**Channels in Macro Papers Project Documentation (20220107)**

Definition of a ‘Macro Paper’

* Has a JEL code in "E – Macroeconomics and Monetary Economics" (e.g., E40, E51) or/and
* Contains one of the keywords: "macro", "macroeconomic", "macroeconomics", "fiscal policy", or "monetary policy" in any part of the paper (title, abstract, keywords, any paragraph in the body of paper)

Summary of Main Processing Pipeline

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Task | Codes | Input | Output |
| **1** | **Download Raw Data** | | | |
| 1.1 | Download citations of top 5 macro journals (AER, QJE, Econometrica, JPE, RES) from 2018-2021 [L] | - | - | O1: [citation].ris  (1 for each journal) |
| 1.2 | Export each .ris to .csv using Zotero and convert to .xlsx using Excel [L] | - | [O1](#conf_call_pdf) | O2: [identifiers].csv  O3: [identifiers].xlsx |
| 1.3 | Download papers from top 5 macro journals from 2018-2021 [L] | - | [O3](#conf_call_txt) | O4: [identifiers].xlsx (updated)  O5: [paper].pdf |
| 1.4 | Record papers from top 5 macro journals with JEL classification of E [L] | - | [O4](#conf_call_csv) | O6: [identifiers].xlsx (updated) |
| **2** | **Process Raw Data for Further Analysis** | | | |
| 2.1 | Convert pdf to txt [M] | convert\_pdf\_to\_txt\_walkdir.sh | [O5](#ciqcompany_sas7bdat) | O7: [paper].txt |
| 2.2 | Process the xlsx file containing identifiers [L] | process\_identifiers\_xlsx.ipynb | [O6](#ciqcountrygeo_sas7bdat) | O8: [identifiers \_processed].xlsx |
| **3** | **Identify Macro Papers** | | | |
| 3.1 | Identify papers that contain any of the pre-specified keywords, and together with the JEL:E classification, identify macro papers. [L] | identify\_macro\_papers.ipynb | [O7](#wrds_gvkey_sas7bdat), [O8](#hassanfile_raw) | O9: macro\_papers.xlsx |
| **4** | **Identify Paragraphs Containing ‘Channel’** | | | |
| 4.1 | Identify macro papers that contain ‘channel’. For each of these papers, break the paper down into paragraphs and identify paragraphs containing ‘channel’. Compile paragraphs with identifiers into a dataframe. [L] | get\_paragraphs\_containing\_channel.ipynb | [O7](#wrds_gvkey_sas7bdat), [O9](#O9) | O10: paragraphs\_channel\_full.xlsx  O11: paragraphs\_channel\_truncated.xlsx |
| 4.2 | Write the dataframe into excel, with the keyword ‘channel’ bolded. [L] | bold\_keywords.ipynb | [O11](#O11) | O12: paragraphs\_channel\_truncated\_bold.xlsx |

\* M = Mercury, L = Local.

Count of Macro Papers

|  |  |  |  |
| --- | --- | --- | --- |
| **Journal** | **Total Number of Papers** | **Number of Macro Papers** | **Share of Macro Papers** |
| AER | 446 | 234 | 52.5% |
| RES | 306 | 170 | 55.6% |
| JPE | 319 | 147 | 46.1% |
| Econometrica | 259 | 109 | 42.1% |
| QJE | 168 | 92 | 54.8% |
| **Total** | **1498** | **752** | **50.2%** |

Count of Paragraphs

* Note: The paragraph count is volatile and depends on:
  + What lines of a paper are kept or removed
  + How the processed paper is broken into paragraphs
* Top 10 Papers:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Title** | **Author** | **Journal** | **Date** | **Count of Paragraph** |
| Recovery from the Great Depression: The Farm Channel in Spring 1933. | Hausman, Joshua K.; Rhode, Paul W.; Wieland, Johannes F. | American Economic Review | 2019-02 | 46 |
| Monetary Policy and the Redistribution Channel. | Auclert, Adrien | American Economic Review | 2019-06 | 33 |
| Informational Channels of Financial Contagion. | Trevino, Isabel | Econometrica | 2020-01 | 32 |
| Regional Heterogeneity and the Refinancing Channel of Monetary Policy \*. | Beraja, Martin; Fuster, Andreas; Hurst, Erik; Vavra, Joseph | Quarterly Journal of Economics | 2019-02 | 31 |
| Credit Market Disruptions and Liquidity Spillover Effects in the Supply Chain. | Costello, Anna M. | Journal of Political Economy | 2020-09 | 28 |
| Understanding HANK: Insights From a PRANK. | Acharya, Sushant; Dogra, Keshav | Econometrica | 2020-05 | 27 |
| Historical Antisemitism, Ethnic Specialization, and Financial Development. | D'Acunto, Francesco; Prokopczuk, Marcel; Weber, Michael | Review of Economic Studies | 2019-05 | 23 |
| Survival and Long-Run Dynamics with Heterogeneous Beliefs under Recursive Preferences. | Borovička, Jaroslav | Journal of Political Economy | 2020-01 | 21 |
| Monetary Policy According to HANK. | Kaplan, Greg; Moll, Benjamin; Violante, Giovanni L. | American Economic Review | 2018-03 | 18 |
| Evaluating Long-Term-Care Policy Options, Taking the Family Seriously. | Barczyk, Daniel; Kredler, Matthias | Review of Economic Studies | 2018-04 | 18 |
| Rate-Amplifying Demand and the Excess Sensitivity of Long-Term Rates. | Hanson, Samuel G; Lucca, David O; Wright, Jonathan H | Quarterly Journal of Economics | 2021-08 | 18 |

Histogram:

* Most paper have 1-10 paragraphs.

Algorithm to Identify Paragraphs

* For a given paper, read in the txt file by line, so the output is a list of lines.
* Process the paper: Remove references, headers, footers, section headers, junk characters, and watermarks
* Only keep lines with length > 0 (non-empty strings)
* A line is labelled as the start of a paragraph if:
  + Its length is > 50, and
  + The length of the line after is > 50, and
  + The length of the line before is at least 20 characters shorter than the current line.
* A line is labelled as the end of a paragraph if:
  + Its length is >5, and
  + The line ends in any of the following punctuation marks: ['.', '!','?', ':'], and
  + The length of the line before is >50, and
  + The length of the line before is at least 20 characters greater than the current line, and
  + The length of the line after is < 30.