

# From Data to Information

Big Data in Finance

Roberto Rigobon

MIT Sloan, NBER, CSAC



Big Distance!



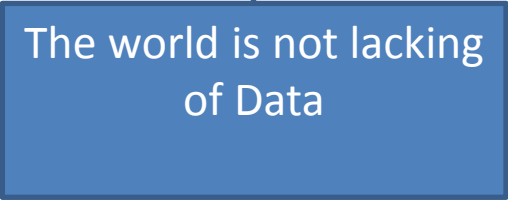
Data



Information




Knowledge



The world is not lacking  
of Data

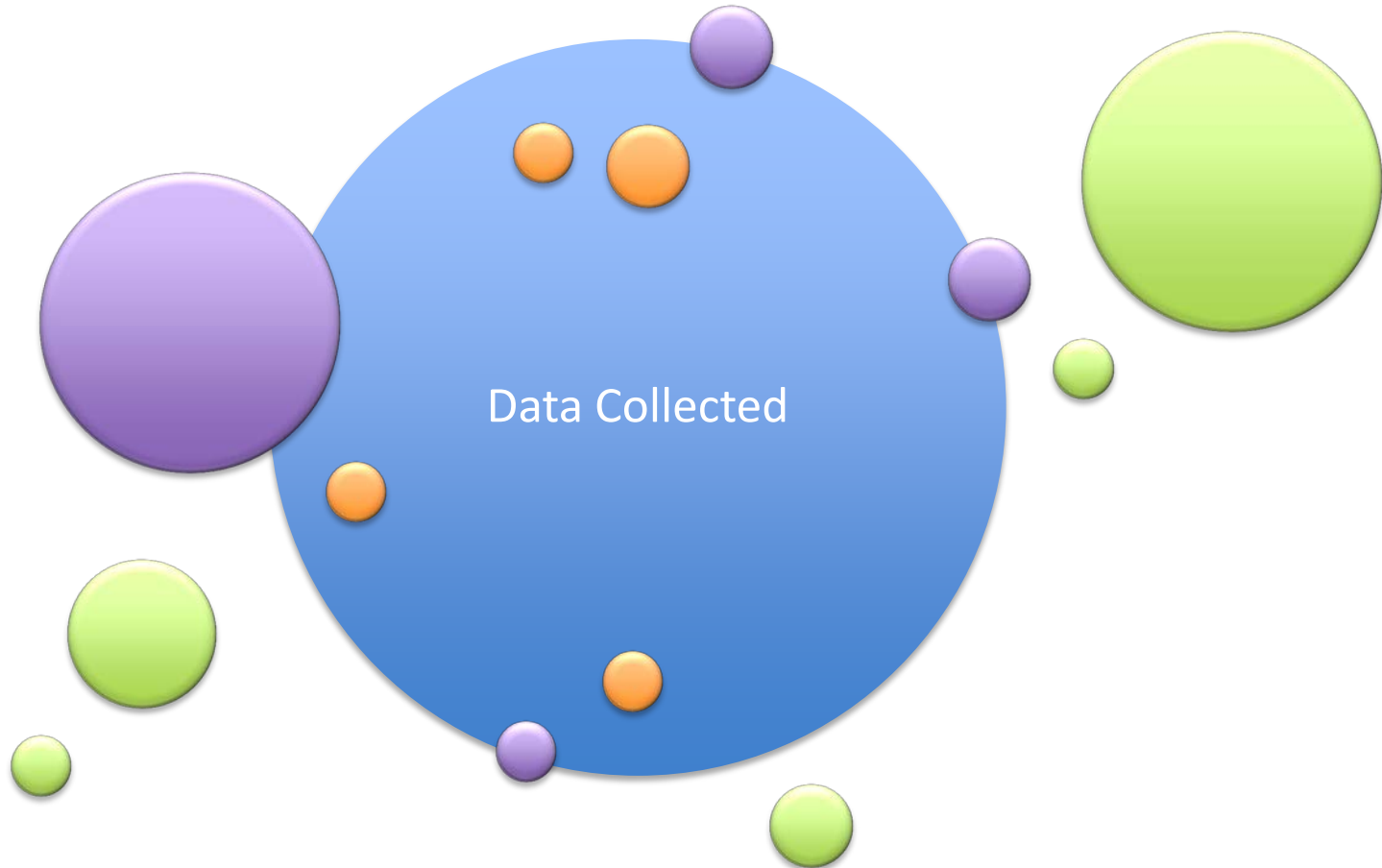


Lacking of Careful  
Empirics

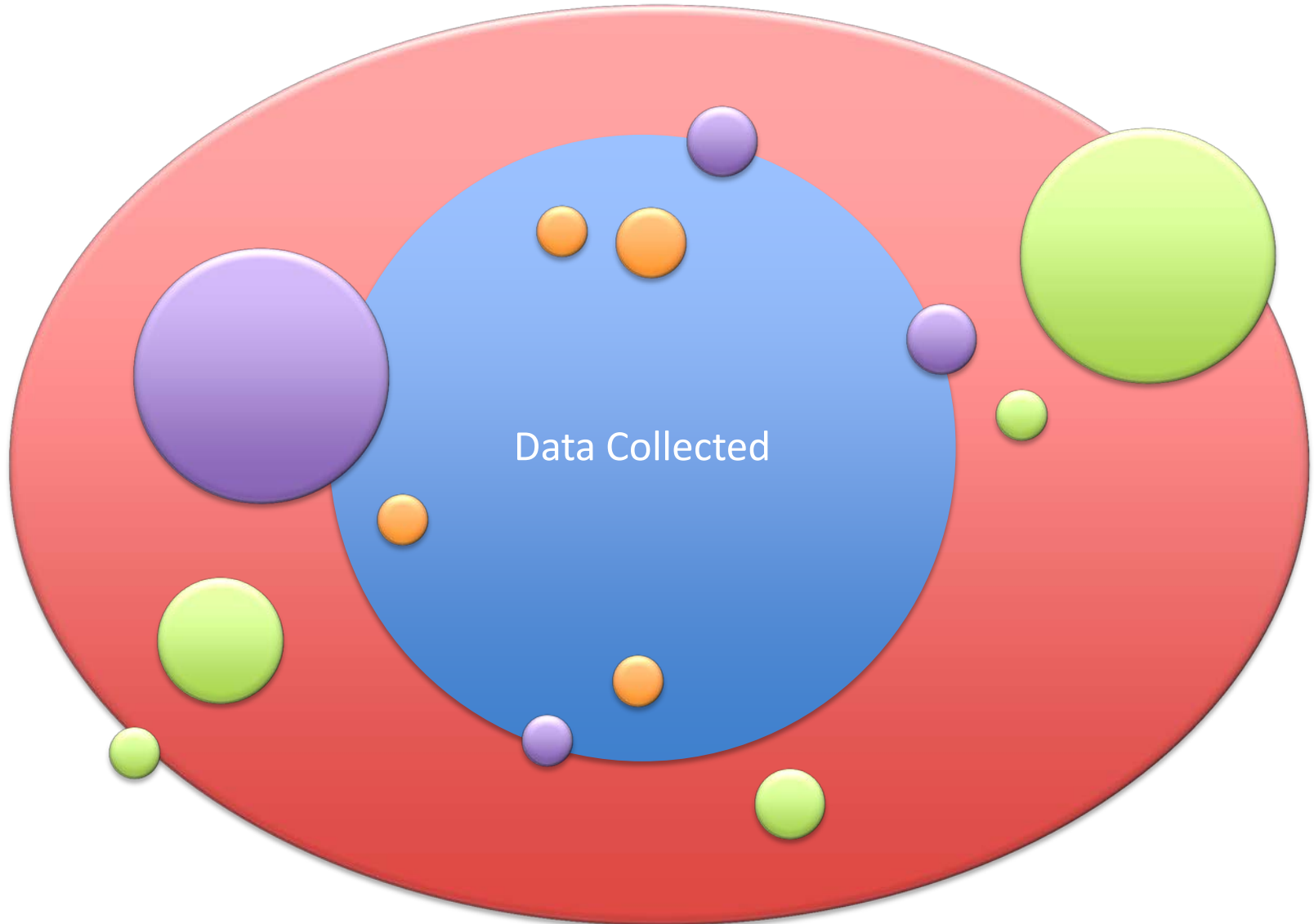


Lacking of Managerial  
Data Analysis

# Official Data Sources



# Big Data



# Data Types

Type	Advantages	Disadvantages
<ul style="list-style-type: none"><li>• Official</li></ul>	<ul style="list-style-type: none"><li>• Representative</li><li>• High Quality</li></ul>	<ul style="list-style-type: none"><li>• Limited</li><li>• Collection is Costly</li></ul>
<ul style="list-style-type: none"><li>• Big Data</li></ul>	<ul style="list-style-type: none"><li>• Size, details, frequency</li><li>• Might solve some identification problems</li><li>• Might improve measurement</li></ul>	<ul style="list-style-type: none"><li>• Statistics are harder</li><li>• Not representative</li><li>• Sample Selection</li></ul>

