sciendo

Economics and Business Review
https://www.ebr.edu.pl

Vol. 9 (2), 2023: 27–40
https://doi.org/10.18559/ebr.2023.2.733

Big data in monetary policy analysis—a critical assessment

Alexandra Bogner[1] **Jürgen Jerger**[2]

Abstract

Over the last years the use of big data became increasingly relevant also for macroeconomic topics and specifically the conduct and analysis of monetary policy. The aim of this paper is to provide a survey of these applications and the relevant methods. The rationale for doing so is twofold. First, there is no straightforward definition of "big data". Since macroeconomics and monetary policy analysis has a long tradition in quite sophisticated and data-intensive empirical applications the nature of the innovation big data is indeed bringing to the field is reflected upon. Second, concerning statistical / empirical methods the analysis of big data necessitates the use of different tools relative to traditional empirical macroeconomics which are in some cases a complement to more traditional methods. Hence big data in monetary policy is not just the application of well-established methods to larger data sets.

Keywords

- big data
- monetary policy
- text analysis
- nowcasting

JEL codes: C55, C82, E52

Article received 2 May 2023, accepted 16 June 2023.

Suggested citation: Bogner, A., & Jerger, J. (2023). Big data in monetary policy analysis—a critical assessment. *Economics and Business Review*, *9*(2), 27–40. https://doi.org/10.18559/ebr.2023.2.733

[1] Universität Regensburg, Chair for International and Monetary Economics, 93040 Regensburg, Germany, alexandra.bogner@ur.de, https://orcid.org/0009-0003-8779-2206.

[2] Universität Regensburg, Chair for International and Monetary Economics, 93040 Regensburg, Germany, corresponding author: j.jerger@ur.de, https://orcid.org/0000-0003-4221-3947.

Introduction

"Big data", "data science", "machine learning" and "artificial intelligence" are only a selection of buzzwords that describe deep methodological changes in the recent past that will continue for some time to come. Although the relevance of these changes varies across scientific disciplines and certainly depends on the precise research question there is hardly any subfield—or any researcher—that is not affected by this development.

In this survey paper the applications of big data and methods that deal with big data in the field of monetary economics and more precisely the analysis of monetary policy are looked at. Section 1 starts with a reflection on the conceptual underpinning of big data. Economics and certainly monetary economics is a field in which empirical applications have a long-standing tradition; it could even be argued that apart from developing abstract formal models to look into economic/social questions the use of sophisticated empirical work is the hallmark of economics relative to most other social sciences. But what exactly is "big data" relative to "normal" or even "small data"? Section 2 tries to shed some light on this question. After this clarification the Section sets out to review concrete applications of big data analysis in monetary economics. In doing so not only the new sorts of questions that can be tackled with big data, but also the methods necessary to do so will be looked at. The last Section offers some concluding remarks.

1. What is "big data" (in macroeconomics)?

The Oxford Dictionary defines big data as "data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges (…)" (https://www.oed.com/). The vagueness of this definition already makes clear that there is no helpful universally accepted definition. Still the "significant logistical challenges" hint at one important aspect—the necessity of new methods to deal with big data relative to more traditional empirical work. This will be addressed later.

The discussion of the conceptual issues is started with attempts to characterise big data. Ademmer et al. (2021, p. 10) provide a useful starting point by enlisting "5 Vs" that distinguish big data from more conventional data sets; see also Doerr et al. (2021), Gandomi & Haider (2015) and Hammer et al. (2017):

- Volume: Big data are simply large data sets.

- Velocity: Big data can often be observed at very high frequencies. This is nothing entirely new—just think about high-frequency trading data on financial markets. But especially in macroeconomic research, monthly or even quarterly observations (used to) classify as "high frequency". Clearly high data frequency also requires the rapid, subsequent analysis of these data up to a point of permanent updates.
- Variety: This describes both the potentially large spectrum and the complexity of big data types and sources.
- Veracity: Unconventional and technologically specific data sources bring about the increased risk of diminished data quality relative to traditional official statistics.
- Volatility: Big data may come from rapidly changing sources that prevent the long-run availability of comparable information on specific issues.

