



IMF Working Paper

Strategy Policy and Review Department

License to Spill: How Do We Discuss Spillovers in Article IV Staff Reports

Prepared by Jelle Barkema, Borislava Mircheva, Mico Mrkaic, and Yuanchen Yang

Authorized for distribution by Rupa Duttagupta

May 2021

IMF Working Papers describe research in progress by the author(s) and are published to elicit comments and to encourage debate. The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

Abstract

This paper dives into the Fund’s historical coverage of cross-border spillovers in its surveillance. We use a state-of-the-art deep learning model to analyze the discussion of spillovers in all IMF Article IV staff reports between 2010 and 2019. We find that overall, while the discussion of spillovers decreased over time, it was pronounced in the staff reports of some systemically important economies and during periods of global spillover events. Spillover discussions were more prominent in staff reports covering advanced and emerging market economies, possibly reflecting their role as sources of global spillovers. The coverage of spillovers was higher in the context of the real, financial, and external sectors. Also, countries with larger economies, higher trade and capital account openness and lower inflation are more likely to discuss spillovers in their Article IV staff reports.

Keywords: Spillovers, Surveillance, Article IV, Staff Reports

Author’s E-Mail Address Jelle.Barkema@bankofengland.co.uk; BMircheva@imf.org; MMrkaic@imf.org; and YYang6@imf.org

CONTENTS	PAGE
Abstract _____	1
I. INTRODUCTION _____	3
II. THE FUND’S MANDATE TO COVER SPILLOVERS _____	4
III. LITERATURE REVIEW _____	5
IV. METHODOLOGY AND DATA _____	6
A. The BERT Model _____	6
B. Comparison to Other Existing Models _____	8
V. BERT MODEL RESULTS _____	8
A. Assessment of the BERT Model Performance. _____	8
B. Discussion of the General Spillover Patterns _____	9
C. How are Major Spillover Events Covered in Article IV Staff Reports? _____	11
VI. THE PROBABILITY OF DISCUSSING SPILLOVERS IN ARTICLE IV STAFF REPORTS _____	13
VII. CONCLUSION AND POLICY RECOMMENDATIONS _____	18
 FIGURES	
1. Number of Countries in Database by Income Group by Year _____	6
2. Overview of Training, Testing, and Prediction Processes _____	7
3. Spillover Mentions Over Time _____	9
4. Spillover Mentions by Income Groups _____	10
5. Spillover Mentions by Region _____	10
6. Spillover Mentions by Sector _____	11
7. Spillover Mentions over the Period of 2010–19 _____	12
 TABLES	
1. BERT Model Performance _____	9
2. The Probability of Discussing Spillovers _____	14
3. The Probability of Discussing Spillovers in Sectors _____	16
4. The Probability of Discussing Spillovers in Sectors using SUR Models _____	17
5. The Percentage of Spillover Discussion _____	18
References _____	23
 ANNEXE	
Technical Annex _____	20

I. INTRODUCTION¹

Work on cross-border spillovers at the IMF is developed and communicated using a wide range of surveillance vehicles.² The coverage of spillover analysis in Article IV staff reports was mainstreamed by the 2012 Integrated Surveillance Decision (ISD), where country teams are required to discuss outward spillovers induced by policies employed by systemic economies. Consequently, country teams are expected to cover both inward and outward spillovers in their surveillance work.

The post-pandemic landscape could expose Fund member countries to a sustained period of macro-financial spillovers. While the entry into the global recession triggered by the pandemic was simultaneous, the recovery is increasingly becoming multi-speed and differentiated (see IMF, April 2021, World Economic Outlook and Global Financial Stability Reports). This reflects differences in the intensity of initial shocks, pre-existing structural characteristics, and countries' policy positions, and, relatedly, strength of policy support across countries. Subsequently, economies will likely simultaneously generate and face spillovers stemming from different policy changes. Therefore, discussing the scope and effects of such spillovers, as countries navigate out of the crisis, will need to be a key component in the Fund's surveillance in the post-COVID environment.

Against this background, the paper dives into the Fund's coverage of spillovers in its surveillance in the period since the implementation of the ISD. In order to prepare policy recommendations related to spillovers in a post-COVID world, it would be important to first understand past trends of the coverage of spillovers in Article IV staff reports. Our goal is two-fold. First, we investigate whether there are any systematic trends in the coverage of spillovers in Article IV staff reports, considering the increasing global interconnectedness in the past decade (see 2021 CSR background paper on the surveillance priority Preempting and Mitigating Spillovers). Second, we assess whether spillover coverage is strongly correlated with country characteristics, or higher for systemically important countries.

Over the past decade, there has been no shortage of events, induced both by monetary and non-monetary policies of large economies, marking the spillover landscape. A few notable such events, stemming from the monetary policies in systemic advanced economies, include the Federal Reserve (FED)'s monetary easing from 2009, the European Central Bank (ECB)'s first deployment of unconventional monetary policy (UMP) in 2010, the FED's "Operation Twist" in 2011 and 2012, the FED's announcement of a future tapering of its policy of quantitative easing in 2013 and subsequent "taper tantrum", the FED's ending of Quantitative Easing (QE) in 2014 and onset of normalization in interest rates in 2015, and the ECB's expanded asset purchase program (APP) in 2015 and 2016. On the non-monetary

¹ The authors gratefully acknowledge the useful discussions and suggestions by Rupa Duttagupta and Fabian Bornhorst as well as the valuable technical assistance provided by Chengyu Huang.

² This includes Article IV reports, Financial System Stability Assessment reports, regular coverage and special features in WEO chapter 1 and other flagships, annual WEO spillover chapters, Spillover Notes, Regional Economic Outlooks, informal briefings for the Board, the External Sector Report (ESR), and the G-20 SSBG Report.

policy side, Brexit-related risks and global trade tensions that heightened in 2018-19 also dominated spillover concerns in the global platform.

With the use of a state-of-the-art deep learning model, we analyze the discussion of spillovers in all IMF Article IV staff reports during 2010 and 2019. We find that while the discussion of spillovers has overall declined in this period, it spiked during specific years that overlapped with the major monetary policy moves by the Fed and ECB as well as other non-monetary policy events. Furthermore, the coverage of spillovers in the staff reports of systemically important economies such as the US, euro area, and China, has remained prominent. Overall, spillovers are predominantly covered in staff reports of advanced economies (AEs) and more in the context of the real, financial, and external sectors.

Additional econometric analysis confirms these findings, and further shows that staff reports covering countries with high trade openness and lower inflation are more likely to cover spillovers.

The paper is organized as follows. Section two and three provide an overview of the Fund's coverage of spillovers and the literature, respectively. Section four discusses the data and deep learning methodology that we use. Section five outlines the results from the deep learning model and section six provides additional econometric analysis. Section seven concludes.

II. THE COVERAGE OF SPILLOVERS AT THE FUND

Work at the Fund related to spillovers is covered under the Integrated Surveillance Decision (ISD)³. The ISD lays out the legal framework for surveillance. Specifically, it clarifies the importance of focusing on global economic and financial stability in the context of multilateral surveillance to fill the gaps of bilateral surveillance. To this aim, the ISD requires the IMF to cover member countries' outward spillovers to the extent that they may significantly influence the effective operation of the international monetary system. In practice, this means that the IMF's Article IV consultations need to include a discussion of the spillover effects of a member's exchange rate and domestic economic and financial policies that may affect global economic and financial stability.

