Text-mining IMF country reports - an original dataset

David Mihalyi[†]
Akos Mate[‡]

September 24, 2018

Abstract

This article introduces an original panel dataset based on the text of country reports by the International Monetary Fund. It consists of a total of 2594 Article IV consultation and program review documents. The reports were published between 2004 and 2017 and cover 189 countries. The text of these reports provides a unique in-depth window into the IMF 's assessment of the most important macroeconomic issues. They provide indications of the perceived policy weaknesses, economic risks, ongoing reforms and implemented or neglected policy advice. Thus the content of IMF reports are widely used for qualitative and quantitative analysis in the economics, political science and IR literature. To our knowledge this is the first comprehensive dataset that aggregates these country reports.

The paper gives a detailed account on the data acquisition and management process. It also provides an overview and key descriptive statistics of the corpus of country reports. We then present three illustrative examples in applying text analytic techniques on the dataset. First, we compare conventional measures of resource dependence with a metric based on term frequency in reports. Second, we analyze mentions preceding reform events as a way to study reform intent. Finally, we show how mentions of keywords describing opposite fiscal policy stances mimic changes in IMF policy advice during the global financial crisis.

Taken together, this paper contributes an original dataset of IMF country reports and demonstrates how it can be a useful foundation for further research into the role of international financial institutions.

Keywords: economic policy, IMF, text analytics, original dataset.

*We would like to thank David Deritei for python code to scrape the website.
[†]Central European University-Natural Resource Governance Institute. Email: davidmihalyi@gmail.com
[‡]Central European University. Email: mate$_a$kos@*phd.ceu.edu*

1 Introduction

In this paper we are presenting an original panel dataset which contains the corpus of the country reports published by the International Monetary Fund (IMF) between 2004 and 2017 for 189 countries.[1] The IMF country reports are one of the main go-to sources for economists and social scientists to gain insight on the latest economic developments and discussions of policy reforms under way. These reports prepared by IMF staff roughly once year provide a unique insight on their 189 member states. In the past, many cross-country studies have built on the content of these country reports, analyzing the policy advice they provide (Ortiz et al., 2015; Rodrik, 2006; Roy and Almeida Ramos, 2012), the conditionalities linked to the loans (Kentikelenis et al., 2016) and critical assumptions underpinning the analysis (Blanchard and Leigh, 2013). Research by Shin and Glennerster (2003) found that countries face lower borrowing costs when they opt to make the content of their IMF reports public. In addition to the economics literature, the IMF country reports are also being used in the political science and international relations fields as well. Lombardi and Woods (2008) looks at the various outputs of the IMFs (including country reports) through an IR theory lens and examines whether they promote learning and socialization. Using data from IMF's Monitoring of Fund Arrangements database (Dreher et al., 2015) analyzes the connection between IMF conditionalities and a country's political importance. This non-exhaustive list of research demonstrates that IMF reports serve as an important data source.

With the advances in computer assisted text analysis it became possible to quantitatively assess large bodies of text, which previously would have required vast amount of hand-coding. It was shown that quantitative content analysis is a viable (and often better) technique when compared to qualitative coding by experts (Laver et al., 2003; Laver and Garry, 2000). Political science research has been experimenting with quantitative content analysis for some years now and developing novel methods to exploit the huge amount of text data available (Grimmer and Stewart, 2013; Lowe, 2008). The economics literature similarly started to make use of texts as data, which is surveyed in a recent paper, where Gentzkow et al. (2017) reviews the possible techniques and use cases for economic analysis.

Applications include using central bank communications to predict changes in policy rates (Apel and Grimaldi, 2012), fluctuations in Treasury securities (Lucca and Trebbi, 2009) and identifying home bias by analyzing the tone of the speeches of the members of the Governing Council of the Eurozone (Bennani and Neuenkirch, 2017). Similar approaches were used to forecast trends in unemployment by examining Google search queries (Choi and Varian, 2009). Other approaches used newspaper articles to measure policy uncertainty in the US (Baker et al., 2016) and forecast stock prices using the sentiment of newspaper articles relating to particular companies (Tetlock, 2007). Finally, Gehring and Lang (2018) used the tone of credit rating agency statements to evaluate the impact of IMF programs.

As this brief overview shows, the explosion of technical and methodological advances gave way to a wide range of research applications that provide important insights for social scientists. With our novel original dataset we aim to contribute to this growing body of research by providing a new and exciting way to look at the possible impact of the IMF's country reports. The paper is structured into four main sections. In

[1] The data set is available here: https://ourdataset.com

the first section we briefly cover the context of the IMF's country reports and why they are important data for research. The second section introduces the dataset. It provides details on the methodology of the data acquisition and processing. We also cover the basic descriptive qualities of the panel there and some discussion on the limitations and missing data. The third section provides some cursory glance at possible use cases for the data, such as using a dictionary to look up word frequencies of keywords of interest and associations between such frequencies and policy actions. Finally, in the fourth section we conclude our paper and discuss further avenues for refining the dataset and using it for research.

As Gentzkow et al. (2017, 50) notes in their review in the quantitative text analytics literature "virtually all of the methods applied to date, including those we would label as sophisticated or on the frontier, are based on fitting predictive models to simple counts of text features" . This is the method we follow in the illustrative examples we present in the use cases section.

2 IMF country reports and their contents

The IMF is one of the most influential international financial institution. It engages in monitoring economic and financial policies, offers technical assistance on economic affairs, and provides loans to countries in need. The monitoring of country policies is carried out as part of the consultations based on the Article IV of its Articles of Agreement. In the case of countries receiving IMF financial assistance, additional monitoring takes place through regular program reviews.

IMF country reports are drafted by IMF teams. A team of IMF economists visits the country in-person (the "IMF mission") to gather data, information and hold discussions with mainly government and central bank officials, but also sometimes private investors, labor representatives, members of parliament, and civil society organizations. Upon its return to headquarters, the staff prepares a report, which forms the basis for discussion by the Executive Board. The Board's views are subsequently summarized and transmitted to the country's authorities. The views expressed in these report are those of the IMF staff team. The views of the Executive Board are summarized in a Public Information Notice (PIN) which is attached to the Article IV report. Comments by the authorities on the staff report are also attached, if any were submitted at the time of the Executive Board discussion. The policy for publication of Article IV staff reports allows for the deletion of market sensitive information.