This list makes it clear that there is no well-defined demarcation line between big data and more traditional empirical work. Therefore, it can be illuminating to simply ask researchers in the relevant field what they consider as big data. This is exactly what the study of the Irving Fisher Committee on Central Bank Statistics (2021) did in central banks.

According to this study an important element of big data is that it encompasses "non-traditional" unstructured data sets the processing of which requires the use of new types of statistical tools. An example for such an unstructured data set are texts e.g. from newspaper articles and press releases or web-scraped images (Irving Fisher Committee on Central Bank Statistics, 2021, p. 6). Traditional statistical methods which are developed for the analysis of numerical data are clearly inappropriate to process these data. Furthermore, "traditional" data such as payment transaction or price data might develop into big data depending on the frequency of observations and the number observation units. Again, the line between traditional and big data is not clearly defined—but it is quite clear that a very much enlarged size changes the characteristics and usability of a data set. Interestingly, only 35% of the central banks in the survey exclusively regard "non-traditional" data as big data whereas 65% have a broader definition of big data that includes structured traditional data (Doerr et al., 2021, p. 4–5; Irving Fisher Committee on Central Bank Statistics, 2021, p. 1). It should be emphasised that the complementary use of traditional and non-traditional data sources is particularly promising (Doerr et al., 2021, p. 4).

The survey cited above also asked about data sources and topics of projects central bank researchers associate with big data. It finds that newspaper and other online articles are an important big data source for central banks. By analysing this text data, it is possible to quantify sentiment or economic uncertainty. In addition, internet-based data such as search queries or data collected through web scraping are also frequently used (Doerr et al., 2021,

p. 4; Irving Fisher Committee on Central Bank Statistics, 2021, p. 9). Besides these unstructured data sets central banks also rely on so-called financial big data sets such as payment transaction data or credit registries. In the rest of this section a short overview of different types of big data used in macroeconomic analyses is provided.

Text data: In recent years automated text analysis has been increasingly employed in many disciplines, including (macro-) economics. On top of newspaper and press articles central bank publications such as the minutes of the Federal Open Market Committee are also used in economic literature (Buono et al., 2017, p. 113). News and press articles contain information that might be relevant e.g. for the current (macro-) economic development, consumers' and investors' sentiments and the perceived relevance of specific economic and/or policy topics in the public debate. A major advantage of this type of data is the more timely availability relative to official statistics on macroeconomic indicators or surveys on sentiments and the like such as documents on central bank board meetings which may inform the public much faster than the actual monetary policy decisions (Hansen et al., 2018, p. 802) and that are in any case well known to take a prolonged period of time to show their full effect.

Internet search queries: Knowing what people look for through Google or other search engines reveals information on what people care and/or worry about. Hence data from internet searches, specifically the dynamics of searches for a particular term, may reveal important information especially for economic forecasts. As search queries are submitted by humans the respective data reflect agents' behaviour (Buono et al., 2017, p. 111). The timely or even immediate availability is an advantageous feature of this kind of big data. Relevant uses include the production of GDP nowcasts that are simply impossible to produce based on traditional data sources (Ademmer et al., 2021, pp. 42–43). Until now the relevant literature has relied exclusively on Google search queries which might be fair enough given the dominance and ubiquitous use this search engine. Data on Google search queries is publicly available through the online platform Google Trends from whence information on how often a specific term was queried relative to the total search volume can be easily obtained (Buono et al., 2017, p. 111).