# New Home Sales

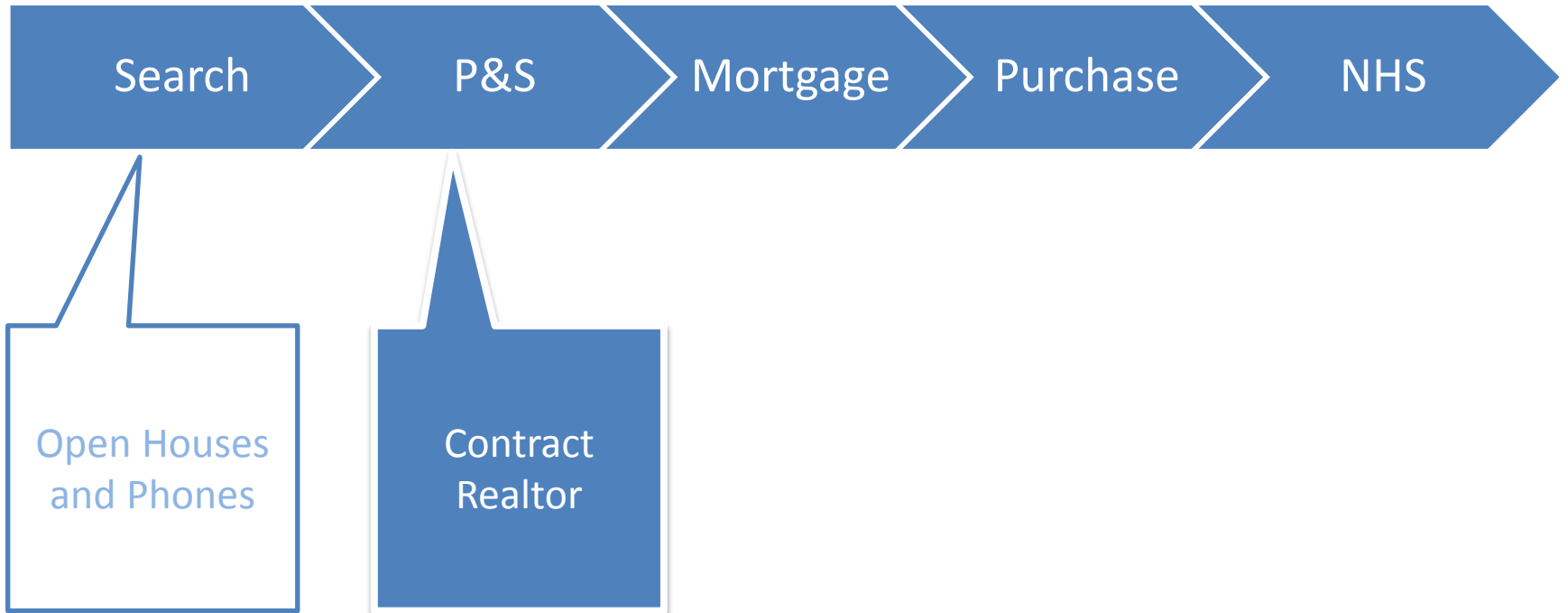


# New Home Sales



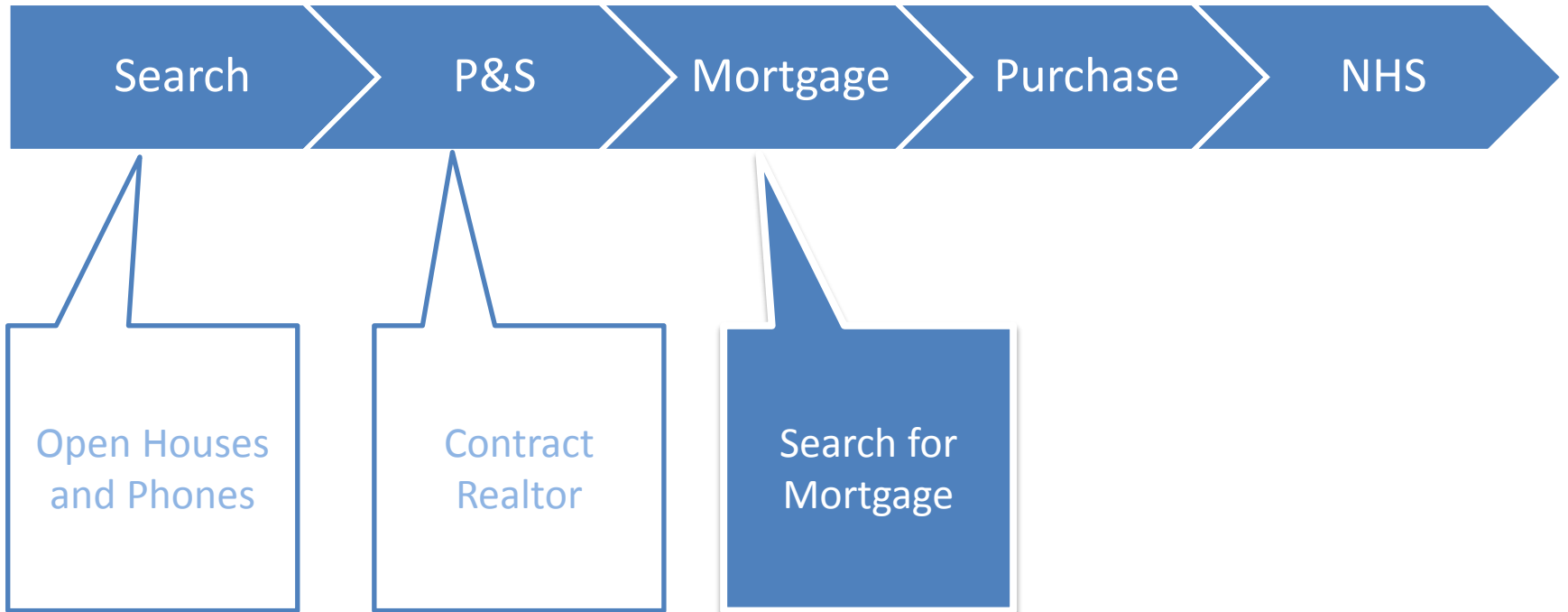
Open Houses  
and Phones

# New Home Sales





# New Home Sales



# Inflation

- Official Consumer Price Index