In principle, Article IV consultations with members takes place annually. The Fund may decide to place a member on an "extended consultation cycle" that is longer than 12 months but not longer than 24 months. This can be done only if the member does not meet any of the following criteria: the member is of systemic or regional importance; the member is perceived to be at macroeconomic risk; the member is facing pressing policy issues of broad interest to the Fund membership; the member has large outstanding credit to the Fund [2]. Countries under IMF program may also be placed on a 24-month consultation cycle, but will generally have more frequent (semi-annual or quarterly) program review reports, which combine a backward-looking

[2]Selected Decisions and Selected Documents of the IMF, Issue 39 - Article IV Consultation Cycles. As updated as of March 31, 2017

assessment with a forward-looking perspective [3].

On April 5, 1999, the IMF Executive Board agreed to a pilot project for the voluntary release of Article IV staff reports. Since February 2004, reports are made public by default unless the country blocks publication. Currently, nine out of ten member countries agree to publication of a Press Release, which summarizes the staff's and the Board's views, and four out of five countries agree to publication of the staff report itself. The availability of country report is even higher for program reviews, 96 percent of them are made public. [4]

These country reports touch on the following key topics: exchange rate, monetary, fiscal, and regulatory policies, macro-critical structural reforms. The main purpose is to surface any risk to domestic and global stability.

3 The IMF country reports dataset

3.1 Scraping, cleaning and constructing the database

We scraped 5380 PDF documents from from the IMF website which were tagged as country reports published between 1st January 2004 and 31st December 2017.[5] We then stored each document alongside the corresponding meta data displayed on the IMF website: the title of the document, its publication date, a series ID, the URL and the filename. This was done using Webscraper.io [6]. We then converted these PDF documents into plain text using the PDFtotext tool[7] and textract [8].

We first divided these documents into two groups based on their title: general and thematic country reports. We kept only the general staff reports: either those labelled Article IV Consultations, IMF program reviews or Post-Program Monitoring. This means that we dropped all thematic country reports, including "ROSC" reports on the compliance with various international standards and codes, Financial Sector Assessments, Poverty Reduction Strategy Papers, and Selected Issues Reports which as its name indicates will focus on only a handful of policy areas.

The reason we dropped these thematic reports is because unlike country reports, they don't conform to similar structure and depth of analysis. The content of a single such thematic report (e.g. 50 pages on Value Added Taxes) may skew the overall body of text for very strongly in one direction. It is also difficult to reconcile how the topics of these reports are selected and their scope defined. While encouraged, publication of thematic reports is voluntary, and there is more variance on whether and when (often years later) they are published. Therefore, unlike general staff reports, the body of text from thematic reports is unlikely to be balanced across major topics of relevance from the perspective of macroeconomic risks. We dropped 2731 such thematic reports and were left with 2649 general reports .

[3] Factsheet. As updated as of March 6, 2018
[4] Transparency at the IMF: http://www.imf.org/en/About/Factsheets/Sheets/2016/07/27/15/35/Transparency-at-the-IMF
[5] http://www.imf.org/en/publications/search?when=After&series=IMF+Staff+Country+Reports
[6] http://webscraper.io/
[7] https://github.com/jalan/PDFtotext
[8] https://textract.readthedocs.io/

From the 2649 general country reports downloaded and converted we dropped 55 entries (less than 2 percent of cases) where we failed to convert the PDF into plain text or it resulted in a very small document likely only covering a fraction of the document (20 kilobyte or smaller).

We also note that there are 44 reports that relate not to a single country rather to a country group, such as a currency or trade union (e.g. Euro, CEMAC, ECCU). The IMF also writes regular country reports on the individual member states within these groups, and we expect those provide more directly relevant information on the country in question. Therefore we do not use the country group reports in country level analysis, but do include them in the dataset.

The final dataset includes 2594 reports, from 2004 to 2017. We chose 2004 as our starting year, because this is the year when reports became published by default. There are much fewer country reports from earlier periods, and even among these the majority are scanned PDF which make text recognition difficult and imprecise. The make-up of the reports is the following: 44 Country group reports and 2550 individual country reports. The variables are the following:

- year: Year of the report being published online
- country: Name of the country being reported (Country reports are included under this variable)
- ccode: ISO 3166-1 alpha-3 country code
- doc_name: The internal document name of the report (serves as unique ID for the reports)
- title: Full title of the country report
- text: The raw plain text content of the report, without any preprocessing or formatting.[9]

[9]We opted to preserve the raw text in the database as each analysis requires a unique set of preprocessing and cleaning and this preserves all of the data contained in the reports.

3.2 Main properties

The panel consists of 2594 observations for the years between 2004 and 2017. The number of reports per year increases rapidly from its 2004 level of 104 to 162 in 2005 with temporary increases in 2009 - 2011, due to the global financial crisis. The detailed breakdown of country reports per year is presented in Table 3.2.

Table 1: Country reports per year, 2004-2017

Year	No. of reports
2004	104
2005	162
2006	173
2007	173
2008	176
2009	202
2010	222
2011	209
2012	183
2013	186
2014	195
2015	184
2016	192
2017	189

The distribution of reports per countries is rather uneven (as shown by Figure 1) as a result of some countries receiving more IMF staff visits than others. There are some notable instances of denying access to IMF staff, such as Venezuela and Argentina. The median reports per country is 13, while the mean is 13.5. As for outliers in the report numbers, there are 15 countries that have 5 or less, and 8 countries with 25 or more reports.[10]

In the Appendix, Table 6 gives a more detailed look at number of reports per countries in the panel, as well as the first and last year the country had an IMF report published and average report per year for the period the country participated in the reporting process. However, our dataset covers most of the countries (with some notable exceptions, such as Venezuela) as shown in Figure 2. It is also visible that our panel contains more reports for the Central Eastern European, and certain African and Latin American countries.[11]

The descriptive statistics for the corpus in Table 3.2 shows that the mean word count and the standard deviation of the word counts are stable over the panel. However, the minimum and maximum word counts for some country reports display considerable fluctuations over the years. The 2012-2014 maximum values with over a 110 000 words in the reports are all reports on the Greek economy, while the minimum values correspond to brief program reviews or staff notes.

The distribution of word counts is shown in Figure 3, with the 5 outliers highlighted. Due to the verbosity

[10] Countries with 5 or less reports: Anguilla, Argentina, Bhutan, Brazil, Ecuador, Guyana, Macau SAR China, Montserrat, Nauru, Serbia and Montenegro, Somalia, South Sudan, Syria, Tuvalu, Uzbekistan. Countries with 25 or more reports: Armenia, Ireland, Jamaica, Mexico, Pakistan, Romania, Rwanda, Uganda

[11] With high report count for Ireland and Greece as well, due to the global financial crisis.