Electronic payment transaction data are another relevant big data source for macroeconomic analysis (Ademmer et al., 2021, p. 50–54). This data allows a continuous trace (part of) of the current course of private consumer spending. Therefore, electronic payment transactions are a particularly relevant basis for short-term forecasting or even nowcasting of economic activity (Aprigliano et al., 2019, p. 55). There are several options to access and use these data. One important source are debit cards (Buono et al., 2017, p. 99) and credit card providers (Ademmer et al., 2021, p. 51), e.g., American Express, Mastercard and Visa. Others include the TARGET 2 and BI-Comp platforms (Aprigliano et al., 2019, p. 61). Moreover, the use of transaction data by the messaging service SWIFT has been proposed (Hammer et al., 2017, p. 15). Data quality from electronic payment transactions is remarkably high since there is no potential for measurement errors and they are—in principle, at least—available at high frequency since they are instantaneously recorded (Ademmer et al., 2021, p. 51). However, it is clear that this kind of data is both proprietary and confidential by its very nature. Therefore, these data do not come for free and need to be anonymized, e.g. by aggregating individual transactions but he use of disaggregated (but still anonymized) data on electronic payment transactions certainly provides the basis for interesting and relevant research (Aprigliano et al., 2019, pp. 61, 77).

Price data: The collection of many individual price data for the construction of reliable price indices and inflation rates of different sorts is a very old branch of economic statistics. More recently, the steady growth and huge relevance of online retailing made online price data a very promising new basis for the measurement of price dynamics (Buono et al., 2017, p. 108). Moreover, this data can be obtained easily and in real-time using web scraping. Its most obvious use is the nowcast and short-term forecast of consumer price inflation. The Billion Prices Project (BPP) pioneered the use of online price data in macroeconomic research, see Cavallo and Rigobon (2016). In 2010 no less than five million prices were recorded daily by the BPP, originating from more than 300 retailers in 50 countries (Cavallo & Rigobon, 2016, p. 152). Half a million of these prices were compiled on a daily basis for the US alone. In contrast the US Bureau of Labor Statistics collects about 80.000 prices at monthly or bi-monthly frequency (Cavallo & Rigobon, 2016, pp. 152–153). Potential advantages of online price data are their inexpensive collection, precision, and speedy availability. Prices quoted online are not necessarily the prices at which transactions are concluded and not all relevant prices are necessarily available online. Hence these data will not eliminate the need for more traditional methods of price measurement but are nonetheless a valuable complement.

2. Big data studies in monetary economics

As argued above big data is quite an elusive term but the generation and size of data sets as well as required methods sets them apart from more traditional empirical studies. In this section both a couple of topics in monetary economics as well as methods that are needed and that have been used in these studies are presented and discussed. We do so in increasing order of innovativeness with respect to the data sets and methods involved.

2.1. Online price data

The price data mentioned at the end of the last section were the basis of the first big data studies in monetary economics or even more generally in macroeconomics. The big data label here is warranted primarily by the use of the non-traditional data collection method of web scraping. Otherwise, the analysis of these price data can be conducted using quite well-established methods. Cavallo (2013) uses data from the Billion Prices Project in order to compute online price indices that can be compared to traditional official price indices. The explicit goal of this analysis was to gauge the reliability of the latter. This, of course, rests on the justifiable, but essentially untestable hypothesis that the online price index gives an adequate picture of price developments. He could show that both inflation levels and inflation dynamics from the online price data in Brazil, Chile, Columbia, and Venezuela were quite in line with the patterns seen in official data. This can be interpreted as evidence that the web scraped prices indeed paint an accurate picture. By contrast, the online price index calculated for Argentina by the author deviates substantially and permanently from the corresponding official figures. Cavallo takes this result as a clear indication that Argentina's national statistics institute manipulates the official inflation estimates (Cavallo, 2013, p. 163). This suspicion has been around for quite some time, of course. If one is ready to accept the validity of online prices also for Argentina, the study by Cavallo (2013) proves this suspicion to be justified. Clearly, similar studies for other countries, in which some degree of mistrust in official price statistics is present—as, e.g., also in the Eurozone during the recent period of high inflation—might be a welcome contribution to judge the validity of this mistrust.