To further strengthen the ISD implementation, the 2014 Triennial Surveillance Review (TSR) recommends a more systematic analysis of outward spillovers and spillbacks in systemic countries, greater quantification of the impact of spillovers on recipient countries based on global risk scenarios, and discussion of policy implications. The 2015 Guidance Note for Surveillance under Article IV Consultations further underlines the need to examine the most significant actual and potential outward spillovers, irrespective of the transmission channels, which members are obliged to discuss with the Fund and provide relevant data. The 2018 Interim Surveillance Review (ISR, see IMF, 2018) highlights the increase in depth of global spillover coverage in flagship reports but also points out the scope for more prominent coverage of outward spillovers in Article IV consultations. To adequately capture new economic realities, especially the changing sources of spillovers during COVID-19, the 2020

³ IMF, 2012.

Comprehensive Surveillance Review (CSR) proposes new modalities and procedures for conducting bilateral and multilateral surveillance.

III. LITERATURE REVIEW

In economics, a spillover effect refers to the impact that events in one economy can have on other economies. Spillovers can be either positive or negative, arising from a domestic event, such as an earthquake, a stock market crisis, a macro policy change, etc. that spreads to other parts of the world.

The discussion of spillovers began with Alfred Marshall (1890), who famously noted the externalities endogenously generated by co-located firms. The concept was given an analytical content by Pigou (1920), in “The Economics of Welfare”, where he analyzed the divergence between marginal private interest and marginal social interest in the case of a negative externality, and proposed a Pigouvian tax on polluting activities that is equal to the cost of the harm to others. Since then, the concept, termed by Milton Friedman as “neighboring effects” or “spillovers”, has morphed into a variety of operational definitions and analytical expressions throughout the different fields of economics (monetary economics, finance, development, ecological economics, etc.). The theory was further extended and developed by scholars such as Coase (1960), Arrow (1962, 1973), Romer (1986), Porter (1990), and Krugman (1991a, b). Coase (1960), among others, focused on partial equilibrium framework and provided a Coasean solution to the problem of social cost. Romer (1986) assumed knowledge exhibits increasing returns due to spillovers and proposed an endogenous long-run growth model.

As economies across the globe become increasingly connected and more analytical tools become available, researchers zoomed out of the firm-level or regional perspective. Specifically, they started to focus on a wider view of spillovers, at the global scale, as well as assess spillovers stemming from policy decisions in addition to those originating from external shocks. For example, as part of the IMF’s spillover notes series, IMF (2016) examines spillovers from international migration and China’s economic transition. Using empirical analysis and model simulations, the study finds that a one percent increase in the migrant share of the adult population leads to a two percent higher GDP in the long run, and that China’s transition can generate positive long-run effects if managed well. Blagrove et al. (2017) assess fiscal spillovers from five major advanced economies (France, Germany, Japan, United Kingdom, United States) on 55 advanced and emerging market economies. The study finds that fiscal spillovers are stronger in the presence of slack and/or accommodative monetary policy and among countries with fixed exchange rates. Diamond et al. (2018) look at the spillovers from easy liquidity and the implications for multilateralism. They argue that capital-receiving countries may feel indignant at having to assume full responsibility for managing the spillover effects of liquidity expansion in source funding countries, while source countries may see reserve build-up in capital-receiving countries as unfair exchange rate manipulation.

Despite the extensive literature dedicated to spillover analysis, the conceptual definition of spillover effects remains vague within the context of economics research and the measurement of spillovers, although essential to good policy, still poses a challenge. To

complicate matters further, the COVID-19 pandemic has created not only new sources of spillovers, such as those from uncoordinated confinements, but also the emergence of nontraditional channels of transmission, such as travel and supply chains.

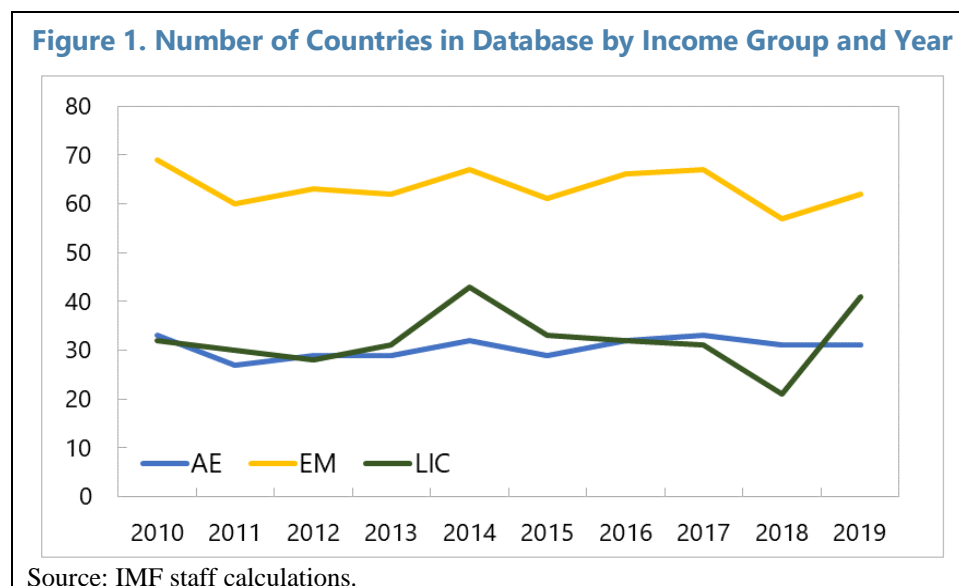
IV. METHODOLOGY AND DATA

A. The BERT Model

Our objective is to gauge how the coverage of cross-border spillovers has evolved over time in IMF Article IV staff reports. To do so, we train a state-of-the-art deep learning model to recognize when a paragraph discusses the topic of spillovers, based not only on a key word search but also on more complicated phrases such as “The ECB’s tightening of monetary policy would have widespread effects across Europe”, for example. Subsequently, we deploy this model on a database containing IMF staff reports over the period 2010 - 2019.

Concurrently, we use the topic model developed in Fayad et al (2020) to assign a sector to those same paragraphs. This allows us to classify a paragraph as covering the external, fiscal, financial, monetary, or real sector. The output comprises a comprehensive dataset containing paragraphs from Article IV staff reports, accompanied by metadata on whether the paragraph discusses cross-border spillovers, its sector, and the country on which the staff report is written, allowing us to identify the level of development, geographic region, and the year in which it was published. We use this rich output to generate summary statistics, visualize trends, and carry out econometric analysis.

