



ESSAYS ON THE USE OF MACHINE  
LEARNING IN ECONOMICS:  
THE CASE OF BREXIT

By

DUIYI DAI

A thesis submitted to  
the University of Birmingham  
for the degree of  
DOCTOR OF PHILOSOPHY

UNIVERSITY OF  
BIRMINGHAM

**University of Birmingham Research Archive**

**e-theses repository**

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

## ABSTRACT

This thesis comprises three papers on Brexit, focusing on uncertainty and non-verbal media bias, utilizing cutting-edge machine learning techniques. The first paper constructs a set of indices to measure Brexit-related uncertainty by analyzing newspaper coverage from 2013 to 2022. Employing machine learning techniques, these indices provide a real-time view of uncertainty trends, including topic-specific indices for key areas such as trade, immigration, and employment. The paper also distinguishes between Brexit-related uncertainty and the impact of the COVID-19 pandemic. The second paper examines visual media bias during the 2016 Brexit referendum by analyzing how UK newspapers depicted politicians in images. Using machine learning and computer vision tools, the analysis uncovers clear partisan bias, particularly in tabloid newspapers, which tended to portray politicians aligned with their stance more positively and placed them in more favorable contexts. This bias was especially pronounced in front-page images and among a few key politicians but diminished after the referendum. The third paper explores the economic consequences of Brexit uncertainty using a proxy structural vector autoregressive (proxy-SVAR) model. To identify Brexit uncertainty shocks, we construct an external instrument based on a high-frequency Brexit uncertainty index developed in the first paper. Our results indicate a short-term boost in economic activity following the uncertainty shocks, driven by a temporary stimulus to the tradable sector. Collectively, these papers enhance our understanding of Brexit-related uncertainty, its economic consequences, and visual media slant during the referendum, providing valuable insights for policymaking efforts and media analysis.

## ACKNOWLEDGMENTS

I would like to express my sincerest gratitude to my supervisors, Prof. Robert Elliott and Dr. Wanyu Chung, for their invaluable guidance, unwavering support, and constant encouragement throughout my PhD journey. I am deeply appreciative of the countless hours they have invested in mentoring me, providing constructive criticism, and facilitating access to the necessary resources that enabled me to pursue my research objectives effectively.

I am profoundly thankful to my family for their unconditional love, patience, and unwavering belief in me. To my parents, who have always encouraged me to pursue my dreams and provided a stable foundation, thank you for your endless support and sacrifices. My siblings and extended family have also played a crucial role in maintaining my morale and motivation throughout this journey, providing laughter, advice, and a sense of normalcy amidst the rigors of academic life.

Lastly, I would like to thank everyone who has helped me along the way, including those who offered words of wisdom, assisted with practical aspects of my research, or simply provided a listening ear. Your contributions, no matter how small, have collectively played a significant role in the successful completion of this PhD. To all those who have been part of this journey, thank you for your support, guidance, and belief in me.

# Contents

	Page
<b>Introduction</b>	<b>1</b>
<b>1 Machine Learning for Texts with Brexit Uncertainty Application</b>	<b>7</b>
1.1 Introduction	9
1.2 Related Literature	16
1.2.1 Research on Text-based Uncertainty Measurement	16
1.2.2 Brexit Uncertainty Measurements	19
1.3 Background on Brexit	20
1.3.1 Brexit Uncertainty	21
1.4 Measuring Brexit Uncertainty	25
1.4.1 News Data and Pre-processing	25
1.4.2 Measuring Aggregate BUI: Expanding Search Words with the Word2Vec Model	28
1.4.3 Measuring Topic-level BUIs: An Application of a LDA Model	31
1.5 Brexit Uncertainty Index	35
1.5.1 Aggregate BUI	35
1.5.2 Aggregate BUI Excluding the Effects of COVID	36
1.5.3 Topic Specific BUIs	38
1.6 Validation Exercises	44
1.6.1 Comparison with the BBD-EPU Index	44

1.6.2	A New EPU Index Using a ML Approach . . . . .	45
1.6.3	A New BUI Using the X (Twitter) API . . . . .	46
1.6.4	Political Leanings in Newspaper Coverage of Brexit Uncertainty . . . . .	47
1.6.5	BUI and Stock Market Volatility . . . . .	48
1.7	Application: Public Sentiment and BUIs . . . . .	50
1.8	Conclusions . . . . .	55
<b>A Appendix for Chapter 1</b>		<b>58</b>
A.1	Word sets . . . . .	58
A.2	Why We Choose the LDA Model for Topic Modeling . . . . .	59
A.3	LDA: Model and estimation . . . . .	59
A.3.1	Model . . . . .	60
A.3.2	Estimation . . . . .	61
A.4	Daily Brexit Uncertainty Index . . . . .	62
A.5	Correlation of Topics in Articles . . . . .	63
A.6	Example Tweets of Brexit Topics and their Sentiment Scores . . . . .	64
<b>2 A Picture is Worth a Thousand Words: An Application of Machine Learning and Computer Vision Image Processing Applied to the Brexit Referendum</b>		<b>69</b>
2.1	Introduction . . . . .	71
2.2	Background . . . . .	75
2.2.1	UK Newspapers: Broadsheets and Tabloids . . . . .	77
2.3	Related Literature . . . . .	79
2.3.1	Image as Data with Computer Vision Techniques . . . . .	80
2.3.2	Visual Media Bias . . . . .	83
2.3.3	Media Bias in the UK . . . . .	85
2.4	Data and Methodology . . . . .	86

2.4.1	News Data . . . . .	86
2.4.2	Measuring Facial Expression Emotions . . . . .	87
2.4.3	Politician Sample . . . . .	90
2.4.4	Newspaper Partisanship on Brexit Vote . . . . .	91
2.5	Visual Media Bias in Politician Facial Expression Emotions . . . . .	92
2.5.1	Empirical Evidence . . . . .	95
2.5.2	Robustness . . . . .	102
2.5.3	Trends in Visual Media Bias during the Brexit Vote . . . . .	102
2.6	Bias in Contextual Images . . . . .	105
2.6.1	Measuring Contextual Image Sentiment . . . . .	105
2.6.2	Empirical Evidence . . . . .	109
2.7	Conclusion . . . . .	111
<b>B Appendix for Chapter 2</b>		<b>113</b>
<b>3 Modelling the Effects of Brexit Uncertainty on the UK Economy</b>		<b>136</b>
3.1	Introduction . . . . .	138
3.2	Literature Review . . . . .	142
3.2.1	Brexit Effects . . . . .	142
3.2.2	Brexit Uncertainty Effects . . . . .	143
3.2.3	VAR Models and Uncertainty . . . . .	145
3.3	Empirical Framework . . . . .	147
3.4	Data . . . . .	150
3.4.1	Brexit Uncertainty Measure . . . . .	150
3.4.2	Daily Data for Instrument Construction . . . . .	151
3.4.3	Data for Monthly VAR Estimation . . . . .	152
3.5	Instrument Construction . . . . .	154
3.5.1	Evaluation of the Instrument . . . . .	156

3.6	Results . . . . .	158
3.6.1	Baseline Results . . . . .	159
3.6.2	Additional Results . . . . .	161
3.6.3	Alternative Identification Strategies . . . . .	162
3.7	Sensitivity Analysis: Alternative Instrument . . . . .	170
3.8	Conclusion . . . . .	171
<b>C</b>	<b>Appendix for Chapter 3</b>	<b>174</b>
C.1	Identification of the Proxy-SVAR . . . . .	174
C.2	Recursive Identification . . . . .	177
C.3	Sectoral GDP . . . . .	180
C.4	Proxy SVAR using Monthly Index . . . . .	182
C.4.1	Evaluation of the Instrument . . . . .	183
	<b>Concluding Remarks</b>	<b>186</b>

# List of Figures

1.1	A Flowchart to Illustrate How We Construct Our Measures of Aggregate and Topic-Specific BUIs Using the Two ML Models, Word2Vec and LDA . . . . .	26
1.2	Example Outputs of the LDA Model . . . . .	34
1.3	BUI, Lower Bound BUI (COVID Excluded), and Brexit Uncertainty Series from the DMP Survey . . . . .	37
1.4	Topic Specific BUIs . . . . .	42
1.5	Our BUI and the BBD-EPU Index . . . . .	45
1.6	Our EPU Index and the BBD-EPU Index . . . . .	46
1.7	Our BUIs and the X (Twitter) BUI . . . . .	47
1.8	Aggregate BUI, Currency BUI, and FTSE 100 Volatility Index (VFTSE) . . . . .	49
1.9	Brexit Sentiment and Frequency of Brexit-Related Tweets . . . . .	53
A.1	Daily BUI and Lower Bound BUI (COVID Excluded). . . . .	63
2.1	Example Outputs of the Amazon Rekognition API . . . . .	89
2.2	Distribution of Emotion Scores: Pre-Referendum Period . . . . .	90
2.3	Newspaper Partisanship from Levy et al. (2016) . . . . .	92
2.4	Visual Media Bias During the Brexit Vote . . . . .	94
2.5	Dynamics of Visual Media Bias During the Brexit Vote . . . . .	104
2.6	Examples of Positive and Negative Contextual Images . . . . .	108
B.1	Bias regarding How Politicians are Covered (Tabloids) . . . . .	115
B.2	Correlation between Emotions for Amazon API . . . . .	116

B.3	Visual Media Bias During the 2016 Brexit Vote (Broadsheets)	117
B.4	Position of the Brexit Referendum Articles Over Time from Levy et al. (2016)	118
B.5	Dynamics of Visual Media Bias During the Brexit Vote - Pro-Leave Politicians	119
B.6	Dynamics of Visual Media Bias During the Brexit Vote - Pro-Remain Politicians	120
3.1	BUI and Lower Bound BUI (COVID Excluded)	151
3.2	Impulse Responses to Brexit Uncertainty Shocks: Baseline Results	160
3.3	Impulse Responses to Brexit Uncertainty Shocks: Additional Results	163
3.4	Impulse Responses to Brexit Uncertainty Shocks: Recursive Identification Scheme	165
3.5	Impulse Responses to Brexit Uncertainty Shocks: Proxy-SVAR with Low Frequency Instrument	166
3.6	Impulse Responses to Brexit Uncertainty Shocks: Bayesian VAR	168
3.7	Impulse Responses to Brexit Uncertainty Shocks: VAR with Sign Restrictions	169
3.8	Impulse Responses to Brexit Uncertainty Shocks: Using DMP Series to Construct Instrument	172
C.1	Impulse Responses to Brexit Uncertainty Shocks: Sectoral GDP	181

# List of Tables

1.1	Brexit Timeline . . . . .	22
1.2	Brexit Uncertainty by Topic and Time Period, 2013-2022 . . . . .	43
1.3	Brexit Uncertainty and Brexit Sentiment . . . . .	54
A.1	Word Sets . . . . .	58
A.2	Correlation of Topics in Articles . . . . .	68
2.1	Brexit Vote Main Events . . . . .	78
2.2	Comparisons between Text and Image Data . . . . .	80
2.3	Newspaper Partisanship and Politician Emotions . . . . .	97
2.4	Newspaper Partisanship and Politician Emotions: Major Politicians . . . . .	99
2.5	Newspaper Partisanship and Politician Emotions: Excluding Major Politicians . . . . .	101
2.6	Newspaper Partisanship and Politician Contextual Image Sentiment . . . . .	110
B.1	Politicians and Their Roles During the Referendum Period (21 February – 31 August 2016) . . . . .	121
B.2	Number of Politician Faces . . . . .	122
B.3	Average Emotion by Politician, Pre-Referendum, Tabloids . . . . .	123
B.4	Average Emotion by Politician, Pre-Referendum, Broadsheets . . . . .	124
B.5	Type, Partisanship and Counts by Newspapers: Whole Period . . . . .	125
B.6	Newspaper Partisanship and Politician Emotions . . . . .	126
B.7	Newspaper Partisanship and Politician Emotions: Tabloids . . . . .	127

B.8 Newspaper Partisanship and Politician Emotions: Excluding Surprised Emotion, Tabloids, Pre-Referendum . . . . .	128
B.9 Newspaper Partisanship and Politician Emotions: Excluding Each Emotion Individually, Tabloids, Pre-Referendum . . . . .	129
B.10 Newspaper Stance (0 for Pro-Remain and 1 for Pro-Leave) and Politician Emotions: Tabloids . . . . .	130
B.11 Newspaper Partisanship and Politician Emotions: Tabloids, April 15 - Referendum date . . . . .	131
B.12 Newspaper Partisanship and Politician Emotions: Broadsheets . . . . .	132
B.13 Newspaper Partisanship and Politician Emotions: All Newspapers . . . . .	133
B.14 Newspaper Partisanship and Politician Emotions by Politicians . . . . .	134
B.15 Newspaper Partisanship and Politician Contextual Image Sentiment: Broadsheets . . . . .	135
3.1 Variables in Monthly VAR . . . . .	154
3.2 Validating the Exogenous Condition of the Instrument . . . . .	158
3.3 Forecast Error Variance Decomposition . . . . .	161
C.1 Constructing Instrument . . . . .	184
C.2 Validating the Exogenous Condition of the Instrument . . . . .	185

# Introduction

Brexit stands as one of the most consequential political and economic events in recent British history, exerting far-reaching effects on the UK (Dhingra et al. 2016, Broadbent et al. 2023), the European Union (Felbermayr et al. 2017), and global markets (Hassan et al. 2024, Graziano et al. 2020). Growing Euroscepticism, along with concerns over national sovereignty, economic policies, and immigration, pressured the Conservative government to hold a referendum on EU membership (Goodwin & Heath 2016). On June 23, 2016, the results of the referendum showed a majority of voters in favor of leaving the EU (51.9% voted to leave versus 48.1% voted to remain). This decision presented the UK with substantial challenges, including negotiating new trade agreements with the EU and other countries, establishing an independent immigration system separate from EU free movement laws, and developing new regulatory frameworks in sectors such as fisheries and environmental standards. Empirical studies have demonstrated that Brexit has reduced goods exports (Du et al. 2023), posed difficulties for small manufacturing firms (Bailey et al. 2023), increased UK investments in the EU as firms set up European subsidiaries to access EU markets, decreased EU investments in the UK, and lowered net migration from EU countries (Bailey et al. 2023, Di Iasio & Wahba 2023).