Another topic for which price data of the Billion Prices Project have been employed is the empirical characterisation of price rigidities—a central element for potential effects of monetary policy on the real economy. Cavallo (2018) shows that the relatively low frequency, the use of averages, and imputation of missing price data in the construction of traditional price indices tends to overestimate the degree of price rigidity. The concentration of the distribution of price changes at 0% is much less pronounced in his BBP price data. Arguably, this discrepancy has to do with the different points of sale. In traditional shops and markets, customers might indeed react differently to frequent price changes than on online marketplaces. Still, the result is indeed relevant for the judgement of the effectiveness of monetary policy.

2.2. GDP nowcasts using payment data and Google Search queries

Economic activity inevitably requires transactions and digital transactions can be observed in real time. Therefore, payment data are an obvious information source for the timely measurement of economic activity. Traditional official data in this area is available with significant delays that are an important obstacle for economic policy decisions that have to be taken in real-time. Especially for the conduct of monetary policy this problem has long been recognized, see, e.g., Orphanides (2001). Aprigliano et al. (2019) use payment data to calculate Italy's GDP and its domestic components such as household consumption, gross fixed investments and value added in the service sector. Specifically, they look at whether payment data can contribute to an improvement in forecasting accuracy. They use data from TARGET 2 and BI-Comp and employ a dynamic factor model that incorporates the payment data among other indicators for economic activity. Out-of-sample-forecasting simulations are carried out both including and excluding payment data. Subsequently the root mean square forecasting error is computed to assess whether forecasts became more precise. Not surprisingly the authors' results confirm that payment data indeed contain additional information value. Especially regarding the nowcast of GDP the findings point to a significant improvement in forecast accuracy (Aprigliano et al., 2019, p. 73). In a similar vein, Chapman and Desai (2022) show that payment data improve the nowcasting of macroeconomic variables such as GDP, retail and wholesale trade.

Another potential data source that might contain information for GDP nowcasts and forecasts are Google search queries. Bantis et al. (2022) use search activities in the main categories that are distinguished at Google Trends. These data are used on top of a set of traditional indicators in order to forecast GDP growth rates for the USA and Brazil. The procedure is similar to Aprigliano et al. (2019), i.e., a dynamic factor model is estimated and used for a pseudo-real-time-out-of-sample exercise. The authors find that the search data indeed improve the forecasts relative to models that rely exclusively on more traditional variables. The conceptual link between search data and economic activity might not be as clear as with payment data but search data is available in real time and—unlike payment data—come for free.

2.3. Text analysis

Arguably the most important recent innovation associated with the use of big data in monetary economics is the exploitation of information from a variety of different text corpora. Again, the general idea is not entirely new since so-called dictionary-based approaches have been in use for quite some time. A short description of this method will start this section followed by a review of more recent machine learning approaches.

2.3.1. Dictionary-based approaches

In the search for leading business cycle indicators the "R-word index"—a simple count measure of the occurrence of the term "recession" in general media such as newspapers and magazines—is quite well-known. It was as a potential indicator for a looming slump of economic activity, see Bandholz and Funke (2003) and *The Economist* (2011). The results concerning its value added for forecasting have been rather mixed. More recently a number of studies also applied the dictionary-based approach for the measurement of uncertainty about economic policy in general (Baker et al., 2016) and monetary policy in particular (Husted et al., 2017).

Baker et al. (2016) construct indices of economic uncertainty for the USA. The digital archives of leading American newspapers are searched for occurrences of the combination of "uncertainty", "economic" and "policy" from which a time series is produced that is used in subsequent traditional regression analysis that looks at the effects of uncertainty on economic outcomes. They conclude that (their measure of) uncertainty indeed is associated with a negative effect on investment and employment and an increase in stock-price volatility. In a similar vein Husted et al. (2017) create an index of monetary policy uncertainty and find that a higher degree of monetary policy uncertainty has contractionary effects and is associated with rising credit costs (Husted et al., 2017, p. 23). An increase in uncertainty thus creates the same dynamic response as a (certain) restrictive monetary policy shock (Husted et al., 2017, pp. 11, 23).