Figure 1: Distribution of country reports per country

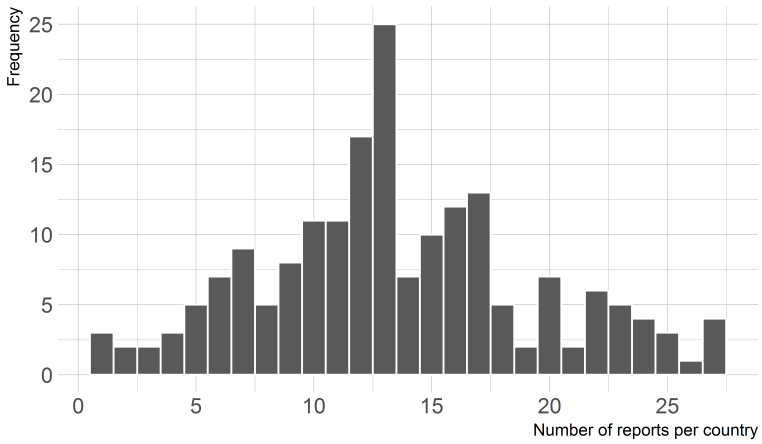
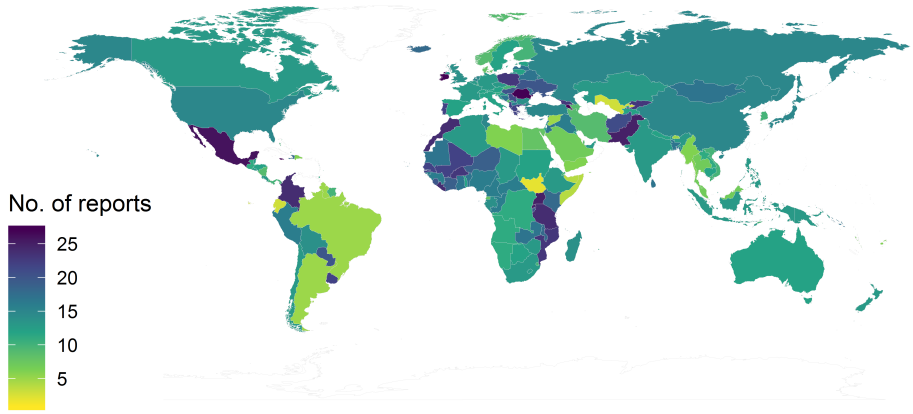


Figure 2: Country coverage of the dataset



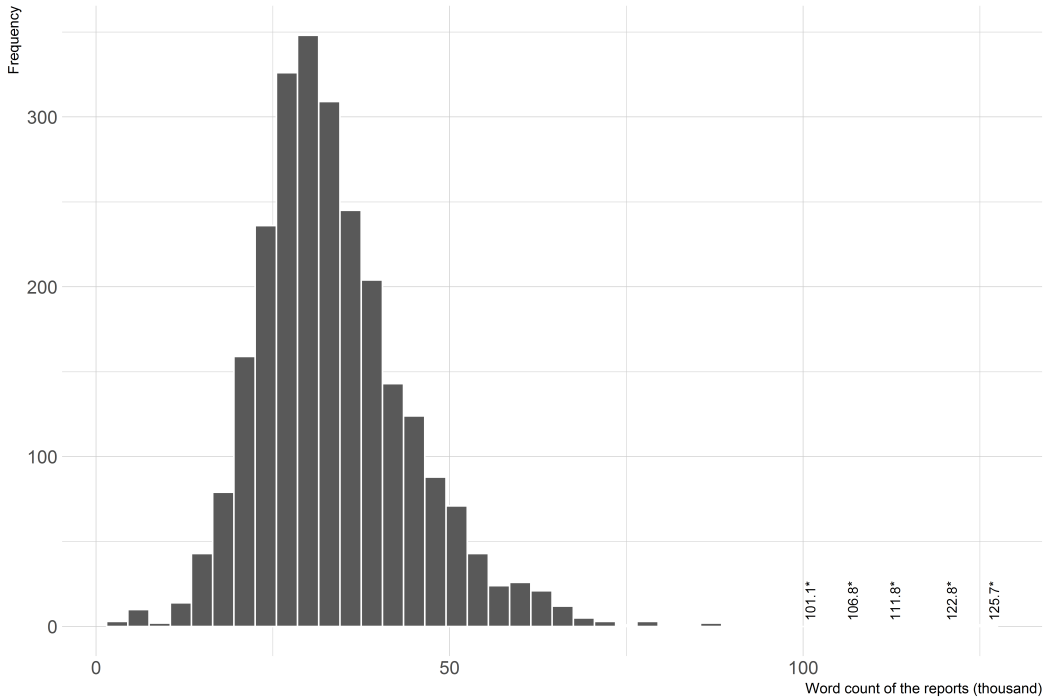
induced by the global financial crisis, the distribution is somewhat skewed. The distribution and descriptive statistics for the sentence count in the corpus tracks the word counts.

Table 2: Descriptive statistics for the corpus

Year	No. of reports	Mean word count	Std.deviation	Minimum	Maximum
2004	104	35.0	11.0	17.3	78.1
2005	162	35.5	11.4	13.7	87.3
2006	173	34.4	10.0	13.1	64.8
2007	173	31.1	9.6	5.2	60.5
2008	176	31.5	9.2	5.3	54.0
2009	202	31.0	10.2	4.8	56.6
2010	222	32.4	10.7	9.3	70.7
2011	209	31.5	11.5	2.3	82.9
2012	183	34.1	12.4	12.7	111.8
2013	186	35.4	14.9	10.9	125.7
2014	195	34.3	12.0	13.6	122.8
2015	184	34.5	10.4	5.5	78.1
2016	192	37.3	10.9	3.0	75.2
2017	189	36.2	11.2	15.7	87.9

Note: The statistics are in thousands.

Figure 3: Distribution of the word count in the corpus



*These outliers represent country reports for Greece

3.3 Limitations of the dataset

One key issue that research should pay attention is that data (country reports) are not missing at random in the data set. The reasons behind the missing country reports are manifold. Not all countries get the same amount of surveillance and smaller and less accessible countries (either geography or language) may be monitored in less detail given resource constraints of the Fund. On the other hand, countries with higher macroeconomic vulnerabilities or undergoing IMF program may get more in depth monitoring.

