dbnomics

Stata client for DBnomics, the world's economic database

Simone Signore

EIF (European Investment Bank Group), s.signore@eif.org

2019 London Stata Users Group meeting London, September 5–6, 2019

This version: September 2019

Disclaimer

This presentation should not be referred to as representing the views of the European Investment Fund (EIF) or of the European Investment Bank Group (EIB Group). Any views expressed herein, including interpretation(s) of regulations, reflect the current views of the author, which do not necessarily correspond to the views of the EIF or of the EIB Group. Views expressed herein may differ from views set out in other documents, including similar research papers, published by the EIF or by the EIB Group. Contents of this presentation, including views expressed, are current at the date of publication set out above, and may change without notice. No representation or warranty, express or implied, is or will be made and no liability or responsibility is or will be accepted by the EIF or by the EIB Group in respect of the accuracy or completeness of the information contained herein and any such liability is expressly disclaimed. Nothing in this presentation constitutes investment, legal, or tax advice, nor shall be relied upon as such advice. Specific professional advice should always be sought separately before taking any action based on this presentation. Reproduction, publication and reprint are subject to prior written authorisation.

Outline

- The DBnomics platform
- Stata's dbnomics command
- Use cases
- Why should you use dbnomics?

The DBnomics platform The birth of the project

Ways to improve data handling processes in macroeconomic research:

- Simplifying retrieval of economic data from multiple (public) data sources;
- Automatically providing updated data;
- Allowing reproducible results.

In 2016: the idea of DBnomics, born out of a partnership between the CEPREMAP and France Stratégie (and the financial support of the Investments for the Future Programme).

The DBnomics platform

Selected economic database aggregators

Fred:

- Simple user Interface;
- Archive system (ALFRED);
- Limited coverage of public data sources (2/3 of the 589,000 series come from the U.S. Census, BLS, BEA, FED);
- Free, but not open-source.

Quandl:

- Simple user Interface;
- Wide(r) coverage of public data source (UNO, BIT, national institutes, etc.), but not systematic within each data provider;
- Free for some services only.

Datastream:

- Heavy user interface;
- Good (i.e., automatised) updating system, especially for financial series;
- Costly.

The DBnomics platform Objectives

Goal: Create a free, *open-source* (and Euro-centric) server to aggregate publicly-available data series provided by national and international statistical institutions.

Four important principles of the project:

- Data series are taken directly from providers and kept unchanged;
- Data series are stored in a tree similar to the provider's;
- Data series are automatically updated via provider-specific functions;
- Archive system: each revision of the data series is archived.

Value added:

- A unique economic database with wide, systematic coverage of economic data (605 million series at present);
- The *free and open-source* nature¹ aims at facilitating the creation of a community: https://forum.db.nomics.world/.

¹GNU Affero General public License.

Figure 1: Public data providers covered, by quarter

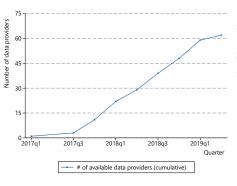
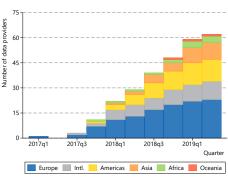


Figure 2: Data providers by geographic scope



The DBnomics platform

A free platform designed for every user

How to use it?

- A website: https://db.nomics.world/ with different search capabilities;
- A RESTful²API (Application Programming Interface) based on the JSON (JavaScript Object Notation) data exchange format: https://api.db.nomics.world;
- The API allows for automated database access from statistical packages like:
 - Python/R;
 - Julia;
 - Gretl;
 - Stata.

²REpresentational State Transfer, a set of recommendations defining a flexible and lightweight architectural style for machine-to-machine communication (e.g., server to client).

- Stata routine to browse, find and extract DBnomics data series;
- Composed of seven sub-commands, handling the various endpoints of its API;
- Mata back-end to parse JSON responses, using the libjson library (Lindsley, 2012b);
- Available from SSC (ssc install dbnomics);
- Source code available on github.com.

Explore and search for data across multiple providers:

- Browse available data providers (dbnomics providers);
- ► Load the "table of contents" for a provider of interest (dbnomics tree);
- ► Search for data across providers (dbnomics <u>find</u>).