This thesis has two main goals: First, to better understand the unstructured data associated with Brexit using machine learning techniques; and second, to explore how the

application of machine learning can help us understand the impact of Brexit on the real economy. The thesis is organized into two parts. The first part (Chapters 1 and 2) aims to measure Brexit-induced uncertainty and visual media bias during the referendum using unstructured text and image data. Unstructured data makes up about 95% of all big data generated today, including text and images (Gandomi & Haider 2015). Unlike structured data, such as numerical information that fits neatly into tables and databases, unstructured data lack the structural organization required for traditional machine analysis, making them challenging to analyze using conventional methods in economics and social science research. We overcame this challenge by using textual analysis and computer vision techniques from machine learning.

We intentionally focus on the two aspects, uncertainty and visual media bias. Uncertainty surged immediately after the Brexit referendum, significantly affecting various aspects of the UK economy (Bloom et al. 2019). While uncertainty has been widely recognized as a critical factor influencing investment, consumption, and policy formulation (Bloom 2009, Jurado et al. 2015), continuous and precise measures of Brexit-induced uncertainty, especially its topical components, have been little researched. This thesis provides timely and cost-effective indicators of uncertainty at both the aggregate and topic levels using continuous text data.

More recently, academics have become interested in the use of visual content as a means of imparting information. Visual content has become increasingly prevalent in the current digital media landscape and is thought to play an important role in political communication (Verser & Wicks 2006, Grabe & Bucy 2009). Although the majority of the existing research measures and analyzes media bias through text (e.g., Groseclose & Milyo 2005, Gentzkow & Shapiro 2010, Groeling 2013, Martin & Yurukoglu 2017), it is worth noting that images are often considered more attention-grabbing and memorable than words (Sul-

livan & Masters 1988, Graber 1990, 1996). This thesis addresses this gap by uncovering the bias embedded in news images using machine learning and computer vision techniques. A deeper understanding of the sources of media bias is essential for exploring the impact of the media on public behaviour.

The second part of the thesis examines how applying machine learning techniques to uncertainty measurement can enhance our understanding of the real economic impacts of Brexit (Chapter 3). Brexit-induced uncertainty has been documented to disrupt trade (Graziano et al. 2021, Crowley et al. 2018), immigration (Wadsworth et al. 2016), employment (Javorcik et al. 2020), and broader economic conditions (Steinberg 2019, Faccini & Palombo 2021). Understanding these effects is crucial for policymakers and businesses as they navigate the post-Brexit economy. Building on the uncertainty measures developed in Chapter 1, the second part of thesis introduces a continuous and daily Brexit uncertainty index that covers both the pre- and post-referendum periods to examine the macroeconomic implications of this uncertainty. The chapter provides novel insights into how uncertainty shapes key economic outcomes such as trade, investment, and employment. The subsequent paragraphs provide a concise overview of each chapter.

Chapter 1 introduces a novel set of Brexit uncertainty indices (BUIs) using UK newspaper coverage from 2013 to 2022, measuring both aggregate uncertainty and specific topic-level uncertainties, such as trade, immigration, and Northern Ireland. Machine learning algorithms are employed to identify Brexit-related articles and topics that reflect different dimensions of uncertainty. We further disentangle Brexit-related uncertainty from the uncertainty due to COVID-19, which may have masked or exaggerated existing measures based on survey evidence.

Our results suggest that the main spikes in Brexit related uncertainty occurred around

the time of the Brexit referendum, the three failed meaningful votes, when the Brexit withdrawal agreements were rejected in the House of Commons in early 2019, and the period just before the final agreement was reached. In terms of average values, Brexit uncertainty was relatively high between the announcement of the referendum date (February 2016) and the UK's formal departure from the EU (December 2019). Even after the Brexit deal in 2021, uncertainty levels did not significantly decline. The COVID pandemic further amplified Brexit uncertainty by up to 50%, particularly affecting employment, government spending and budgets, and supply chains, while sectors like fishing rights and Northern Ireland were less impacted.

Chapter 2 examines visual media bias in UK online news articles during the 2016 Brexit referendum using computer vision and machine learning techniques. We analyze images published by major UK media outlets to explore whether politicians were portrayed differently based on the political leanings of the newspapers. Specifically, we compile a dataset of over 64,000 images from seven UK newspaper websites, for around six months surrounding the referendum. Computer vision methods are used to detect and recognize faces, assess facial expressions, and analyze the sentiment of the broader visual context in images.

Our results highlight significant partisan bias, especially among the tabloid press. Newspapers tend to depict politicians who align with their stance more favorably, in terms of facial expressions and contextual settings, while portraying opposing politicians more negatively. This bias was particularly noticeable in front-page images and in the portrayal of key political figures during the campaign. An exception to this was Boris Johnson, who was depicted with consistent negative emotions in both pro-Leave and pro-Remain newspapers.

As a follow-up to Chapter 1, Chapter 3 examines the macroeconomic implications of

Brexit-induced uncertainty on the UK economy using a proxy structural vector autoregressive (proxy-SVAR) model. The period from 2016 to 2020 was marked by significant uncertainty following the UK’s decision to leave the European Union ([Bloom et al. 2019](#)). In this chapter, we quantify the dynamic causal effects of uncertainty shocks on key economic indicators such as GDP, trade openness, and hours worked from 2016 to 2022.

The chapter begins by using the BUI (developed in Chapter 1) to construct an external instrument for recovering the uncertainty shocks. Our methodological approach involves a two-step process. First, we estimate a high-frequency daily VAR model that includes the daily BUI and other economic variables to recover the daily Brexit uncertainty shock series. Recognizing that financial markets and economic agents can respond rapidly to uncertainty shocks, this high-frequency estimation captures immediate reactions that might be missed in lower-frequency analyses ([Alessandri et al. 2023](#)). Second, we aggregate the daily shock series to a monthly frequency and use it as the external instrument in the proxy-SVAR model.

Our key findings reveal that Brexit-induced uncertainty shocks had a short-term positive effect on the UK economy, such as a temporary boost to GDP, hours worked, and trade openness in goods towards the EU. This result aligns with the findings of [Broadbent et al. \(2023\)](#), who suggest that the Brexit vote temporarily stimulated the tradable sector as firms accelerated production and exports in anticipation of future productivity reductions.

The remainder of this doctoral thesis is organized as follows. The first paper titled “Machine Learning for Texts with Brexit Uncertainty Application” is presented in Chapter 1. The second paper on “A Picture is Worth a Thousand Words: An Application of Machine Learning and Computer Vision Image Processing Applied to the Brexit Referendum” is presented in Chapter 2, followed by the third paper, “Modelling the Effects of Brexit Uncertainty on the UK Economy” in Chapter 3. The thesis concludes with final remarks,

including limitations and potential pathways for future research.

# Chapter One

## Machine Learning for Texts with Brexit Uncertainty Application

A version of this chapter is publicly available as: Chung, W., Dai, D. & Elliott, R. (2022), *‘Measuring Brexit Uncertainty: A Machine Learning and Textual Analysis Approach’*, CEPR Discussion Papers 17410.

# Abstract

In this paper we develop a series of Brexit uncertainty indices (BUI) based on UK newspaper coverage from January 2013 to April 2022. Using unsupervised machine learning (ML) methods to automatically select topics, our main contribution is to generate timely and cost-effective indicators of uncertainty. In further analysis we are able to distinguish between Brexit-related uncertainty and uncertainty due to COVID-19. We show that the pandemic accounted for around 50% of the overall uncertainty index post 2020, and this magnification effect varies across policy areas. Our indices can be used to quantify and compare the impact of Brexit uncertainty on different areas of the economy. We undertake a series of validation exercises and use sentiment analysis as an example of how our generated BUIs can be used in practice.

JEL Codes: D80, E32, E66, F50.

**Keywords:** Brexit; uncertainty; machine learning; textual analysis.

## 1.1 Introduction

The term “Brexit” was first used in May 2012, eight months before the then British Prime Minister David Cameron announced he would hold a referendum on whether the UK should leave the EU.<sup>1</sup> The term “Brexit” is now in popular usage especially after the largely unanticipated result that the UK was to leave the EU. An immediate reaction to the vote was a substantial and protracted increase in uncertainty that affected different aspects of the UK economy with some sectors and groups of people more heavily impacted than others. For example, the [Bank of England \(2019\)](#) points out that Brexit uncertainty has driven delays in firm investment and depressed productive capacity.<sup>2</sup>

The purpose of the paper is to construct aggregate and topic-specific news-based indices of Brexit uncertainty for the UK economy that from a policymakers perspective can be updated cost effectively and in close to real time. Using techniques from computational linguistics we derive our Brexit uncertainty indices (BUIs) from the frequency of relevant news coverage from eleven leading UK newspapers.<sup>3</sup> More broadly, a growing body of literature has emerged that documents the impact of uncertainty on macroeconomic activity.<sup>4</sup> Explanations for a negative uncertainty effect include that firms often employ a wait-and-see strategy when there is reduced visibility of the future ([Bernanke 1983](#)), increased financing costs ([Gilchrist et al. 2014](#)), and search frictions in the labor market and nominal rigidities ([Leduc & Liu 2016](#)). The usual definition of uncertainty used in macroeconomics is the

---

<sup>1</sup>“The rise of the word Brexit,” BBC news 2016.

<sup>2</sup>See “Monetary Policy Report - November 2019: In focus - Uncertainty and Brexit” published by the Bank of England.

<sup>3</sup>The Brexit uncertainty indices that are generated in this paper (and regularly updated) are freely accessible online at <https://sites.google.com/view/brexituncertainty/bui>.

<sup>4</sup>See, for example, [Bloom \(2009\)](#), [Fernández-Villaverde et al. \(2011\)](#), [Bachmann et al. \(2013\)](#), [Basu & Bundick \(2017\)](#), [Fajgelbaum et al. \(2017\)](#), [Bloom et al. \(2018\)](#), [Hassan et al. \(2019\)](#) and [Arellano et al. \(2019\)](#).

conditional volatility of economic or policy shocks that cannot be predicted (Jurado et al. 2015). In the context of this paper, the term Brexit uncertainty captures a range of possible uncertainties faced by the UK economy that were triggered by the decision to withdraw from the EU and subsequently reflected in UK newspaper coverage.

The contribution of the paper is threefold. First, we provide a close to real-time method for measuring historic and ongoing Brexit uncertainty. Recent studies have tended to rely on proxies for aggregate Brexit uncertainty that encompass different parts of the Brexit process. Examples include a proxy for pre-referendum uncertainty using public expectations of the Brexit referendum outcome in prediction markets (Graziano et al. 2020, 2021), and for post-referendum uncertainty using a firm-level survey (Bloom et al. 2019). Our news-based index, that starts in 2013, provides consistent and comparable indices before, during, and after the referendum. Second, in addition to an aggregate Brexit uncertainty index (BUI), for the first time to the best of our knowledge, we provide indices that capture the uncertainty associated with specific sectors and broader topics of interest. For example, in addition to the broader macroeconomy, we include uncertainty indices for trade policy, immigration, Northern Ireland, supply chains, energy & climate, and employment. Third, we disentangle the uncertainty induced by COVID-19 from Brexit for both the aggregate BUI and the topic based indices. This is important as the UK economy experienced, and is experiencing, considerable uncertainty as a result of the pandemic (Altig et al. 2020). Making a distinction between COVID and Brexit induced uncertainty allows us to quantify how the COVID and Brexit interact and hence show how COVID may have masked or exaggerated alternative measures of Brexit-related uncertainty.

