Implementing Reproducibility and Transparency in Central Banks

Sylvérie Herbert
Banque de France

Banco de Portugal
December 15, 2021

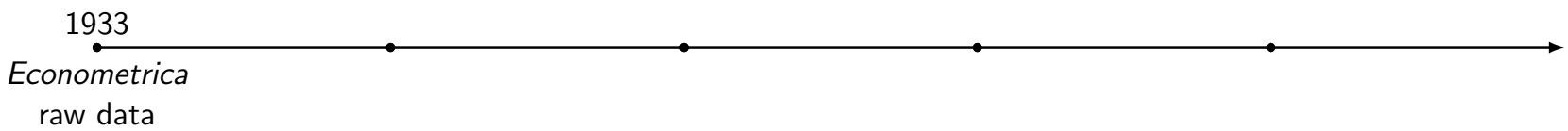
Credibility of research in economics

- Transparency and openness are essential to the credibility of economics research
- Even more important for policy making because research informs economic policies
- Evidence-based policy making with reliance on academic research
- Research needs to be trustworthy, hence reliable

Peer review and reproducibility

- Peer review ensures high quality and original research, but zero obligation to reproduce results from the codes and data

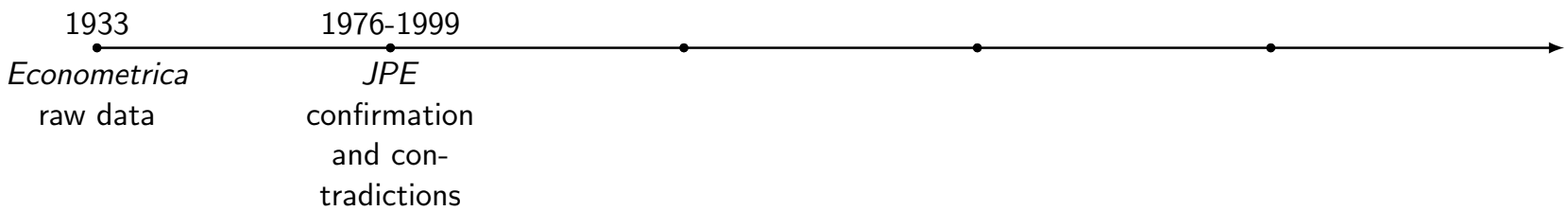
- Data availability policies to promote transparency



Peer review and reproducibility

- Peer review ensures high quality and original research, but zero obligation to reproduce results from the codes and data

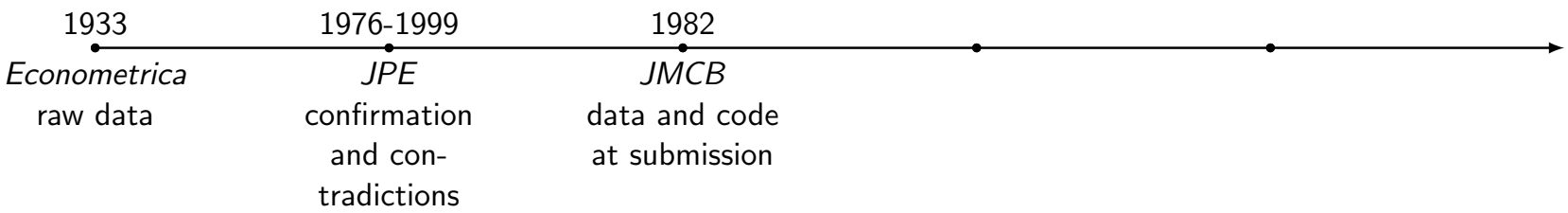
- Data availability policies to promote transparency



Peer review and reproducibility

- Peer review ensures high quality and original research, but zero obligation to reproduce results from the codes and data