# Online Information and Indexes

## Our Approach to Daily Inflation Statistics

1

*Use scraping technology*



2

*Connect to thousands of online retailers every day*



3

*Find individual items*



4

*Store and process key item information in a database*

- *Date*
- *Item*
- *Price*
- *Description*

5

*Develop daily inflation statistics for ~20 countries*

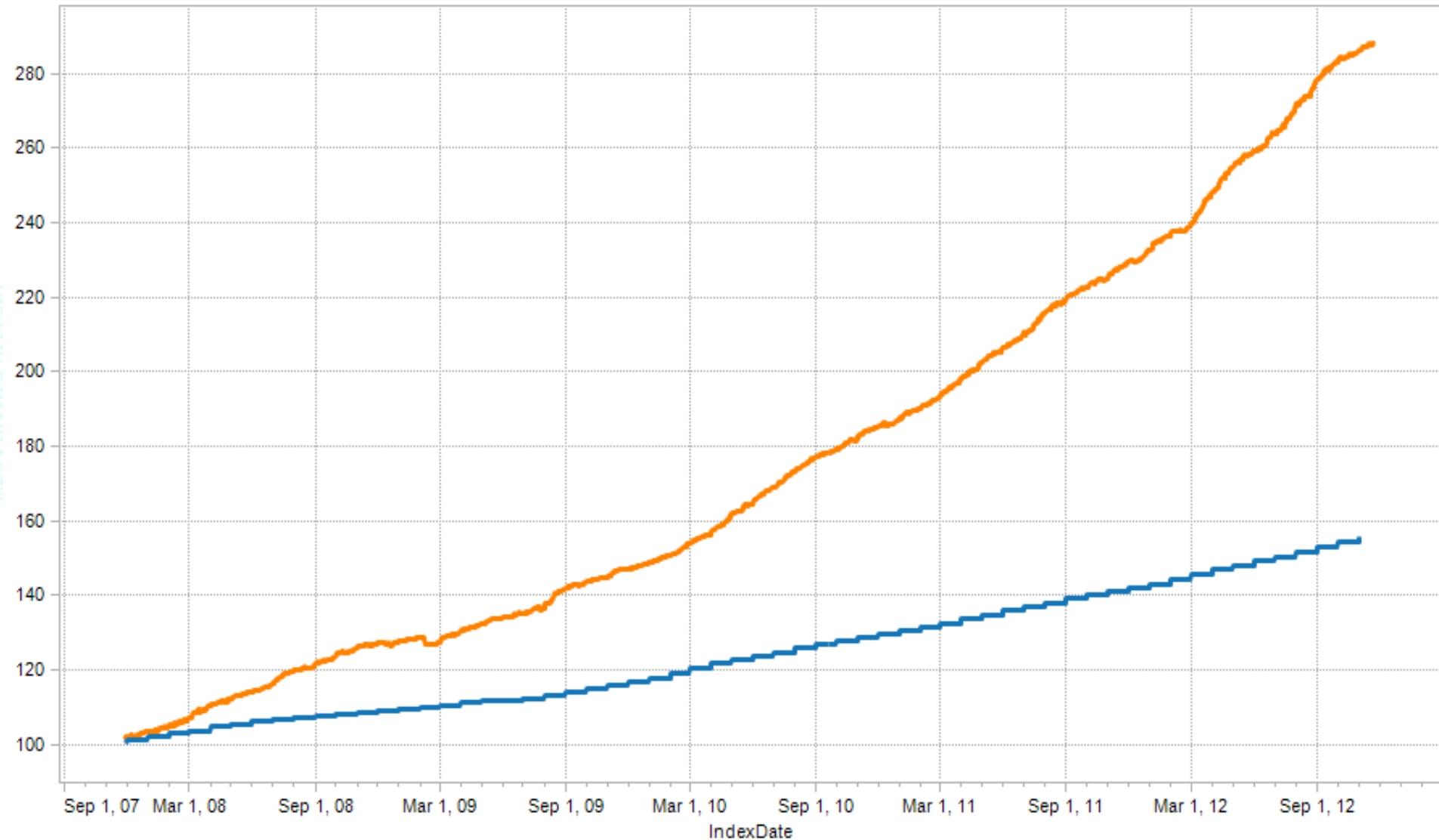