Our methodological approach is to use machine learning algorithms to identify the news coverage of different aspect of Brexit uncertainty. For example, our aggregate BUI captures the frequency of newspaper articles that contain the word “Brexit,” a word sug-

gesting “uncertainty,” and a word indicating the country “UK.” To do this we adapt the Word2Vec algorithm developed by [Mikolov et al. \(2013\)](#) to find semantically similar words to “uncertainty” and “UK” in a context where Brexit uncertainty is being described. The algorithm captures semantic similarities between words by learning from a sub-sample of the news data. To this end, inputting “uncertain” and “UK” as seed words, the algorithm outputs semantically related words which we then prune using informed judgment to expand the word sets.

To capture how Brexit-related uncertainty impacts different topics, we rely on the Latent Dirichlet allocation (LDA) model, a probabilistic topic model, introduced by [Blei et al. \(2003\)](#) to decompose the content of articles to obtain the probability or the “share” of each article that is related to each Brexit topic, given that multiple topics could be covered in any one article. For this analysis, all articles containing the word “Brexit” and at least one word in the “uncertain” and “UK” sets are curated, processed, and composed into a corpus. This model automatically learns the corpus and extracts multiple topics in the form of probability distributions over words. For each article, a probability distribution over topics is obtained after fitting the LDA model. We therefore obtain the proportion of newspaper coverage of our corpus linked to each topic at each point in time. The subsequent time-varying proportions then constitute topic-specific BUIs. As a result, our topic-specific BUIs capture a series of uncertainty issues associated with the Brexit process including supply chain concerns, the UK-France dispute over fishing rights before the end of transition period, and Northern Ireland at the time when the Northern Ireland Protocol was enacted.

Although measuring Brexit-related uncertainty using our machine learning approach was relatively straight forward for the pre- and early post referendum result periods, the onset of the COVID-19 pandemic presented a particular challenge given the uncertainty the global pandemic induced. Our solution is to disentangle the uncertainty driven by Brexit

from that driven by COVID for the post-2020 period. Post 2020, both COVID and Brexit contribute to the uncertainty faced by the UK with many of the same issues being impacted such as international trade, labour markets, and supply chains.<sup>5</sup> An important concern is therefore the extent to which the COVID-19 pandemic has a distortionary effect on our BUI indices. To this end, we re-compute our BUIs excluding articles that mention COVID. We argue that such an approach provides a lower bound for our BUIs.

In practical terms, our approach is cost-effective in both human and financial terms (there is no costly survey to manage), and it enables us to examine both the pre- and post-referendum periods, and is not subject to the survey sample size and the response rate of those asked to complete a survey. Moreover, by examining different topics it means we can provide a level of disaggregation not previously considered in studies of Brexit uncertainty.

After generating our BUIs, the next stage of our analysis is to undertake a series of validation exercises to provide reassurance that we are capturing Brexit-related uncertainty with a high degree of accuracy. First, we investigate whether our aggregate BUI demonstrates strong co-movements with the well known survey-based Brexit uncertainty series from [Bloom et al. \(2019\)](#) which captures the proportion of managers who rate Brexit as the top source of uncertainty for their companies. Second, we compare our index against a similarly news-based measure of economic policy uncertainty from [Baker et al. \(2016\)](#). As a third sense check we construct an alternative BUI based on tweets on twitter and compare it with our news-based index. Third, we investigate whether the political leaning of the newspapers used to generate our indices effect the magnitude of the recorded uncertainty. Finally, we

---

<sup>5</sup>A number of newspaper articles discuss the relationship between COVID and Brexit uncertainty. For example, “COVID pandemic masks Brexit impact on UK economy” published by Financial Times in 2021 and “Impact of Brexit on economy ‘worse than COVID’” published by BBC in 2021. On average, over three-quarters of businesses viewed COVID as their top source of uncertainty after March 2020. Data from Decision Maker Panel (DMP): <https://decisionmakerpanel.co.uk/>.

examine how the evolution of our BUIs compares to that of two measures of UK stock market volatility over the same period.

As a final exercise we provide a small application of how our topic-specific BUIs can be used in future research by investigating whether public sentiment towards Brexit, as represented by tweets on Twitter, responds differently to different Brexit-related topics in the wake of the UK decision to leave the EU.<sup>6</sup> First, we perform a sentiment analysis with Brexit-related tweets and compute monthly sentiment scores from our sample of tweets from January 2016 to April 2022. We then regress these sentiment scores on our topic-specific BUIs to see which topics caused the largest positive and negative reactions from the general public expressed through social media.

Our paper relates to two main strands of literature. The first strand of literature looks explicitly on the effects of Brexit uncertainty which has been shown to impact investment and productivity ([Bloom et al. 2019](#)), employment ([Javorcik et al. 2020](#)), the EU-UK bilateral trade in goods ([Crowley et al. 2018](#), [Douch et al. 2020](#), [Graziano et al. 2021](#)) and trade in services ([Ahmad et al. 2023](#)). Some of these studies also look at how uncertainty impacts certain sectors. For example, [Crowley et al. \(2018\)](#) and [Douch et al. \(2020\)](#) examine the cross-sectional variation in trade policy-related Brexit uncertainty utilizing the contingent gaps in trade terms for different Brexit outcomes. Such measures, however, do not capture the evolution over time although [Bloom et al. \(2019\)](#) do show how uncertainty evolves by asking managers the same question each month and recording the number of managers who report Brexit as a major source of uncertain.

---

<sup>6</sup>Sentiment analysis has been used in studying patterns of information diffusion on social media with Twitter data ([Gorodnichenko et al. 2021](#)), measuring news sentiment with news articles ([Shapiro et al. 2022](#)) and firm's concerns towards epidemic diseases with transcripts of earnings conference call ([Hassan et al. 2020](#)).

The second strand of the literature conducts text analysis in an economics setting in which news-based measures of uncertainty have received considerable attention. The most widely used text-based uncertainty metric is the so-called dictionary method, which identifies a set of terms of interest for researchers and then calculates the frequency of those terms across text corpus to generate an uncertainty index. Examples include [Baker et al. \(2016\)](#) for economic policy uncertainty, [Caldara & Iacoviello \(2022\)](#) for geopolitical risk, [Husted et al. \(2020\)](#) for monetary policy uncertainty, and [Caldara et al. \(2020\)](#) for trade policy uncertainty. Leveraging machine learning tools, our method seeks to reduce subjective identification of terms and to allow for a larger feature space, i.e., accounting for many more words rather than merely those in the dictionary. Further, in a news-based dictionary method, the smallest unit is typically an article. More specifically, each qualifying article contributes equally to the topic-specific uncertainty measurement, despite the fact that the fraction of content concerning the topic of interest varies considerably from article to article.

In this paper, we employ the LDA model to decompose articles into a distribution of topics, and weight each article accordingly. [Hansen et al. \(2018\)](#), to our knowledge, was the first to introduce the LDA model into economics research. Previous studies have also used the LDA model to examine different types of uncertainty, for example, economic policy uncertainty ([Azqueta-Gavaldón 2017a](#)), and general business uncertainty ([Larsen 2021](#)).<sup>7</sup> In addition to being the first to introduce the LDA model into a Brexit uncertainty context we use the Word2Vec model to expand search term sets. By doing so, we reduce the subjectivity in the identification of term sets. Previous applications of this model include [Burn et al. \(2019\)](#), and [Davis et al. \(2020\)](#). There is also a growing literature on forecasting with text (see [Kalamara et al. 2022](#) for a discussion).

---

<sup>7</sup>Other applications of the LDA model in an economics setting includes [Hansen & McMahon \(2016\)](#), [Hansen et al. \(2019\)](#), [Larsen & Thorsrud \(2019\)](#) and [Bybee et al. \(2020\)](#).

To briefly summarise our results we find that the main spikes in Brexit-related uncertainty occurred around the time of the Brexit referendum, the three failed meaningful votes, when the Brexit withdrawal agreements were rejected in the House of Commons in early 2019, and the period just before the final agreement was reached. In terms of average values, Brexit uncertainty was relatively high between the announcement of the referendum date (February 2016) and the UK's formal departure from the EU (December 2019). Our results also show that even after the Brexit deal was reached, Brexit uncertainty did not fall significantly (post-2021) with a magnitude roughly four-fifths of that during the transition period. By taking into account of COVID uncertainty we show that Brexit uncertainty was exaggerated by as much as fifty percent in the post-2020 period due to the pandemic, with the most pronounced magnification effect related to employment (COVID effect accounting for 77.9% of the underlying uncertainty), government spending & budgets (74.3%), and supply chains (69.0%). Results suggest that fishing rights and Northern Ireland were the least affected by a COVID effect accounting for 29.1% and 30.1%, respectively.

Validation exercises reveal a correlation between our aggregate BUI and the DMP index of 0.82 rising to 0.93 when we drop the first two years of their index (due to a relatively small number of panel members and survey respondents (see [Bloom et al. 2019](#)). Our lower bound BUI (excluding COVID pandemic effects) results in an even higher overall correlation of 0.86. Our aggregate BUI also has a positive correlation with the UK economic policy uncertainty (EPU) index constructed by [Baker et al. \(2016\)](#) with a recorded correlation of 0.73 for the post-2016 period, during which Brexit emerges as the main source of economic policy uncertainty in the UK economy. In a further robustness check, we find that the Twitter-based BUI correlates highly with the news-based BUI (with a correlation coefficient of 0.91 with the lower bound BUI) and that, post-referendum, there was significant co-movement between our currency-related BUI and measures of UK stock market volatility. Finally, our sentiment analysis reveals that the general public's engagement with social media

elicits the most negative sentiment when it comes to supply chains and Northern Ireland and the most positive when it comes to the broader macroeconomy and trade policy.

The remainder of this paper is organized as follows. Section 2 describes the data and describes the methodological approach. Section 3 presents our different indices and discusses the policy implications while Section 4 shows the results of our validation exercises. Section 5 analyzes how public sentiment on Twitter evolves in response to Brexit uncertainties. Section 6 concludes.

## 1.2 Related Literature

### 1.2.1 Research on Text-based Uncertainty Measurement

The most widely used text-based uncertainty metric is the so-called dictionary method, which identifies a set of terms of interest for researchers and then calculates the frequency of those terms across a text corpus to generate an uncertainty index. For example, [Baker et al. \(2016\)](#) construct an Economic Policy Uncertainty (EPU) index by capturing the frequency of newspaper articles containing researcher-identified words such as “Economics,” “Policy,” and “Uncertainty.” In particular, they first collect articles containing the two term sets: “economic” and “uncertain” (or “economy” and “uncertainty”). Thousands of human hours are then spent selecting appropriate policy term sets from these articles for the aggregate economic policy and sub-categories (e.g., policy uncertainties in health care and national security). Constructed this way, their index surges during significant events, including the September 11 attacks and presidential elections. Following this approach, policy uncertainty has been measured in Japan ([Saxegaard et al. 2022](#)), Spain ([Ghirelli et al. 2019](#)), Flanders,

Wallonia, and Belgium ([Algaba et al. 2020](#)), and China ([Huang & Luk 2020](#)).

Using the same news-based approach, [Husted et al. \(2020\)](#) quantify monetary policy uncertainty with a self-defined monetary policy dictionary. Similarly, [Caldara & Iacoviello \(2022\)](#) develop geopolitical risk dictionaries and quantify industry exposure by regressing portfolio returns for each sector on the variation in the geopolitical risk index. [Caldara et al. \(2020\)](#) also propose a trade policy uncertainty dictionary to quantify uncertainty in trade policy. The dictionary method is by far the most intuitive text-based measure of uncertainty, and its simplicity has earned it widespread usage. However, the downsides of the method are non-trivial: it offers only a limited feature space, meaning only a limited number of terms can be considered. The reliability of the method is compromised by the reliance on human judgment, which fails to distinguish the sentiment that is reflected in the text. Other studies incorporating machine learning methods have attempted to improve the uncertainty metric by reducing human involvement ([Davis et al. 2019](#), [Larsen 2021](#)).

The most widely used machine learning and natural language processing approaches for uncertainty metrics include keyword expansion, supervised learning, and unsupervised learning algorithms. Keyword expansion techniques learn text data and automatically output words with similar semantics to the seed words. Human judgment is then applied to these synonyms to define a desirable dictionary. This automated approach helps reduce the subjective bias of the dictionary method in term selection. For example, [Davis et al. \(2019\)](#) used Chinese newspapers to measure economic policy uncertainty in China by manually defining economic and uncertainty-related words and used the TextRank algorithm to define policy-related words. However, keyword-matching measures with these ML-based dictionaries may fail to capture all the ways in which humans express uncertainty ([Keith et al. 2020](#)) and may be subject to self-selection measurement bias ([Tobback et al. 2018](#)).