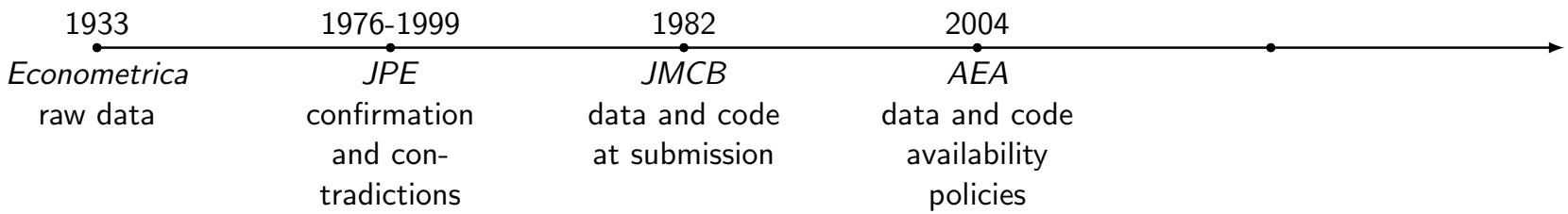
- Data availability policies to promote transparency



Peer review and reproducibility

- Peer review ensures high quality and original research, but zero obligation to reproduce results from the codes and data

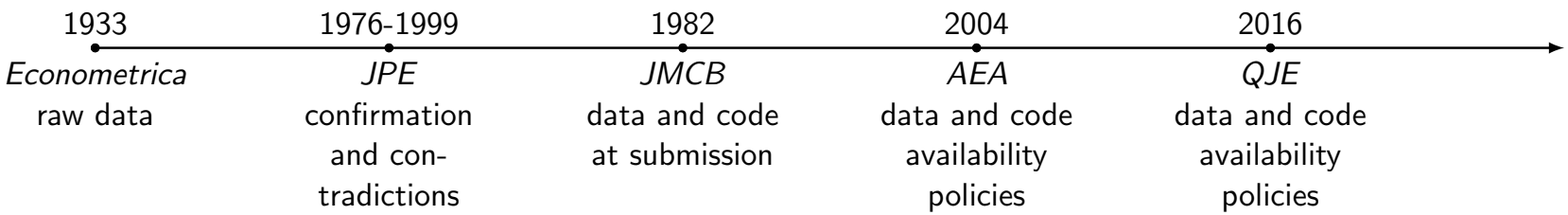
- Data availability policies to promote transparency



Peer review and reproducibility

- Peer review ensures high quality and original research, but zero obligation to reproduce results from the codes and data

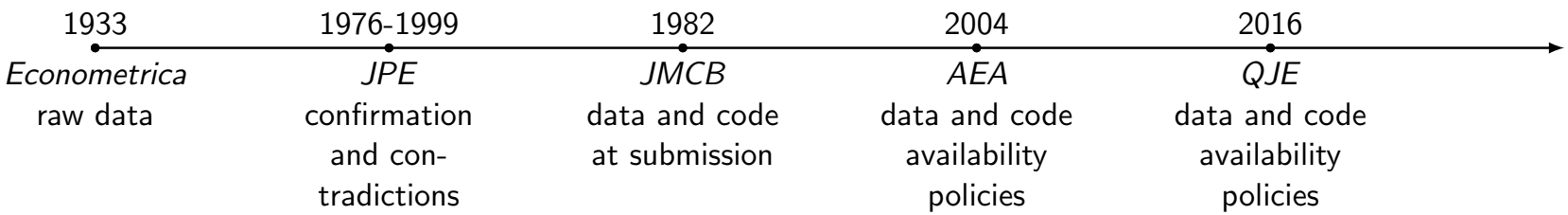
- Data availability policies to promote transparency



Peer review and reproducibility

- Peer review ensures high quality and original research, but zero obligation to reproduce results from the codes and data

- Data availability policies to promote transparency

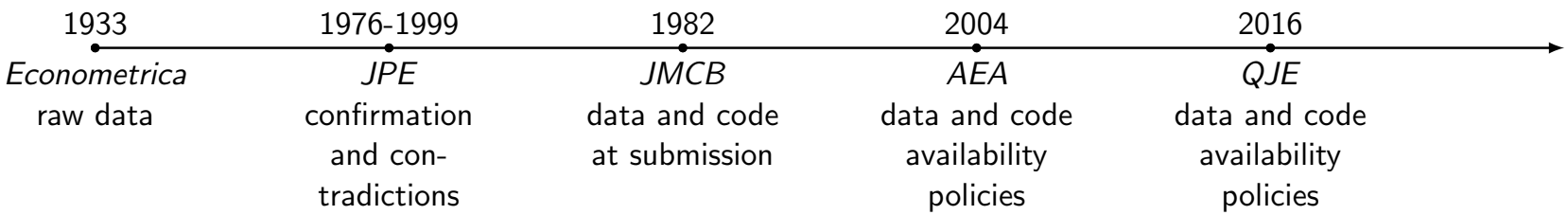


- Between 8.1 % and 54% of economics journal have it, depending on sample (Duvendack et al., 2015; Vlaeminck and Herrmann, 2015; Hoffler 2017)