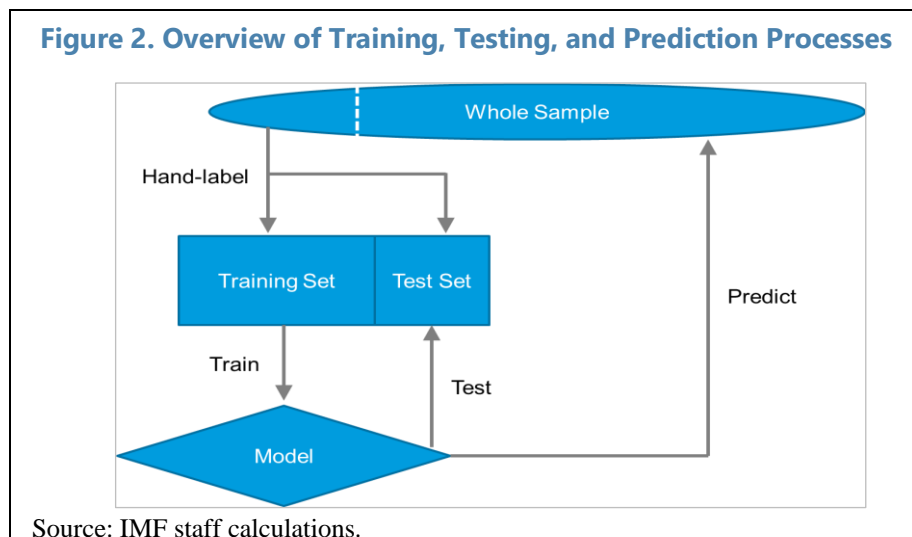
The initial dataset consists of all paragraphs from published IMF staff reports during 2010-2019. The database contains more than 80,000 paragraphs, across 1280 IMF staff reports, for 193 countries/territories. Coverage varies across years due to several factors (see Figure 1). One is that due to confidentiality reasons some staff reports are not published and are therefore not in our database. Second, while most countries are on a 12-month Article IV consultation cycle, some are on a 24-month one. For example, countries that have IMF-supported programs are temporarily moved to a longer, 24-month consultation cycle.



We use Google’s Bidirectional Encoder Representations from Transformers (BERT) model to capture the discussion of spillovers in our database of paragraphs from Article IV staff reports. The BERT model has received wide praise for delivering state-of-the-art results in a wide variety of complex natural language processing (NLP) tasks.

BERT is a supervised deep learning model, which means that it relies on prepared user input to learn a specific task. In our case, we feed the model a set of paragraphs for which we have already (manually) indicated whether they discuss spillovers or not. In order to identify these paragraphs, we read through thousands of paragraphs and identify a subset that discuss spillovers, per the ISD definition, to create our training set. Using this training set, BERT applies its neural network architecture to learn the textual context in which discussions of spillovers generally take place (a process called training). BERT can then use this “knowledge” on the whole sample of paragraphs to identify if a given paragraph mentions spillovers. Due to the complexity of the operation as well as the amount of data, the task cannot be performed on an IMF machine or server. Therefore, we use Amazon Web Services (AWS) to carry out the exercise.

Following the training phase, we also evaluate the accuracy of the model. Specifically, we evaluate the model on a different collection of paragraphs with spillover labels (called the test set), and conclude that it has an accuracy of 90 percent—that is for 9 out of 10 paragraphs, the model assigns the correct label (see Figure 2 for an overview of the entire process). Though in some cases labeling a paragraph as mentioning spillovers might seem straightforward (Iceland experienced severe spillovers following the change in risk appetite), others represent more complex tasks (US monetary policy could have widespread ripple effects across EMs). Similar challenges arise when identifying a paragraph that does not mention cross-border spillovers: some instances are relatively clear-cut (Fiscal policy has relaxed over the past couple of years), while others require a deeper contextual understanding (There were significant spillovers to the banking sector). Therefore, an accuracy rate of 90 percent is an impressive feat and we are confident in the results presented in the next section. For comparison, alternative models that we use for robustness purposes generate accuracy rates between 79 and 88 percent.



B. Comparison to Other Existing Models

BERT is a leading multilayer deep learning model that is currently considered state-of-the-art for natural language processing (NLP). Before the introduction of BERT, logistic regression (LR) was considered one of the simplest and easiest models for textual analysis. Therefore, we use LR as our baseline model, complemented by the support-vector-machine (SVM) model and the random forest (RF) model.

The LR is a probabilistic classifier that relies on supervised machine learning. Its goal is to train a classifier that can make a binary decision about the class of a new input observation, which in our case is to decide whether a paragraph is about spillovers or not. Consider an input paragraph x , which is typically vectorized and represented as $[x_1, x_2, \dots, x_n]$. The classifier output y can be 1 (meaning the paragraph is about spillovers) or 0 otherwise. The LR estimates the probability $P(y = 1|x)$ by learning, from a training set, a vector of weights. Each weight w_i is a real number that represents how important that input feature is to the classification decision. To create a probability, we pass the weighted sum through the sigmoid function, or so called the logistic function. It takes a real-valued number and maps it into the range $[0,1]$. The paragraph is classified as a spillover paragraph if the probability $P(y = 1|x)$ is larger than 0.5.

By contrast, an SVM training algorithm is a non-probabilistic binary linear classifier that assigns input observations to one class or the other. It maps training observations to points in space so as to maximize the width of the gap between the two classes. The observations are then mapped into that same space and predicted to belong to a class based on which side of the gap they fall.

An RF model comprises a set of decision trees each of which is trained using random subsets of features. The prediction by the RF is obtained via majority voting of the predictions of all the trees in the forest, with different values for the features contributing differently to the predictions.

To perform these tests, as a first step we turn the text into a bag of words and transform the words into numeric vectors. We then create a 70/30 train-test split of the dataset and use a cross-validation to obtain an unbiased performance of the models. Next, we apply the LR, SVM, and RF models. The LR classifier uses the weighted combination of the input features and passes them through a sigmoid function, which transforms any real number input to a number between 0 and 1. Finally, we use accuracy scores to assess model performance, where results for the LR, SVM and LR models are 0.88, 0.84 and 0.79, respectively. The model performance of all comparative models is lower than the BERT model.

V. BERT MODEL RESULTS

A. Assessment of the BERT Model Performance.

The BERT model achieves an accuracy of 0.9, which indicates that it correctly identifies 90 percent of the paragraphs in the test dataset as discussing spillovers. Meanwhile, the evaluation loss, which captures the difference between the predicted and actual values,

constitutes 0.29. The model boasts a mean F1-score of 0.78. The F1-score is an often-used performance measure for classification models and combines the precision (ratio of true positives to all predicted positives) and recall (ratio of true positives to all actual positives) statistics. In general, the closer these statistics are to 1 (or 0 in the case of the evaluation loss), the better the performance of the model. However, it is crucial to take them into context of the task at hand. For relatively easy tasks, BERT models can achieve accuracies of 0.99, while more complex objectives often come with much lower precision.

Table 1. BERT Model Performance

Measure	Score
Accuracy	0.9
F1-score	0.78
Evaluation loss	0.29

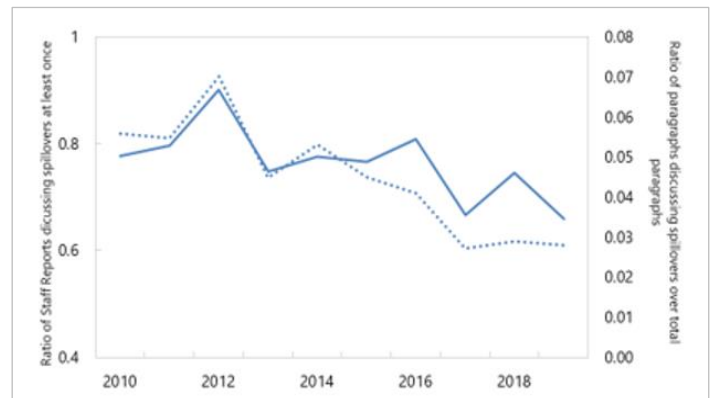
Source: IMF staff calculations.