In some cases the country authorities may refuse to have the Article IV published. IMF publications on transparency reveal this happened in about 20 percent of cases in 2004-2005, when our dataset starts and gradually declined to 5 percent of cases in 2014-16 [12]. An earlier study by Edwards et al. (2011) finds that more democratic governments are more likely to release reports, as well as a strong variation in regional patterns (most notably less report in Latin America).

These matter for the research design. The likelihood of a key word appearing at all in reports will depend on the likelihood of the report being published and its depth. The frequency of certain themes may also affect whether the report gets approval from authorities to be published.

We build on Edwards et al. (2011) in the selection of explanatory variables and review how the following variables affect report availability and length of reports.

- Population - in log form (logpop)
- GDP per capita - in log form (loggdppc)
- IMF program in place for at least 5 months in the year (Dreher, 2006) updated) - dummy variable (IMFprogram)
- Polity IV score - on a -10 to 10 scale (polity2).

We first review the relationship between the population and income variables and the number of reports, the average length of reports (thousands of words) and total length of reports (thousand of words) by country in Table 3.3 using linear regressions. We than present the likelihood of having a public report in a given year in a given country in Table 3.3 using probit regressions depending on whether the country is going through an IMF program and its polity IV score. We also present marginal effects at mean of the probit regression in 3.3.

We find that there are somewhat fewer reports available in our dataset for smaller and richer countries. The reports available also tend to be shorter on average for these countries. We also find that we there is a higher likelihood of having a report in instances where countries are undergoing IMF programs at the time of assessment. An IMF program increases likelihood of coverage in any given year by 17 percentage point. We also have somewhat higher coverage in country years with higher institutional scores as measured by

[12] Source: Key Trends in Implementation of the Fund's Transparency Policy, IMF

Table 3: Linear regression results on the number of reports and word count across reports by country

	Dependent variable:		
	reports	total_thousand_words	avg_thousand_words
	(1)	(2)	(3)
loggdppc	-0.777***	-54.945***	-1.943***
	(0.277)	(11.522)	(0.277)
logpop	0.589***	26.520***	0.777***
	(0.185)	(7.720)	(0.186)
Constant	11.270***	525.732***	37.352***
	(4.009)	(166.888)	(4.018)
Observations	183	183	183
R^2	0.103	0.180	0.295
Adjusted R^2	0.093	0.171	0.287
Residual Std. Error (df = 180)	5.389	224.365	5.402
F Statistic (df = 2; 180)	10.305***	19.731***	37.718***
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4: Probit regression results on the availability of report for a given country-year

	Dependent variable:	
	yreport	
	(1)	(2)
polity2	0.024***	
	(0.005)	
IMFprogram		0.517***
		(0.126)
Constant	0.591***	0.578***
	(0.038)	(0.030)
Observations	1,855	2,149
Log Likelihood	-1,023.741	-1,245.885
Akaike Inf. Crit.	2,051.483	2,495.771
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 5: Marginal effect at mean of Polity score

	dF/dx	Std. Err.	z	P>|z|
polity2	0.008	0.002	4.666	0.00000

Table 6: Marginal effect at mean of IMF program

	dF/dx	Std. Err.	z	P>|z|
IMFprogram	0.170	0.041	4.133	0.00004

polity: a 1 point increase in the polity score at the median value (on 21 point scale) increases likelihood of having a report by 0.8 percentage point.

Because rare keywords are more likely to appear in countries with more reports and longer text, measures of keyword appearance may be somewhat biased towards larger, less wealthy and more democratic countries and years of financial difficulty. These factors need to be taken into account when designing research using the dataset. One avenue to address the bias is to measure the relative frequency of appearance of chosen keywords.

4 Selected use cases

The below section presents illustrations at how the dataset may be applied in research. These short demonstrations are not intended at providing substantive new empirical contributions. Rather they are meant to showcase how simple text analytic metrics capture meaningful and important characteristics of IMF country diagnosis and advice.

- Oil dependence
- Fiscal rules - The sequencing of reform intent vs actual reform event
- Stimulus vs consolidation - Patterns in IMF policy priorities

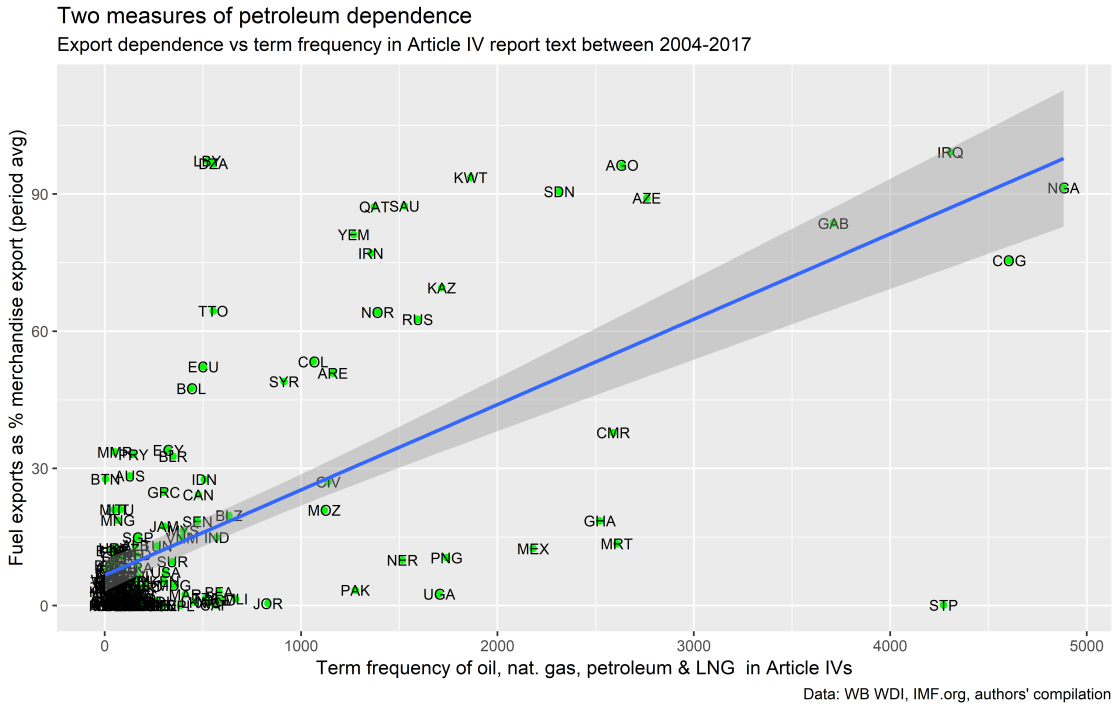
For the analysis we used the R package "quanteda" (Benoit, 2018) to create a document-feature matrix (dfm) from our corpus, which allows for extracting various word frequency measures. We created a dfm with single words as tokens and with bi-grams (two word combinations) as tokens.