Explore dataset structure, contents and format:

- Browse dataset structure (dbnomics <u>datastructure</u>);
- ▶ Load list of series related to a dataset (dbnomics series).

Load data into Stata:

- Import data from DBnomics (dbnomics import).
- Show recently updated data (dbnomics <u>news</u>).

- The design of dbnomics is heavily inspired by sdmxuse (Fontenay, 2018) for the sub-command structure and Stata's freduse for the handling of metadata;
- **However,** Fontenay's sdmxuse relies on the SDMX standard, which conveniently allows to "slice" and filter data through "SDMX masks", e.g.:

```
. sdmxuse data ECB, clear dataset(SAFE) dimensions(H.T2.SME.A.O.O.O.Q4.FFAC..AL.WP)
```

- Not all DBnomics providers support SDMX (many do!);
- How to make the following query Stata-friendly?

```
https://api.db.nomics.world/v21/series?provider_code=ECB&dataset_code=SAFE&limit=500&offset=0&dimensions="REF_AREA":["T2"], "SAFE_QUESTION":
["Q4"], "FIRM_SIZE":["SME"], "FIRM_SECTOR":["A"], "FIRM_AGE":["O"],
"SAFE_DENOM":["WP"], "SAFE_FILTER":["AL"], "SAFE_ITEM":["FFAC", "FEQI",
"FBLN", "FLEH", "FOVD", "FTCR"]
```

 Solution: turn dimension labels into option names and dimension filters into contents of the options:

• **Advantage:** queries are stored (e.g., in a do-file) in a human-readable format and the command retains a Stata-like syntax.

Can Stata parse the following JSON payload? Hint: yes.

```
" meta": {
  "python_project_name": "DBnomics-API",
  "python_project_version": "0.21.6"
"series": {
  "data": [
      "Firm age (SAFE)": "All ages included",
      "Firm economic activity (SAFE)": "All sectors",
      "period": [
        "2009-S1",
        "2009-S2".
        ... (omitted)
      "SAFE question": "Q4. Financing structure",
      "FIRM_SIZE": "SME",
      "series_name": "Half-yearly ...",
      "FIRM_TURNOVER": "O",
      "FIRM SECTOR": "A".
      "Frequency": "Half-vearly".
      "SAFE answer": "Not applicable to the firm",
       "period_start_day": [
         "2009-01-01",
         "2009-07-01".
        ... (omitted)
      "FIRM_AGE": "O",
      "series_code": "H.T2.SME.A.O.O.O.Q4.FBLN.N7.AL.WP",
      "Reference area": "Euro area countries ...".
```

```
"SAFE_FILTER": "AL",
    "@frequency": "bi-annual",
    "SAFE_QUESTION": "Q4",
    "SAFE_ITEM": "FBLN",
     "value": [
       3.858787.
       10.14487,
      ... (omitted)
    "SAFE question related item": "Bank loan".
    "SAFE filter - applicable answer": "Including ...
    "Firm other breakdowns (ownership, export) (SAFE)
    "dataset code": "SAFE".
    "SAFE DENOM": "WP".
    "Firm turnover (SAFE)": "All turnover ...".
    "REF AREA": "T2".
    "Denomination in SAFE context": "Weighted ...".
    "provider_code": "ECB",
    "SAFE ANSWER": "N7".
    "FIRM OWNERSHIP": "O".
    "dataset_name": "Survey on the Access ...",
    "Firm size (SAFE)": "Small and medium-sized ...".
    "FREQ": "H"
  7
"limit": 1,
"num_found": 36,
"offset": 0
```

Ι,

The dbnomics command Challenge #2: parsing JSON (2/3)

- Lindsley's insheetjson unfortunately far too limited to handle DBnomics' JSON;
- Buchanan's jsonio offers plenty of capabilities, but requires external dependencies (i.e. the Jackson Java library), a rather laborious set-up and a non-trivial RegEx-based syntax for JSON consumption;
- In line with the project's spirit, I wished to build an open-source tool, minimising the number of necessary external dependencies;
- **Solution?** Lindsley's libjson Mata library more than meets the eye:
 - Mata's libjson class comfortably handles most JSON read operations;
 - Custom parsing Mata functions (e.g. json2table, fetchjson) help navigate through and convert JSON payloads;
 - dbnomics's parsing yields a table structure that replicates the output of DBnomics' own official Python client (pip install dbnomics);
 - ▶ On the flip side, libjson is not the fastest JSON library out there...
- The command moss (Picard and Cox, 2011) deals with the necessary residual string cleanup (e.g. encoding).

