Datasheet*

Abdul Ahad Qureshi

April 20, 2024

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*

   - The dataset was created to analyze how changes in the Bank of Canada's interest rate policy affect consumer debt levels in Canada, aiming to understand the implications of monetary policy decisions on household indebtedness.

2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*

   - The dataset is compiled from Statistics Canada and Federal Reserve Economic Data (FRED), by researchers such as Abdul Ahad Qureshi for academic and policy analysis purposes.

3. *Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*

   - There is no specific information provided on funding or grants for the creation of the datasets from Statistics Canada and FRED; these datasets are maintained by governmental and institutional budgets.

4. *Any other comments?*

- The study utilizes established, publicly available economic data to explore significant economic questions, aiding in policy formulation and economic understanding.

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*

---

*Code and data are available at: https://github.com/abdulahadq29/debt.

- The instances in the dataset represent economic and financial metrics (such as consumer debt levels and interest rates), not individuals or interactive entities.

2. *How many instances are there in total (of each type, if appropriate)?*

   - The exact number of instances (data points) is not specified but would typically include time-series data points spanning several years for each variable.

3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*

   - The dataset comprises comprehensive data from national sources (Statistics Canada and FRED), representing macroeconomic trends rather than a sample from a larger dataset.

4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*

   - Each instance consists of processed numerical data representing economic indicators like debt-to-income ratios and interest rates.

5. *Is there a label or target associated with each instance? If so, please provide a description.*

   - The primary target variable in the dataset is the Credit Market Debt to Disposable Income ratio, analyzed against interest rates.

6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*

   - There is no specific mention of missing information; economic datasets typically undergo processing to handle any gaps or inconsistencies.

7. *Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.*

   - Relationships are not between individual entries but are statistical correlations between different economic indicators (debt levels and interest rates) analyzed over time.

8. *Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.*

- Not applicable for this econometric study, as it uses complete datasets for comprehensive analysis without splitting into training/validation/testing sets.

9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*

   - No specific errors, noise, or redundancies are detailed, but economic datasets typically undergo cleaning to minimize these issues.

10. *Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.*

    - Relies on external data from Statistics Canada and FRED; these sources are stable and maintained but no absolute guarantee of constancy.

11. *Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.*

    - The dataset does not contain confidential information as it is based on aggregated economic data.

12. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*

    - The data is neutral, consisting of economic indicators that are not offensive or personally threatening.

13. *Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*

    - The dataset does not focus on sub-populations by demographics like age or gender, it's aggregated on a national level.

14. *Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.*

    - It is not possible to identify individuals directly or indirectly as the data is aggregated and anonymized.

15. *Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.*

    - The dataset does not contain sensitive data as it exclusively deals with macroeconomic indicators and financial rates.

16. *Any other comments?*

    - The dataset is essential for understanding economic impacts and policy implications but does not delve into individual-level data or sensitive personal details.

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.*

    - Data were derived from governmental economic databases (Statistics Canada, FRED), focusing on observable economic indicators.

2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?*

    - Collected through official national data recording systems validated by respective government standards and protocols.

3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*

    - The dataset is not a sample; it comprehensively includes all available data from the designated time periods.

4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*

    - Data were collected by employees of government agencies without external compensation issues.

5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old*

news articles)? If not, please describe the timeframe in which the data associated with the instances was created.

- The data spans multiple years, reflecting ongoing economic conditions and policy changes as they occurred.

6. *Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*

- As governmental data, collection and dissemination are typically subject to governmental review and regulations ensuring ethical standards.

7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*

- Data was obtained directly from established governmental economic databases, not from individuals directly.

8. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.*

- No individual notifications necessary as the data involves aggregated economic indicators without personal data.

9. *Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.*

- Consent not applicable; the data comprises macroeconomic data publicly released for analysis and policy-making.

10. *If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).*

- Not applicable as the data does not involve personal or sensitive information requiring consent.