Short illustrations:

Oil dependence We review whether the countries which are considered resource dependent based on conventional economic metrics are the ones where oil is most frequently discussed in country reports. We calculated the frequency of appearance of the word oil across reports, using term frequency with inverse document frequency weighting. We contrast that with oil dependence measures as fuel exports as percentage of all merchandise export (from World Bank). Note that this includes oil and gas and that there are many known oil producers with no fuel export data. We plotted both on figure 5.

We find a strong positive correlation, with an R-squared of nearly 50 percent. Most countries export no oil, or oil export are tiny in fraction to other goods, and in their reports oil is infrequently mentioned. The

Figure 4: Two measures of oil dependence



5 countries where oil is most frequently mentioned are in order: Chad, Nigeria, Republic of Congo, Iraq, Gabon, all extremely oil dependent countries (note that no export data for Chad, hence not displayed on plot). This reaffirms, that IMF reports focus heavily on the key sector in these countries.

Another interesting insight comes from the countries below the trend line, where oil is mentioned more than the level of resource dependence would directly imply. Among these are 4 new or prospective oil exporters Uganda, Ghana, Sao Tome, Mauritania, where oil export is low, but IMF reports discuss the expected rapid ramping up of oil production and exports in the future.

This example highlights how text search can be used to capture the perceived salience of various macroeconomic risk factors across reports.

Comparing actual reform vs reform intent. We have conducted bigram search on the term 'fiscal rule' and related terms ('expenditure rule', 'debt rule', 'balance rule', 'deficit rule', 'revenue rule'). This allows to analyze whether fiscal rules are being more frequently discussed in IMF country reports prior to adoption or major changes to existing fiscal rules (using the IMF FAD Fiscal Rule dataset).

Fiscal rules are mentioned in 25 percent of all reports. Figure 6 shows that there has been a gradual increase in the number of fiscal rules across countries, and they are also being mentioned in increasing frequency across reports especially since 2008.

Figure 7 shows the frequency of search term appearance depending on whether the country has a fiscal rule,

Figure 5: Percentage of countries with fiscal rules and reports with fiscal rule mentions by year

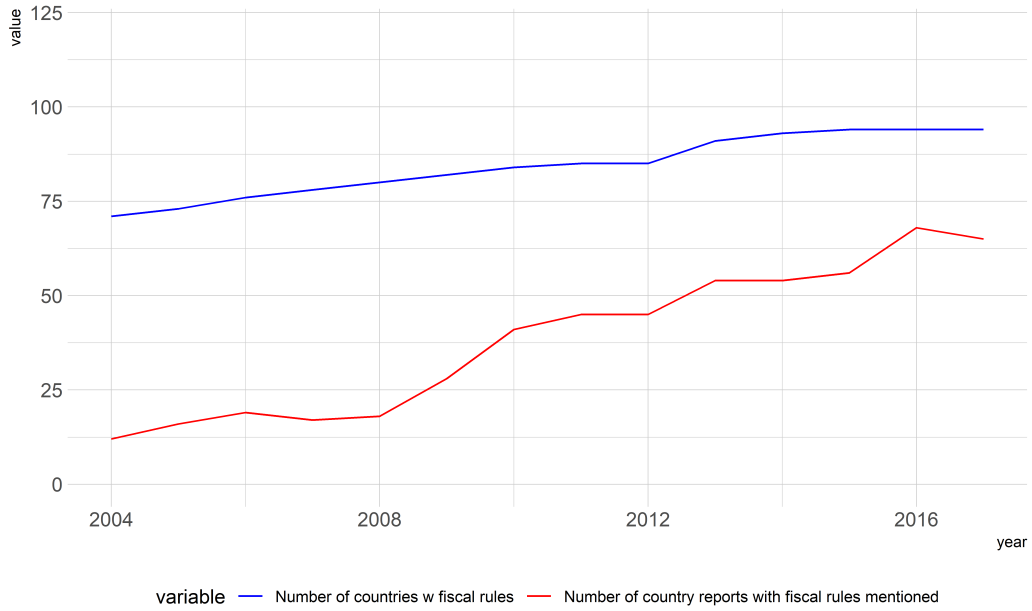
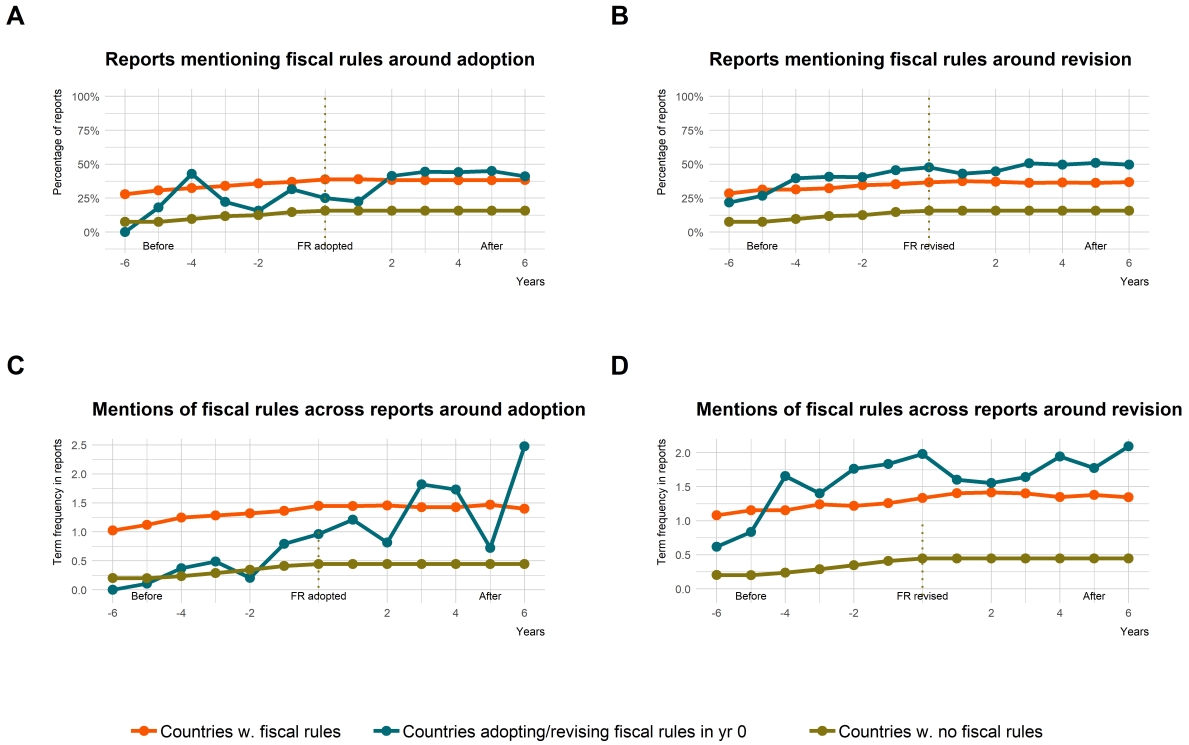