The dbnomics command Challenge #2: parsing JSON (3/3)

36 series found and imported

. browse period-firm_size in 1/20

period	period_start_day	value	firm_age_safe	firm_economic_ activity_safe	safe_question	firm_size
2009-S1	2009-01-01	3.858787	All ages included	All sectors	Q4. Financing structure	SME
2009-S2	2009-07-01	10.14487	All ages included	All sectors	Q4. Financing structure	SME
2010-S1	2010-01-01	35.15501	All ages included	All sectors	Q4. Financing structure	SME
2010-S2	2010-07-01	34.84572	All ages included	All sectors	Q4. Financing structure	SME
2011-S1	2011-01-01	37.09875	All ages included	All sectors	Q4. Financing structure	SME
2011-S2	2011-07-01	35.04251	All ages included	All sectors	Q4. Financing structure	SME
2012-S1	2012-01-01	35.26051	All ages included	All sectors	Q4. Financing structure	SME
2012-S2	2012-07-01	34.46809	All ages included	All sectors	Q4. Financing structure	SME
2013-S1	2013-01-01	32.81848	All ages included	All sectors	Q4. Financing structure	SME
2013-S2	2013-07-01	32.96257	All ages included	All sectors	Q4. Financing structure	SME
2014-S1	2014-01-01	42.37593	All ages included	All sectors	Q4. Financing structure	SME
2014-S2	2014-07-01	45.33803	All ages included	All sectors	Q4. Financing structure	SME
2015-S1	2015-01-01	47.79228	All ages included	All sectors	Q4. Financing structure	SME
2015-S2	2015-07-01	44.68392	All ages included	All sectors	Q4. Financing structure	SME
2016-S1	2016-01-01	45.37493	All ages included	All sectors	Q4. Financing structure	SME
2016-S2	2016-07-01	46.51465	All ages included	All sectors	Q4. Financing structure	SME
2017-S1	2017-01-01	46.38163	All ages included	All sectors	Q4. Financing structure	SME
2017-S2	2017-07-01	47.19019	All ages included	All sectors	Q4. Financing structure	SME
2018-S1	2018-01-01	48.41314	All ages included	All sectors	Q4. Financing structure	SME
2018-S2	2018-07-01	49.46145	All ages included	All sectors	Q4. Financing structure	SME

A brief walk-through

How can dbnomics simplify the retrieval of macroeconomic data? (1/4)

- The European Small Business Finance Outlook (ESBFO, Kraemer-Eis et al., 2019) gives a bi-annual snapshot of the access to finance conditions of European SMEs;³
- The report compiles vast amount of data from public data sources. For instance, the ESBFO tracks the sources of external financing for Euro area SMEs. Let's use dbnomics to find relevant information;

```
. dbnomics find "external financing", clear
Searching for external financing in datasets and series...5 results found.
```

Type	Provider name	Name	Code	Nb series	Nb match~g	Indexed at
dataset dataset	Banco Central do Brasil Banque De France	Survey on the Access to Fina- V.2 - Balance of current tra-i Bank Lending Survey Access to Finance of SMEs		189501 6 187 4	59565 6 3 2	01sep2019 29aug2019 24ju12019 14jun2019
	Statistics Canada n a highlighted link to loa	Central government operation~bd related data)	10100133	24	1	30 jul2019

- Looks like our first result hits the mark. However, 189,501 series are probably more than we need, so we first browse the dataset structure in search for useful filters;
- By the way, you can click on the SAFE link to directly load the dataset structure.

³Small and Medium-sized Enterprises.

