, they used unusual and highly questionable statistical procedures; and finally, yes, they made an Excel coding error. Correct these oddities and errors, and you get what [other researchers have found](#): some correlation between high debt and slow growth, with no indication of which is causing which, but no sign at all of that 90 percent “threshold.”

# AUSTERITY AND EXCEL

---

Table 1. Real GDP Growth as the Level of Government Debt Varies:  
Selected Advanced Economies, 1790-2009  
(annual percent change)

Country	Period	Central (Federal) government debt/ GDP			
		Below 30 percent	30 to 60 percent	60 to 90 percent	90 percent and above
Australia	1902-2009	3.1	4.1	2.3	4.6
Austria	1880-2009	4.3	3.0	2.3	n.a.
Belgium	1835-2009	3.0	2.6	2.1	3.3
Canada	1925-2009	2.0	4.5	3.0	2.2
Denmark	1880-2009	3.1	1.7	2.4	n.a.
Finland	1913-2009	3.2	3.0	4.3	1.9
France	1880-2009	4.9	2.7	2.8	2.3
Germany	1880-2009	3.6	0.9	n.a.	n.a.
Greece	1884-2009	4.0	0.3	4.8	2.5
Ireland	1949-2009	4.4	4.5	4.0	2.4
New Zealand	1952-2009	2.5	2.5	3.5	3.6
Norway	1880-2009	2.9	4.4	n.a.	n.a.
Portugal	1851-2009	4.8	2.5	1.4	n.a.
Spain	1850-2009	1.6	3.3	1.3	2.2
Sweden	1880-2009	2.9	2.9	2.7	n.a.
United Kingdom	1830-2009	2.5	2.2	2.1	1.8
United States	1790-2009	4.0	3.4	3.3	-1.8
Average		3.7	3.0	3.4	1.7
Median		3.9	3.1	2.8	1.9
Number of observations =	2,317	866	654	445	352

**Debt:GDP = 90%+ → 2.2% growth**

# GENES AND EXCEL

---

Septin 2

Membrane-Associated  
Ring Finger (C3HC4) 1

2310009E13

Actual value	What Excel turns it into
SEPT2	2-Sep
MARCH1	1-Mar
2310009E13	2.31E+19

20% of genetics papers between 2005–2015

# GENERAL GUIDELINES

---

**Don't touch the raw data**

If you do, explain what you did!

**Use self-documenting code**

R Markdown!

**Ensure code is reproducible**

R Markdown!

**Use open formats**

Use .csv, not .xlsx

# R MARKDOWN IN REAL LIFE

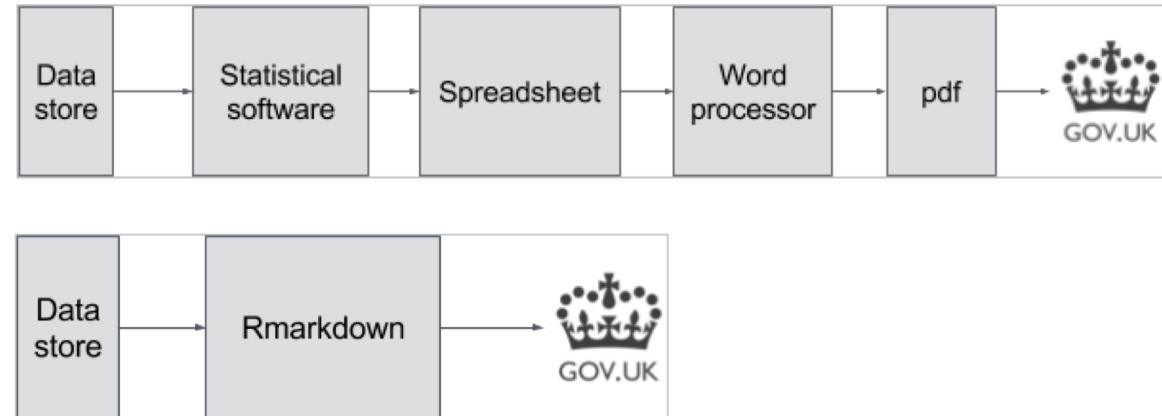
---

### 3.1.2 Data Visualization

We use `ggplot2` as our main package to create ad-hoc exploratory graphics as well as polished-looking customized visualizations. When combined with tools to clean and transform data, `ggplot2` allows analysts to quickly translate insights into high quality, compelling visualizations. In addition to the static graphics of `ggplot2`, we often make interactive visualizations or dashboards using R packages such as `plotly` (Sievert et al. 2017), `leaflet` (Cheng et al. 2017), `dygraphs` (Vanderkam et al. 2017), `DiagrammeR` (Sveidqvist et al. 2017), and `shiny` (Chang et al. 2017).