Figure 6: Mentions of fiscal rules vs actual fiscal rules



has no fiscal rule, and when the fiscal rule is being implemented or being reformed. Fiscal rules are mentioned at all about twice as frequently in countries which currently have fiscal rules (35 percent of reports), than in countries without fiscal rules (18 percent of reports). They are also being discusses with increasing frequency over time across all groups. But the graph also shows a sharp increase in mentions (and especially term frequency, which takes into account how much it is mentioned in a signle report) before first implementation of a fiscal rule and major reform of the fiscal rule (as per IMF FAD dataset). This suggests that IMF is closely monitoring reforms to fiscal rules and probably in many cases providing its own advice.

This example highlights how text search can be used to capture the sequencing of IMF advice and reform events.

Changes in perceived policy priorities The dataset also enables to monitor changes to perceived policy priorities. We have constructed dictionaries that describe opposite fiscal policy directions.[13]

- Fiscal consolidation: 'fiscal consolid*','fiscal discipl*', 'restor* fiscal', 'fiscal slip*', 'fiscal solv*','fiscal adjust'
- Fiscal stimulus: 'fiscal stimul*', 'stimul* package', 'fiscal expan*'.

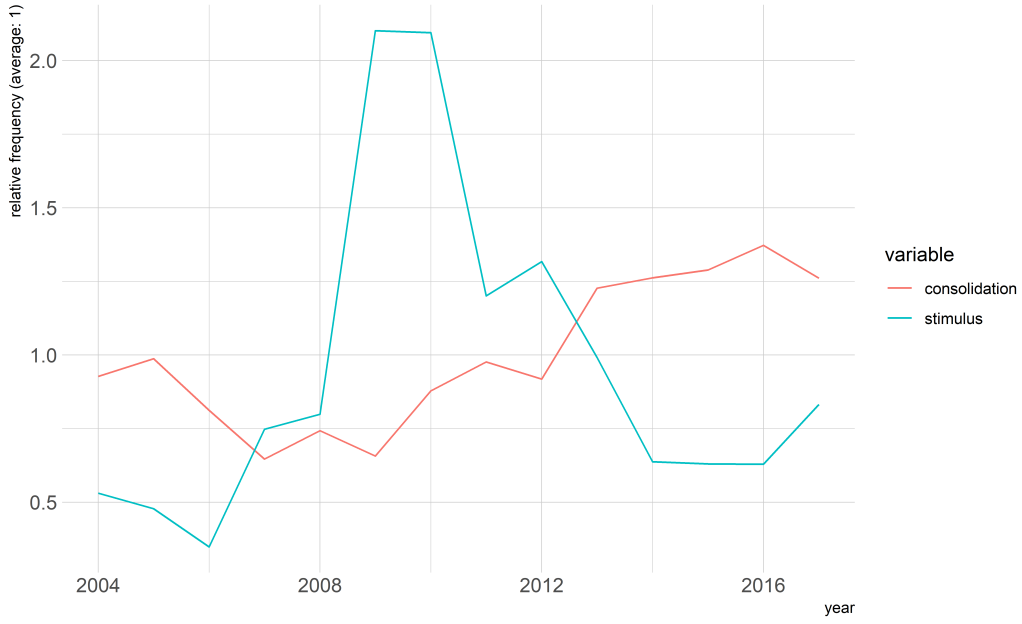
The longer list of key words associated with consolidation appears much more frequently than the words associated with stimulus. But when evaluating relative frequency of their appearance, we can discern trends over time. We find a sharp increase in mentions of stimulus in 2009 and 2010, then followed by a rapid decline. In the meantime fiscal consolidation mentions have increased starting in 2010.

This mirrors the patterns described in (IMF, 2014) and (Dhar, 2014) which discusses IMF policy advice in response to the financial crisis. In fact, the so-called Triennial Surveillance Review, manually reviewed their country reports to map the policy advice it had provided on a sub-sample of 24 countries. They found that across these 24 countries short-term stimulus was recommended in 3/4 of IMF article IVs in 2009 (IMF, 2014).

This illustration provides an example on how text search can be used to map whether changes in global policy pronouncements are being reflected in country reports.

[13]The * denotes a wildcard, e.g.: restor* will find restor*ing* as well as restor*ation*

Figure 7: Relative frequency of mentions of keywords associated with consolidation and stimulus by year



5 Conclusion

IMF country reports are a treasure trove of information on the economic and policy developments of countries across the world. They also provide a window into the policy priorities and advice that the IMF provides. We presented a new dataset which builds on the content of these country reports and show how simple text analytic techniques can be used to gain new insight. Subsequent refinement of the dataset may decompose reports into its chapters, enrich the meta data with authors and exact dates of drafting. Subsequent research may use the dataset to explore critical determinants of report content and sentiment as well as the link between report content and policy outcomes.

