Datasheet for 'Commodity price index'*
the Bank of Canada commodity price index (BCPI)

Xinqi Yue

2024-12-03

Extract of the questions from Gebru et al. (2021).

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 Bank of Canada Commodity Price Index (BCPI) was created to support the Bank of Canada's analysis of commodity price movements and their impact on the Canadian economy, particularly in areas such as inflation, exchange rates, and economic growth. This dataset serves as a key input in the formulation of monetary policy by providing a comprehensive measure of commodity price trends relevant to Canada. The BCPI fills the gap in providing a standardized, weighted index of Canadian commodity prices denominated in USD, facilitating economic modeling and decision-making.

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

   - The data set was created by the Bank of Canada, the nation's central bank. It was developed by the Bank's economic research and analysis team as part of its efforts to provide data for monitoring economic indicators and supporting monetary policy objectives.

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

   - The creation of the data set was funded internally by the Bank of Canada. No external grants or funding sources are associated with this dataset, as it is developed and maintained as part of the Bank's regular operations.

Composition

---

*Code and data are available at: https://github.com/xinqiyue/exchange-rate-analysis..

1. *What do the instances that comprise the data set 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.*

   - The instances in the BCPI data set represent commodity price data for various goods produced in Canada and sold in global markets. These commodities are grouped into categories such as Energy, Metals and Minerals, Forestry, Agriculture, and Fisheries. Each instance in the data set corresponds to a specific commodity and its associated price data, which includes the commodity's price, its unit of measurement, and the date the price was recorded.

- The data set can be thought of as containing multiple types of instances based on commodity categories (e.g., Energy, Metals, Agriculture) and the data points associated with each commodity (e.g., price, volume, or index). There are also temporal aspects involved, as the data includes weekly, monthly, and annual price records for each commodity.

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

   - There are 26 commodities grouped into 5 main categories:
   - Energy: Includes coal, oil (three benchmarks), and natural gas.
   - Metals and Minerals: Includes gold, silver, nickel, copper, aluminum, zinc, potash, lead, and iron.
   - Forestry: Includes pulp, lumber, and newsprint.
   - Agriculture: Includes potatoes, cattle, hogs, wheat, barley, canola, and corn.
   - Fisheries: Includes finfish and shellfish.
   - This adds up to 26 individual commodities.
   - If using weekly data, the number of records per commodity equals the number of weeks in the dataset's time span.
   - If using monthly data, the number of records per commodity equals the number of months in the time span.
   - For annual data, the number of records per commodity equals the number of years in the time span.

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).*

   - **Larger Set**: The data set is a subset of global commodity transactions, focusing on 26 key commodities produced in Canada and priced in USD, relevant to Canada's export economy.

- **Representativeness**: The data set is representative of Canadian commodities, selected based on economic significance and market relevance. Categories (e.g., energy, metals, agriculture) ensure broad sectoral coverage.

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.*

- **Features**:

  - **Date**: The specific week of observation.
  - **Weekly Bank Rate**: Interest rates set by the Bank of Canada.
  - **Weekly Avg USD/CAD**: Average exchange rate of the Canadian Dollar against the US Dollar.
  - **Weekly Total BCPI**: Aggregated index for 26 key export commodities.
  - **Weekly Energy BCPI**: Sub-index focusing on energy-related commodities.
  - **Weekly Metals BCPI**: Sub-index for metals and minerals.

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

  - No

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.*

- the `raw_data` does not contain any missing values in its current form.

- The data has been curated and cleaned to ensure all required fields are present for analysis.

- The absence of missing values allows for direct analysis without the need for imputation or handling of gaps.

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.* The data set includes temporal relationships between instances, as each row corresponds to a specific week.

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.*

  - No

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

    - **Errors**: No explicit errors have been identified in the data set. However, errors could arise from data entry inaccuracies or revisions in the source data.

- **Sources of Noise**: Exchange rates and commodity prices may exhibit volatility due to market fluctuations, external shocks, or geopolitical events, introducing noise unrelated to the variables of interest.

- **Redundancies**: The dataset does not appear to contain redundant rows or duplicate entries; each instance represents a unique weekly observation.

- **Considerations**: Careful preprocessing, such as smoothing techniques or outlier detection, may help mitigate noise when analyzing trends or building models.

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.*

    - The data set is **self-contained**, as it primarily consists of internal data collected from publicly available sources.