The supervised learning approach is theoretically capable of addressing these problems as all the words in the corpus can be considered, and subjective selection is not necessary. These methods do not require pre-determined keywords. The models learn from a subsample of articles pre-labeled by the researcher as relevant to certain uncertainties and then they automatically predict whether the remaining articles address those uncertainties. For example, [Noailly et al. \(2022\)](#) introduce a news-based index of US environmental and climate policy uncertainty (EnvPU), using supervised machine learning algorithms to analyze news articles from ten US newspapers over the period 1990-2019. First, they use a supervised support vector machine (SVM) model, trained on 2,500 hand-labeled articles, to define environmental and climate policy-related articles from the entire set of articles. Second, they labeled 622 articles, randomly selected from environmental policy-related articles, as uncertainty relevant or irrelevant. This labeled data was then used to train an SVM classifier, which was subsequently used to identify uncertainty-related articles. [Saltzman & Yung \(2018\)](#) also employ supervised machine learning models to quantify a variety of uncertainties from the Federal Reserve Beige Books from 1970 to 2018, integrating a support vector machine model, deterministic algorithms, and graph pattern rules to identify context-specific instances of uncertainty.

Although supervised machine learning has been applied to uncertainty metrics, a significant drawback is the requirement for substantial manpower needed to pre-label article sets. In contrast, labeled datasets are not required in unsupervised machine learning. The most commonly used unsupervised algorithm for uncertainty measures is the topic model, which detects hidden patterns in textual datasets to automatically cluster articles. Topics can then be defined for each cluster based on the features (e.g., keywords) of each cluster. For example, [Larsen \(2021\)](#) measures uncertainty by applying the unsupervised LDA model to classify a large set of news articles into various themes without predefining specific keywords. They then quantify uncertainty in these categorical topics by counting the

frequency of uncertainty-related terms appearing in the articles. Other examples include [Azqueta-Gavaldón \(2017a\)](#) and [Azqueta-Gavaldon \(2017b\)](#).

## 1.2.2 Brexit Uncertainty Measurements

There is literature that considers Brexit-related uncertainty and how to measure it. [Bloom et al. \(2019\)](#) introduce a survey-based measure of Brexit uncertainty, known as the ‘decision maker panel’ survey, and utilize the proportion of firms reporting Brexit uncertainty as a significant source of uncertainty to construct their Brexit Uncertainty Index. The survey provides a monthly index of Brexit uncertainty that evolves over time, starting in September 2016. [Graziano et al. \(2020\)](#) and [Graziano et al. \(2021\)](#) measure pre-referendum Brexit uncertainty using the average daily price of a contract on PredictIt.org, an online prediction market where users can buy and sell shares based on the outcomes of political and financial events. This contract pays one dollar if Brexit occurs by December 2016, reflecting the market’s perceived probability of the event.

[Hassan et al. \(2024\)](#) measure companies’ exposure to risks associated with Brexit by tracking how often ‘risk,’ ‘uncertainty,’ or their synonyms appear near the term ‘Brexit’ in the transcripts of quarterly earnings calls between companies and financial analysts. This is a firm-level metric. To study the impact of the Brexit vote on the labor market, [Javorcik et al. \(2020\)](#) employ four Brexit uncertainty measures, including the survey-based Brexit Uncertainty Index proposed by [Bloom et al. \(2019\)](#) and the average of [Hassan et al. \(2024\)](#)’s firm-level Brexit risk as a measure of aggregate uncertainty. They also construct their own Brexit-related trade policy uncertainty measure based on the dictionary method, defining Brexit, uncertainty, and trade-related terms. Their fourth metric utilizes an index of Google searches, i.e., the frequency with which people search for relevant terms on Google, to indicate

the level of Brexit uncertainty.

Some studies focus explicitly on subdomains of Brexit uncertainty. In [Crowley et al. \(2018\)](#), trade policy-related Brexit uncertainty is assessed by examining the potential changes in future tariff schedules. Specifically, the uncertainty arises from the potential shift between two distinct trade policy scenarios: (1) maintaining tariff-free access to the EU Customs Union or (2) reverting to the EU's WTO tariff schedule. By analyzing how the increased risk of higher tariffs influences firm entry and exit in the EU market, researchers quantify the uncertainty stemming from the renegotiation process. Similar metrics are also used in [Douch et al. \(2020\)](#).

### 1.3 Background on Brexit

The entire process of Brexit can be divided into five main periods. These five periods are: (1) the pre-referendum period, from January 2013 to January 2016, (2) the referendum period, from the announcement of the referendum date in February 2016 to May 2017, (3) the negotiation period, from the start of the UK-EU negotiations in June 2017 to December 2019, (4) the transition period, from January 2020, when the UK formally left the EU and entered the transition period, to December 2020, when the Brexit deal was reached and the transition period ended, and (5) the post-Brexit period, from January 2021 to April 2022. [Table 1.1](#) presents a timeline of the main events surrounding Brexit.

The first period commenced with Prime Minister David Cameron's announcement that he was in favor of holding a referendum on leaving the European Union. In the second period, the referendum on leaving the EU was held in June 2016, and the unanticipated result saw a majority of voters in favor of leaving the EU (51.9% voted to leave versus 48.1%

voted to remain). The unexpectedness of the result was captured by the odds in the betting market, where the predicted probability of voting to leave the EU in the referendum was around 0.3, and 0.25 on the day before the referendum.<sup>8</sup>

The third period lasted two and a half years and was marked by political instability, involving two general elections, three votes that rejected the withdrawal agreement proposed by the prime minister, multiple motions of no confidence, and three postponements of the leaving date. This period ended when the UK officially left the EU. The uncertainty in the fourth period was largely due to the uncertainty about the future relationship between the EU and the UK, and it ended when the Brexit deal, ‘The EU-UK Trade and Cooperation Agreement,’ was agreed upon and the transition period ended.

After the UK-EU deal was achieved, there was a clearer understanding of the future UK-EU relationship. However, a series of disagreements remained, such as the dispute over the Northern Ireland protocol and issues within the fishing industry. Beyond this, the UK’s relationship with the rest of the world needs to be redefined as an independent entity rather than as a member of the EU.

### **1.3.1 Brexit Uncertainty**

The UK formally leaving the EU single market and customs union means that it is free to pursue trade agreements with the EU and third countries, to set immigration policy, and to legislate independently. Here, we discuss a few of the key dimensions related to Brexit uncertainty.

---

<sup>8</sup>Data Source: “Brexit at the Bookies”, Centre on Constitutional Change.  
<https://www.centreonconstitutionalchange.ac.uk/opinions/brexit-bookies>

**Table 1.1:** Brexit Timeline

<b>The Pre-Referendum Period</b>	
23 January 2013	Prime Minister David Cameron announced his backing for the EU referendum
14 April 2015	Conservative Party Manifesto for the 2015 General Election promises referendum
7 May 2015	The 2015 General Election is held - David Cameron as the leader of the Conservative Party is re-elected Prime Minister.
<b>The Referendum Period</b>	
22 February 2016	Prime Minister declares the referendum date
The Referendum	
23 June 2016	Brexit Referendum
24 June 2016	Prime Minister David Cameron declares his intention to resign
13 July 2016	Theresa May serves as the new Prime Minister
29 March 2017	Article 50 of the Treaty on European Union is triggered
<b>The Negotiation Period</b>	
8 June 2017	The 2017 General Election is held. The Conservative Party wins, and Theresa May remains as Prime Minister
19 June 2017	The UK-EU Brexit negotiations starts
24 July 2018	Government White Paper on UK-EU Future Relationship is published, confirming the leaving date - March 29, 2019
14 November 2018	The Brexit Withdrawal Agreement is published
15 January 2019	First failed “meaningful vote” – the Withdrawal Agreement is rejected in the UK House of Commons; Opposition leader raises a motion of no confidence in the Government
16 January 2019	The Government wins a vote of confidence
12 March 2019	Second failed “meaningful vote” on the Withdrawal Agreement
13 March 2019	MPs vote to exclude a ‘no-deal Brexit’
21 March 2019	Theresa May asks for extension of Article 50 period – postponing the date of leaving the EU - (to 22 May 2019 if the Withdrawal Agreement is agreed and otherwise, to 12 April 2019)
29 March 2019	Third failed “meaningful vote”
10 April 2019	The UK and the EU27 agreed to a Brexit extension until 31 October 2019
24 May 2019	Theresa May announces her intention to resign
24 July 2019	Boris Johnson becomes the UK Prime Minister
25 July 2019	Boris Johnson refused to exclude the possibility of a “no-deal Brexit”
19 October 2019	A new Brexit deal is rejected in the Commons; Boris Johnson requests Brexit extension
28 October 2019	Brexit is extended to 31 January 2020
12 December 2019	The 2019 General Election is held with Boris Johnson’s re-election
<b>The Transition Period</b>	
31 January 2020	The UK officially exits the EU and moves into a transition period.
24 December 2020	The Brexit deal, i.e., The EU-UK Trade and Cooperation Agreement, is finally reached
31 December 2020	Transition period ends with UK leaving EU single market and customs union
<b>The Post-Brexit Period</b>	

Notes: This table presents the timeline of main Brexit events.

**Trade:** The EU is the nearest and largest trading partner of the UK and accounted for 44% of UK exports and 53% of imports in 2015. Once a member of the EU, the UK enjoyed frictionless trade with other EU member states. After leaving the EU, the possible outcomes of trade relations between the two sides during the UK-EU negotiations ranged from, in the worst case, trading under WTO rules to, in the best case, maintaining the original trade provisions. Trade in goods without tariff and quota restrictions was largely retained after the trade deal was reached. However, newly imposed red tape and regulatory procedures at the border resulted in considerable uncertainty. The need to sign new trade agreements with third countries also introduced unpredictability into the trade environment.

**Immigration:** Before Brexit, the UK was bound by EU free movement of people laws that allowed citizens of EU/EEA member states to move freely into and reside in its territory. Withdrawal from the EU implies that the UK can set its own immigration guidelines. Post-Brexit, the UK introduced a points-based immigration system that claims to treat immigrant applicants from EU and non-EU member states equally and is targeted at high-skilled immigrants. However, shortages of heavy goods vehicle drivers and food industry workers intensified controversy over the new immigration policy that blocks low-skilled immigrants from the EU (except for those from Ireland).

**Fishing:** The impact of Brexit on the fishing industry is characterized by a high degree of indeterminacy. Post-Brexit uncertainty for UK fisheries mainly stems from possible changes to the fishing rights of EU vessels in UK waters and those of UK vessels in EU waters and new customs regulations. As fish are live animals, their timely processing and transportation across borders are critical. However, post-Brexit, the introduction of new border checks and health certification requirements led to delays. The implementation of new customs regulations added to these concerns, making what was once seamless trade between the UK and EU more complex and unpredictable. Additionally, the frequent crossing of fishing vessels between UK and EU waters, coupled with unclear regulations and changing

access rights, could spark disputes and increase tensions, further challenging an industry already under significant pressure in the post-Brexit era.

**Northern Ireland:** The Northern Ireland issue became a focal point of the UK-EU negotiations. It is the only area where the UK shares a land border with another EU member state. Brexit means that the external border of the EU lies through the island of Ireland. This brings about the so-called “Irish Trilemma,” which means that one of the following three scenarios is bound to happen: staying in the EU Single Market and Customs Union, a hard border on the island of Ireland, or a customs border on the Irish Sea. The first scenario was ruled out by the first Theresa May government. The Northern Ireland Protocol, agreed in October 2019, guarantees no customs checks or controls on the island of Ireland, i.e., no hard border, but rather a de facto customs border built over the Irish Sea, separating Northern Ireland from Great Britain. The dispute in the post-Brexit period mainly arose from the implementation of the Protocol, for example, the extent to which goods traded should comply with EU or UK regulations, and the threat of renegotiating the protocol.

**Supply chains:** The impact of Brexit on supply chain uncertainty is mostly through potential trade barriers with the EU, worker shortages, and labour costs due to immigration restrictions.

**Environment and Energy:** Brexit means that the UK is no longer required to comply with EU environmental laws, raising concerns that environmental standards may be lowered, despite government claims that regulations would be strengthened.

