

The Role of Data Supplements in Reproducibility:

Curation Challenges



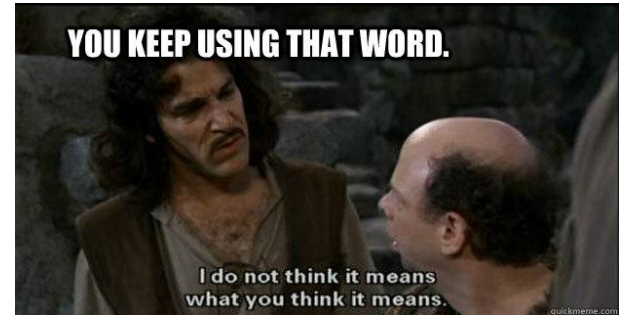
Part 1: Curation, Reproducibility, and Third-Party Data

(Oh my!)



Curation

- Data curation is “the active and ongoing management of data through its life cycle of interest and usefulness to scholarship, science, and education. Data curation activities enable data discovery and retrieval, maintain its quality, add value, and provide for reuse over time...”¹
- For this presentation, includes:
 - Knowledge and information management considerations
 - Workflows, policies, procedures
 - Technical set-up
 - Organizational change management



Primary research object

- Useful for research in and of itself
(no dependencies)
- If associated files were not included, object would still be useful
 - Papers
 - Survey data



Paper example

- Primary object
 - Thorough description
 - Findable, citable, etc.
 - Expectation paper will be reused

FEDERAL RESERVE BANK of KANSAS CITY • Denver • Oklahoma City • Omaha


Home | Banking | Research & Data | Community Development | Newsroom | Publications | Education | About Us


Home > Publications > Research & Data Publications > Research Working Papers > A Model of Monetary Policy shocks for Financial Crises and Normal Conditions

Research Working Paper

A Model of Monetary Policy shocks for Financial Crises and Normal Conditions

By John W. Keating, Logan J. Kelly, A. Lee Smith and Victor J. Valcarcel



 **A Model of Monetary Policy Shocks for Financial Crises and Normal Conditions**

Download paper, RWP 14-11, February 16, 2018

Deteriorating economic conditions in late 2008 led the Federal Reserve to lower the target federal funds rate to near zero, inject liquidity into the financial system through novel facilities, and engage in large scale asset purchases. The combination of conventional and unconventional policy measures prevents using the effective federal funds rate to assess the effects of monetary policy beyond 2008. This paper develops an approach to identify the effects of monetary policy shocks in such instances. We employ a newly created broad monetary aggregate to elicit the effects of monetary policy shocks both prior to and after 2008. Our model produces plausible responses to monetary policy shocks free from price, output, and liquidity puzzles that plague other approaches. It also produces a series of monetary policy shocks which aligns well with major changes in the Fed's asset purchase programs.

JEL Classification: E3; E4; E5

Article Citation

- Keating, John W., Logan J. Kelly, A. Lee Smith, Victor J. Valcarcel. 2018. "A Model of Monetary Policy Shocks for Financial Crises and Normal Conditions," Federal Reserve Bank of Kansas City, working paper no. 14-11, February, available at <https://doi.org/10.18651/RWP2014-11>

Related Research

- Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. "Monetary policy shocks: What have we learned and to what end?." *Handbook of macroeconomics* 1 (1999): 65-148.
- Belongia, Michael T., and Peter N. Ireland. "Interest rates and money in the measurement of monetary policy." *Journal of Business & Economic Statistics* 33.2 (2015): 255-269.

Additional Files

- Appendix
- Replication Code and Data

Data example

- Primary object
 - Thorough description
 - Findable, citable, etc.
 - Expectation data will be reused

The screenshot displays the ICPSR (Inter-university Consortium for Political and Social Research) website interface. The top navigation bar includes links for FIND DATA, START SHARING DATA, MEMBERSHIP, SUMMER PROGRAM, TEACHING & LEARNING, and DATA MANAGEMENT & CURATION. The main header features the ICPSR logo and the text 'Find & Analyze Data'. Below this, a search bar and several tabs (FIND DATA, SEARCH/COMPARE VARIABLES, DATA-RELATED PUBLICATIONS, RESOURCES FOR STUDENTS, HELP) are visible.

The main content area is titled 'Economic Behavior of the Affluent, 1964 (ICPSR 7429)'. This title is highlighted with a red box. Below the title, the publication date (Mar 2, 2010) and links to 'Cite this Study' and 'Share this Study' are shown. The 'Principal Investigator(s)' section, also highlighted with a red box, lists Robin Barlow, Harvey Brazer, James N. Morgan, and a DOI link (https://doi.org/10.3886/ICPSR07429.v2). The 'Version V2' label is also present.