A brief walk-through

How can dbnomics simplify the retrieval of macroeconomic data? (2/4)

- . dbnomics data, pr(ECB) d(SAFE) clear
- Survey on the Access to Finance of SMEs 189501 series found. Order of dimensions: (FREQ.REF_AREA.FIRM_SIZE.FIRM_SECTOR.FIRM > _TURNOVER.FIRM_AGE.FIRM_OWNERSHIP.SAFE_QUESTION.SAFE_ITEM.SAFE_ANSWER.SAFE_FILTER > .SAFE_DENOM)
 - Browsing through the available series, we identify the following as relevant:
- . (intermediate ouput omitted)
- . list dimensions-seriesnr if filter > 0, abbrev(50) noobs sep(0)

++					
dimensions	values	labels	seriesnr		
FIRM_AGE	0	All ages included	116921		
FIRM_SECTOR	A	All sectors	131393		
FIRM_SIZE	SME	Small and medium-sized enterprises	30639		
REF_AREA	U2	Euro area (changing composition)	79854		
SAFE_DENOM	WP	Weighted percentage of responses	138041		
SAFE_FILTER	AL	Including not applicable responses	149632		
SAFE_QUESTION	Q4	Q4. Financing structure	15432		
SAFE_ITEM	FBLN	Bank loan	10230		
SAFE_ITEM	FEQI	Equity investments in your firm	7579		
SAFE_ITEM	FFAC	Factoring	1144		
SAFE_ITEM	FLEH	Leasing or hire-purchase	7579		
SAFE_ITEM	FOVD	Credit line, bank overdraft or credit cards overdraft	10230		
SAFE_ITEM	FTCR	Trade credit	10230		

• Before downloading the data (and to make sure we are downloading the right stuff), we can create a "shopping cart" of series using the command dbnomics series:

```
. dbnomics series, pr(ECB) d(SAFE) SAFE_QUESTION(Q4) SAFE_FILTER(AL) ///
FIRM_SIZE(SME) FIRM_SECTOR(A) FIRM_AGE(O) SAFE_DENOM(WP) REF_AREA(T2) ///
SAFE_ITEM("FFAC", "FEQI", "FBLN", "FLEH", "FOVD", "FTCR") clear
```

• We are happy with our selection, we can now proceed to loading the data in memory:

36 of 189501 series selected. Order of dimensions: ... (ouput omitted)

```
. dbnomics import, pr(ECB) d(SAFE) SAFE_QUESTION(Q4) SAFE_FILTER(AL) ///
FIRM_SIZE(SME) FIRM_SECTOR(A) FIRM_AGE(O) SAFE_DENOM(WP) REF_AREA(T2) ///
SAFE_ITEM("FFAC","FEQI","FBLN","FLEH","FOVD","FTCR") clear
```

36 series found and imported

A brief walk-through

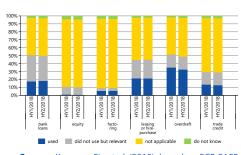
How can dbnomics simplify the retrieval of macroeconomic data? (4/4)

 At last, our chart (Kraemer-Eis et al., 2019, p.18): download do-file:

```
. /* Download data */
. dbnomics import, pr(ECB) d(SAFE) REF AREA(U2) SAFE QUESTION(Q4) ///
     FIRM_SIZE(SME) FIRM_SECTOR(A) FIRM_AGE(O) SAFE_DENOM(WP) ///
     SAFE FILTER(AL) ///
     SAFE ITEM("FFAC", "FEGI", "FBLN", "FLEH", "FGVD", "FTCR") clear
36 series found and imported
. /* Few lines of code to parse and prepare the data */
. quietly dbnomics_chart_prepare
. /* Graph twoway bar and rbar */
. oui mylabels 0(10)100, myscale(@/100) suffix("%") local(vyal)
. oui levelsof count if !missing(nperiod), local(xlab1) clean
. qui levelsof showitem if !missing(itemstr), local(xlab2) clean
. qui levelsof tcount, local(xtick1) clean
. oui levelsof tcount if !missing(itemstr[ n+1]), local(xtick2) clean
. #delimit :
. twoway (bar share0 count, barw(0.75) color("51 51 153")
     fintensity(100)) (rbar share1 share0 count, barw(0.75) color("102
     102 178") fintensity(100)) (rbar share2 share1 count, barw(0.75)
     color("255 192 0") fintensity(100)) (rbar share3 share2 count,
     barw(0.75) color("166 166 166") fintensity(100)) (scatter nothing
     showitem, xaxis(2)) , scheme(eif4 official) xscale(axis(2) alt
     noline outergap(*-0.5)) xlab('xlab1', valuelab angle(vertical)
     axis(1) noticks) xlab('xlab2', valuelab angle(hor) axis(2) noticks
     labsize(*0.85)) xsize(7) vlabel(`vval', angle(hor)
     glpattern(solid)) legend(order(4 "do not know" 3 "not applicable" 2
     "did not use but relevant" 1 "used" ) symxsize(*0.35) pos(3)
     cols(1) region(lcolor(none))) xtitle("", axis(1)) xtitle("".
     axis(2)) plotregion(margin(sides)) xtick('xtick1' 18.5, nolabels
     tpos(outside) tlength(*12) axis(1)) xtick("xtick2" 18.5, add
     nolabels tpos(cross) tlength(*8) axis(2)) name(figure15_esbfo,
```