Shapiro and Wilson (2022) use transcripts of FOMC deliberations in order to measure central bankers' sentiments in the US and estimate a central bank loss function from these text data. They find that the implied inflation target in the US was 1.5 % from 2000-2011. This is significantly below the commonly assumed but by then not officially declared target of 2%.

2.3.2. Machine learning based text analysis

In recent years the analysis of text corpora developed well beyond the simple dictionary- based approaches outlined above. More specifically machine learning approaches are suitable to extract structured information on the semantic nature of texts. This information can then be used in subsequent analyses—in the same manner as shown above for the dictionary-based studies—that look at the effects on some outcome variables. Furthermore, this information can also be used for a formalized comparison of different texts.

In order to employ machine learning techniques the text must be converted into a formal representation. Suppose some analysis is based on p different texts in which n different terms are of potential relevance. Then the so-called term-document matrix X is given by

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

where each entry x_{ij} denotes some count measure for term $i \in [1, \ldots, n]$ occurring in document $j \in [1, \ldots, p]$. We focus here on so-called unsupervised learning (or latent-variable only) techniques that do not establish a link between the elements of matrix X and some output variables that might be influenced by X. Thus, unsupervised learning algorithms are "only" intended to structure the given data (Chakraborty & Joseph, 2017, p. 7). This is done by categorising the observations into groups (Chakraborty & Joseph, 2017, p. 7). Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA) are two methods employed in research on monetary policy (Bholat et al., 2015, pp. 11–13). These methods are now briefly presented.

Both methods require a number of pre-processing steps. Clearly any text analysis needs to start with the determination of the relevant text corpus. Then text cleaning has to be carried out, i.e. the selection of allegedly relevant terms from the total vocabulary in the text (Hansen et al., 2018, p. 818). Obviously irrelevant elements such as punctuation marks, numbers, HTML tags or percentage signs and currency symbols are deleted (Benchimol et al., 2022, p. 3; Gentzkow et al., 2019, p. 538). Needless to say that these steps already imply a good deal of decisions that are necessarily subjective to some extent. Next a researcher might choose to join those words that convey a certain content only in conjunction with each other - a procedure that is called "identify collocations" (Hansen et al., 2018, p. 818). For example, the term "labour market" conveys a different meaning relative to the two words of which this term exists. Further elements of the text, so-called stop words without information content such as "and", "they" or articles like "a", "the" etc. can be discarded (Benchimol et al., 2022, p. 3). Usually, the removal of the stop words is conducted based on a predefined list (Gentzkow et al., 2019, p. 538). However, there is no stop list that is standard. Hence there is some degree of subjectiveness also in this step. Lastly, the remaining vocabulary is treated to the so-called stemming (Benchimol et al., 2022, p. 3). This means that only the respective word stems are considered. For example, in “banking” the affix “ing” is removed and the stem “bank” is retained (Bholat et al., 2015, pp. 7–8). Likewise words such as “economic“, “economics“ and “economically“ would all be represented by their corresponding stem “economic“ (Gentzkow et al., 2019, p. 538). The preparatory measures are important to generate a certain degree of uniformity within the initially unstructured text corpus. Hence, the dimension of the text data is reduced. This is significant because the inherently high dimension of texts is a substantial challenge for their formal analysis (Gentzkow et al., 2019, p. 538).