On the right side, there are two circular statistics: '311 downloads * Used Report' and '2 related publications'. The '2 related publications' circle is highlighted with a red box.

The 'At A Glance' section includes tabs for Data & Documentation, Variables (highlighted with a red box), Publications, and Export Metadata. The 'Project Description' section contains a 'Summary' (highlighted with a red box) which describes the study's focus on high-income individuals and their investment behaviors. Below the summary, the 'Citation' section provides the full citation for the study. The 'Export Citation' section offers options to export the citation in various formats. The 'Funding' section lists the National Science Foundation, Ford Foundation, and Brookings Institution. The 'Subject Terms' section (highlighted with a red box) displays a grid of tags related to the study, including 'affluence', 'capital gains', 'economic behavior', 'financial management', 'income', 'income tax', 'investment returns', 'investments', 'investors', 'personal finances', 'personal wealth', 'savings', 'taxes', and 'wealth'. The 'Scope of Project' section includes a 'Methodology' tab (highlighted with a red box) and a 'Version(s)' section.

On the right side of the page, there is a 'Notes' section with a bullet point stating: 'Data in this collection are available only to users at ICPSR member institutions.' Below this is the ICPSR logo and a brief description of the organization's mission.

What does that have to do with reproducibility?

- Reproducibility has been used to mean:
 - Completely re-doing a study, including data collection
 - Re-calculation using fixed code and a fixed set of data
 - Two independent studies producing similar results
 - Etc.
- In our experience:
 - Researchers think reproducibility means exact recreation of results using fixed code and data to check for validity



REPLICATION \neq REPRODUCIBILITY



More is needed

To have a scientific “result”:

- Others have to “know” about it
- Others have to be able to validate it
 - Reproduce the method and achieve the same result
 - Achieve the same result via a different method
 - Reuse the result in a new method



Broad definition of validation

Type	Example Files	Use
Verification*	Original Dataset Original Code	Original Dataset Original Code
Reanalysis*	Original Dataset	Original Dataset New or Altered Computer Code
Reproduction*	Original Code	Original Code New Dataset within Original Population
Extension*	Original Code	Original Code New Dataset within New Population
Reuse	Original Dataset	Original Dataset Combined with New Dataset New Code
Exact Reproduction	Table/Graph/Chart File	Reproduced in Review Section

*Clemens, Michael A. (2015). "The meaning of failed replications: A review and proposal." Journal of Economic Surveys, 31(1): 326-342. <https://doi.org/10.1111/joes.12139>



What's that got to do with data supplements?

- What is the data?
- Do I need specific software or versions to run any of it?
- Are there any restrictions on use?
- Is it being updated? If so, where can I find updates?
- How do I cite it in my work?
- **Not well described, not findable, not citable
no expectation for reuse**

FEDERAL RESERVE BANK of KANSAS CITY • Denver • Oklahoma City • Omaha

Home | Banking | Research & Data | Community Development | Newsroom | Publications | Education | About Us

Home > Publications > Research & Data Publications > Research Working Papers > A Model of Monetary Policy shocks for Financial Crises and Normal Conditions

Research Working Paper

A Model of Monetary Policy shocks for Financial Crises and Normal Conditions

By John W. Keating, Logan J. Kelly, A. Lee Smith and Victor J. Valcarcel

📄 📧 📘 🌐

A Model of Monetary Policy Shocks for Financial Crises and Normal Conditions

Download paper, RWP 14-11, February 16, 2018

Deteriorating economic conditions in late 2008 led the Federal Reserve to lower the target federal funds rate to near zero, inject liquidity into the financial system through novel facilities, and engage in large scale asset purchases. The combination of conventional and unconventional policy measures prevents using the effective federal funds rate to assess the effects of monetary policy beyond 2008. This paper develops an approach to identify the effects of monetary policy shocks in such instances. We employ a newly created broad monetary aggregate to elicit the effects of monetary policy shocks both prior to and after 2008. Our model produces plausible responses to monetary policy shocks free from price, output, and liquidity puzzles that plague other approaches. It also produces a series of monetary policy shocks which aligns well with major changes in the Fed's asset purchase programs.

JEL Classification: E3; E4; E5

Article Citation

- Keating, John W., Logan J. Kelly, A. Lee Smith, Victor J. Valcarcel. 2018. "A Model of Monetary Policy Shocks for Financial Crises and Normal Conditions," Federal Reserve Bank of Kansas City, working paper no. 14-11, February, available at <https://doi.org/10.18651/RWP2014-11>

Related Research

- Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. "Monetary policy shocks: What have we learned and to what end?," *Handbook of macroeconomics* 1 (1999): 65-148.
- Belongia, Michael T., and Peter N. Ireland. "Interest rates and money in the measurement of monetary policy," *Journal of Business & Economic Statistics* 33.2 (2015): 255-269.