# Countries covered

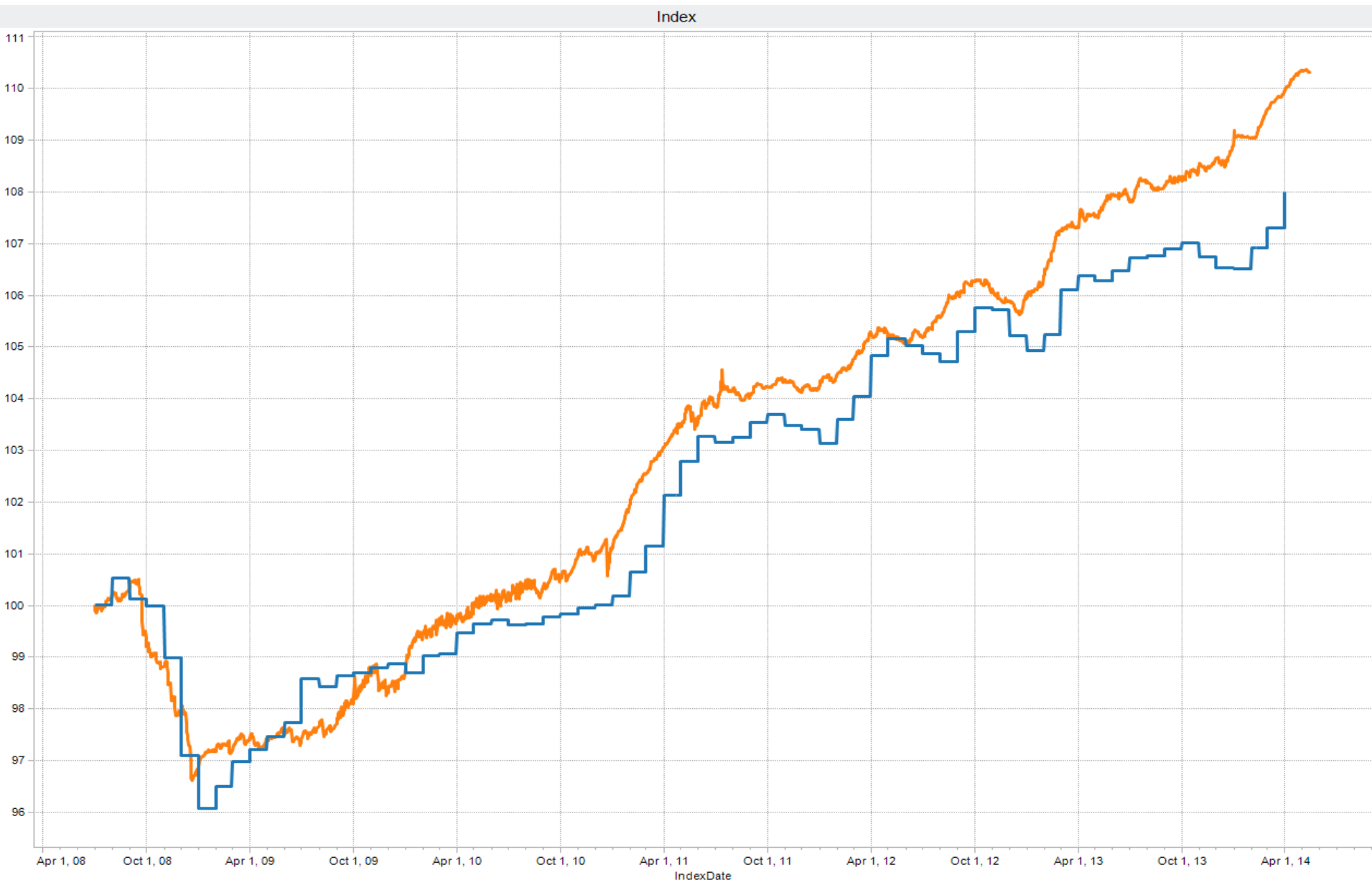


# Argentina <http://www.inflacionverdadera.com>

Index



# USA (<http://bpp.mit.edu/usa/>)

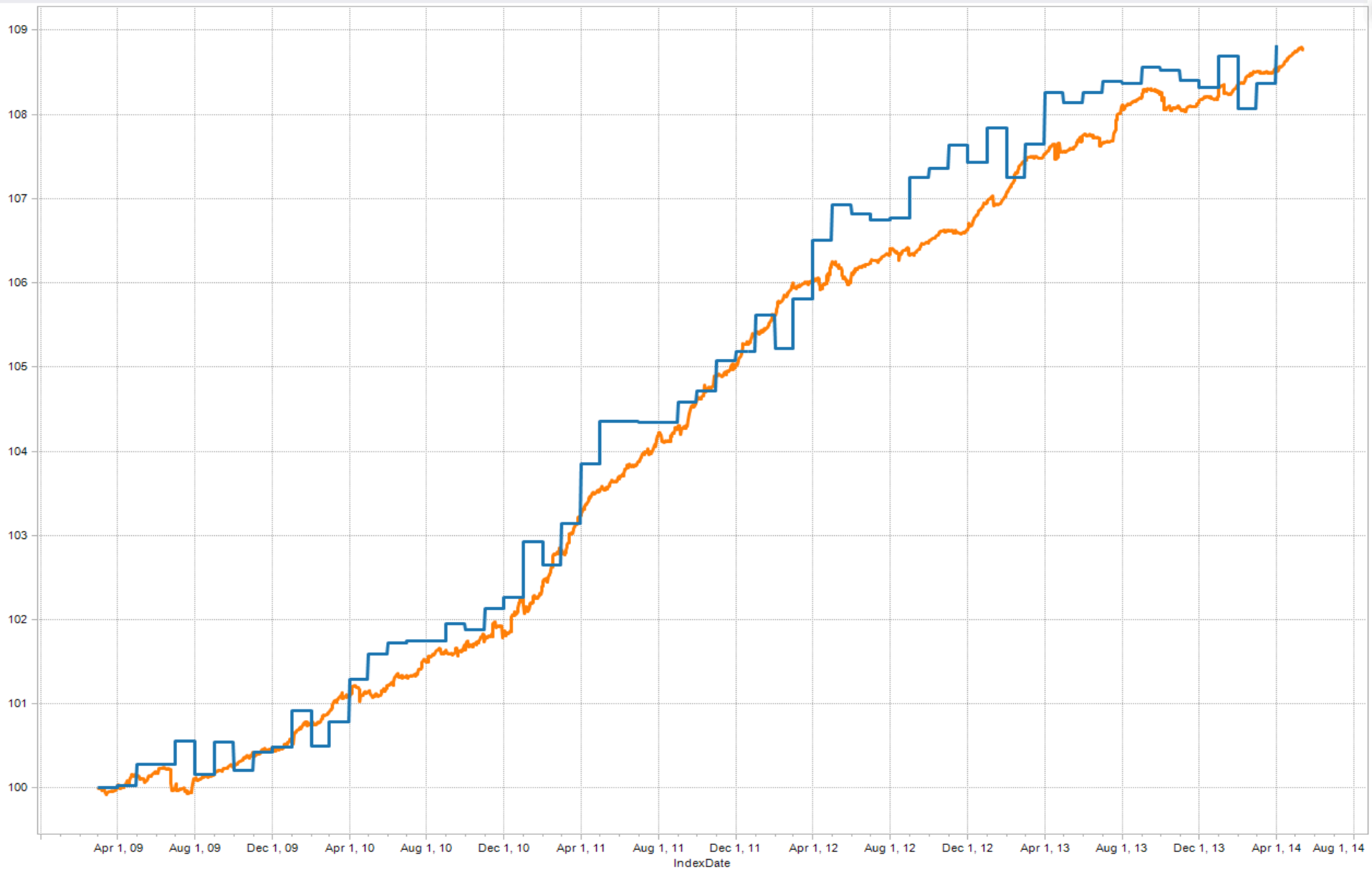