The result of this procedure is the term-document matrix introduced above and therefore a vector space representation of the documents to be analysed (Benchimol et al., 2022, p. 4). This representation of the text data in a matrix is also called “bag-of-words method”. It needs to be stressed that this numerical formalisation of texts takes away their original message. To put it differently, looking at any column of X will not suffice to even vaguely restore the original meaning of the codified text. Instead of simple word counts or relative frequencies for the elements x_{ij} of X a suitable weighting scheme can be applied (Benchimol et al., 2022, p. 4). This takes care of a potential over-representation of words that occur very frequently and that may not be helpful to differentiate between documents in terms of content. A weighting scheme which tackles the problem described above is referred to as the term frequency-inverse document frequency (Bholat et al., 2015, p. 9) and can be calculated as $x_{ij} = (1 + \log f_{ij}) \cdot \log \frac{p}{d_{ij}}$, where f_{ij} denotes the simple count of some term i in document j, p is the total number of documents and d_{ij} denotes the number of documents in which term i is used. Hence, those terms are weighted more heavily that occur frequently in some document but are rarely used in the other documents of the text corpus (Ademmer et al., 2021, p. 21).

LSA now aims at detecting a latent semantic structure in the text data of (Deerwester et al., 1990). This structure is partially disguised since word choice is random to some extent, leading to what is called “obscuring noise”.

This is achieved by a singular value decomposition of X as follows $X = T\Sigma D'$ (Zong et al., 2021, p. 147). T and D both contain orthonormal columns. Σ is a diagonal matrix containing singular values (Deerwester et al., 1990, p. 397). An advantage regarding the described singular value decomposition of the matrix X is that an optimal approximation using matrices of lower rank is easily computable (Deerwester et al., 1990, p. 398). The singular values of the matrix Σ are ranked according to size keeping the largest singular values and setting the remaining smaller singular values to zero (Deerwester et al., 1990, p. 398). By multiplying this adjusted diagonal matrix with both remaining matrices the matrix $\hat{X}$ is obtained which is an approximation of the term-document matrix X of rank k. This leads to a reduction of the dimension of X but retains the most important information (Bader & Chew, 2010, p. 22). This method is used by Acosta (2015) in order to characterise and compare the contributions of FOMC deliberations before and after a change in the transparency rules. He found out that this change indeed led to a measurable and significant change in the behaviour of FOMC members.

Hansen et al. (2018) look at a very similar question and use a text corpus consisting of documents from FOMC meetings. However, they employ the LDA (Latent Dirichlet Allocation pioneered by Blei et al., 2003) method which might be thought of as a further elaboration of LSA. The goal of LDA is to narrow down the dimension of the matrix X or $\hat{X}$ even further by identifying the grouping of single terms into so-called topics. In Hansen et al. (2018) the text corpus of the FOMC documents could thus be narrowed down to just fourty topics. Figure 1 gives a rough description of this procedure and also indicates the drastic reduction of the dimension.

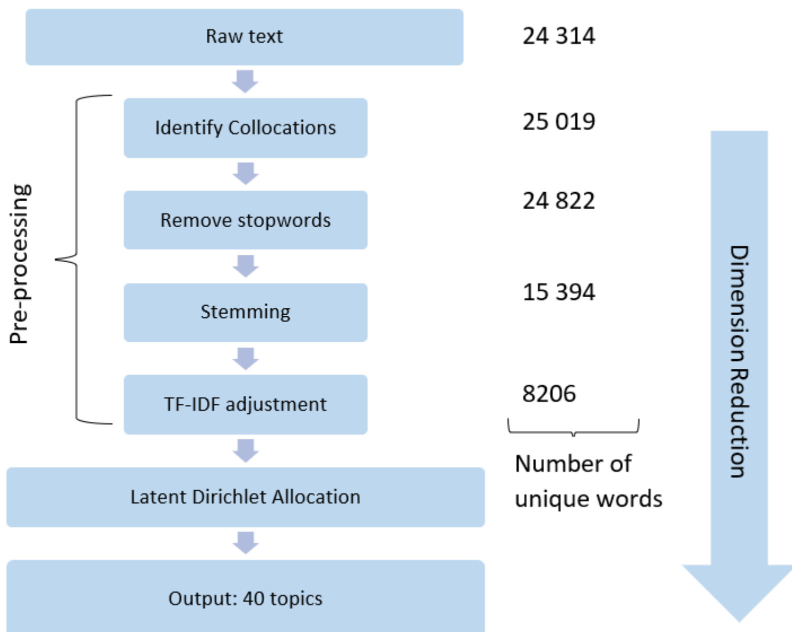

Figure 1. Own representation based on Hansen et al. (2018)

Source: Own work.