Peer review and reproducibility

- Peer review ensures high quality and original research, but zero obligation to reproduce results from the codes and data

- Data availability policies to promote transparency



- Between 8.1 % and 54% of economics journal have it, depending on sample (Duvendack et al., 2015; Vlaeminck and Herrmann, 2015; Hoffler 2017)

- Even less have a systematic check (AEA, 2018) with data editors

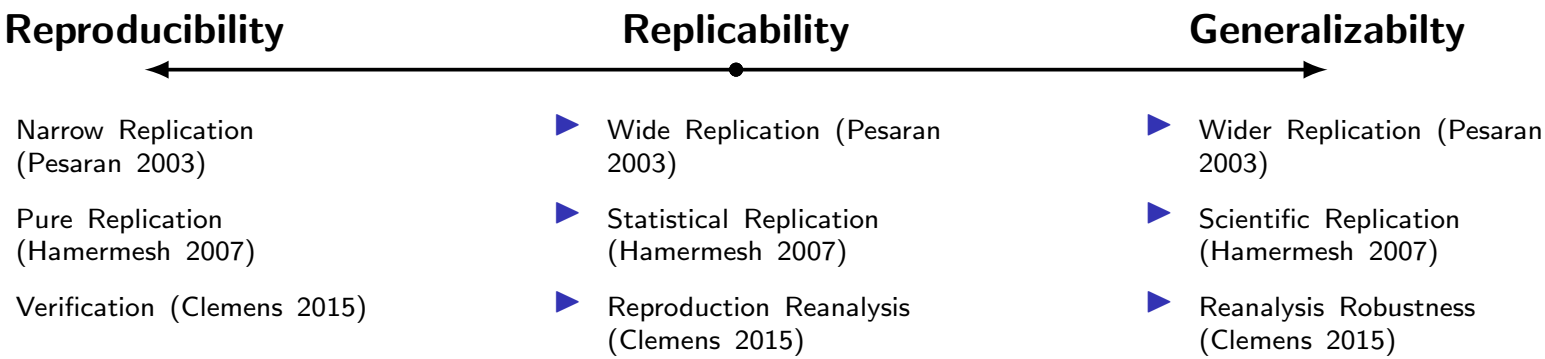
Research and policy work in central banks

1. Central banks' research through the peer-review process can benefit from the aforementioned checks

2. Policy work much more short-term, with many changes of authors for a given model or indicator

- **Today**: good practices for (1) and (2)

Variety of replication concepts



- Reproducible research: same materials + same code as original investigator = duplicate results

Reproducibility exercise

- Herbert, Kingi, Stanchi and Vilhuber (2021): large scale reproducibility exercise of *AEJ:Applied* between 2009-2018
- 25% replicable, 42% conditional on non-confidential data
- Overall good documentation (75% complete documentation) with positive correlation between documentation and replicability
- Even replicable papers required code changes:
  - 1/2 required at least a directory change
  - 1/3 required complex code changes
- Reason for failure: (1) different values for no reason, (2) missing or corrupted data, and (3) code error

Good practices for research and policy work in central banks

Coding practices for a reproducible workflow

3 principles (Orozco et al. 2020):

1. Organize
2. Code for others
3. Automate

```
project_title/
        ├── notes/
        │   ├── binding_site_notes.md
        │   └── neural_network_notes.md
        ├── code/
        │   ├── 1_cleaning.py
        │   ├── 2_analyze.py
        │   └── 3_visualize.py
        ├── data/
        │   ├── input/
        │     └── all_sites_raw_13Feb2018.py
        │   └── output/
        │     ├── thresholds_bcd_17Feb2018.py
        │     └── thresholds_kru_17Feb2018.py
        └── README.md
```