# Catching up!



# EuroZone

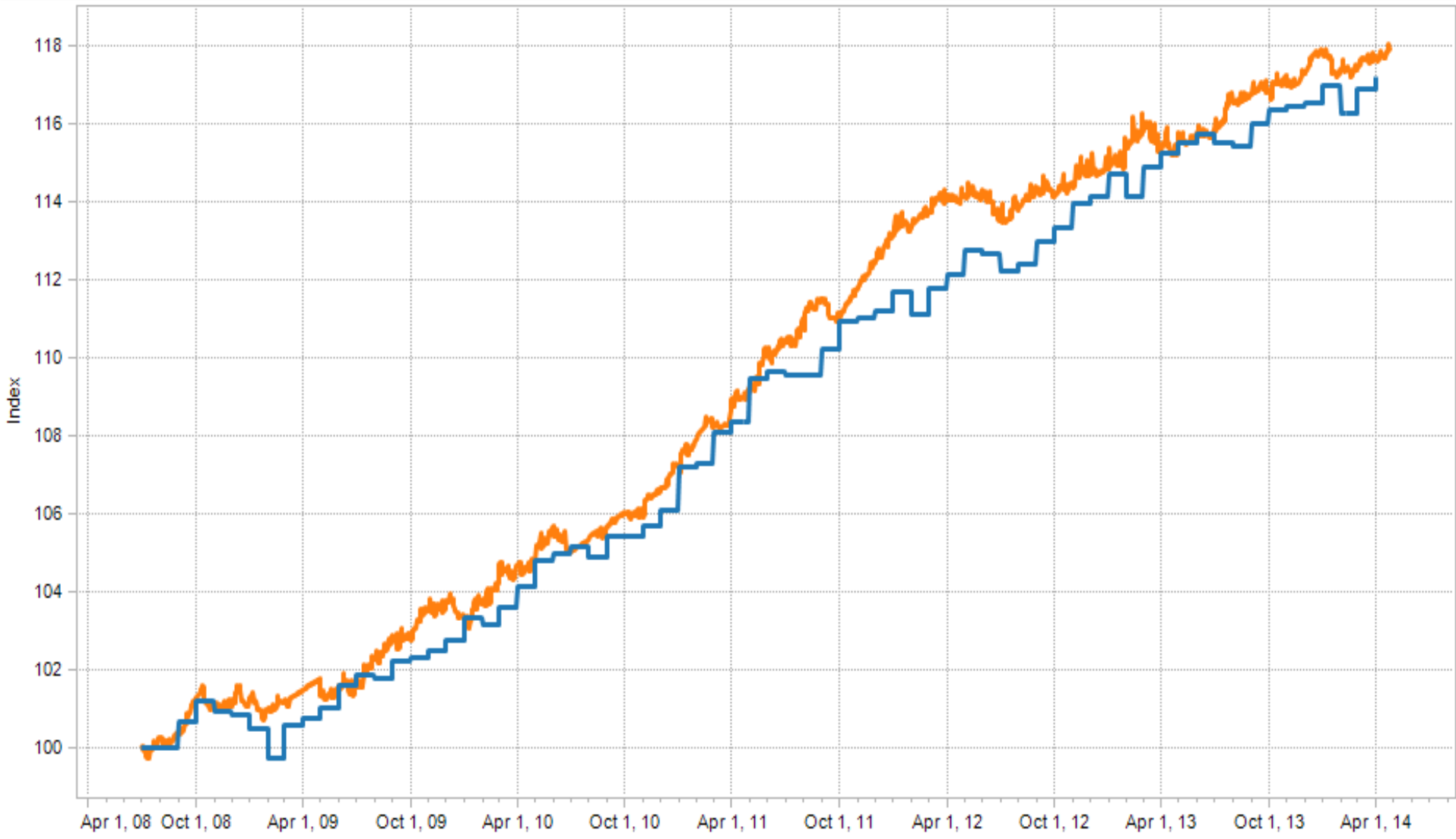
Index



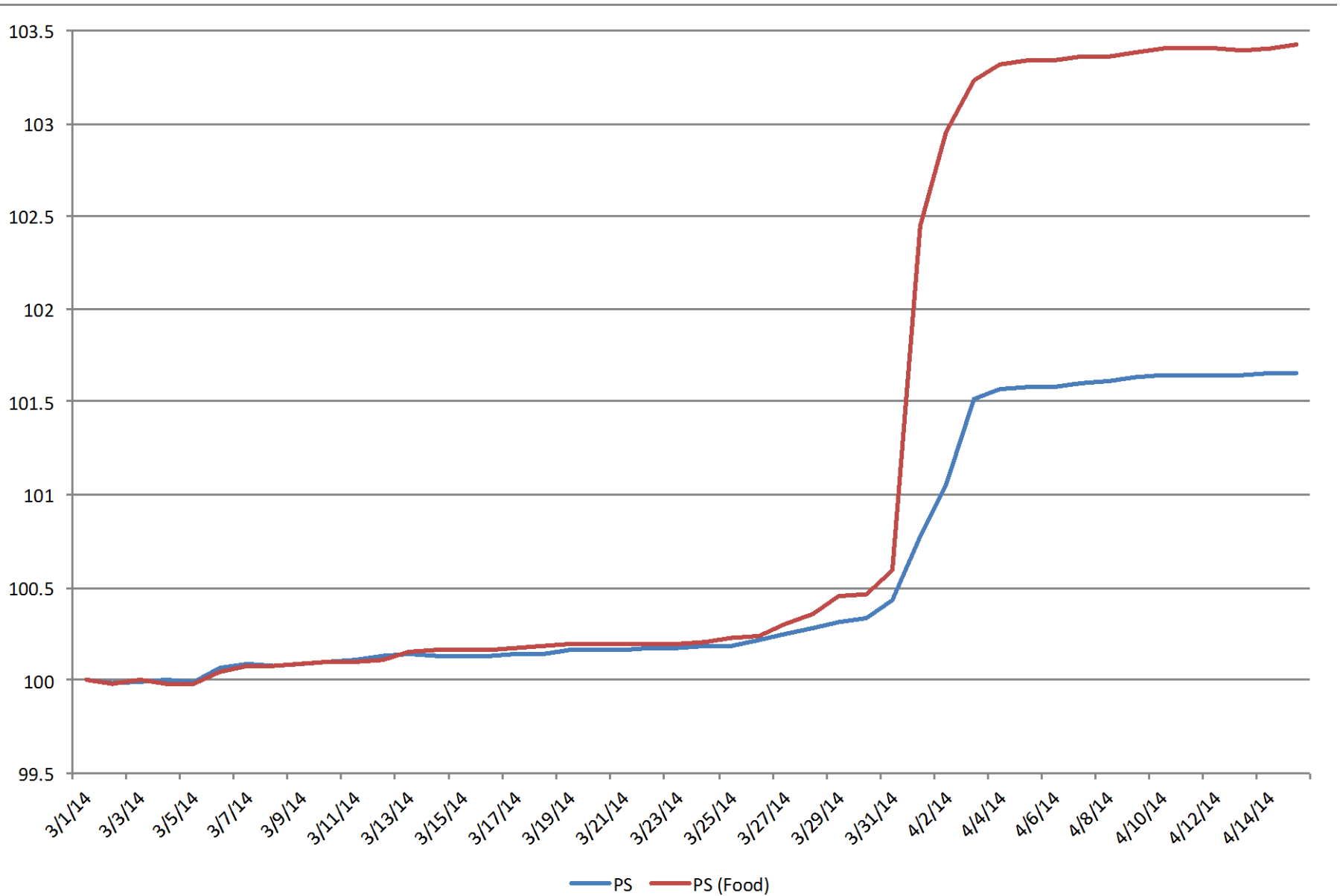


# UK

Index



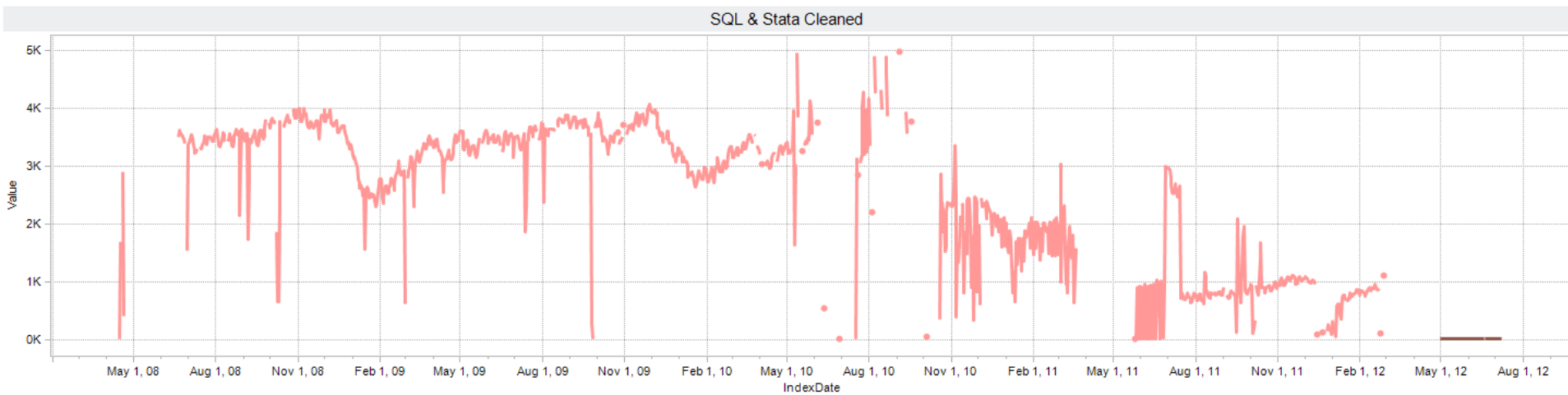
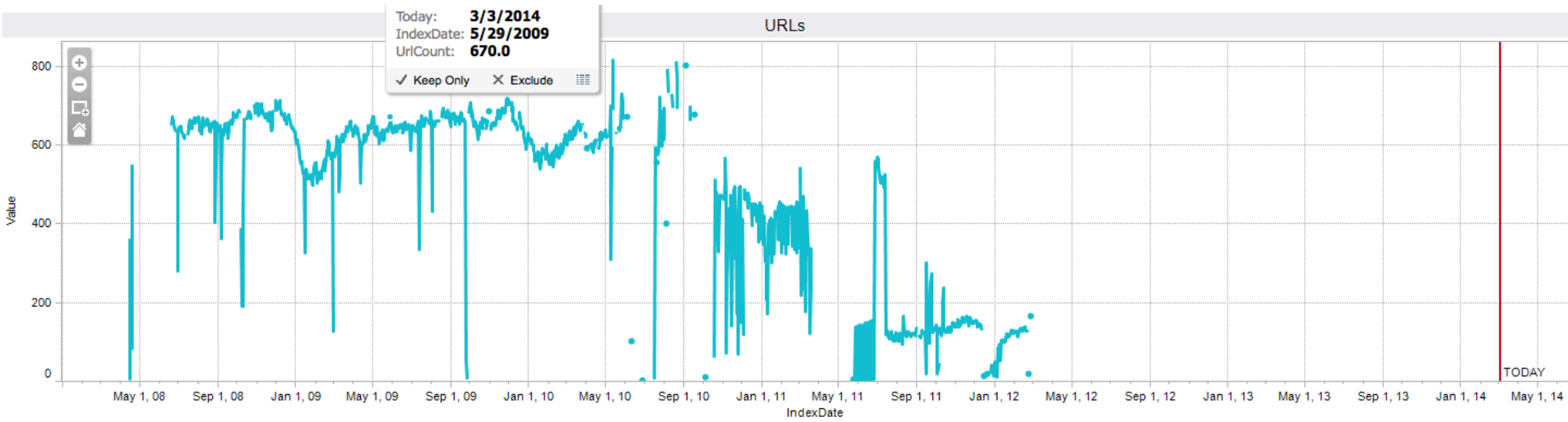
# Japan Tax