6 Appendix

Table 7: Number of reports per country in the panel

Country	ccode	First report	Last report	No. of reports	Mean
Afghanistan	AFG	2004	2017	21	1.6
Angola	AGO	2005	2017	11	0.9
Anguilla	AIA	2012	2012	1	-
Albania	ALB	2004	2017	22	1.7
Netherlands Antilles	ANT	2005	2017	7	0.6
United Arab Emirates	ARE	2004	2017	13	1.0
Argentina	ARG	2004	2017	5	0.4
Armenia	ARM	2004	2017	27	2.1
Antigua & Barbuda	ATG	2004	2015	7	0.6
Australia	AUS	2004	2017	12	0.9
Austria	AUT	2004	2017	12	0.9
Azerbaijan	AZE	2005	2016	9	0.8
Burundi	BDI	2005	2015	20	2.0
Belgium	BEL	2005	2017	13	1.1
Benin	BEN	2004	2017	17	1.3
Burkina Faso	BFA	2005	2017	23	1.9
Bangladesh	BGD	2004	2017	17	1.3
Bulgaria	BGR	2004	2017	14	1.1
Bahamas	BHS	2005	2017	10	0.8
Bosnia & Herzegovina	BIH	2004	2016	17	1.4
Belarus	BLR	2004	2017	17	1.3
Belize	BLZ	2004	2017	13	1.0
Bolivia	BOL	2004	2017	14	1.1
Brazil	BRA	2012	2017	5	1.0
Barbados	BRB	2004	2016	11	0.9
Bhutan	BTN	2007	2016	5	0.6
Botswana	BWA	2004	2017	12	0.9
Central African Republic	CAF	2005	2017	16	1.3
Canada	CAN	2004	2017	13	1.0
Switzerland	CHE	2005	2016	12	1.1
Chile	CHL	2004	2016	13	1.1
China	CHN	2004	2017	15	1.2
Côte d'Ivoire	CIV	2007	2017	19	1.9
Cameroon	CMR	2005	2017	16	1.3
Congo - Kinshasa	COD	2004	2015	11	1.0

Continued on next page

Country	ccode	First report	Last report	No. of reports	Mean
Congo - Brazzaville	COG	2004	2015	16	1.5
Colombia	COL	2004	2017	24	1.8
Comoros	COM	2004	2016	14	1.2
Cape Verde	CPV	2004	2016	16	1.3
Costa Rica	CRI	2004	2017	13	1.0
Cyprus	CYP	2005	2017	16	1.3
Czechia	CZE	2004	2017	13	1.0
Germany	DEU	2004	2017	13	1.0
Djibouti	DJI	2004	2017	10	0.8
Dominica	DMA	2004	2017	16	1.2
Denmark	DNK	2004	2017	8	0.6
Dominican Republic	DOM	2010	2017	6	0.9
Algeria	DZA	2005	2017	13	1.1
Ecuador	ECU	2006	2016	3	0.3
Egypt	EGY	2005	2017	8	0.7
Spain	ESP	2005	2017	12	1.0
Estonia	EST	2004	2017	12	0.9
Ethiopia	ETH	2005	2016	13	1.2
Finland	FIN	2005	2017	10	0.8
Fiji	FJI	2010	2016	6	1.0
France	FRA	2004	2017	13	1.0
Micronesia	FSM	2005	2017	7	0.6
Gabon	GAB	2005	2017	14	1.2
United Kingdom	GBR	2004	2016	14	1.2
Georgia	GEO	2004	2017	23	1.8
Ghana	GHA	2004	2017	17	1.3
Guinea	GIN	2004	2017	16	1.2
Gambia	GMB	2004	2017	20	1.5
Guinea-Bissau	GNB	2005	2017	16	1.3
Equatorial Guinea	GNQ	2005	2016	8	0.7
Greece	GRC	2005	2017	22	1.8
Grenada	GRD	2005	2017	15	1.2
Guatemala	GTM	2005	2016	11	1.0
Guyana	GUY	2010	2017	5	0.7
Hong Kong SAR China	HKG	2005	2017	9	0.8
Honduras	HND	2006	2017	12	1.1
Croatia	HRV	2004	2016	13	1.1
Haiti	HTI	2005	2017	18	1.5
Hungary	HUN	2004	2017	20	1.5

Continued on next page

Country	ccode	First report	Last report	No. of reports	Mean
Indonesia	IDN	2004	2017	13	1.0
India	IND	2005	2017	13	1.1
Ireland	IRL	2004	2017	27	2.1
Iran	IRN	2004	2017	9	0.7
Iraq	IRQ	2005	2017	16	1.3
Iceland	ISL	2005	2017	18	1.5
Israel	ISR	2004	2017	12	0.9
Italy	ITA	2005	2017	12	1.0
Jamaica	JAM	2004	2017	27	2.1
Jordan	JOR	2004	2017	17	1.3
Japan	JPN	2004	2017	15	1.2
Kazakhstan	KAZ	2004	2017	13	1.0
Kenya	KEN	2008	2017	18	2.0
Kyrgyzstan	KGZ	2004	2017	23	1.8
Cambodia	KHM	2004	2017	12	0.9
Kiribati	KIR	2009	2017	7	0.9
St. Kitts & Nevis	KNA	2007	2017	15	1.5
South Korea	KOR	2005	2016	10	0.9
Kuwait	KWT	2004	2017	12	0.9
Laos	LAO	2005	2017	11	0.9
Lebanon	LBN	2006	2017	15	1.4
Liberia	LBR	2005	2017	24	2.0
Libya	LBY	2005	2013	6	0.8
St. Lucia	LCA	2006	2017	8	0.7
Sri Lanka	LKA	2004	2017	17	1.3
Lesotho	LSO	2004	2016	13	1.1
Lithuania	LTU	2005	2017	12	1.0
Luxembourg	LUX	2004	2017	10	0.8
Latvia	LVA	2004	2017	17	1.3
Macau SAR China	MAC	2014	2017	2	0.7
Morocco	MAR	2004	2017	24	1.8
Moldova	MDA	2005	2017	17	1.4
Madagascar	MDG	2004	2017	14	1.1
Maldives	MDV	2009	2017	7	0.9
Mexico	MEX	2004	2017	26	2.0
Marshall Islands	MHL	2006	2016	6	0.6
Macedonia	MKD	2004	2017	16	1.2
Mali	MLI	2004	2017	22	1.7
Malta	MLT	2005	2017	9	0.8