Organizing codes

- Organize source code in logical units or building blocks

- Modular program composed of reusable blocks (functions)

```
project
|   README.md
└──data/
|       └──raw_data/
|           |   data_orig.csv
|       └──processed_data/
|           |   data_clean.csv
|       └──results/
|           |   model_results.csv
└──documents/
|       meeting_notes.md
|       data_dictionary.md
└──code/
|       └──exploration/
|           |   01_data_exploration.Rmd
|           |   02_model_results.Rmd
|       └──scripts/
|           |   01_do_clean_data.R
|           |   02_do_model_data.R
|           └──functions/
|               |   01_funs_clean_data.R
|               |   02_funs_model_data.R
```

  - Exploration notebook (Jupyter, Knitr) or a main file
  - Action script
  - Function script: custom functions in dedicated files
  - Configuration file with directories set up

Writing for others

- Writing in relative paths enhances portability

```
/**** Stata EXAMPLE ****/
  /**** Definition of the useful path ****/
  local CodeFolder "c:/ApplePie/Progs"

  /**** Positioning ****/
  cd "`CodeFolder'"

  /**** Using data that is in another folder ****/
  use ../Raw_Data/Sugar.dta, replace
  append using ../Raw_Data/Apple.dta

  save ../Final_Data/ApplePie.dta, replace
  qui log close
```

Source: Orozco et al. (2020)

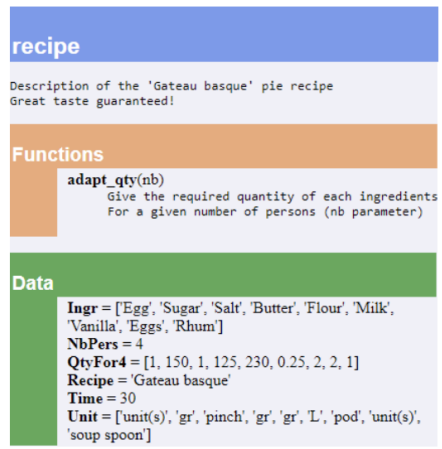
- Consistent variable labeling to improve readability, with explicit names
  - gender variable taking 1 for women and 2 for men, could be coded as a dummy with 1 for women, 0 for men
  - *HICP_* with underscore for the country, *i_* for dummy variables
- Be liberal on comments

Writing for others

- Comment before each block, explaining the purpose of the block, infrequent end of line comments

- Python comments can be extracted using *pydoc*

```
# -*- coding: utf-8 -*-
"""
Description of the 'Gateau basque' pie recipe
Great taste guaranteed!
"""

RECIPE = 'Gateau basque'
TIME = 30 #minutes

INGR = ['Egg', 'Sugar', 'Salt', 'Butter', 'Flour',
        'Milk', 'Vanilla', 'Eggs', 'Rhum']

QTY_FOR_4 = [1, 150, 1, 125, 230,
             0.25, 2, 2, 1]

UNIT = ['unit', 'gr', 'pinch', 'gr', 'gr',
        'L', 'pod', 'unit', 'soup spoon']

def adapt_qty(nbpers):
    """Give the required quantity of each ingredients
    For a given number of persons (nb parameter)
    """
    for i, j, k in zip(INGR, QTY_FOR_4, UNIT):
        print('Ingredient', i, ': ',
              float(j)*nbpers/4, k,
              'required for a ', nbpers, 'person pie')

adapt_qty(4)
adapt_qty(6)
```



Source: Orozco et al. (2020)

- Also able to extract the set of variables defined in the code

- Embedded documentation facilitates documentation-updating when modifying code

Automate

- Automate: call all the action scripts into a master file or write a bash script

```
60
61                          * Master file
62                          * ============
63
64
65      * this program merges the trade database with gravity dataset
66      * it only keeps one observation per country*year due to size constraints
67      run 1_merge.do
68
69
70      * label of the variables
71      run 2_labels.do
72
73
74      * First stylized facts
75      run 3_stylized_facts.do
76
77      * Regressions
78      run 4_regression.do
79
```