# What is next?

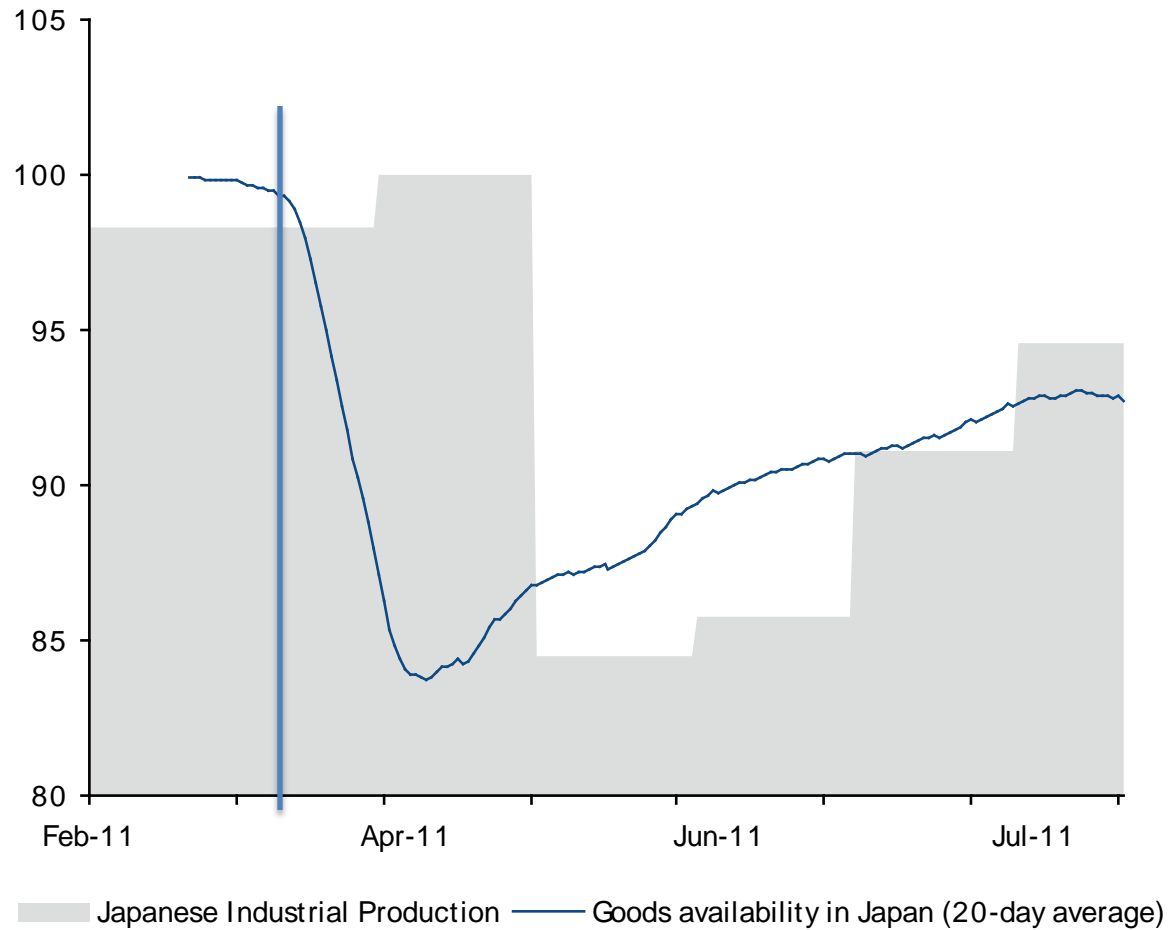
- Economic Activity
- Purchasing Power Parity
- Labor Market Conditions
- Real Estate

# Scarcity



# Japan

Gauging shortages in Japanese goods



# PPP indicators

Compare prices for a bottle of Coke  
across countries

- The objective of the PPP series is to help determine whether currencies are under or overvalued over time
- Online prices represent an effective tool to measure PPP fluctuations
  - Identical items sold around the world
  - Detailed descriptions to achieve a nearly perfect matching
  - Daily Prices

# Methodology

EXAMPLES OF PRODUCT CATEGORIES

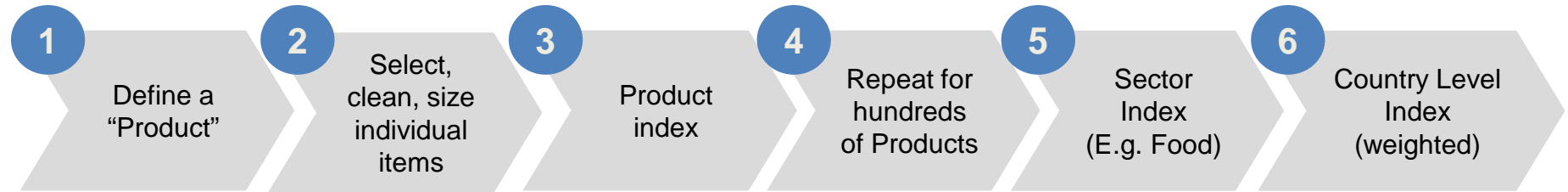
1

Define a  
“Product”