A topic is formally defined as "a distribution over a fixed vocabulary" (Blei, 2012, p. 78). Less formally a topic is a weighted list of words (Hansen et al., 2018, p. 817).

Those words that express the same idea in terms of content are assigned to the respective list (Hansen et al., 2018, p. 817). However, the topic structure of the text corpus to be analysed is hidden and only the words of the documents are observable.

To obtain the topic structure the posterior distribution must be determined. This is the conditional distribution of the hidden variables given the observed documents (Blei, 2012, pp. 80–81). The computational details are beyond the scope of this short survey so the focus will be on the findings of Hansen et al., (2018). They take advantage of the fact that in 1993 the Fed changed its procedure for publishing the transcripts of FOMC meetings. Since the mid-1970s nearly verbatim written records of the FOMC meetings have been produced and archived. The FOMC members were unaware of this detailed documentation and assumed that their contributions would not be made public. Due to political pressure, it was decided in October 1993 to publish this material both from the past and also in the future. Hence, there is a natural experiment that allows to identify the impact of this change in transparency on the behaviour of FOMC members. Maybe unsurprisingly a higher degree of transparency lead to an increasingly conform manner of the discussion contributions of FOMC members. This specifically applies to less experienced members who are more likely to avoid controversial statements if these are made public. Moreover, evidence for a more disciplined preparation of the meetings has been found (Hansen et al., 2018, pp. 841–844). Overall, the findings of this study are thus very well in line with the those of Acosta (2015).

Klejdysz and Lumsdaine (2023) are the first to apply LDA to transcripts of ECB press conferences. They identify communication patterns corresponding to the ECB monetary policy stance and find that the press conferences are indeed informative for stock markets. More precisely, switches in ECB communication regimes lead to an increase of stock market volatility.

Conclusions

In this paper a still relatively new development in applied research in monetary economics was looked at, the use of big data. It started out by providing some thoughts on the demarcation line between big data on the one hand and more traditional empirical work on the other hand. It became clear that big data is much more than just large (or larger) data sets. Still there is no clear and unequivocal definition of this demarcation line. Arguably there is not even a need for this. In any case in concrete applications there are frequently mixtures of quite traditional techniques and more recent big data approaches that prove to be useful.

A couple of concrete examples of research in monetary economics that can be associated with big data were also looked at. By doing so it became clear that there are many topics on which big data indeed can provide relevant and interesting insights that would not be feasible with more traditional techniques. It was also shown that to some extent big data might also lead to a larger role of subjective decisions not only in the choice of methods but also their application. Especially in formal text analysis a plethora of decisions both in pre-processing the data and specifying the use of the chosen method is necessary. This gives a huge and quite non-transparent role to the individual researcher. Similarly, the use of web scraped data comes necessarily at the cost of some lack of control concerning the quality and the representativeness of the data obtained. It was therefore concluded that big data will definitely contribute also to general macroeconomics and monetary economics by extending the set of answerable questions and also the way old questions can be answered. But as always, this innovation does not come without new problems and potential pitfalls.

References

Acosta, M. (2015). *FOMC responses to calls for transparency*. Finance and Economics Discussion Series 2015-060. Board of Governors of the Federal Reserve System.

Ademmer, M., Beckmann, J., Bode, E., Boysen-Hogrefe, J., Funke, M., Hauber, P., Heidland, T., Hinz, J., Jannsen, N., Kooths, S., Söder, M., Stamer, V., & Stolzenburg, U. (2021). *Big Data in der makroökonomischen Analyse* (32ed.). Institut für Weltwirtschaft.

Aprigliano, V., Ardizzi, G., & Monteforte, L. (2019). Using payment system data to forecast economic activity. *International Journal of Central Banking*, *15*(4), 55–80.