Automate

- Bash scripts: bash script.sh

```
#!/bin/bash
alias python='/usr/bin/python3.5'
pip install numpy
wd=/Users/Research/CB_persuasion/code
python $wd/progs/01.pre_processing.py
python $wd/progs/02.data_cleaning.py
python $wd/progs/03.LDA_analysis.py
stata /e do $wd/progs/04.OutputCode.do
done
```

- List all libraries/modules or packages and provide a set up to install modules/packages

```
pip freeze > requirements.txt
pip install -r requirements.txt
conda env export > environment.yml
```

- Document version of the software, eventually specify version of the software in a bash script

Dynamic documents

- Dynamic document:
  - *Sweave* for R and LaTeX, or *RMarkdown*
  - *dyndoc*, *putdocx*, *putpdf*, *markstat* for Stata
- Include tables by linking to a file, instead of a static image
- Include number by linking to a value calculated by an analysis file, instead of a static number typed manually with *S_expr*
- Automatically update tables and numbers
- Produce entire paper/policy note with 1 click

Literate programming tools

Language		Tool	Source extension	Output format	Chunk usage	Chunk syntax
Code	Text					
Sweave-like tools						
R	LaTeX	Sweave	.Rnw	TeX, Beamer, PDF	code	<<chunckname>>= *R code* @
R, Python, SAS, SQL, ...	Markdown	R Markdown	.Rmd	HTML, PDF, MS Word, Beamer, ...	code	```r *R code* ```
SAS	LaTeX	SASWeave	.SAStex	TeX, PDF	code	\begin{SAScode} *SAS code* \end{SAScode}
R			.Rtex			
SAS + R			.SASRtex, .RSAStex			
R	noweb		.Rnw			
SAS + R			.SASnw, .nwSAS			
R, SAS, Matlab, Stata, ...	LaTeX, OpenOffice	StatWeave	.snw	TeX, ODT	code	\begin{Statacode} *Stata code* \end{Statacode}
Stata	Markdown	Markstat	.stmd	TeX, PDF, HTML	code	```s *Stata code* ```
SAS	LaTeX	StatRep	.tex	TeX, PDF	code	\begin{SAScode} *SAS code* \end{SAScode}
Matlab	plain text markup	Publish	.m	MS Word, HTML, PDF, TeX	text	%%title %text
R, Stata, Matlab, Python, ...	plain text markup	Org-mode	.org	TeX, PDF, HTML, ODT, ...	text	#+BEGIN_SRC <language> code #+END_SRC
Notebooks						
Python, R, SAS Stata, Matlab, Julia, ...	Markdown	Jupyter Notebook	.ipynb	HTML, rST, PDF	code & text	
Mathematica	Wolfram language	Mathematica Notebook	.nb	HTML, PDF, TeX, ...	code & text	
R, Python, SAS, SQL, ...	Markdown	R Notebook	.Rmd	HTML, rST, PDF	code	```r *R code* ```
Matlab	Formatted text	Live Scripts	.mlx	HTML, PDF	code & text	

Source: Orozco et al. (2020)