B. Discussion of the General Spillover Patterns

Our analysis, using the BERT model as described in the previous section, finds that even though the discussion of spillovers in IMF Article IV staff reports seems to be on a declining trend over the last 10 years, there are notable spikes during the periods of global spillover events discussed earlier. This pattern is clearly illustrated by Figure 3. The solid line represents the ratio of staff reports discussing spillovers at least once over the total number of staff reports in the sample. The dashed line represents the average, over the whole sample, of the ratio of number of paragraphs discussing spillovers over the total number of paragraphs in a staff report. Both ratios are included as an illustration as well as for robustness. However, we find the second measure to be more indicative as it measures, on average, the coverage of spillover issues in a given staff report.

At the income group level, based on the ratio of number of paragraphs discussing spillovers over the total number of paragraphs (Figure 4), our analysis shows that spillovers are discussed more in the staff reports covering advanced economies (AEs). Even though these discussions have decreased over time, there are noticeable upticks over the period. Spillovers are discussed least in the staff reports covering low income countries (LICs), and this trend has remained stable. However, if we take a look at the ratio of staff reports discussing spillovers at least once over the total

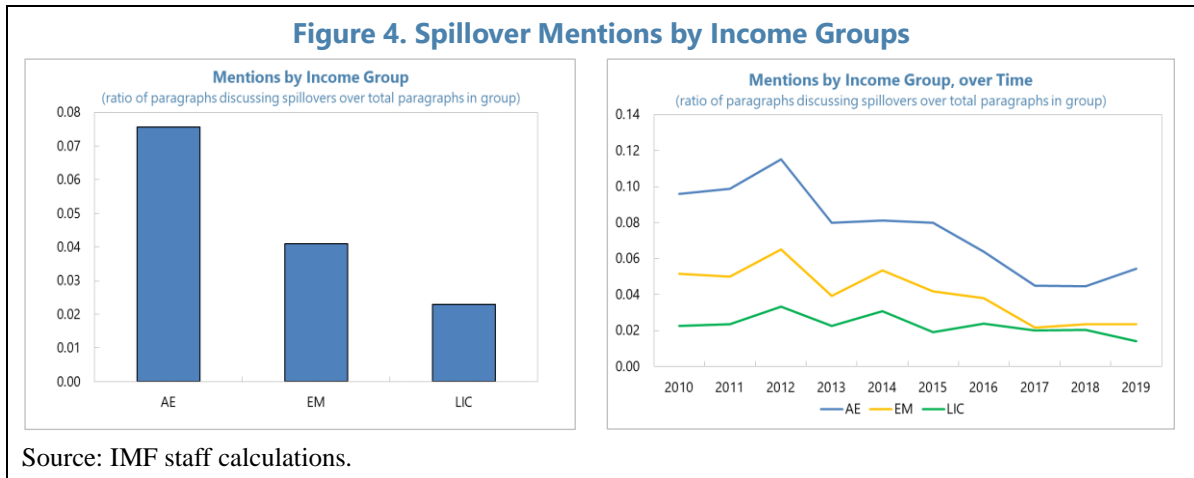
Figure 3. Spillover Mentions Over Time



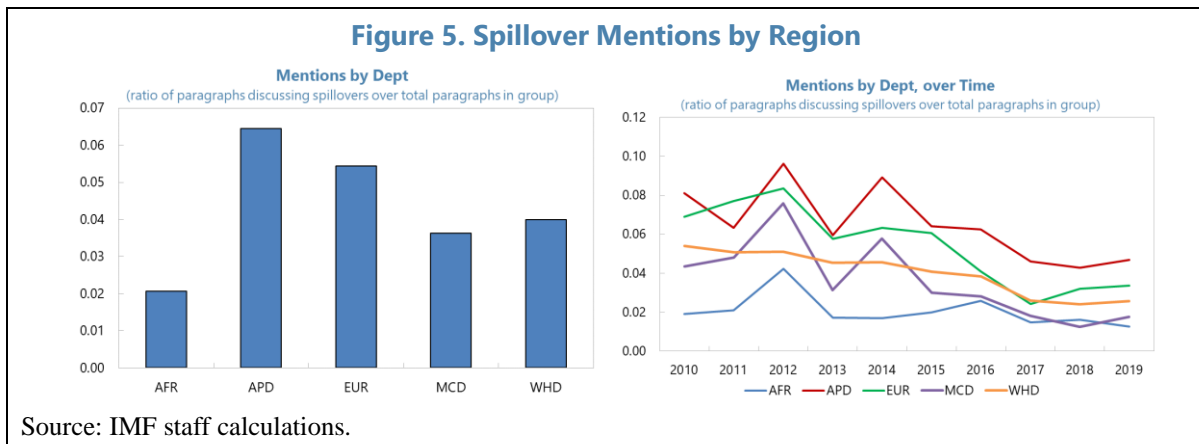
Source: IMF staff calculations.

Note: The solid line (LHS) refers to the ratio of staff reports discussing spillovers at least once over the total number of staff reports (within groups). The dashed line (RHS) refers to the average (within groups) of the ratio of number of paragraphs discussing spillovers over the total number of paragraphs in a staff report. Both are included for robustness.

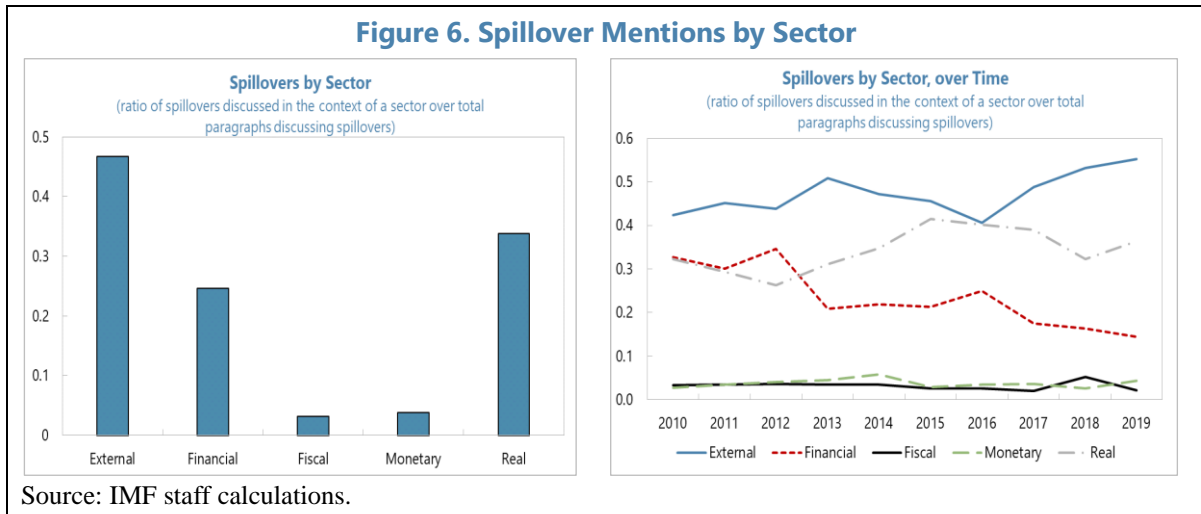
number of staff reports in the group (not shown on Figure 4), we see that in the case of LICs it has increased somewhat over time.