## 1.4 Measuring Brexit Uncertainty

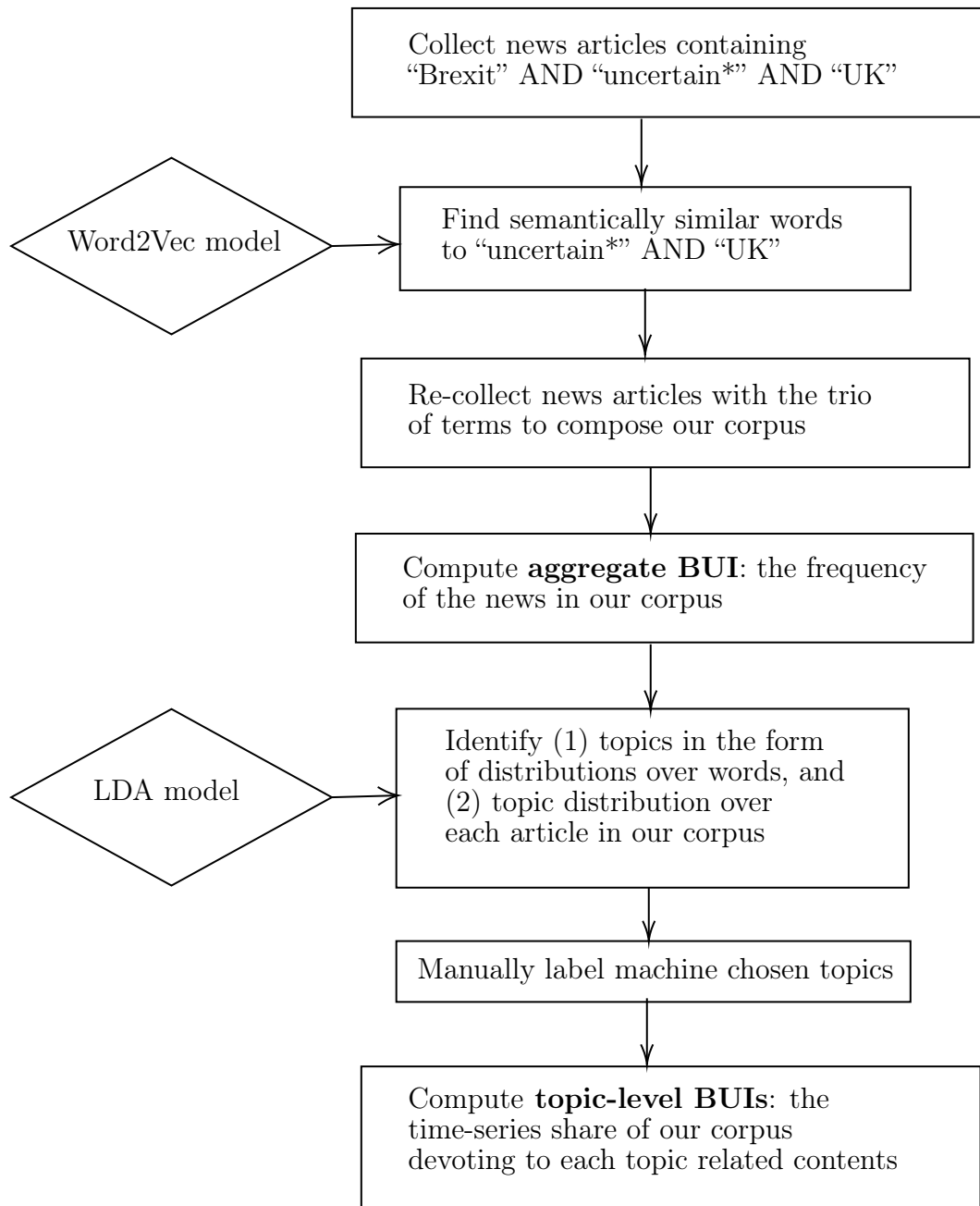
In this section we describe the data and methodology used to measure aggregate and our topic specific Brexit uncertainty indices. Put simply, our Brexit uncertainty indices are based on the frequency of newspaper coverage of different aspects of uncertainty related to the decision to hold a referendum on EU membership in June 2016 and the impact of the largely unanticipated decision to leave the EU.

In terms of methodology, two machine learning (ML) models are used to assist in identifying Brexit-related news content. A flowchart that illustrates our approach is shown in Figure 1.1. First, from the thousands of news articles each day we select those articles that touch on the topic of Brexit uncertainty based on the inclusion of the word “Brexit” and words referring to “uncertainty” and “UK.” The Word2Vec model helps by outputting the groups of words that capture “uncertainty” and “UK.” A detailed description of the Word2Vec model can be found in Section 2.2 where we explain the construction of the aggregate BUI.

The second ML algorithm, the LDA model, serves to match newspaper coverage to Brexit topics by identifying topics in the form of distributions over words, and subsequently, topic distribution over each news article in our news corpus. We provide more details of how topic-specific BUIs are calculated in Section 2.3.

### 1.4.1 News Data and Pre-processing

Our BUIs are based on articles from eleven leading British newspapers which are *The Financial Times*, *The Times*, *The Sunday Times*, *The Daily Telegraph*, *The Daily Mail and Mail on Sunday*, *The Daily Express*, *The Guardian*, *The Mirror*, *The Sun*, *The Northern Echo*, and *The Evening Standard*. It was a deliberate decision to choose the same newspaper



**Figure 1.1:** A flowchart to illustrate how we construct our measures of aggregate and topic specific BUIs using the two ML models, Word2Vec and LDA.

list used by [Baker et al. \(2016\)](#) to measure UK economic policy uncertainty (this also allows us to compare our results with those of [Baker et al. \(2016\)](#) later in the paper). Our ML methodology is applied to newspaper articles compiled and made accessible to the public by *Nexis* which is an online database that archives articles from a broad range of news sources including our eleven UK based newspapers. We chose January 2013, when the Brexit referendum was first announced by the Prime Minister, as the starting point for our data collection. Prior to this date, the term “Brexit” was not in popular usage. Data was collected up until April 2022.

Once the data is collected it needs to be pre-processed and cleaned before it can be used with our ML algorithms. First, a pattern matching technique is employed to automatically extract publication date and the content of each news article from the downloaded documents. In this case, we look for two common patterns. The first pattern is to locate and extract the news body by the fixed words or sentences that appear before and after the news body, that is, “Body” and “End of Document.” The second pattern is to locate the common expression of publication date which is invariably in the order of the month, e.g., January, the one or two-digit date, e.g., 18, a comma or space, and then the four-digit year, e.g., 2013. We take the first eligible date in each document as the publication date.

The next step is to remove duplicated articles by excluding those that have identical body parts. Duplication is mostly attributed to various editions of the same piece of news. We then clean the news content data by converting uppercase letters to lowercase, removing stop words that do not carry information, e.g., “the,” “a,” and “is,” punctuation marks, white spaces and numbers. The fourth step is tokenization, i.e., splitting the text data into individual words. Finally, the words are stripped down to their stems, for example, the words “manufacturing,” “manufacture,” “manufacturer” and “manufacturers” are all reduced to “manufactur”. The word stems are not necessarily English words.

## 1.4.2 Measuring Aggregate BUI: Expanding Search Words with the Word2Vec Model

The main assumption underpinning our analysis is that articles describing Brexit uncertainty in the UK contain the term “Brexit,” one or more terms indicating “uncertainty” and at least one term suggesting the country “UK.”<sup>9</sup> If one was to use the dictionary method, those semantically similar terms are defined manually and hence, are to some extent subjective. As such, it can be difficult to obtain uniform term sets across multiple pieces of research which could potentially generate inconsistent results. To reduce subjective judgement we draw on a ML algorithm to identify the words that are included in the “uncertainty” and “UK” term sets.

More specifically, we adopt the Word2Vec model developed by [Mikolov et al. \(2013\)](#) to look for words that are semantically similar to “uncertain” and “UK.” We choose this technique for its relative simplicity and long-term good performance in ML applications. To train the model around 10,000 news articles including “Brexit,” “uncertain\*” and “UK” are collected and cleaned.<sup>10</sup>

The key to the model is word embedding, also known as word vectorization. The outputs from the model consists of word vectors that capture the association between words by performing dimensionality reduction of vector space. In other words, mapping all words to low-dimensional vectors. In general, the simplest vectorization is to create a vector space in which each dimension corresponds to a unique word. Suppose that the corpus involves  $V$

---

<sup>9</sup>Including the terms “EU referendum,” “European Union referendum,” “UK referendum,” “UK’s withdrawal,” “EU Exit,” “EU withdrawal,” “leave the EU,” “exit from the EU,” “Exit the EU,” “Withdrawal from the EU,” to indicate Brexit does not change the results. Indeed, the correlation between indices with and without these terms is effectively one.

<sup>10</sup>When we use the term “\*” it means it includes all the words with that root. For example, “uncertain\*” includes the terms “uncertain,” “uncertainty,” “uncertainties,” and “uncertainly.”

unique words in total, then the vector space would have  $V$  dimensions. Word vectors would only have one non-zero valued element corresponding to its particular dimension, and the other elements would be 0, i.e., one-hot encoded vector. These vectors are orthogonal with zero cosine similarity between them, and thus, not conveying the similarity between words. Using the Word2Vec model, the vector space is reduced (to less than  $V$ ), and similar words are assigned to vectors that are close on the vector space. The similarity between words can be predicted based on the context in which the words appear, that is, the model defines words that appear in similar contexts as similar words.

We apply the model using a continuous bag-of-words (CBOW) architecture. In essence, the model with CBOW architecture aims to predict the center words, conditional on their context words. The algorithm works as a deep learning classification model, and self-generates the observation-label pairs, that is, the context-center word pairs. Given that the context of the center word can be lengthy, it is necessary to set a window size  $C$  when training the model, signifying that the preceding and following  $C$  words of the center word are taken as the context. Suppose that the corpus involves  $V$  unique words in total, then the input layer consists of  $V$ -dimensional one-hot encoded vectors  $(X_{-c}, \dots, X_{-1}, X_1, \dots, X_c)$  of contextual words with size  $2C$ . The output layer is the one-hot encoded vector of the center word  $X_0$ . In the process of predicting the central word or in the hidden layer of the neural network, the model computes a  $N$ -dimensional vector representation of it, in which the size of vector space  $N$  is pre-defined. Then, the distance, i.e., the cosine similarity, between all word vectors can be calculated.<sup>11</sup> These values range between -1 and 1, with larger values corresponding to greater similarity. The distance between word vectors captures the contextual similarity of words, and can also represent semantic relationships (Mikolov et al. 2013).

We fix the vector space to 200 dimensions ( $N = 200$ ), set the window size to five

---

<sup>11</sup>See Mikolov et al. (2013) for more details.

words before or after ( $C = 5$ ), and follows [Davis et al. \(2020\)](#) by assuming a cosine similarity of 0.4 as the cut-off. This is a common setting for applications of the model. When we expand the uncertainty word set, we end up with 53 stems. The three stems most similar to the two seed stems “uncertain” and “uncertaini” are “unstabl\*,” “unpredict\*,” “unsettl\*” and “jitter,” “uncertainli,” “anxieti,” respectively.<sup>12</sup> The model-generated term set is then manually trimmed, dropping terms that would most likely suggest other notions. In the final stage, eighteen stems, as synonyms for “uncertain,” are selected. The top three analogous stems to the term “UK” are “Britain,” “British” and “Country”. We incorporate the first two into the set as they directly refer to the country. A full list of selected terms is provided in [Appendix A.1](#).

News containing the trio of term sets are collected and this gives us a total of 114,734 articles (after removing duplicates).<sup>13</sup> The monthly frequency of those news articles contributes to our aggregate BUI. As there is variation in the volume of news over time, we scale the raw frequency by the total number of articles in the eleven newspapers over the same period. Following [Azzimonti \(2018\)](#), we take the number of articles containing the word “today” as a proxy for the total number of news items. The index is then normalized to a maximum value of 100.

---

<sup>12</sup>Using a narrow “uncertainty” term set, i.e., including only “uncertain\*,” gives a correlation of 0.96 with our aggregate BUI. We use the extended term set to be consistent with previous studies (see, for example, [Caldara et al. 2020](#), [Hassan et al. 2020](#), and [Caldara & Iacoviello 2022](#)). This approach also has the advantage of generating a bigger sample of news data for subsequent topic-level studies.

<sup>13</sup>A potential concern is that our article-level co-occurrence criterion might over-count uncertainty references that are unrelated to Brexit. As a robustness check, we re-estimated the uncertainty by requiring uncertainty-related terms to appear within ten words of Brexit in each article. The correlation between this narrower measure and our baseline index is 0.98, indicating that the results are robust to this alternative specification.

### 1.4.3 Measuring Topic-level BUIs: An Application of a LDA Model

The Latent Dirichlet Allocation (LDA) model is an unsupervised topic modelling ML algorithm and is one of the most popular models used in textual analysis. More broadly, supervised and unsupervised learning are the two main tools in ML. Supervised learning refers to the process of inferring a function that maps observations to labels, i.e., classes, by learning the observation-label pairs. With this function, the classes for unlabeled observations can be predicted. In this respect, pre-determined search words are not required, and all the words in the corpus can be taken into consideration. Yet, in our case, using these algorithms requires a substantial number of manually labeled content-topic pairs, making it highly labour-intensive, especially considering the variability of the English language and lexical usage. Such pitfalls can be avoided by using unsupervised learning models where a labeled dataset is not required. These models can self-learn the underlying structure of the data and look for hidden patterns, i.e., classes within the data and can automatically match the patterns to observations. Intuitively, in our setup, these algorithms can self-uncover latent topics in the corpus, and map news content to these topics. Appendix A.2 provides a more detailed discussion about why we choose the LDA model for topic modeling.

LDA is a Bayesian probabilistic model. Suppose there is a corpus with  $V$  unique words and  $D$  documents. The first objective of the LDA model is to extract a predefined number of  $K$  latent patterns, or so-called “topics,” from the corpus, with each topic  $k$  being a probability distribution over  $V$  unique words from our corpus denoted by vector  $\varphi_k$ . Intuitively, a topic is a grouping of words, each of which contributes differently to that topic. The LDA model also estimates the topic distribution of each article. In this sense, each article is represented by a mixture of topics with different “weights.” Technically speaking, each document  $d$  can be interpreted as a probability distribution over  $K$  topics denoted by vector  $\theta_d$ .