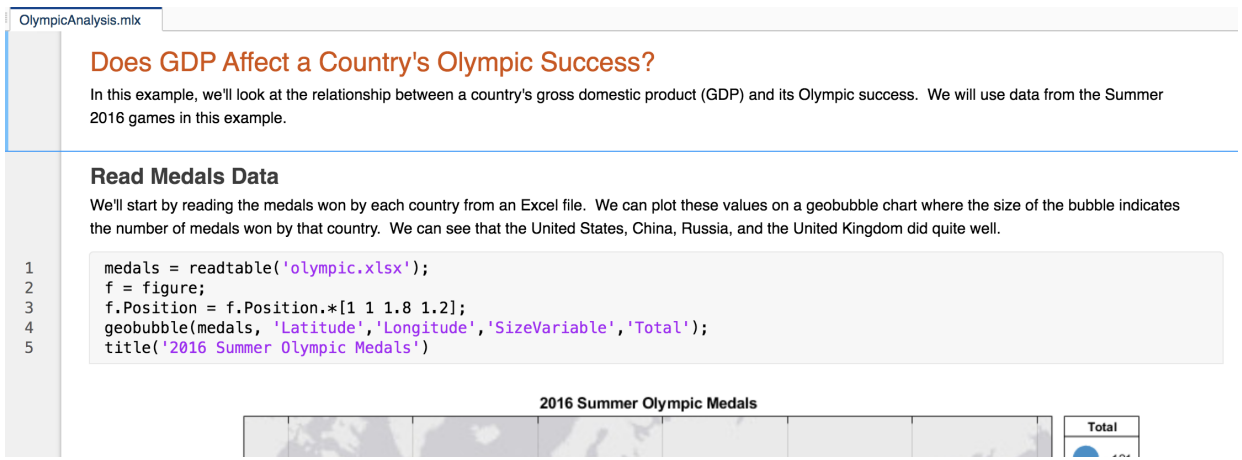
Code example

- R Sweave

```
218  %==================== Results ======================
219
220  \section{Are data policies enough to ensure full reproducibility?}\label{sec:results}
221  <<results1, child='results1.Rnw'>>=
222  @
223  <<results2, child='results2.Rnw'>>=
224  @
225
226  %==================== Discussion ======================
227
228  \section{Conclusion}\label{sec:discussion}
229  <<results, child='conclusion.Rnw'>>=
230  @
```

```
90  <<tab4_1,warning=FALSE,echo=FALSE,results='asis',cache=FALSE>>=
91   # Print table
92   stargazer(tab_absence,
93             type="latex",title = "Was Data Provided?",label="tab:absence",style="aer",
94             flip=FALSE,summary=FALSE, rownames = FALSE,font.size = fs,column.sep.width = cw)
95   @
96   Most papers had programmed codes and stored data in proprietary softwares format. The vast majo
     Stata programming language for at least some portion of analysis (Table~\ref{tab:prog}). This p
```

- Matlab notebook



Source: Matlab

READMEs

- Readme not only for the data but also for the code: it should contain operating systems, analytical software used and software dependencies

```
Program for estimating topics
Goal: Generate the topics distribution
Date: 2021/01/21
Author: Sylverie Herbert
Running under Python version 3.8.2 (2015-08-14)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Input files: mpc_minutes.txt
Output: topics_FED.csv, share_topics.csv
Version 4 of the program: + wordclouds
```

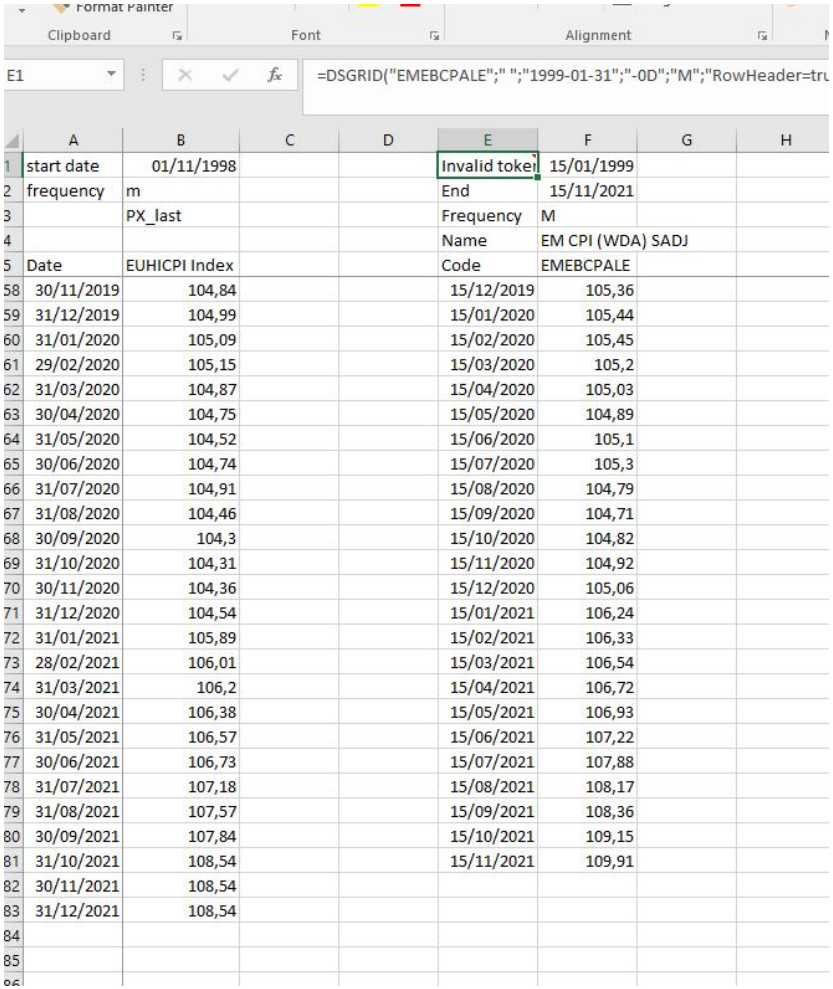
- Update README.md with vital information about repo or directory