.
gr export "\\${WORK}/chapters/03_results/figures/figure15_esbfo.pdf", ///
as(pdf) name(figure15_esbfo) replace

Figure 3: Sources of external financing of Euro area SMEs



Source: Kraemer-Eis *et al.* (2019), based on ECB SAFE (ECB, 2019).

replace); . #delimit cr

⁽file C:\Users\SIGNORE\Desktop\Presentations\Stata_UG_meeting_UK2019/
> chapters/03_results/figures/figure15_esbfo.pdf written in PDF format)

• Over 605 million time series from 62 data sources at your fingertips:

Provider Name	Website	Geographic area
African Development Group	https://www.afdb.org/en/	Africa
Annual macro-economic database of the European Commission's Directorate General for Economic and Financial Affairs	http://ec.europa.eu/economy_finance/ameco/user/serie/SelectSerie.cfm	Europe
Banco Central do Brasil	http://www.bcb.gov.br	BR
Banque Centrale des Etats de l'Afrique de l'Ouest	http://www.bceao.int	Africa
Banque De France	http://webstat.banque-france.fr/fr/home.do	FR
U.S. Bureau of Economic Analysis	http://www.bea.gov	US
Bank Indonesia	https://www.bi.go.id/en/	ID
Bank for International Settlements	https://www.bis.org/	International
U.S. Bureau of Labor Statistics	https://www.bls.gov/	US
Bank of England	http://www.bankofengland.co.uk/	GB
Bank of Japan	https://www.boj.or.jp/en/index.htm/	JP
Bundesbank	https://www.bundesbank.de/	DE
Congressional Budget Office macro economic database	https://www.cbo.gov/about/products/budget-economic-data	US
Centre d'études prospectives et d'informations internationales	http://www.cepii.fr/	International
Central Statistics Office of Ireland	https://www.cso.ie/	IE
Direction de l'Animation de la Recherche des Etudes et des Statistiques	https://dares.travail-emploi.gouv.fr/	FR
Federal Statistical Office Germany	https://www-genesis.destatis.de/genesis/online	DE
Direction de la recherche, des études, de l'évaluation et des statistiques	http://solidarites-sante.gouv.fr/ministere/organisation/directions/article/drees-direction- de-la-recherche-des-etudes-de-l-evaluation-et-des-statistiques	FR
European Central Bank	https://www.ecb.europa.eu/	Europe
U.S. Energy Information Agency	https://www.eia.gov/	US
Helenic Statistical Authority	http://www.statistics.gr/en/home/	GR
Economic and Social Research Institute, Cabinet Office, Government of Japan	http://www.esri.cao.go.jp/index-e.html	JP
Eurostat	http://ec.europa.eu/eurostat/home	Europe
Food and Agriculture Organization of the United Nations	http://www.fao.org/faostat/en/	International
Federal Reserve Board of Governors	https://www.federalreserve.gov/	US
Federal Housing Finance Agency	https://www.fhfa.gov/	US
Groningen Growth and Development Center, University of Groningen	https://www.rug.nl/ggdc/productivity/pwt/	International
International Labour Organization	https://www.ilo.org	International
International Monetary Fund	https://www.imf.org/	International
Instituto Nacional de Estadistica y Censos	https://www.indec.gob.ar/	AR
Instituto Nacional de Estadistica	http://www.ine.es/	ES

Let's recap

Why should you use dbnomics? (2/3)

(list of providers continued)