LDA is a generative statistical model in which the estimation process generates our

corpus with two distributions  $\varphi_k$  and  $\theta_d$ . The probability of the  $i$ th word appearing in document  $d$  is  $p(d_i) = \sum_k \varphi_k^i \theta_d^k$ , where  $\varphi_k^i$  is the probability of the  $i$ th word appearing in the topic  $k$ , and  $\theta_d^k$  is the probability of topic  $k$  in document  $d$ . Then, the probability of accurately generating our corpus is  $p(C) = \prod_{d=1}^D \prod_{i=1}^V p(d_i)^{n_{di}}$ , where  $n_{di}$  represents the number of occurrences of  $i$ th word in document  $d$ .

The aim in the generative process is therefore to find the parameters for the two distributions that maximize  $p(C)$ . An important step in the LDA process is to place Dirichlet priors on the two probability vectors, that is,  $\varphi_k \sim \text{Dirichlet}(\beta)$  and  $\theta_d \sim \text{Dirichlet}(\alpha)$ , where  $\alpha$  and  $\beta$  are the hyperparameters that decide the concentration of the two distributions. A low  $\alpha$  results in a steep topic distribution for each article, and the model with a low value of  $\beta$  provides a steep word distribution over each topic. We apply a popular Gibbs Sampling estimator for model estimation (Griffiths & Steyvers 2004).

Essentially, LDA reduces the dimensionality of our corpus, with  $V$  dimensions ( $V$  unique words) for the original data, and afterwards,  $K$  dimensions ( $K$  topics). It is worth noting that each topic is the probability distribution over  $V$  words. Hence, the LDA accounts for the full  $V$  dimension, i.e., all words. Appendix A.3 provides a more detailed description.

As for model selection, we set the number of topics  $K$  to 80.<sup>14</sup> In this process, perplexity, a statistical measure of how well the generative model predicts samples, is often used to evaluate the performance of the LDA model. Yet, good statistical indicators may coexist with low interpretability of the output topics. Hence, models should be evaluated on the basis of real-world performance in specific tasks but not the technical criteria (see Chang et al. 2009). Accordingly, what we value most in the process of parameter selection

---

<sup>14</sup>We rely on the hyperparametric optimization techniques under LDA model estimation using the popular ‘MALLET’ package to achieve self-tuned  $\alpha$  and  $\beta$  that are optimized every ten iterations. Robustness checks for different hyperparameter values are computationally expensive so, in line with other studies, we only perform one set of hyperparameters (Hansen & McMahan 2016, Hansen et al. 2018, Larsen et al. 2021).

is the extent to which the topics of interest are reasonably grouped. Fewer topics will result in topics of interest being mixed up, and a greater number of topics can lead to topics that are difficult to interpret.

Qualified news articles, described in Section 2.2, are cleaned according to the process outlined in Section 2.1 and used as a training corpus in the LDA model estimation. Before feeding the corpus into the LDA model, it is usual to exclude frequent and rare words to obtain a more interpretable model fit, i.e., grouping of topics (Gentzkow et al. 2019). To this end, we filter extreme words by removing word stems that occur in less than 300 documents and more than 70% of all the documents. The result is that we end up with 8,380 unique word stems. The model estimation usually stabilizes within a few hundred iterations. We therefore set the number of iterations in our estimations to 2,000.

The LDA model generates two distributions of interest, the distribution of words in each topic and the distribution of topics in each article. Notably, the estimated LDA model fixes the first distributions, i.e., topics. The model then serves to infer the topic distribution of the articles, either from the original training corpus or from new, unseen texts. This feature allows us to update the indices quickly as new articles appear.<sup>15</sup>

Figure 1.2 provides a visualization of the outputs from the first distribution. Each word cloud represents a topic with the size of the words reflecting their relative weight in the topic. We define each topic based on the most important words contained in that topic. For example, Figure 1.2 (a), (b), (c) and (d) are example topics (word mixtures), that we identify as Northern Ireland, supply chain, employment, and macroeconomy-currency relevant topics, respectively.<sup>16</sup> Then, based on the second distribution, we compute the share of each topic

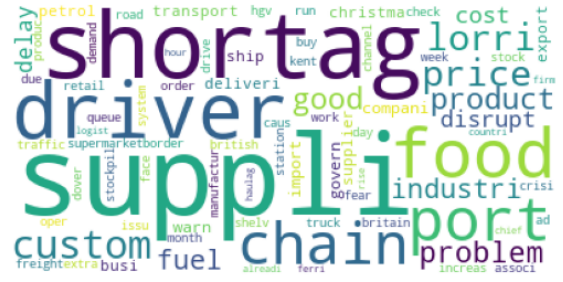
---

<sup>15</sup>We estimate the LDA model using news articles from our corpus up to November 2021 giving us 111,797 articles in total. If a new compelling topic of Brexit uncertainty arises it would be possible to re-estimate the LDA model.

<sup>16</sup>Word cloud graphs for the remaining 76 topics are available from the authors upon request.



(a) Northern Ireland



(b) Supply Chain



(c) Employment



(d) Macroeconomy-Currency

**Figure 1.2:** Example outputs of the LDA model. Each word cloud represents a topic, i.e., a probability distribution of words. The size of a word indicates the probability of the word appearing in the topic. Each word cloud shows only the 80 words with the highest probability of occurrence. The labels of the topics are manually defined.

in all news published in a given month and subsequently compute the topic frequencies. This gives rise to  $K$  monthly time series. i.e., topic-specific BUIs. To keep the topic-level and the aggregate BUIs comparable, the topic-specific indices are scaled in the same way that we scale the aggregate index in the normalization process.<sup>17</sup>

<sup>17</sup>For example, if the maximum value of the aggregate index before normalization is 200, then the normalization of the aggregate BUI is to divide all index values by two in order to scale the index to a maximum value of 100. Subsequently, the topic BUIs are normalized via a division by the number that normalizes the total index to a maximum value of 100 (two in this case).

## 1.5 Brexit Uncertainty Index

In this section we present our Brexit uncertainty indices and our analysis of the amplification effect of the COVID pandemic on our BUIs. We present, in turn, the aggregate BUI, the aggregate BUI after removing the effect of COVID, topic-level BUIs and finally, topic-level BUIs excluding COVID effects.

### 1.5.1 Aggregate BUI

Figure 1.3 provides a visual representation of our aggregate BUI shown here at the monthly level. The first observation is that the index shows obvious spikes around the Brexit referendum date in June 2016, the general election and the start of UK-EU negotiations in June 2017, failed votes on the withdrawal agreement in early 2019, the Brexit extension in late 2019, and the end of the transition period in late 2020. In contrast, the economy experienced lulls in the level of uncertainty following the Brexit deal, i.e., the UK-EU Trade and Cooperation Agreement, that was signed in December 2020. The low and flat curve before 2016 shows that there was little concern or mention of the term Brexit prior to this date. It is also worth noting that in mid-2020, before the agreement had been reached, there is a clear trough in the index. This can be attributed to the outbreak of COVID-19. As mentioned before, our index is scaled by the total number of news items, i.e., it measures the relative share of Brexit uncertainty in all news. As such, our index rightly falls as attention turned to reporting on the pandemic. Appendix A.4 presents a similar figure for a daily BUI which, unsurprisingly, is noisier.

As part of our validation exercises that we undertake to ensure that our index is a reasonable proxy of Brexit-related uncertainty, we compare our index with the Brexit uncertainty index generated from the DMP survey. The DMP is a survey that targets

business managers and is available from September 2016. In this survey, subjects respond to whether, and to what extent, Brexit is a source of uncertainty for their business. In this case, we compare our BUI with the time evolution of the proportion of managers that rate Brexit as their largest driver of uncertainty. The results reveal that the responses of the macro variables are in line with the baseline results.

As Figure 1.3 shows, there is a clear common movement in the BUI and the DMP series with a correlation coefficient of 0.82. This correlation climbs to 0.93 after September 2018 (before which the DMP had a relatively small number of respondents, and the question about Brexit uncertainty was not asked every month).<sup>18</sup> These high correlations suggest that our index closely matched the reported concerns of UK businesses and gives us confidence that, to some extent at least, we have an accurate measure of the degree of Brexit uncertainty at any given point in time.

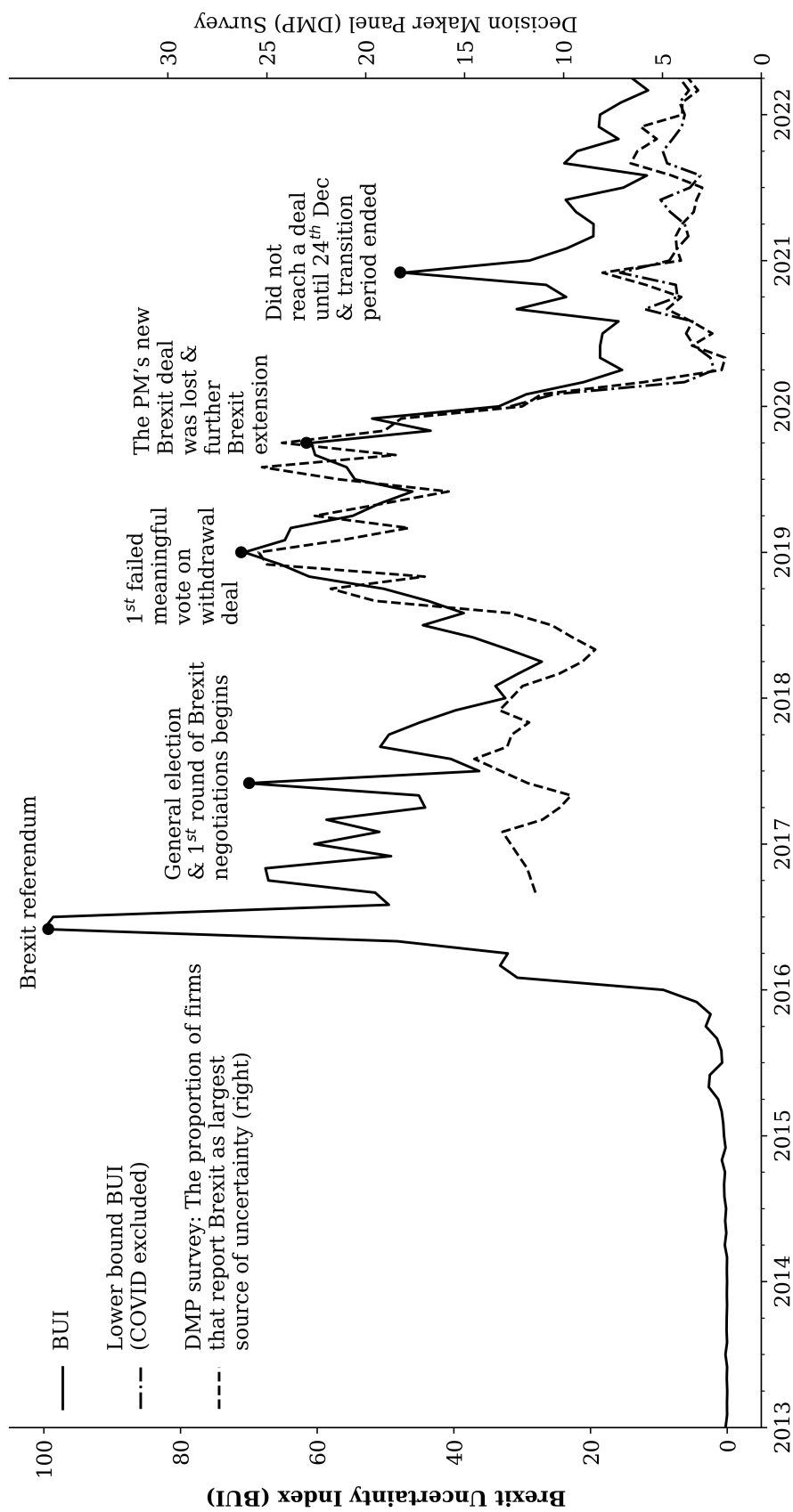
## 1.5.2 Aggregate BUI Excluding the Effects of COVID

Note that while the BUI and DMP series in Figure 1.3 move in roughly the same direction in the post-2020 period, Brexit uncertainty under DMP experiences a relatively sharper decline in early 2020 and remains at a lower level compared to our BUI. One possible explanation is that our BUI is exaggerated by the COVID pandemic and led to the next stage of our investigation.

Hence, a possible concern one might have with the interpretation of our BUI is whether the rapid spread of COVID-19 in late 2019 and early 2020 biased our uncertainty indices, given that the pandemic also generated a significant degree of uncertainty for the UK economy and would most likely work in the same direction on some issues, for example, supply

---

<sup>18</sup>The DMP's membership panel was below 2,000 before July 2017, exceeded 4,000 in May 2018 and is now over 9,000.