Regionally, based on the ratio of number of paragraphs discussing spillovers over the total number of paragraphs (illustrated in the figure below), spillovers are discussed most in staff reports covering countries in the Asian-Pacific (APD) and European (EUR) regions. At the same time, over the last two years in the sample, the discussion of spillovers has increased slightly in staff reports covering Middle-East and Central Asia (MCD), Western Hemisphere (WHD), and APD countries. If we look at the ratio of staff reports discussing spillovers at least once over the total number of staff reports in the group (not shown in the picture below), we see that it has increased somewhat for WHD, EUR, and African (AFR) countries.



In terms of sectoral coverage, spillovers are mentioned most in relation to the external and real sectors. In addition, the discussion of spillovers has increased over time in the case of both sectors, while in other sectors the coverage has stayed flat or even fallen (e.g., the financial sector).



C. How are Major Spillover Events Covered in Article IV Staff Reports?

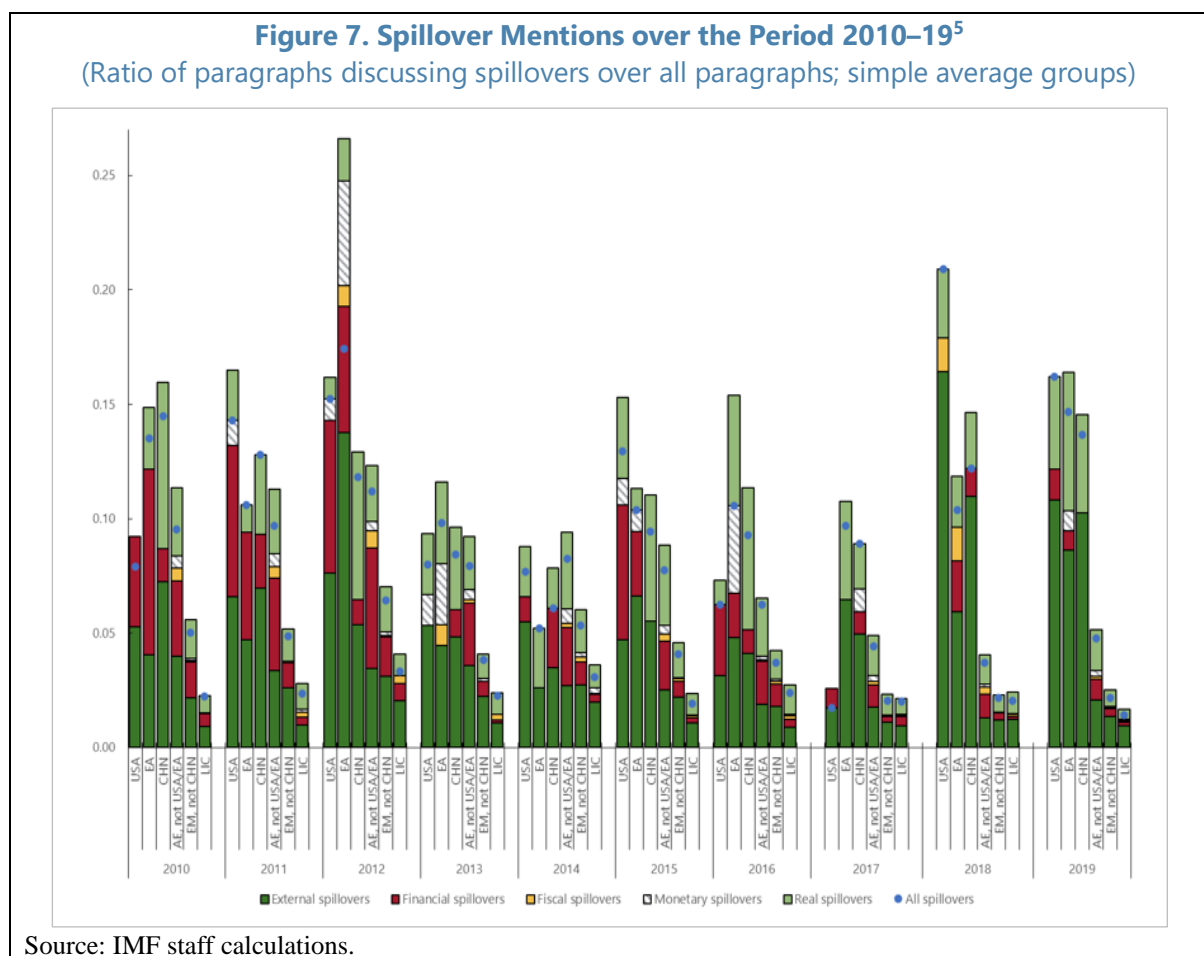
The monetary policy actions of the FED and the ECB over the period 2010–16 induced major spillover events during these years, which have been studied extensively in the literature and IMF publications.⁴ For example, the global economy experienced spillovers stemming from the FED’s efforts, starting in 2009, to stimulate the US economy post Global Financial Crisis (GFC). In 2010 the ECB started to conduct UMP. In 2011 and 2012, the FED implemented “Operation Twist” to help stimulate the economy. The FED exited from its QE in 2014 and started interest rate hikes in 2015. On the other side of the Atlantic, the ECB announced an expanded APP in January of 2015, adding the purchase of government and public sector securities to those earlier programs, with the APP further expanded in April 2016.

Non-monetary policy induced spillovers dominated the discussions on the global platform as well. For example, Brexit related concerns and global trade tensions heightened concerns about spillovers during 2018 and 2019. Indeed, spillovers from global trade tensions were discussed extensively in Chapter 4 of the April 2018 WEO as well as in the 2018 and 2019 Article IV consultation staff reports for China and the United States (US).

We next zoom in on the 2010-2019 time period for the systemic economies (US, euro area, China) and compare spillover discussions in these reports relative to those in other AEs, emerging markets, and LICs. As expected, spillovers are mentioned more often in the Article IV staff reports covering these systemic economies, compared to other countries, with a notable increase during the years of major spillover events. In addition, the staff reports covering AEs contain more mentions to spillovers compared to the ones covering EMs. Staff reports covering LICs contain a much lower share of paragraphs discussing spillovers. In terms of sectors, the discussions are concentrated in the financial and external sectors as well as the real sector.

⁴ Spillover reports, WEO spillover chapter (published annually since 2016), chapter 1 of the WEO, the GFSR, and other products (e.g., ESR, REOs).

These findings further suggest that the discussion of spillovers in the IMF Article IV staff reports of systemic economies has remained elevated, with notable increases during periods of major spillover events. Therefore, the overall decreasing trend that we observe in the discussion of spillovers, is likely driven by the considerable heterogeneity across the countries and regions of the IMF membership. While identifying the factors that determine greater spillover coverage in AE staff reports (and conversely least so in LICs) is outside the scope of our paper, it is reasonable to venture that this is related to the more systemic relevance of AEs (and thereby greater scope for generating spillovers) and the relatively limited integration of LICs to global markets. The coverage of spillovers in emerging market staff reports (excluding China) typically falls between the two groups. A more systematic regression analysis in the next section confirms these priors.



⁵ “All spillovers” do not match the sum of sector spillovers as a paragraph labeled as discussion spillovers can be labeled to belong to more than one sector.