- What are the files? Where did they come from? How were they created?

Transparent data management

1. Organize
   - All output data is kept in data/output and tagged with date of creation
   - Raw data is kept raw, with specific folder for processed data, output
2. Document
   - Time stamps for frequent indicator updates, to keep track of vintages
   - How to access data if cannot be provided
3. Automate: excel links, Stata/Matlab/Python workflow as mentioned

Importance of time stamps



Format Painter | Clipboard | Font | Alignment

E1 | =DSGRID("EMEBCPALE";" ";"1999-01-31";"-0D";"M";"RowHeader=tru

	A	B	C	D	E	F	G	H
1	start date	01/11/1998			Invalid toker	15/01/1999		
2	frequency	m			End	15/11/2021		
3		PX_last			Frequency	M		
4					Name	EM CPI (WDA) SADJ		
5	Date	EUHICPI Index			Code	EMEBCPALE		
58	30/11/2019	104,84			15/12/2019	105,36		
59	31/12/2019	104,99			15/01/2020	105,44		
60	31/01/2020	105,09			15/02/2020	105,45		
61	29/02/2020	105,15			15/03/2020	105,2		
62	31/03/2020	104,87			15/04/2020	105,03		
63	30/04/2020	104,75			15/05/2020	104,89		
64	31/05/2020	104,52			15/06/2020	105,1		
65	30/06/2020	104,74			15/07/2020	105,3		
66	31/07/2020	104,91			15/08/2020	104,79		
67	31/08/2020	104,46			15/09/2020	104,71		
68	30/09/2020	104,3			15/10/2020	104,82		
69	31/10/2020	104,31			15/11/2020	104,92		
70	30/11/2020	104,36			15/12/2020	105,06		
71	31/12/2020	104,54			15/01/2021	106,24		
72	31/01/2021	105,89			15/02/2021	106,33		
73	28/02/2021	106,01			15/03/2021	106,54		
74	31/03/2021	106,2			15/04/2021	106,72		
75	30/04/2021	106,38			15/05/2021	106,93		
76	31/05/2021	106,57			15/06/2021	107,22		
77	30/06/2021	106,73			15/07/2021	107,88		
78	31/07/2021	107,18			15/08/2021	108,17		
79	31/08/2021	107,57			15/09/2021	108,36		
80	30/09/2021	107,84			15/10/2021	109,15		
81	31/10/2021	108,54			15/11/2021	109,91		
82	30/11/2021	108,54						
83	31/12/2021	108,54						
84								
85								

- Importance of providing at least the template of data downloads

The age of "big data"

- In the age of big data, even more important to document download methods and timing as some companies like Factiva require deleting the raw data
- Addition of journals over time so running the same code would not give the same sample if no documentation on the time on downloads and sample coverage
- Version control tools: Git available within the Central Bank to keep all the versions in branches in an organized manner

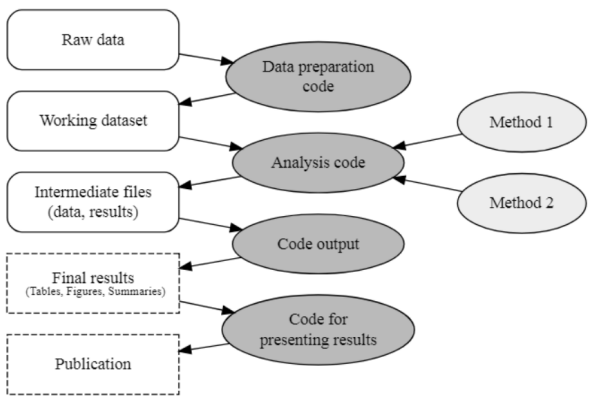
Enhancing reproducibility in central banks