Provider Name	Website	Geographic area
Instituto Nacional de Estadística y Geografía	https://www.inegi.org.mx/	MX
Statistics Portugal	https://www.ine.pt	PT
National Institute of Statistics and Economic Studies	https://insee.fr/	FR
Institute for Supply Management	https://www.instituteforsupplymanagement.org	US
Italian National Institute of Statistics	https://www.istat.it/en/	IT
The London Bullion Market	http://www.lbma.org.uk	GB
Ministry of Economy, Trade and Industry	http://www.meti.go.jp/english/	JP
Ministry of Statistics and Programme Implementation	http://www.mospi.gov.in	IN
U.S. National Association of Realtors	https://www.nar.realtor/	US
National Bank of Belgium Online statistics	http://stat.nbb.be/	BE
National Bureau of Statistics of China	http://data.stats.gov.cn/english/	CN
Organisation for Economic Co-operation and Development	http://www.oecd.org/	International
Office for National Statistics	https://www.ons.gov.uk	GB
Pôle Emploi	http://www.pole-emploi.org/opendata/	FR
Reserve Bank of Australia	https://www.rba.gov.au/	AU
Russian Federation Federal State Statistics Service	http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/en/main/	RU
State Administration of Foreign Exchange	http://www.safe.gov.cn/en/	CN
China Africa Research Initiative	http://www.sais-cari.org/	Africa
Saudi Arabian Monetary Authority	http://www.sama.gov.sa/	SA
South African Reserve Bank	https://www.resbank.co.za/	ZA
Statistics Sweden	https://www.scb.se/	SE
Surveys of Consumers, University Michigan	https://data.sca.isr.umich.edu/	US
State Secretariat for Economic Affairs, Switzerland	https://www.seco.admin.ch/seco/en/home.html	CH
Statistics Canada	https://www.statcan.gc.ca/	CA
Statistics Japan, Statistics Bureau, Ministry of Internal Affairs and Communication	http://www.stat.go.jp/english/	JP
Statistics Poland	https://stat.gov.pl/en/	PL
Türkiye Cumhuriyet Merkez Bankasi	http://www.tcmb.gov.tr/	TR
United Nation Conference on Trade and Development	https://unctad.org/en/Pages/Home.aspx	International
United Nations	http://data.un.org/	International
World Bank	http://www.worldbank.org/	International
World Trade Organization	https://www.wto.org/	International

Let's recap

Why should you use dbnomics? (3/3)

A command that adapts to the specificities of each data provider...

 Store queries (e.g., in a do-file) in a human-readable format, using Stata's native option syntax;

...requiring minimal dependencies

ssc install libjson and ssc install moss;

Limitations:

dbnomics is not a bulk download tool. There is a hard-coded limit of 500 maximum downloadable series (which can be lifted via the limit() option). In any case, the server only yields a maximum of 1000 series per query; here are some work-arounds

To dos:

- Update the dbnomics command to support version 22 of the DBnomics API (released December 2018);
- Add option to request specific versions of time series (both from the front-end and back-end);
- Bugs? s [dot] signore at eif [dot] org (or signoresimone at yahoo [dot] it);
- Want to help? Fork https://dreameater89.github.io/dbnomics/.

References I

- Buchanan, B. (2015). JSONIO: Stata module for I/O operations on JSON data. Statistical Software Components, Boston College Department of Economics.
- ECB (2019). Survey on the access to finance of enterprises in the Euro Area. October 2018 to March 2019. [data file and code book]. Retrieved from https://www.ecb.europa.eu/stats/ecb_surveys/safe/html/index.en.html.
- Fontenay, S. (2018). sdmxuse: Command to import data from statistical agencies using the SDMX standard. Stata Journal, vol. 18, no. 4, pp. 863–870.
- Kraemer-Eis, H., Botsari, A., Gvetadze, S., Lang, F. and Torfs, W. (2019). European small business finance outlook, June 2019. EIF Working Paper 2018/57, EIF Research & Market Analysis.
- Lindsley, E. (2012a). INSHEETJSON: Stata module for importing tabular data from JSON sources on the internet. Statistical Software Components, Boston College Department of Economics.
- ——— (2012b). LIBJSON: Stata module to provide Mata class library for obtaining and parsing JSON strings into object trees. Statistical Software Components, Boston College Department of Economics.
- Picard, R. and Cox, N.J. (2011). MOSS: Stata module to find multiple occurrences of substrings. Statistical Software Components, Boston College Department of Economics.