- Very narrow product definition
- Must be available in multiple countries
- Branded and Unbranded categories

FOOD	ELECTRONICS
Coffee_Illy_Ground (excluding decaf)	Television_Samsung_LED 32inch basic
Coffee_Regular_Ground (excluding decaf)	Television_Samsung_LED 32inch All Other (including Full HD, Smart, 3D)
Coffee_Regular_Beans (excluding decaf)	Television_Samsung_LED 40-43inch basic
Coffee_Decaf	Television_Samsung_LED 40-43inch All Other (including Full HD, Smart, 3D)
Coffee_All Other	Television_Sony_LED 32inch basic
Ketchup_Heinz_Regular	Television_Sony_LED 32inch All Other (including Full HD, Smart, 3D)
Ketchup_Heinz_Low Sodium, no salt	Television_Sony_LED 40-43inch basic
Ketchup_Heinz_All Other (e.g., flavored)	Television_Sony_LED 40-43inch All Other (including Full HD, Smart, 3D)
Ketchup_All Other_Regular	Television_Sony_LED 44-47inch All Other (including Full HD, Smart, 3D)
Ketchup_All Other_All Other (e.g., flavored)	Television_LG_LED 32inch basic
Soy Sauce_All Other_Regular	Television_LG_LED 32inch All Other (including Full HD, Smart, 3D)
Soy Sauce_All Other_Low Sodium, no salt, light	Television_LG_LED 40-43inch basic
.....	.....

# Methodology



- Very narrow product definition
- Must be available in multiple countries
- Branded and Unbranded categories

- Dozens of items per "product" in each country
- Different retailers, brands, and sizes

- Product availability varies across countries and time
- If a good is not available in the US, it will not appear in our series

- Some sectors are cheaper, others more expensive

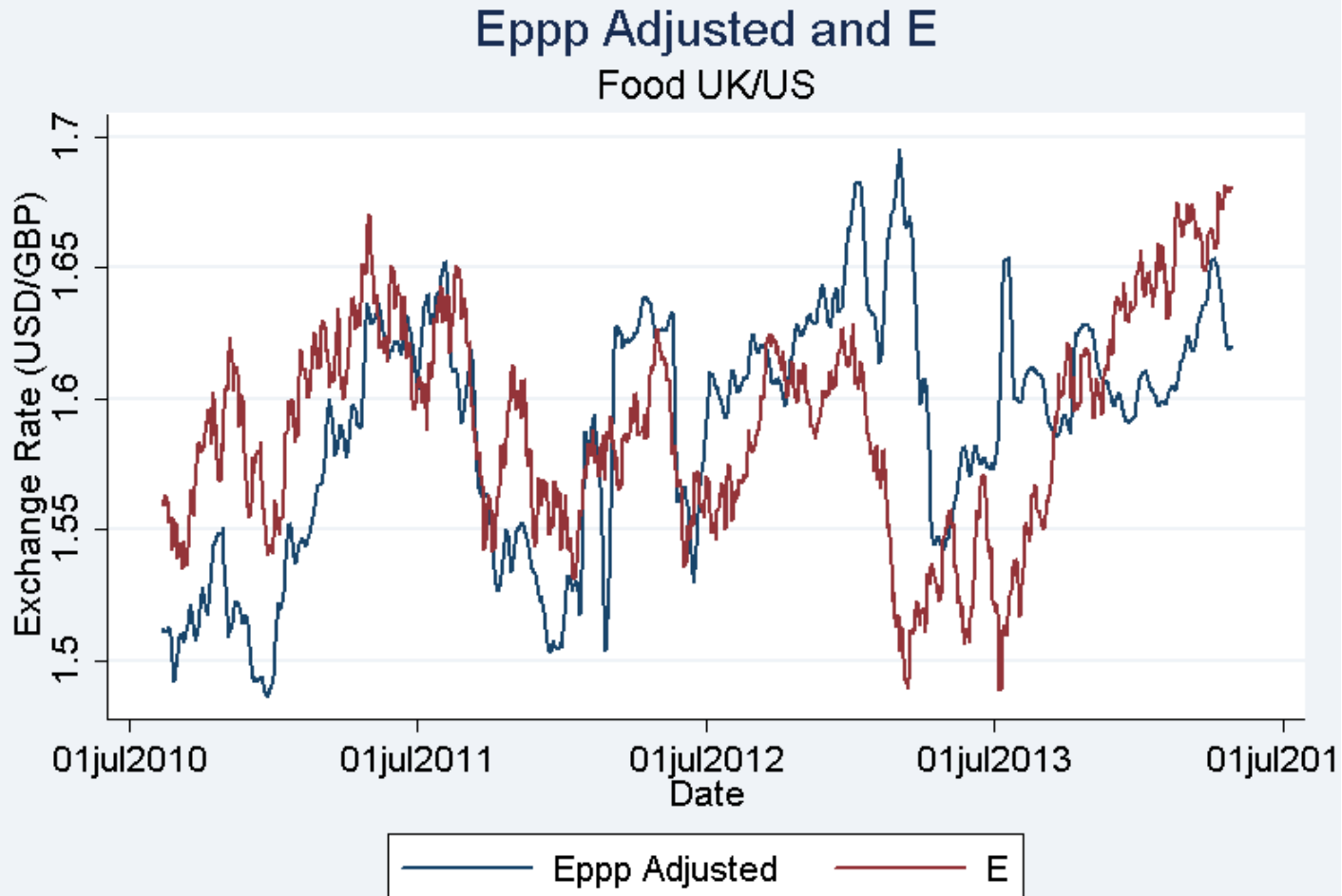
- Use to compare Eppp and E



# PPP-Exchange Rate (Adjusted)

6

Country – Level Index



Eppp Adjusted by substracting mean difference Eppp and E

# What is my (our?) objective?

- Change the way statistical offices measure economic indicators
  - Price is the first step
  - Scarcity and PPP is the second one
  - International Trade, Economic activity, Labor Markets, Consumer Confidence, Real Estate... and GDP....
  - Well, I just need... one billion prices... Well, actually...



*One trillion!!!!*