VI. THE PROBABILITY OF DISCUSSING SPILLOVERS IN ARTICLE IV STAFF REPORTS

In this section, we use regression analysis to further explore the probability of discussing spillovers in IMF Article IV staff reports. Specifically, we use the dataset generated by the BERT model to investigate to what extent spillovers are discussed in staff reports and whether they depend on the income level of a country or the sector that is being covered. We also look at the links between specific country-level characteristics, such as trade openness, capital account openness, public debt level, size, growth, inflation, etc. and the discussion of spillovers in staff reports.

To do so, we examine the probability of spillover discussions using the following probit model:

$$Y_{i,t}^* = \beta_1 + \beta_2 X_{i,t} + \beta_3 I_{i,t-1} + \gamma_t + \varepsilon_{i,t} \quad (1)$$

$$Y_{i,t} = \begin{cases} 1, & \text{if } Y_{i,t}^* > 0 \\ 0, & \text{if } Y_{i,t}^* \leq 0 \end{cases}$$

In equation (1), $Y_{i,t}$ is a dummy variable that equals 1 if spillovers are discussed in country i 's Article IV staff report in year t , and 0 otherwise. X is a dummy variable that indicates whether the country is an advanced economy or an emerging economy. The vector I captures an array of country-specific factors that may affect spillover discussions, including the level of GDP and the GDP growth rate. All control variables are lagged by one year to potentially reduce issues with endogeneity. Time fixed effects are also included to control for secular changes. Robust standard errors are reported to cover the possibility that the model's errors may be heteroskedastic. The results are provided in Table 2.

We first investigate whether the income level affects the probability of discussing spillovers in staff reports. To do so, in regression 1, we include two dummy variables. The variable AE takes the value of 1 if country i is an advanced economy and the variable EM takes the value of 1 if country i is an emerging market economy⁶. The control group is thus the one representing low-income economies. We also include the level of GDP, measured by the natural logarithm of a country's GDP in US dollars and the GDP growth rate as control variables. The data for both variables are obtained from the World Economic Outlook (WEO) Database. The results indicate that spillovers are covered more in countries with

⁶ The list of emerging market economies include: Albania, Algeria, Angola, Antigua, Argentina, Armenia, Azerbaijan, Bahamas, Bahrain, Barbados, Belarus, Belize, Bolivia, Bosnia, Botswana, Brazil, Brunei, Bulgaria, Cape, Chile, China, Colombia, Costa Rica, Croatia, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Fiji, Gabon, Georgia, Grenada, Guatemala, Guyana, Hungary, India, Indonesia, Iran, Iraq, Jamaica, Jordan, Kazakhstan, Kosovo, Kuwait, Lebanon, Libya, Macedonia, Malaysia, Maldives, Marshall Islands, Mauritius, Mexico, Micronesia, Mongolia, Montenegro, Morocco, Namibia, Nauru, Oman, Pakistan, Palau, Panama, Paraguay, Peru, Philippines, Poland, Qatar, Romania, Russia, Samoa, Saudi Arabia, Serbia, Seychelles, South Africa, Sri Lanka, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Suriname, Swaziland, Thailand, Tonga, Trinidad, Tunisia, Turkey, Turkmenistan, Tuvalu, Ukraine, United Arab Emirates, Uruguay, Vanuatu.

larger economies. Also, compared to low-income countries, spillovers are more likely to be discussed in the staff reports covering advanced and emerging economies, which lends further credence to our observed trends in the previous section.

We also include other control variables, such as trade openness, capital account openness, public debt level, and inflation rate on spillover discussions in regression 2. Trade openness is measured as the sum of a country's exports and imports as a share of GDP. Capital account openness is calculated based on Chinn-Ito Financial Openness Index. Public debt level is the ratio of general government gross debt over GDP, and the change in the consumer price index is used as a proxy for inflation. The data for capital account openness is obtained from the Chinn-Ito Index website, while the other variables are directly available or can be derived from the WEO database. The results show that trade openness and inflation significantly affect the discussion of spillovers. More specifically, countries with greater trade openness and lower inflation have a higher probability of discussing spillovers in their staff reports.

In un-tabulated results, we explore whether US, EU, and China, which are important spillover producers, are driving the pattern, by adding their respective dummies. However, the results turn out to be statistically insignificant. We also model the relationship between spillover discussions and status of IMF-supported programs (captured with a dummy for such programs, as provided in the IMF Monitoring of Fund Arrangements (MONA) Database). We do not find the program dummy to be statistically significant, which may reflect the main focus of IMF-supported programs on reducing domestic economic imbalances.⁷

We confirm the results of the probit models using OLS models with the same set of variables. The specific coefficients are available in regressions 3 and 4 of Table 2.

Variables	Probit Model 1 Spillover	Probit Model 2 Spillover	OLS Model 3 Spillover	OLS Model 4 Spillover
AE	0.685*** (0.139)	0.106 (0.206)	0.192*** (0.039)	0.058 (0.055)
EM	0.444*** (0.096)	0.338*** (0.122)	0.153*** (0.033)	0.128*** (0.040)
Trade Openness		0.004*** (0.001)		0.001*** (0.000)
CapAcct Openness		0.144		0.047

⁷ We perform additional testing to investigate whether specific spillover years drive the results. To do so, we use 2019 as the base year and create year dummies for the period 2010-2018. The results indicate that spillovers are more heavily discussed during the years 2010-2016, which is consistent with the existence of major global spillover events discussed earlier in the paper. This conjecture is confirmed when we replace the year dummies with two event dummies, one for the period 2010-2012 and the other for 2014-2016. Specifically, we find that the coefficients are significant for both events. Therefore, we opt for using time fixed effects in the analysis.

Table 2. The Probability of Discussing Spillovers (Concluded)

		(0.117)		(0.031)
Debt		0.002		0.000
		(0.002)		(0.000)
Inflation		-0.023**		-0.006*
		(0.009)		(0.003)
Ln GDP	0.120***	0.214***	0.030***	0.049***
	(0.018)	(0.026)	(0.005)	(0.006)
GDP Growth	-0.015**	-0.011	-0.005**	-0.003
	(0.008)	(0.013)	(0.002)	(0.004)
Constant	0.039	-0.567***	0.542***	0.426***
	(0.144)	(0.210)	(0.043)	(0.056)
Obs.	1,272	1,006	1,272	1,006
Year FE	YES	YES	YES	YES
Pseudo R2	0.117	0.170	0.119	0.164
Note: * p<0.1, ** p<0.05, *** p<0.01				
Source: IMF staff calculations.				

As a second step, we investigate the likelihood of discussing spillovers in specific sectors of the staff reports. This is illustrated in Table 3, regressions 1 through 5, where the dependent variables are represented by the real sector, fiscal sector, financial sector, monetary sector, and external sector spillover dummies, respectively.

We find that staff reports covering advanced and emerging economies are more likely to discuss spillovers related to the real, financial, and external sectors. This is consistent with our findings previously in Table 2 as well as in the previous section. The results also point to the sensitivity of spillover discussions to the economies' openness, which is consistent with the prior that greater openness would expose an economy to cross-border shocks. In particular, trade openness is an important determinant of spillover discussions in the real, financial, monetary, and external sectors, whereas greater capital account openness predicts a higher probability of discussing spillovers in the real and fiscal sectors.