Additional Files

- Appendix
- Replication Code and Data**

Replication example

Replication Datasets

Replication datasets are datasets and other materials deposited by an investigator that allow another researcher to replicate the analysis in a published article, book, or dissertation. Often these datasets are subsets of the original datasets, containing only those variables necessary for replication of published findings. Many journals, publishers, funding agencies, and graduate departments require investigators to deposit replication datasets in a public archive. Depositing data with the ICPSR fulfills these requirements.

- Minimal description, mostly about the paper
- Less focus on file formats, naming, and other curation standards
- Overall, less thought given to being “usable” – aka reproducible

The 2001 Recession: How Was It Different and What Developments May Have Caused It? (ICPSR 1292)

Published: Oct 9, 2003 [Cite this Study](#) | [Share this Study](#)

Principal Investigator(s): [Kevin L. Kleson](#), Federal Reserve Bank of St. Louis

<https://doi.org/10.3886/ICPSR01292.v1>

Version V1

Download [Download](#) | [Analyze Online \(R\)](#)

[A-Z A-Z](#) [Data & Documentation](#) [Variables](#) [Publications](#) [Export Metadata](#)

▼ Project Description

Summary

The 2001 recession was unique in several respects. For instance, the peak-to-trough decline in real Gross Domestic Product was one of the smallest on record and its duration was slightly shorter than average. This article examines some of the other unique features of the 2001 recession compared with the “average” post-World War II recession. The article also shows that forecasters were surprised by the onset of the recession, perhaps because of incomplete data available to them in real time. Finally, the article examines the errors from a well-known macroeconomic forecast and finds that forecasters were surprised by the declines in real business and household fixed investment, as well as real net exports, before the March 2001 business cycle peak.

Citation

Kleson, Kevin L. The 2001 Recession: How Was It Different and What Developments May Have Caused It? Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2003-10-09. <https://doi.org/10.3886/ICPSR01292.v1>

Export Citation:

- [BIB](#) (generic format for RefWorks, EndNote, etc.)
- [EndNote](#)

Subject Terms

[Forecasting](#) [Gross Domestic Product](#) [Investments](#) [Recession](#)

Geographic Coverage

[United States](#)

▼ Scope of Project

Data Collection Notes

(1) The file submitted is the Excel data file 0309kdd.xls. (2) These data are part of ICPSR's Publication-Related Archive and are distributed exactly as they arrived from the data depositor. ICPSR has not checked or processed this material. Users should consult the investigator(s) if further information is desired.

▼ Version(s)

231 downloads + [Usage Report](#) (past three years)

1 related publications

Notes

- These data are flagged as [replication datasets](#) and are distributed exactly as they arrived from the data depositor. ICPSR has not checked or processed this material. Users should consult the investigator(s) if further information is desired.
- The public-use data files in this collection are available for access by the general public. Access does not require affiliation with an ICPSR member institution.

ICPSR

This study is provided by ICPSR. ICPSR provides leadership and training in data access, curation, and methods of analysis for a diverse and expanding social science research community.

When they're already hesitant...

Collected data

- Funder mandates support sharing
- Repositories facilitate public access
- Metadata standards (e.g., DDI) promote usability

Purchased data

- Agreement terms restrict sharing
- Restricting access requires trust
- Provenance sometimes unclear (collection methods, etc.)

Mechanisms in place for reuse

Disincentivises reuse



Third-party outputs get the short shrift

- “Supplement” terminology implies secondary object status
- Cannot assume output files will not be used for novel analysis
- Should be just as usable as other primary objects



Part 2: Challenges in Curating Third-Party Outputs

(and lessons learned)



Restrictions

- At a minimum, understand the terms
- Even better, negotiate better terms!
- Don't be afraid to ask for additional permissions later
 - *Note: this is **not** a “better to ask for forgiveness” situation*



Restrict and Restrict Alike

- Specify restrictions in documentation
 - Proprietary vendors will be appreciative!
- Do not give permissions that you do not have the authority to give

*Any data provided herein is provided pursuant to a license between the data provider and the Federal Reserve Bank of Kansas City, is proprietary, and may not be redistributed. The Federal Reserve Bank of Kansas City does not own the data and cannot authorize its redistribution to any other party. Please contact the data provider for any questions concerning redistribution of the included data.



Obtaining permissions

- Vendor
- Data Librarian
- Legal/Privacy Officers
- Researcher
- Data Manager
- Information Security
- Records Management
- And more!



Citations

- For the output file
- For the third-party sources
- Some data citation components less clear cut than traditional citations

Owner. Publication Year(s). "Supplement or Subsection." *Dataset*, Date Range. Accessed through Third-Party Publisher or Distributor. Available at URL or DOI



You can't reproduce what you can't find

- Force11 Principle #1:

“Data should be considered legitimate, citable products of research. Data citations should be accorded the same importance in the scholarly record as citations of other research objects, such as publications.”