**Figure 1.3:** BUI, lower bound BUI (COVID excluded), and the proportion of firms that report Brexit as their largest source of uncertainty from the DMP survey. BUI is a scaled monthly index based on news articles that include “Brexit” and one or more words in the “uncertainty” word set, and one or more terms in the “UK” term set listed in Appendix A.1. The index is normalized to a maximum value of 100. Lower bound BUI stands for scaled monthly amount of news articles including the identical triplet of terms as BUI but excluding all articles containing one or more words suggesting “COVID” that are issued after January 2020. The DMP index reflects the proportion of firms that report Brexit as their largest source of uncertainty.

chain disruptions and labor shortages. In fact, the majority of our eligible post-2020 Brexit uncertainty news items also discuss the coronavirus pandemic (13,420 out of 22,594 articles). It is therefore hard to say to what extent the uncertainty being discussed in the news truly stems from Brexit, rather than the pandemic. Accordingly, any news-based measure of Brexit uncertainty may overestimate Brexit-related uncertainties. Hence, we exclude those articles that include COVID related words after January 2020 to give us what we call our lower bound BUI.<sup>19</sup>

Figure 1.3 also includes the lower bound BUI. Post-2020, the lower bound BUI and the DMP series move together closely with a correlation coefficient of 0.90 (compared to 0.62 between the aggregate BUI and the DMP series for the same period). The correlation coefficients are 0.86 overall and 0.97 after September 2018. These findings suggest that our lower bound BUI succeeds in disentangling the uncertainty of Brexit from that of the pandemic and captures purely BUI as a lower bound.

### 1.5.3 Topic Specific BUIs

Figure 1.4 plots eight different topic-level BUIs.<sup>20</sup> Note that the topic-level BUI figures have different scales reflecting the degree of uncertainty associated with that topic. Looking at the different figures we see that the Northern Ireland BUI (Figure 1.4(a)) had its greatest peak when the “Irish Backstop” was replaced by the new “Northern Ireland Protocol” that guarantees no customs checks or controls on the island of Ireland, i.e., no hard border, but rather a de facto customs border in the Irish Sea. Similarly, the immigration BUI spiked five times (see Figure 1.4(b)), first around the Brexit referendum, the second at the commence of UK-EU negotiation, the third on the publishing of post-Brexit immigration system white

---

<sup>19</sup>COVID related word stems are “COVID,” “coronavirus,” “pandem,” “vaccin,” and “epidem.”

<sup>20</sup>We only show topic-level BUIs from 2015 onwards, i.e., excluding the 2013 to 2015 period. Prior to 2015 index values are close to zero.

paper, the forth at the deadline of the EU settlement scheme, and finally during the time when migrants were first seen crossing the English Channel in late 2021.

For other topics, Figure 1.4(c) shows that the supply chain index peaked at the end of transition period, when uncertainty on reaching a Brexit deal prompted unprecedented supply chain concerns. More recently, in late 2021, post-Brexit Britain is again facing a significant supply chain crisis. Labor shortages, represented by the shortage of truck drivers, impacted the supply of energy and food and is reflected not only in the supply chain BUI, but also the energy & climate (Figure 1.4(d)) and the employment (Figure 1.4(e)) BUIs.

The trade policy BUI (Figure 1.4(f)) shows considerable volatility with significant responses when the leaving date was confirmed in mid-2017, and periods of major concern surrounding the signing of the Brexit deal including the three failed meaningful votes on the withdrawal deal in early 2019 and in late 2021 when Brexit deal was reached only a week before the end of transition period. The fishing BUI (Figure 1.4(g)) that mainly captures worries about possible changes to the fishing rights of EU vessels in UK waters and those of UK vessels in EU waters surged during the signing of the Brexit deal that set out fishing licence issues, the May 2021 protests over fishing rights in Jersey, and a series of intensified British-French conflicts over fishing licences in late 2021. The broader macroeconomy BUI (Figure 1.4(h)) that covers tax, government spending and the budget, interest rates, inflation rates, currencies, financial markets and economic growth, rose sharply during the Brexit referendum (note the scale of the index is significantly higher than the other seven topic-level BUIs shown in Figure 1.4).

To investigate the performance of our topic-level BUIs further, Table 1.2 reports 14 different BUIs. Index values are reported for five different periods that capture the four main Brexit events. The five periods are (1) the pre-referendum period, from January 2013 to January 2016, (2) the referendum period, from the announcement of the referendum date

in February 2016 to May 2017, (3) the negotiation period, from the start of the UK-EU negotiations in June 2017 to December 2019, (4) the transition period, from January 2020, when the UK formally leaves the EU and enters the transition period, to December 2020, when the Brexit deal has been reached and the transition period ends, and (5) post-Brexit periods, from January 2021 to April 2022.

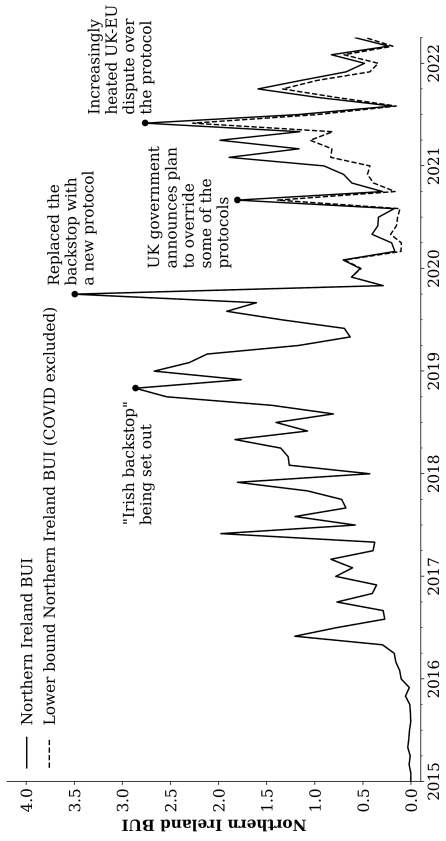
In general, Northern Ireland and trade policy-related Brexit uncertainty received the most attention.<sup>21</sup> During the pre-referendum period, Brexit uncertainty was barely evident. Later, the macroeconomy-currency and trade policy topics led to some of the highest recorded levels of uncertainty in the referendum period. The Northern Ireland issue is the largest source of uncertainty during the negotiation, transition, and post-Brexit periods, as the centre of a dispute between the UK and the EU.<sup>22</sup>

It is also clear that COVID matters as we saw in Figure 1.4, especially for the employment, macroeconomy-government spending & budget, and supply chain issues. This is shown in Table 1.2, where the joint effects of the COVID and Brexit on these three issues are three times that of Brexit uncertainty alone. The role of COVID on energy & climate BUI is also apparent (falling to 46.7% after removing the amplification effect of COVID), reflecting the recent energy crisis exacerbated by both events. In contrast, Northern Ireland and the fishing dispute BUIs appear to be the topics that are influenced the least by the pandemic.

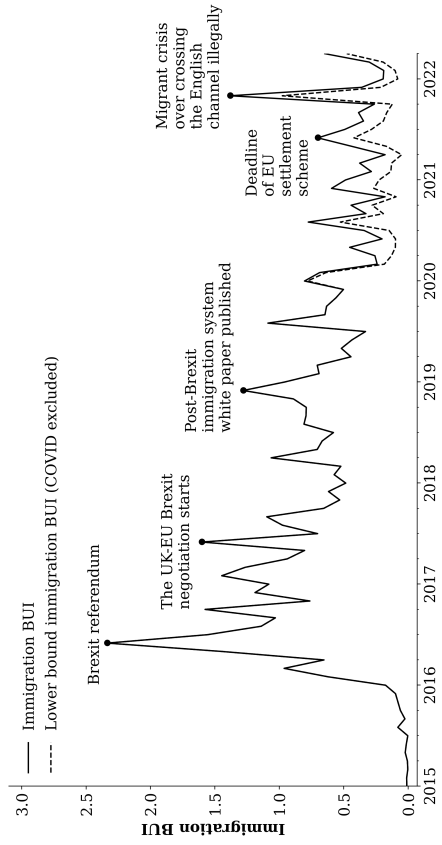
---

<sup>21</sup>Here, we do not consider the macroeconomy topic as a whole, but discuss its subtopics.

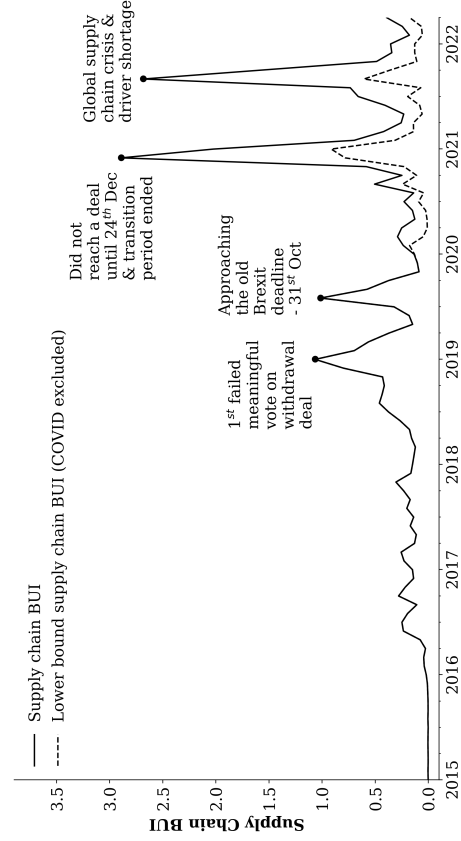
<sup>22</sup>Some topics are more likely to appear at the same time in the same article, with the most frequent topics appearing together being tax and government spending & budget (correlation of .10 in articles), immigration and employment (.08), as well as trade policy and food (.06). For completeness correlations between the different topics are shown in Appendix A.5.