Continued on next page

Country	ccode	First report	Last report	No. of reports	Mean
Myanmar (Burma)	MMR	2012	2017	6	1.2
Montenegro	MNE	2008	2017	9	1.0
Mongolia	MNG	2005	2017	17	1.4
Mozambique	MOZ	2004	2016	23	1.9
Mauritania	MRT	2006	2017	17	1.5
Montserrat	MSR	2012	2012	1	-
Mauritius	MUS	2005	2017	11	0.9
Malawi	MWI	2004	2017	18	1.4
Malaysia	MYS	2009	2017	7	0.9
Namibia	NAM	2005	2016	11	1.0
Niger	NER	2004	2017	19	1.5
Nigeria	NGA	2004	2017	16	1.2
Nicaragua	NIC	2004	2017	9	0.7
Netherlands	NLD	2004	2017	15	1.2
Norway	NOR	2005	2017	9	0.8
Nepal	NPL	2004	2017	12	0.9
Nauru	NRU	2017	2017	1	-
New Zealand	NZL	2004	2017	13	1.0
Pakistan	PAK	2004	2017	25	1.9
Panama	PAN	2006	2017	10	0.9
Peru	PER	2006	2017	16	1.5
Philippines	PHL	2005	2017	13	1.1
Palau	PLW	2006	2016	6	0.6
Papua New Guinea	PNG	2004	2017	13	1.0
Poland	POL	2004	2017	23	1.8
Portugal	PRT	2005	2017	22	1.8
Paraguay	PRY	2004	2017	20	1.5
Qatar	QAT	2008	2017	9	1.0
Romania	ROU	2004	2017	27	2.1
Russia	RUS	2004	2017	15	1.2
Rwanda	RWA	2004	2017	25	1.9
Saudi Arabia	SAU	2011	2017	7	1.2
Serbia and Montenegro	SCG	2004	2006	4	2.0
Sudan	SDN	2005	2017	12	1.0
Senegal	SEN	2006	2017	22	2.0
Singapore	SGP	2004	2017	14	1.1
Solomon Islands	SLB	2004	2016	17	1.4
Sierra Leone	SLE	2005	2017	20	1.7
El Salvador	SLV	2005	2016	13	1.2

Continued on next page

Country	ccode	First report	Last report	No. of reports	Mean
San Marino	SMR	2004	2017	11	0.8
Somalia	SOM	2015	2017	4	2.0
Serbia	SRB	2006	2017	21	1.9
South Sudan	SSD	2014	2017	2	0.7
São Tomé & Principe	STP	2005	2017	18	1.5
Suriname	SUR	2005	2016	10	0.9
Slovakia	SVK	2005	2017	10	0.8
Slovenia	SVN	2004	2017	11	0.8
Sweden	SWE	2004	2017	12	0.9
Swaziland	SWZ	2006	2017	10	0.9
Seychelles	SYC	2008	2017	20	2.2
Syria	SYR	2005	2010	5	1.0
Chad	TCD	2005	2017	13	1.1
Togo	TGO	2007	2017	12	1.2
Thailand	THA	2009	2017	7	0.9
Tajikistan	TJK	2004	2012	12	1.5
Timor-Leste	TLS	2004	2017	10	0.8
Tonga	TON	2006	2016	11	1.1
Trinidad & Tobago	TTO	2005	2017	11	0.9
Tunisia	TUN	2004	2017	15	1.2
Turkey	TUR	2004	2017	15	1.2
Tuvalu	TUV	2011	2016	4	0.8
Tanzania	TZA	2004	2017	24	1.8
Uganda	UGA	2004	2017	25	1.9
Ukraine	UKR	2005	2017	20	1.7
Uruguay	URY	2004	2017	22	1.7
United States	USA	2004	2017	15	1.2
Uzbekistan	UZB	2007	2013	3	0.5
St. Vincent & Grenadines	VCT	2005	2017	13	1.1
Vietnam	VNM	2005	2017	10	0.8
Vanuatu	VUT	2005	2016	7	0.6
Samoa	WSM	2005	2017	8	0.7
Kosovo	XKC	2011	2017	13	2.2
Yemen	YEM	2005	2014	6	0.7
South Africa	ZAF	2004	2017	13	1.0
Zambia	ZMB	2004	2017	17	1.3
Zimbabwe	ZWE	2004	2017	13	1.0

References

Apel, M. and M. Grimaldi (2012). The information content of central bank minutes.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics 131*(4), 1593–1636.

Bennani, H. and M. Neuenkirch (2017). The (home) bias of european central bankers: new evidence based on speeches. *Applied Economics 49*(11), 1114–1131.

Benoit, K. (2018). *quanteda: Quantitative Analysis of Textual Data*. R package version 1.3.4.

Blanchard, O. J. and D. Leigh (2013). Growth forecast errors and fiscal multipliers. *American Economic Review 103*(3), 117–20.

Choi, H. and H. Varian (2009). Predicting initial claims for unemployment benefits. *Google Inc*, 1–5.

Dhar, S. (2014). Imf macroeconomic policy advice in the financial crisis aftermath. *IEO Background Paper No. BP/14/07 (Washington: International Monetary Fund)*.

Dreher, A. (2006). Imf and economic growth: The effects of programs, loans, and compliance with conditionality. *World Development 34*(5), 769–788.

Dreher, A., J.-E. Sturm, and J. R. Vreeland (2015). Politics and imf conditionality. *Journal of Conflict Resolution 59*(1), 120–148.

Edwards, M. S., K. A. Coolidge, and D. A. Preston (2011). Who reveals? transparency and the imf's article iv consultations.

Gehring, K. and V. Lang (2018). Stigma or cushion? imf programs and sovereign creditworthiness.

Gentzkow, M., B. T. Kelly, and M. Taddy (2017). Text as data. Technical report, National Bureau of Economic Research.

Grimmer, J. and B. M. Stewart (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis 21*(3), 267–297.

IMF (2014). 2014 triennial surveillance review - staff background studies. *International Monetary Fund Policy Papers*.

Kentikelenis, A. E., T. H. Stubbs, and L. P. King (2016). Imf conditionality and development policy space, 1985–2014. *Review of International Political Economy 23*(4), 543–582.

Laver, M., K. Benoit, and J. Garry (2003). Extracting policy positions from political texts using words as data. *American Political Science Review 97*(2), 311–331.

Laver, M. and J. Garry (2000). Estimating policy positions from political texts. *American Journal of Political Science*, 619–634.

Lombardi, D. and N. Woods (2008). The politics of influence: An analysis of imf surveillance. *Review of International Political Economy 15*(5), 711–739.

Lowe, W. (2008). Understanding wordscores. *Political Analysis 16*(4), 356–371.

Lucca, D. O. and F. Trebbi (2009). Measuring central bank communication: an automated approach with application to fomc statements. Technical report, National Bureau of Economic Research.

Ortiz, I., M. Cummins, J. Capaldo, and K. Karunanethy (2015). The decade of adjustment: A review of austerity trends 2010-2020 in 187 countries.

Rodrik, D. (2006). Goodbye washington consensus, hello washington confusion? a review of the world bank's economic growth in the 1990s: learning from a decade of reform. *Journal of Economic literature 44*(4), 973–987.

Roy, R. and R. Almeida Ramos (2012). Imf article iv reports: an analysis of policy recommendations. Technical report, Working Paper, International Policy Centre for Inclusive Growth.

Shin, Y. and M. R. Glennerster (2003). *Is transparency good for you, and can the IMF help?* Number 3-132. International Monetary Fund.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance 62*(3), 1139–1168.