Table 3. The Probability of Discussing Spillovers in Sectors

Variables	Probit	Probit	Probit	Probit	Probit
	Model 1	Model 2	Model 3	Model 4	Model 5
	Real	Fiscal	Financial	Monetary	External
AE	0.303* (0.180)	-0.391 (0.251)	0.340* (0.196)	0.054 (0.262)	-0.133 (0.186)
EM	0.277** (0.113)	-0.712*** (0.194)	0.268** (0.130)	-0.323 (0.201)	0.201* (0.115)
Trade Openness	0.002** (0.001)	0.001 (0.001)	0.002** (0.001)	0.003*** (0.001)	0.003*** (0.001)
CapAcct Openness	0.279*** (0.100)	0.685*** (0.172)	0.170 (0.108)	0.087 (0.183)	-0.082 (0.105)
Debt	-0.000 (0.001)	0.003 (0.002)	0.003** (0.001)	0.000 (0.002)	-0.001 (0.001)
Inflation	-0.018** (0.009)	-0.010 (0.015)	-0.021** (0.011)	-0.015 (0.014)	-0.021** (0.009)
Ln GDP	0.094*** (0.023)	0.036 (0.037)	0.161*** (0.026)	0.180*** (0.042)	0.222*** (0.024)
GDP Growth	0.004 (0.011)	-0.021 (0.014)	-0.010 (0.011)	-0.005 (0.015)	-0.019 (0.012)
Constant	-0.831*** (0.184)	-1.906*** (0.265)	-1.276*** (0.201)	-2.392*** (0.314)	-0.741*** (0.187)
Obs.	1,006	1,006	1,006	1,006	1,006
Year FE	YES	YES	YES	YES	YES
Pseudo R2	0.081	0.129	0.162	0.162	0.118

Note: Note: * p<0.1, ** p<0.05, *** p<0.01

Source: IMF staff calculations.

We also run seemingly unrelated regressions (SURs) in order to account for the potential correlation between different sectors. The results confirm our findings obtained through the probit models and are presented in Table 4.

Table 4. The Probability of Discussing Spillovers in Sectors using SUR Models

Variables	SUR	SUR	SUR	SUR	SUR
	Model 1	Model 2	Model 3	Model 4	Model 5
	Real	Fiscal	Financial	Monetary	External
AE	0.114* (0.067)	-0.016 (0.037)	0.133** (0.063)	0.040 (0.037)	-0.011 (0.065)
EM	0.100** (0.042)	-0.073*** (0.023)	0.071* (0.039)	-0.036 (0.023)	0.079** (0.040)
Trade Openness	0.001** (0.000)	0.000 (0.000)	0.001** (0.000)	0.001*** (0.000)	0.001*** (0.000)
CapAcct Openness	0.103*** (0.037)	0.066*** (0.020)	0.052 (0.035)	-0.004 (0.020)	-0.026 (0.036)
Debt	-0.000 (0.000)	0.000* (0.000)	0.001 (0.000)	0.000 (0.000)	-0.001 (0.000)
Inflation	-0.006** (0.003)	-0.000 (0.002)	-0.004 (0.003)	-0.001 (0.002)	-0.005 (0.003)
Ln GDP	0.035*** (0.008)	0.005 (0.005)	0.053*** (0.008)	0.023*** (0.005)	0.074*** (0.008)
GDP Growth	0.001 (0.004)	-0.002 (0.002)	-0.005 (0.004)	-0.001 (0.002)	-0.004 (0.004)
Constant	0.174*** (0.057)	0.025 (0.031)	-0.016 (0.053)	-0.061** (0.031)	0.266*** (0.055)
Obs.	1,006	1,006	1,006	1,006	1,006
R2	0.097	0.053	0.146	0.085	0.121

Note: * p<0.1, ** p<0.05, *** p<0.01

Source: IMF staff calculations.

Last, but not least, we run a few OLS regressions where the dependent variable is represented by the ratio of number of paragraphs discussing spillovers over the total number of paragraphs in a country's staff report (Table 5). The results indicate that the staff reports covering advanced and emerging economies have a significantly higher proportion of paragraphs devoted to spillover discussions (regression 1 in Table 5). At the same time, the staff reports covering countries with an IMF-supported program contain fewer paragraphs that discuss spillovers and the results are significant at the one percent level (regression 2 in Table 5).

Table 5. The Percentage of Spillover Discussion

Variables	OLS	OLS	OLS
	Model 1	Model 2	Model 3
	Spillover Percentage	Spillover Percentage	Spillover Percentage
AE	0.038*** (0.005)		0.016*** (0.006)
EM	0.013*** (0.002)		0.006*** (0.003)
IMF- supported Program		-0.012*** (0.003)	
Trade Openness			0.000*** (0.000)
CapAcct Openness			-0.001 (0.003)
Debt			-0.000* (0.000)
Inflation			-0.001*** (0.000)
Ln GDP	0.004*** (0.001)	0.007*** (0.001)	0.008*** (0.001)
GDP Growth	-0.000 (0.000)	-0.001*** (0.000)	-0.001* (0.000)
Constant	0.024*** (0.004)	0.033*** (0.004)	0.005 (0.007)
Obs.	1,272	1,272	1,006
Year FE	YES	YES	YES
R2	0.238	0.192	0.321

Note: * p<0.1, ** p<0.05, *** p<0.01

Source: IMF staff calculations.

VII. CONCLUSION

We use a state-of-the-art deep learning model to analyze the discussion of spillovers in all IMF Article IV staff reports between 2010 and 2019. We find that overall, even though the discussion of spillovers seems to have decreased over time, it spikes during periods of major global spillover events and remains elevated in the staff reports of systemically important countries. In addition, spillovers are discussed most in staff reports covering advanced and emerging economies and predominantly in the context of the real and external sectors. These trends are confirmed when we zoom in on specific global spillover events.

Additional econometric analysis, using a probit model as well as OLS regressions, confirms these findings. It further shows that staff reports covering countries that are relatively larger in size are more likely to discuss spillovers. At the same time, countries with lower growth, higher trade and capital account openness, and lower inflation are more likely to have spillover discussions in their Article IV staff reports.

Even though the discussion of spillovers in IMF surveillance staff reports seems to have declined over time, it picked up during years with important spillover events and remained prominent for systemically important economies. The COVID pandemic is likely to spur a new period of spillover events. If historical trends continue, the coverage of spillovers will remain high in Staff Reports of the larger countries. Improving and strengthening the coordination between IMF country teams would be an important step in the work on spillovers to ensure that spillover discussions remain prominent not only in source countries, but also those on the receiving side of cross-border spillovers.

Technical Annex

This annex seeks to provide additional details regarding the inner workings of the BERT model, as well as provide more insight into the topic model used to assign sectors to paragraphs.

BERT Model-Specifics

When BERT was first introduced by Google’s Jacob Devlin in 2018, it made quite the splash in the machine learning community for its state-of-the-art performance in NLP tasks. The model combines two key innovations in NLP-based deep learning, namely *transfer learning*, that is to train a model in one context, and then leverage the acquired knowledge to improve the model’s performance in another context, and the *Transformers model*, an attention model that learns contextual relations between words in a text.