Bader B. W., & Chew, P. A. (2010). Algebraic techniques for multilingual document clustering. In M. W. Berry & J. Kogan (Eds.), *Text mining. Applications and theory* (pp. 21–36). John Wiley & Sons.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *13*1(4), 1593–1636.

Bandholz, H., & Funke, M. (2003). In search of leading indicators of economic activity in Germany. *Journal of Forecasting*, *22*, 277–297.

Bantis, E., Clements, M. P., & Urquhart, A. (2022). Forecasting GDP growth rates in the United States and Brazil using Google Trends. *International Journal of Forecasting*, in press.

Benchimol, J., Kazinnik, S., & Saadon, Y. (2022). Text mining methodologies with R: An application to central bank texts. *Machine Learning with Applications*, *8*, 100286.

Bholat, D., Hansen, S., Santos, P., & Schonhardt-Baily, C. (2015). *Text mining for central banks*. Bank of England. https://www.bankofengland.co.uk/-/media/boe/files/ccbs/resources/text-mining-for-central-banks.pdf

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Buono, D., Mazzi, G. L., Kapetanios, G., Marcellino, M., & Papailias, F. (2017). Big data types for macroeconomic nowcasting. *Eurostat Review on National Accounts and Macroeconomic Indicators*, *1*, 93–145.

Cavallo, A. (2013). Online and official price indexes: Measuring Argentina's inflation. *Journal of Monetary Economics*, *60*(2), 152–165.

Cavallo, A. (2018). Scraped data and sticky prices. *The Review of Economics and Statistics*, *100*(1), 105–119.

Cavallo A., & Rigobon R. (2016). The billion prices project: Using online prices for measurement and research. *Journal of Economic Perspectives*, *30*(2), 151–178.

Chakraborty, C., & Joseph, A. (2017, September). *Machine learning at central banks*. Bank of England Staff Working Paper, 674. https://www.bankofengland.co.uk/working-paper/2017/machine-learning-at-central-banks

Chapman, J., & Desai, A. (2022). *Macroeconomic predictions using payments data and machine learning*. Bank of Canada Staff Working Paper, 10. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3907281

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407.

Doerr, S., Gambacorta, L., & Serena, J. M. (2021, March). *Big data and machine learning in central banks*. BIS Working Papers, 930. https://www.federalreserve.gov/econres/ifdp/monetary-policy-uncertainty.htm

Economist, The. (2011). Gauging the gloom. https://www.economist.com/graphic-detail/2011/09/16/gauging-the-gloom

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature, 57*(3), 535–574.

Hammer, C. L., Kostroch, D. C., Quirós, G., & STA Internal Group. (2017). *Big data: Potential, challenges and statistical implications*. IMF Staff Discussion Note.

Hansen, S., McMahon, M., & Prat, A. (2018). Transparency and deliberation within the FOMC: A computational linguistics approach. *The Quarterly Journal of Economics*, *133*(2), 801–870.

Husted, L., Rogers, J., & Sun, B. (2017). *Monetary policy uncertainty*. International Finance Discussion Papers, 1215. https://www.federalreserve.gov/econres/ifdp/monetary-policy-uncertainty.htm

Irving Fisher Committee on Central Bank Statistics. (2021). *Use of big data sources and applications at central banks*. IFC Report, 13. Bank for International Settlements.

Klejdysz, J., & Lumsdaine, R. L. (2023). Shifts in ECB Communication: A textual analysis of the press conference. *International Journal of Central Banking*, *19*(2), 473–542.

Orphanides, A. (2001). Monetary policy rules based on real-time data. *American Economic Review*, *91*(4), 964–985.

Shapiro, A. H., & Wilson, D. J. (2022). Taking the fed at its word: A new approach to estimating central bank objectives using text analysis. *The Review of Economic Studies*, *89*(5), 2768–2805.

Zong, C., Xia, R., & Zhang, J. (2021). *Text data mining*. Springer.