- **External Dependencies**: The data on exchange rates and BCPI are sourced from external institutions like the Bank of Canada and financial platforms (e.g., market data services), but no continuous reliance on external resources is required for the data set itself.

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 data set **does not contain confidential data**.

- It primarily includes **publicly available financial and economic data** such as the Bank of Canada Commodity Price Index (BCPI), exchange rates, and interest rates.

- No personally identifiable information (PII), private communications, or data protected by legal privilege (such as doctor-patient confidentiality) is included in the data set.

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.*

- No.

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.*

    - – Yes, the data set identifies sub-populations based on **industry sectors** within the BCPI.

  - The BCPI is divided into multiple **sub-indices**, each representing a specific industry, including:
    - **Agriculture**
    - **Forestry**
    - **Energy**
    - **Metals and Minerals**
    - **Fisheries**

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.*

    - No, the data set does not contain information that directly or indirectly identifies individuals.

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.*

    - No

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.*

    - The data was collected and reported by the Bank of Canada. It includes transaction prices for various commodities produced in Canada and sold in world markets. The prices are obtained from multiple sources such as futures markets, producer price indexes, and data from organizations like the New York Mercantile Exchange (NYMEX), London Metal Exchange (LME), and Statistics Canada.

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?*

- The data is collected through market transactions, official statistics, and publicly available price sources. The Bank of Canada compiles this data from various exchanges and other sources like Haver Analytics and Bloomberg L.P., and averages the daily or weekly prices for inclusion in the BCPI. Prices for commodities are either daily averages, weekly, or monthly data, depending on the commodity.

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 data set represents actual prices for a fixed set of commodities, collected and averaged across several periods (weekly, monthly, or annually). It is not a sample, but a full collection of available data for the commodities included in the BCPI, collected directly from relevant price sources.

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)?*

- The data collection was performed by analysts and economists at the Bank of Canada. These individuals are responsible for gathering, processing, and updating the commodity price data. They also work with data sources like Statistics Canada, the National Energy Board, and other relevant agencies.

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 is collected over multiple years, with updates available on a monthly, weekly, and annual basis. The timeframe includes long historical data series that are continuously updated, with regular revisions to reflect more accurate or newly available data.The earliest record is on 1972-01-12.

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.*

- The BCPI data set does not contain personal or sensitive information. Since it is composed of publicly available economic data (commodity prices) and compiled by a canadian central bank, ethical review processes are not required.

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

- The data was collected from third-party sources like futures exchanges (NYMEX, LME), commodity price reporting agencies (e.g., Haver Analytics, Bloomberg), and Statistics Canada. The Bank of Canada consolidates this data to generate the BCPI.

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.*

    - There are no individuals involved in the dataset as it pertains to aggregate commodity prices. The dataset is purely economic in nature, with no personal data collection or individual notification needed.

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 is not applicable because the dataset does not involve personal data. The BCPI data is based on publicly available market data from recognized sources.

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).*

    - Since there is no individual data or consent required, this is not applicable.

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.*

    - As the dataset does not involve personal or sensitive data, there is no impact analysis or data protection impact assessment required. The dataset is used for economic modeling and analysis, which does not affect individuals directly.

12. *Any other comments?*

    - The BCPI dataset is an economic indicator used by government agencies, businesses, and researchers. Since it is based on publicly available market prices and does not involve personal information, the use of this dataset does not raise privacy concerns.

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**, preprocessing and cleaning were done on the BCPI data. Some commodities have data available on different time intervals (e.g., daily, weekly, monthly), and the Bank of Canada averages prices over these intervals (daily, weekly, or monthly) for consistency in reporting. Missing values are typically handled by using the most recent available data until actual data is received. For instance, when data for a given period is not available, the most recent data is repeated, and this series is updated once the actual data becomes available. Additionally, the BCPI methodology is updated periodically, including adjustments to data from previous years, and corrections to commodity price sources like coal, barley, and lumber.

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.*

   - No.

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 Bank of Canada does not explicitly provide software tools used for preprocessing or cleaning the BCPI data. However, the methodologies and algorithms used for weighting and aggregating data are outlined in their reports. The data processing procedures, such as using a chain Fisher price index and updating weights based on production volumes and price changes, are explained in the Bank of Canada's methodology documents. For more detailed insights into the methodology, you can refer to their official documentation: BCPI Methodology.