11. *Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.*

- Likely conducted internally by the governmental bodies to ensure compliance with data protection laws, though specifics are not detailed in the document.

12. *Any other comments?*

- The dataset focuses on economic metrics for macroeconomic analysis and policy assessment, without involving personal data collection or ethical concerns typical of datasets involving individual participants.

Preprocessing/cleaning/labeling

1. *Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.*

- Yes, the data underwent cleaning and processing such as removing outliers, handling missing values, and standardizing variable formats for robust econometric analysis.

2. *Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.*

- The raw data is typically maintained by the original sources, Statistics Canada and FRED, which may be accessible for verification and comparison.

3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*

- The preprocessing likely used statistical software like R; however, specific scripts or tools used were not detailed but could be similar to those publicly available for econometric analysis.

4. *Any other comments?*

- Preprocessing ensures the integrity and usability of the data for accurate economic analysis and policy evaluation.

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*

- The dataset has been used to analyze the impact of the Bank of Canada's interest rate policy on consumer debt levels in Canada, focusing on economic research and policy implications.

2. *Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*

- No specific repository linking to papers or systems using the dataset was mentioned, but relevant studies might be found through academic databases or the websites of Statistics Canada and FRED.

3. *What (other) tasks could the dataset be used for?*

    - Beyond economic policy analysis, the dataset could be used for financial market prediction, academic studies in macroeconomics, or sector-specific analyses impacting consumer finance.

4. *Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?*

    - The dataset's aggregation and anonymization ensure no risk of unfair treatment or stereotyping of individuals; future users should continue using the data for aggregated macroeconomic analysis to avoid misinterpretation.

5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*

    - The dataset should not be used for individual financial advice or predictions about specific personal financial outcomes, as it is designed for aggregate economic analysis.

6. *Any other comments?*

    - The dataset is a robust tool for understanding broad economic trends and should be used within the scope of its economic and policy-oriented purposes.

Distribution

1. *Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.*

    - The dataset, derived from Statistics Canada and FRED, is already publicly available and distributed by these entities to anyone interested.

2. *How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*

    - Distributed online through websites and data portals of Statistics Canada and FRED, without specific mention of a DOI for the combined dataset used in the study.

3. *When will the dataset be distributed?*

- The datasets from Statistics Canada and FRED are continually updated and available for access at any time through their respective platforms.

4. *Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/ or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.*

   - Data from Statistics Canada and FRED is generally available under open licenses that allow for wide usage, though specific terms can be found on their respective websites.

5. *Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.*

   - No additional third-party IP restrictions are imposed on the datasets from Statistics Canada and FRED beyond standard copyright notices applicable to government data.

6. *Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.*

   - No export controls or other regulatory restrictions specifically apply to these datasets, as they contain economic data intended for public research and policy analysis.

7. *Any other comments?*

   - Users should ensure compliance with the terms of use provided by Statistics Canada and FRED when using the data for research or other purposes.

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*

   - The datasets are maintained by their respective sources, Statistics Canada and FRED (Federal Reserve Economic Data).

2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*

   - Contact can typically be made through the official websites of Statistics Canada and FRED, where they provide customer service or contact links.

3. *Is there an erratum? If so, please provide a link or other access point.*

- There's no specific erratum link provided; issues are generally handled through updates or notices directly on the respective websites.

4. *Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (for example, mailing list, GitHub)?*

    - Both Statistics Canada and FRED regularly update their datasets; updates are communicated via their websites or through official publications.

5. *If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.*

    - Not applicable, as the dataset consists of aggregated economic data, not individual-level data.

6. *Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.*

    - Older versions of economic data are often archived and accessible for historical research through Statistics Canada and FRED websites.

7. *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.*

    - Extensions or contributions to the official datasets are typically not applicable as these are controlled by national and federal institutions.

8. *Any other comments?*

    - The datasets are crucial resources for economic analysis and are maintained with high standards to ensure reliability and accuracy for public and academic use.