### 3.1.3 Reproducible Research

At Airbnb, all R analyses are documented in `rmarkdown`, where code and visualizations are combined within a single written report. Posts are carefully reviewed by experts in the content area and techniques used, both in terms of methodologies and code style, before publishing and sharing with the business partners. The peer review process is



# TECHNICAL STUFF

# LEFT\_JOIN

`left_join(x, y)`

1	x1	1	y1
2	x2	2	y2
3	x3	4	y4

# LEFT\_JOIN

County	Year	Books
<chr>	<dbl>	<dbl>
Utah	2017	50000
Salt Lake	2017	100000
Weber	2017	30000
Utah	2018	55000
Salt Lake	2018	150000
Weber	2018	35000

6 rows

County	City
<chr>	<chr>
Utah	Provo
Salt Lake	Salt Lake City
Weber	Ogden
Juab	Nephi

4 rows

```
libraries %>%  
  left_join(county_seats, by = "County")
```

# LEFT\_JOIN

<b>County</b> <chr>	<b>Year</b> <dbl>	<b>Books</b> <dbl>	<b>City</b> <chr>
Utah	2017	50000	Provo
Salt Lake	2017	100000	Salt Lake City
Weber	2017	30000	Ogden
Utah	2018	55000	Provo
Salt Lake	2018	150000	Salt Lake City
Weber	2018	35000	Ogden

6 rows

# LEFT\_JOIN

County <chr>	Year <dbl>	Books <dbl>
Utah	2017	50000
Salt Lake	2017	100000
Weber	2017	30000
Utah	2018	55000
Salt Lake	2018	150000
Weber	2018	35000

6 rows

County <chr>	Year <dbl>	Lions <dbl>
Utah	2017	5
Salt Lake	2017	17
Weber	2017	20
Juab	2017	10
Utah	2018	6
Salt Lake	2018	20
Weber	2018	22
Juab	2018	14

8 rows

```
libraries %>%  
  left_join(lions, by = c("County", "Year"))
```

# LEFT\_JOIN

<b>County</b> <chr>	<b>Year</b> <dbl>	<b>Books</b> <dbl>	<b>Lions</b> <dbl>
Utah	2017	50000	5
Salt Lake	2017	100000	17
Weber	2017	30000	20
Utah	2018	55000	6
Salt Lake	2018	150000	20
Weber	2018	35000	22

6 rows

# R I G H T \_ J O I N

`right_join(x, y)`

1	x1
2	x2
3	x3

1	y1
2	y2
4	y4

# RIGHT\_JOIN

County <chr>	Year <dbl>	Books <dbl>
Utah	2017	50000
Salt Lake	2017	100000
Weber	2017	30000
Utah	2018	55000
Salt Lake	2018	150000
Weber	2018	35000

6 rows

County <chr>	Year <dbl>	Lions <dbl>
Utah	2017	5
Salt Lake	2017	17
Weber	2017	20
Juab	2017	10
Utah	2018	6
Salt Lake	2018	20
Weber	2018	22
Juab	2018	14

8 rows

```
libraries %>%  
  right_join(lions, by = c("County", "Year"))
```

# RIGHT\_JOIN

County <chr>	Year <dbl>	Books <dbl>	Lions <dbl>
Utah	2017	50000	5
Salt Lake	2017	100000	17
Weber	2017	30000	20
Juab	2017	NA	10
Utah	2018	55000	6
Salt Lake	2018	150000	20
Weber	2018	35000	22
Juab	2018	NA	14

8 rows

# LEFT\_JOIN

County <chr>	Year <dbl>	Books <dbl>
Utah	2017	50000
Salt Lake	2017	100000
Weber	2017	30000
Utah	2018	55000
Salt Lake	2018	150000
Weber	2018	35000

6 rows

County <chr>	Year <dbl>	Lions <dbl>
Utah	2017	5
Salt Lake	2017	17
Weber	2017	20
Juab	2017	10
Utah	2018	6
Salt Lake	2018	20
Weber	2018	22
Juab	2018	14

8 rows

```
libraries %>%
```

```
  left_join(lions, by = c("County", "Year")) %>%
  left_join(county_seats, by = "County")
```

# LEFT\_JOIN

<b>County</b> <code>&lt;chr&gt;</code>	<b>Year</b> <code>&lt;dbl&gt;</code>	<b>Books</b> <code>&lt;dbl&gt;</code>	<b>Lions</b> <code>&lt;dbl&gt;</code>	<b>City</b> <code>&lt;chr&gt;</code>
Utah	2017	50000	5	Provo
Salt Lake	2017	100000	17	Salt Lake City
Weber	2017	30000	20	Ogden
Utah	2018	55000	6	Provo
Salt Lake	2018	150000	20	Salt Lake City
Weber	2018	35000	22	Ogden

6 rows

# COMMON PROBLEM SET ISSUES

---

.ZIP files

Working directories  
and projects

**ANSWERING YOUR  
OWN QUESTIONS  
WITH DATA**