Uses

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

   - The Bank of Canada Commodity Price Index (BCPI) dataset has been used in various economic and financial analyses, particularly in studies examining the relationship between commodity prices and economic factors such as exchange rates, inflation, and trade balances.

   -Additionally, the BCPI data has been utilized for policy analysis, including understanding the effects of changes in commodity prices on Canada's monetary policy decisions. However, specific tasks and use cases for this dataset might vary depending on the research focus and analysis goals.

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.*

   - The **Bank of Canada** provides access to the BCPI dataset through their official website, but it does not have a dedicated repository linking to all papers and systems that use the data.

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

   - The BCPI dataset can be used for a variety of tasks, including:
   - **Predictive modeling**: Analyzing the relationship between commodity prices and exchange rates, interest rates, inflation, and economic growth.
   - **Market analysis**: Assessing the impact of commodity price fluctuations on Canadian industries, particularly those related to energy, agriculture, and natural resources.
   - **Policy analysis**: Supporting government decisions on monetary and fiscal policy by understanding the effects of commodity price changes on inflation and the trade balance.
   - **Risk management**: Analyzing potential risks in Canadian exports or in sectors highly dependent on commodity prices, such as energy and agriculture.
   - **Commodity price forecasting**: Developing models for predicting future commodity prices, which could inform investment strategies or policy recommendations.

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?*

   - No.

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

   - Microeconomic analysis of specific firms or sectors: The BCPI is a broad indicator of commodity prices and may not reflect the nuanced conditions of specific firms or sectors within Canada or globally.

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.*

   - **Yes**, the BCPI dataset is publicly available through the **Bank of Canada** website and may be accessed by external parties for research and analytical purposes. The data is made available to various stakeholders, including researchers, policymakers, financial analysts, and the general public, to support economic and financial analysis.

- For example, same dataset can be found on government of canada open data (https://open.canada.ca/data/en/dataset/d40af834-ae9c-43f2-82ea-57641910ac6f?_gl=1*76zq9g*_ga*N

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

    - The BCPI data set is distributed via the Bank of Canada website, where users can download it in formats such as CSV or Excel. It is not available through an API for automated access. As of now, the data set does not have a DOI, as it is provided by the Bank of Canada as a public resource rather than a formal academic data set.

3. *When will the dataset be distributed?*

    - The BCPI data set is updated regularly, with data typically released on a weekly, monthly or quarterly basis.

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.*

    - The BCPI data set is provided under the Bank of Canada's Terms of Use. You are permitted to use, copy, and distribute the content freely, with the following conditions:
        - **Attribution**: You must credit the Bank of Canada as the source of the data and indicate if any changes were made.
        - **Accuracy**: You must ensure the accuracy of any content you reproduce.
        - **Notice Requirement**: If you incorporate the content into documents for sale, you must inform buyers that the content is freely available on the Bank of Canada website.

    The data set is provided at no cost for non-commercial use, and no additional fees apply. The full Terms of Use can be found here: Bank of Canada Terms of Use.

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

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.*

- The BCPI data set is not subject to export controls or regulatory restrictions. It is publicly available and can be used internationally without additional regulatory concerns. However, users should ensure compliance with any local laws or regulations when using the data outside of Canada. No specific supporting documentation for regulatory restrictions is required.

Maintenance

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

   - The dataset will be supported and maintained by the **Bank of Canada**, which is responsible for the collection, updating, and publication of the BCPI data. The Bank of Canada ensures the integrity and accuracy of the dataset as part of its role in supporting economic research and policy analysis.

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

   - For inquiries related to the BCPI dataset, you can contact the Bank of Canada through the following options:

- Use the online contact form available on the Bank of Canada Contact Page.(https://www.bankofcanada.ca/ information/general-inquiries/#online)
- by phone: North America (Toll-Free): 1-888-882-8863
- by phone: Outside North America: +80078227800 (dial country code first)
- Operating Hours: Monday to Friday, 7:00 to 17:00 (ET)

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

   - No.

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)?*

   - The dataset is updated regularly with new weekly data, but changes to past data will not be available in older versions. Updates are communicated via the Bank of Canada's website.

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.

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 will not be available once the data set is updated. However, new data will not affect previously recorded data, which remains accessible unless changes are made to historical data.

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.*

   - No.

References

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. "Datasheets for Datasets." *Communications of the ACM* 64 (12): 86–92.