Improving policy work replicability

- Many indicators created, updated at each governing council with different economists in charge

- Organizing the workflow for policy work, e.g. GraphViz or RGraphViz or naming convention

```
01_preprocessing.py
02_datacleaning.py
03_LDA_analysismodel1.py
04_LDA_analysismodel2.py
```



Source: Orozco et al. (2020)

Improving policy work replicability

- Conventions for layout, labeling
  - Limit # of characters per line, indentations
  - Definitions at the top for variables, functions then execution statements
  - Lower case, underscore conventions
- Tools to correct code to follow conventions: *pylint* (python), *check and formatR* (R), *Monkeyproof solutions* (Matlab)

```
$ pylint my_module.py
************* Module my_module
my_module.py:38:0: C0303: Trailing whitespace (trailing-whitespace)
my_module.py:62:28: C0303: Trailing whitespace (trailing-whitespace)
my_module.py:66:0: C0305: Trailing newlines (trailing-newlines)
my_module.py:1:0: C0114: Missing module docstring (missing-module-docstring)
my_module.py:5:0: C0103: Constant name "h5file" doesn't conform to UPPER_CASE naming style (inva

-------------------------------------------------------------------
Your code has been rated at 6.88/10 (previous run: 6.88/10, +0.00)
```

▸ Pylint ▸ R check

Enhancing research replicability

- Data availability policies for central banks working paper series (Herbert, Kingi, Stanchi and Vilhuber (2021) find no citation bonus for replicable papers)
- Cooperation among ESCB for confidential data (e.g., LTROs)
- Access for outside researchers: Open Data Room

Thank you!

References

Clemens, M.A, (2017) "The Meaning of Failed Replications: a Review and Proposal", Journal of Economic Surveys, 31: 326-342

Duvendack, Maren, Palmer-Jones, Richard W. and Reed, W., (2015), "Replications in Economics: A Progress Report", Econ Journal Watch, 12, issue 2.

Hamermesh, D.S. (2007), "Viewpoint: Replication in economics". Canadian Journal of Economics, 40: 715-733.

Herbert, S., Kingi, H., Stanchi, F. and Vilhuber, L. (2021)"Reproducibility of Economics Research: a case study", Banque de France Working paper 853

Orozco, V., Bontemps, C., Maigné, E., Piguet, V., Hofstetter, A., Lacroix, A., Levert, F. and Rousselle, J.-M. (2020), "How to Make a Reproducible Pie: Reproducible Research for Empirical Economics and Econometrics". Journal of Economic Surveys, 34: 1134-1169

Pesaran, H. (2003) "Introducing a replication section". Journal of Applied Econometrics, 18, 111.

VizCode

```
digraph G {
rankdir = RL;
node [width =2, height=0.7];
  subgraph cluster_data {
    style=invis;
    node [shape=box, style = rounded]

    rawdata [label = "Raw data"];
    working [label = "Working dataset"];
    interm [label = <Intermediate files<BR />(data, results)>];

    node [style=dashed]
    final [label=<Final results<BR /><FONT POINT-SIZE="10">
    (Tables, Figures, Summaries)</FONT>>];
    publication [label = "Publication"];

    {rank=same; rawdata; working; interm; publication}
  }

  subgraph cluster_code {
    style=invis;
    node [shape = ellipse, fillcolor=gray73, style="filled"]

    dataprep [label = <Data preparation<BR />code>];
    analysis [label = "Analysis code"];
    codeout [label = "Code output"];
    coderes [label = <Code for<BR />presenting results>];

    node [fillcolor=gray93, style="filled"]
    method1 [label = "Method 1", width=1.5];
    method2 [label = "Method 2", width=1.5];

    {rank=same; dataprep; analysis; codeout; coderes}
    {rank=same; method1; method2}
    //dataprep -> analysis -> codeout -> coderes [invis];
    method1 -> analysis;
    method2 -> analysis;
  }
```

Source: Orozco et al. (2020)

back