In this annex, we will elaborate on both mechanisms, starting with the latter. In its simplest form, Transformer contains two separate elements — an encoder that reads the text input and a decoder that produces a prediction for the task. This structure is supplemented by an attention mechanism, which assigns weights to parts of the text input that are most relevant to the task at hand (that is, it gives *attention* to certain parts of the text). As opposed to directional models, which read the text input sequentially (left-to-right or right-to-left), the Transformer encoder reads the entire sequence of words at once. Therefore, it is considered *bidirectional*. This characteristic allows the model to gauge the context of a word based on all of its surroundings (left and right of the word).

Another key advantage of BERT is that it is already pre-trained on a large corpus of unlabeled text, including all of Wikipedia (2500 million words) and Book Corpus (800 million words). This training is *unsupervised*, and consists of two separate tasks: Masked Language Modeling (MLM), in which it randomly *masks* 15 percent of the words and attempts to predict what those words could be based on the context; and Next Sentence Prediction (NSP), where it seeks to determine whether a random pair of sentences is next to each other in the original document. While MLM allows the model to gain an understanding of the words and context on a sentence-level, NSP teaches BERT to grasp long-term dependencies throughout the text.

The pre-trained model, called BERT Base, thus already comes with a strong understanding of contextualized word embeddings. As mentioned earlier, to use the BERT model for our specific objective, i.e. predict whether a paragraph in a given Article IV discusses spillovers or not, we need to *finetune* the model using actual language from staff reports. We prepare a training dataset with 1356 paragraphs from Article IVs, of which we manually label 379 paragraphs (28 percent) as containing a discussion of spillovers by simply assigning a 1 (spillovers) or a 0 (no spillovers). In addition, we generate a test dataset which includes 717 paragraphs taken from staff reports, of which roughly 27 percent mention spillovers. We use the training dataset to finetune the already pre-trained BERT Base model, and the test dataset to calculate its accuracy.

To further optimize the performance of the model, we tune its hyperparameters, i.e. we tweak different settings of the model to achieve the best result. For the number of epochs, i.e. the amount of times the entire training set is fed through the model, we choose 3 as it generates the lowest loss. We set the batch size at 16, which means the training set moves through the network in 16 subsequent samples. The original BERT paper by Devlin et al. recommends a batch size of 16, and increases in batch size comes with necessary upgrades in computing power. The learning rate dictates the pace with which the model *learns* and is set using the *Adam optimizer*. The Adam optimizer is an adaptive learning rate method (as opposed to a constant), which computes different learning rates for each weight of the neural network. The last task is to deploy the fully trained model to predict for all paragraphs in our entire sample if they mention spillovers. Of the roughly 80,000 paragraphs in our dataset, the model determines that 3495 (4.4 percent) are on the subject of spillovers.

Topic Model

In addition to the above, further text analysis allows us to gauge the context in which spillovers are mentioned in staff reports. Following the methodology introduced by Fayad et al. (2020), we assign macroeconomic sectors to each paragraph in a staff report, using a vector of terms associated with those topics. The sectors are: external, financial, fiscal, monetary, and real. In line with Fayad et al., we use the IMF’s Enterprise Business Vocabulary (EBV), which contains the aforementioned sectors and a number of terms related to those sectors. Though the structure of the EBV allows us to easily assign macroeconomic terms to each sector, the language used is often abstract and the exact words do not generally appear in staff reports. To be sure, the BERT model, or a similar type of machine learning model, could also be used to assign these sectors—though Fayad et al. find that their dictionary approach achieves a similar if not better accuracy.

Due to this, we enter the terms associated to each sector in a Word2Vec model (Mikolov et al., 2013) trained on IMF Staff Reports. Word2vec generates a vector space, typically of several hundred dimensions, with each unique word in our corpus of Article IVs being assigned a corresponding vector in the space. Word vectors are positioned in the vector space such that words that share common contexts in the corpus are located close to one another in the space. We extract the terms most similar to our five sectors. Lastly, we match these terms with the contents of individual paragraphs and determine to which sector the paragraph corresponds to most. Fayad et al. find that this method achieves an 88 percent accuracy rate. However, it should be stressed that the potential for errors remains and we therefore interpret the results with caution.

Looking at our sample, 37 percent of paragraphs discuss the real sector, followed by fiscal (28 percent), financial (22 percent), external (16 percent), and monetary (6 percent).⁸ Specifically looking at paragraphs that discuss spillovers, we find that the majority of mentions of spillovers occur in the context of the external sector (45 percent). This finding is especially significant given that only a sixth of all paragraphs discusses external sector issues. Furthermore, 34 percent of the spillover paragraphs take place in the real sector,

⁸ Note that the sum of these shares surpasses a 100%, as we allow double assigning whereby a paragraph can discuss two topics equally.

followed by 25 percent for the financial sector. Only a small fraction of spillover paragraphs discusses the monetary sector (6 percent), though it is important to keep in mind that a mere 2 percent of the full sample considers monetary issues.

References

- Arrow, K.J., 1962. Economic Welfare and the Allocation of Resources to Invention, *The Rate and Direction of Inventive Activity: Economic and Social Factors*, edited by the Universities-National Bureau Committee for Economic Research and the Committee on Economic Growth of the Social Science Research Councils, 609– 626.
- Arrow, K.J., 1973. Social Responsibility and Economic Efficiency. *Public Policy*, 21, 303-317.
- Blagrove, P., Ho, G., Koloskova, K. and Vesperoni, E., 2017. Fiscal Spillovers : The Importance of Macroeconomic and Policy Conditions in Transmission. *Spillover Notes*, 11, IMF.
- Coase, R.H., 1960. The Problem of Social Cost. *Journal of Law and Economics*, 3(1), 1-44.
- Devlin, J., Chang, M., Lee, K., and Toutanova, K., 2019. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. In North American Association for Computational Linguistics (NAACL).
- Diamond, D.W., Hu, Y. and Rajan, R.G., 2020. The Spillovers from Easy Liquidity and the Implications for Multilateralism. *IMF Economic Review*, 68, 4–34.
- Fayad, G., Huang, C., Shibuya, Y. and Zhao, P., 2019, “How Do Member Countries Receive IMF Policy Advice: Results from a State-of-the-art Sentiment Index”, IMF Working Paper 20/7.
- Krugman, P., 1991a. *Geography and Trade*. Cambridge: MIT Press.
- Krugman, P., 1991b. Increasing Returns and Economic Geography. *Journal of Political Economy*, 99(3), 483-499.
- IMF, 2012. *Modernizing the Legal Framework for Surveillance - An Integrated Surveillance Decision*.
- IMF, 2014. *Triennial Surveillance Review*.
- IMF, 2015. *Guidance Note for Surveillance under Article IV Consultation*.
- IMF, 2016. Spillovers from Migration and China's Transition. *World Economic Outlook*, October.
- IMF, 2018. *Interim Surveillance Review*.
- Marshall, A., 1890. *Principles of Economics*. MacMillan & Co., London and New York.

- Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. *Efficient Estimation of Word Representations in Vector Space*. arXiv: 1301:3781.
- Pigou, A.C., 1920. *The Economic of Welfare*. Macmillan & Co., London.
- Porter, M.E., 1990. The Competitive Advantage of Nations. *Harvard Business Review*, 68(2), 73–93.
- Romer, P.M., 1986. Increasing Returns and Long-Run Growth. *Journal of Political Economy*, 94(5), 1002-1037.