- Take it a step further:

*Data should be considered legitimate, **discoverable** products of research. Data **metadata** should be accorded the same importance in the scholarly record as **metadata** of other research objects, such as publications.*



A principled approach

- Gold (AU)DRIPSS
- Understand trade-offs, make informed decisions
- Especially hard for restricted data – what to prioritize?



You can't reproduce what doesn't exist

- A data/code file sitting on a hard drive is at risk of:
 - Bit rot/file corruption
 - Obsolescence
 - Personal knowledge loss
- Preservation is necessary to ensure reproducibility in 1, 5, 10, 20+ years



You're not the boss of me.

- “The right thing to do” is a little less incentivizing than funder or journal mandates
- Make recommendations, not requirements
 - Data availability statement
 - File types
 - Naming conventions
 - Reproducibility practices (e.g., ReadMe)
- Base guidance on **standards**



How long will this take?!

- Prioritize
- Be transparent
- Manage expectations
- Streamline processes, when possible



Change is hard

- If you build it...
they will defend current practices and/or complain
about the extra work
- Goal is worthwhile, compromise and long-view are key



Start with “the end” in mind

- Good curation is proactive
 - Build relationships with stakeholders
 - Know your users
 - Get involved in agreements
 - Provide education and outreach
 - Streamline, streamline, streamline



Deeper dive

- Kulp, C. & Butler, C. R. (2018). *Starting with "the end" in mind: Data services at the Federal Reserve Bank of Kansas City*. Presentation to the 2018 Annual International Association for Social Science Information Services & Technology (IASSIST) Conference, Montreal, Quebec, Canada.
- Currier, B. D. & Butler, C. R. (2018). *Gold (AU)DRIPSS: A decision-making framework for knowledge management*. Presentation to the 2018 Annual International Association for Social Science Information Services & Technology (IASSIST) Conference, Montreal, Quebec, Canada.
- Butler, C. R. & Currier, B. D. (2018). *Moving from Compliance to Reproducibility: Metadata for Supplementary Research Collections*. Presentation to the 2018 North American Data Documentation Initiative Conference, Washington, D.C. Accessed through LIS Scholarship Archive. Available at <http://doi.org/10.5281/zenodo.1217148>
- Currier, B. D. & Butler, C. R. (2017). *Research Data Reproducibility and the Importance of Attachment Level Metadata*. Presentation to the 11th U.S. Networked Knowledge Organization Systems (NKOS) Workshop at the 2017 International Conference on Dublin Core and Metadata Applications, Washington, D.C. Accessed through LIS Scholarship Archive. Available at <http://doi.org/10.17605/OSF.IO/7KUGA>
- Butler, C. R. & Currier, B. D. (2017). *You can't replicate what you can't find: Data preservation policies in economic journals*. Presentation to the 2017 Annual International Association for Social Science Information Services & Technology (IASSIST) Conference, Lawrence, KS. Accessed through LIS Scholarship Archive. Available at <http://doi.org/10.17605/OSF.IO/HF3DS>
- Butler, C. R. & Currier, B. D. (2017). *Creating Data Citations in LaTeX for Economists*. Poster presented to the 2017 Annual International Association for Social Science Information Services & Technology (IASSIST) Conference, Lawrence, KS. Accessed through LIS Scholarship Archive. Available at <http://doi.org/10.17605/OSF.IO/P64BR>
- Currier, B. D., Kim, B., Edwards, C., & Butler, C. R. (2017). *Research Data Preservation Beyond Data Sharing and Open Science*. Presentation to the 2017 DLF Forum, Pittsburgh, PA. Accessed through LIS Scholarship Archive. Available at <http://doi.org/10.17605/OSF.IO/A8HM2>
- Currier, B. D., Kim, B., Butler, C. R., Edwards, C., & Dayrit, L. (2017). *Research Data Preservation*. Presentation to the National Digital Stewardship Alliance's Digital Preservation 2017, Pittsburgh, PA. Accessed through LIS Scholarship Archive. Available at <http://doi.org/10.17605/OSF.IO/J7MEU>



Our approach

- FRBKC Research Division has established a local digital preservation strategy and platform
- Deposit of data/code is completely voluntary
- All file types are accepted, and migrations occur during/after ingestion
- Legal and security reviews occur prior to ingestion
- Publications and data files are ingested into separate but linked collections (e.g., dc:relation) and have similar but separate metadata schemas
- ReadMe files are required for all data collections and must contain:
 - file inventory with technical information
 - license for reuse
 - citations (for both supplement and data sources); and
 - any other relevant information
- Files may be restricted access, but all metadata is exposable within the Federal Reserve System