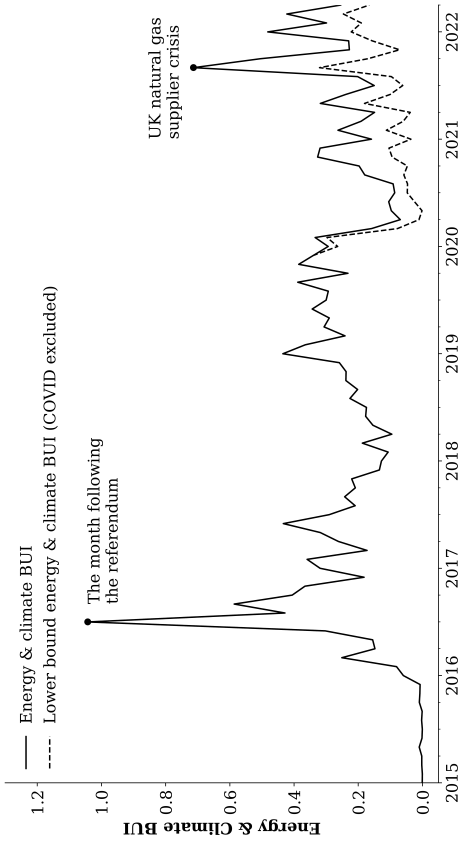
(a) Northern Ireland



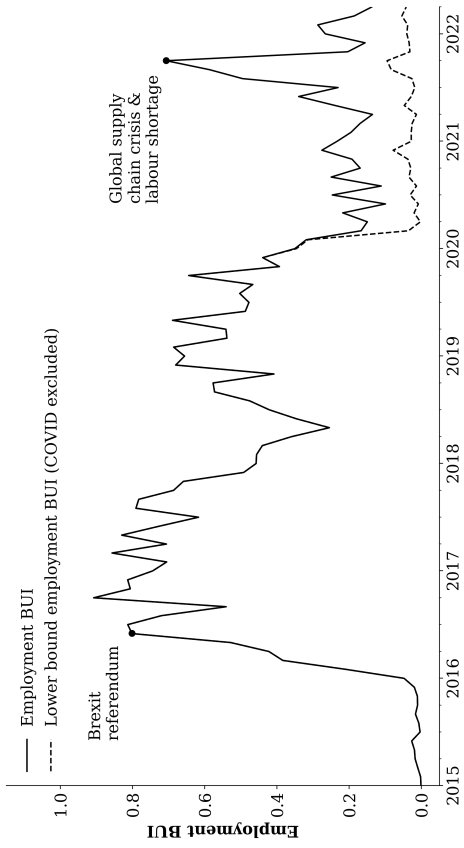
(b) Immigration



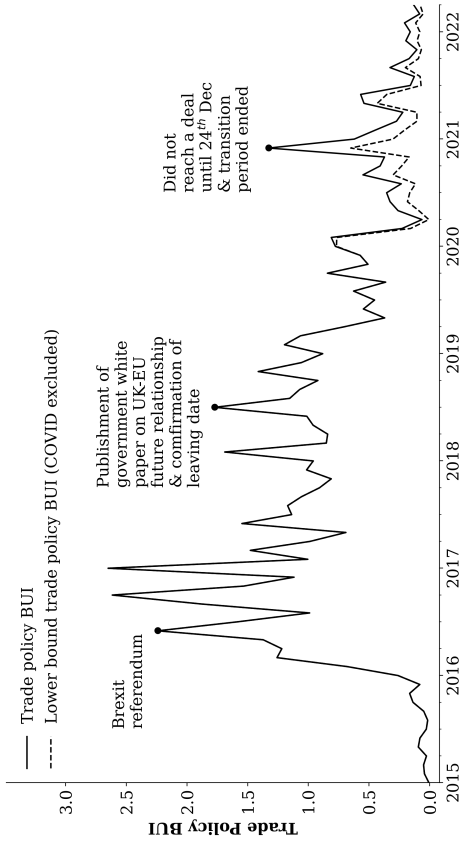
(c) Supply chain



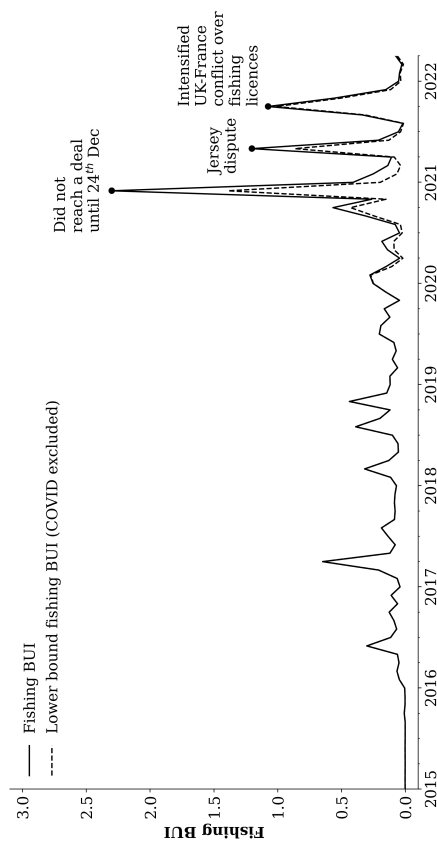
(d) Energy & climate



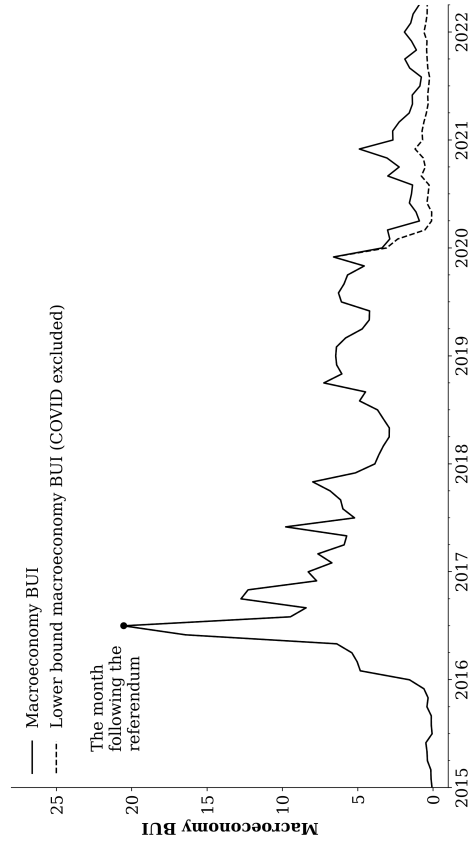
(e) Employment



(f) Trade policy



(g) Fishing



(h) Macroeconomy

**Figure 1.4:** Topic specific BUIs. Topic specific BUIs indicate scaled monthly share of news coverage concerning each topic in question. The indices are normalized by dividing by the number that scales the maximum value of the aggregate BUI to 100. Lower Bound BUIs stands for the generated BUIs after excluding all articles mentioning COVID issues.

**Table 1.2:** Brexit Uncertainty by Topic and Time Period, 2013-2022

Time Period	2013:1-2016:1	2016:2-2017:5	2017:6-2019:12	2020:1-2020:12	2021:1-2022:4	2020:1-2022:4	2013:1-2022:04
	Pre-referendum period	Referendum period	Negotiation period	Transition period (COVID excluded)	Post-Brexit period (COVID excluded)	COVID period	Overall (COVID excluded)
Aggregate Brexit Uncertainty	0.95	55.45	48.6	24.9 (42.23%)	19.03 (37.37%)	21.55 (39.78%)	27.07
Macroeconomy	0.14	8.96	5.37	2.40 (35.43%)	1.57 (27.90%)	1.92 (31.92%)	3.29
Government Spending & Budget	0.01	0.82	0.59	0.34 (29.03%)	0.24 (22.20%)	0.28 (25.70%)	0.35
Tax	0.01	0.64	0.43	0.19 (41.01%)	0.21 (32.44%)	0.20 (35.94%)	0.26
Currency	0.02	1.46	0.51	0.17 (37.80%)	0.06 (20.58%)	0.11 (32.26%)	0.38
Housing Price	0.00	1.02	0.73	0.27 (43.34%)	0.15 (16.22%)	0.20 (31.81%)	0.40
Trade Policy	0.03	1.46	0.95	0.48 (62.99%)	0.27 (56.27%)	0.36 (60.09%)	0.57
Northern Ireland	0.01	0.49	1.45	0.52 (70.41%)	1.11 (70.76%)	0.86 (70.67%)	0.69
Supply Chain	0.00	0.15	0.35	0.48 (32.72%)	0.73 (30.14%)	0.62 (31.00%)	0.27
Energy & Climate	0.00	0.34	0.25	0.19 (47.48%)	0.30 (46.33%)	0.25 (46.69%)	0.18
Fishing Dispute	0.00	0.14	0.14	0.38 (67.08%)	0.29 (74.75%)	0.33 (70.95%)	0.14
Immigration	0.02	1.17	0.74	0.44 (64.61%)	0.44 (55.87%)	0.44 (59.63%)	0.49
Employment	0.01	0.67	0.54	0.21 (37.19%)	0.29 (13.72%)	0.25 (22.12%)	0.31
Food Industry	0.00	0.31	0.37	0.38 (68.86%)	0.29 (61.38%)	0.33 (65.09%)	0.23
Manufacturing	0.01	0.60	0.72	0.30 (44.10%)	0.18 (47.57%)	0.23 (45.63%)	0.34

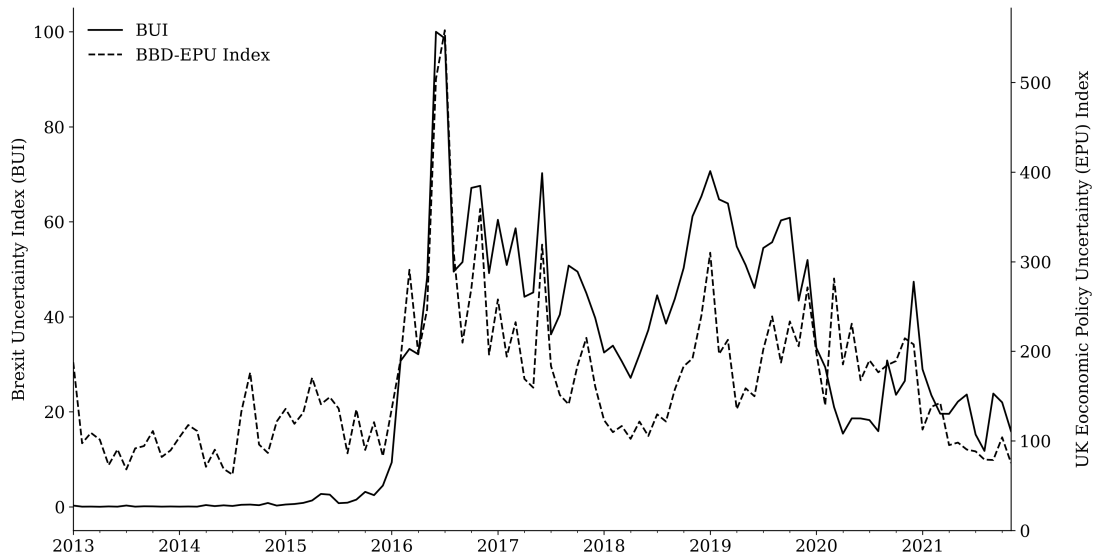
Notes: A summary of the index values for the aggregate BUI and 14 topic-level BUIs of Brexit uncertainty averaged over time. The percentages (in parentheses) are the proportions after removing articles that mention COVID.

## 1.6 Validation Exercises

As we note in Section 1.5.1, our aggregate BUI is closely related to the index derived from business concerns towards Brexit uncertainty from the DMP survey. This section provides a series of additional validation exercises. We begin with a comparison of BUI with an UK economic policy uncertainty (EPU) index constructed by Baker et al. (2016) (henceforth BBD-EPU). Next, we present a new UK EPU index that uses our ML approach and compares it to the previously calculated BBD-EPU index. As a further robustness check, we also introduce a BUI built from Twitter data and compare it with our news-based index. We also examine whether accounting for the political leanings of the newspapers included in our study change the results. Finally, as a simple application and sense check we examine how closely our BUIs are matched to two measures of UK stock market volatility.

### 1.6.1 Comparison with the BBD-EPU Index

The BBD-EPU index is constructed using the dictionary method, i.e., based on the frequency of news items that contain pre-defined sets of words. As can be seen in Figure 1.5, the BBD-EPU index and our aggregate BUI have followed very similar trends since 2016 with a correlation of 0.73. Our BUI, nevertheless, reacts more strongly to the post-referendum Brexit event, benchmarked against the EU referendum. In contrast, the BBD-EPU index shows a stronger response to the COVID-19 epidemic in 2020. While the two indexes encapsulate different elements, the uncertainty associated with Brexit is clearly the most important factor affecting UK economic policy in recent years and hence it is reassuring that the indices follow similar trends.



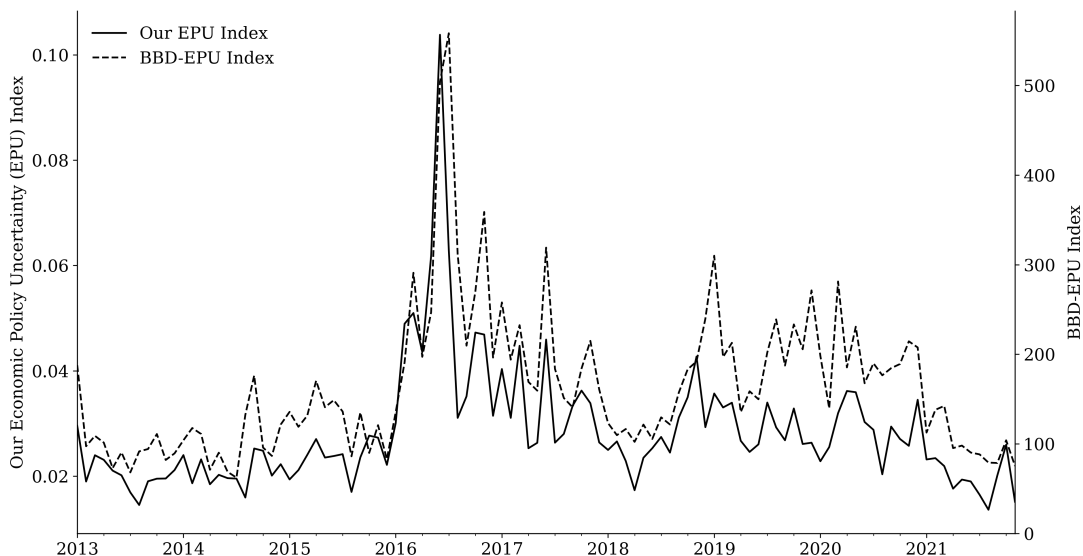
**Figure 1.5:** Our BUI and the BBD-EPU Index. The figure shows our aggregate BUI from Figure 1.3 and the BBD-EPU index for the UK economy.

## 1.6.2 A New EPU Index Using a ML Approach

The BBD-EPU index is based on the frequency of news containing one or more term from the trio sets indicating “Economic,” “Policy” and “Uncertainty,” respectively. We, instead, collect the news articles with their “Economic” term set and our “Uncertainty” term set without using the word “Policy.” We then rely on the LDA model to distinguish economic “policy” relevant content from collected articles. The resulting index, shown in Figure 1.6, also exhibits a clear co-movement with the BBD-EPU index, with a correlation coefficient of 0.83. The correlation coefficient is only 0.65 before the LDA model was used to select economic policy-related content.

The strong correlation between the BBD-EPU index and the EPU index developed with our ML approach reveals three insights. First, the selection of our “uncertainty” term set

is reasonable. Second, the LDA model helps to identify policy-related content. Accordingly, we confirm that the LDA model can help to disentangle the Brexit and its topics from a large number of other elements. Finally, it demonstrates that our indices can be considered to be reasonably reliable proxies for different Brexit uncertainties.

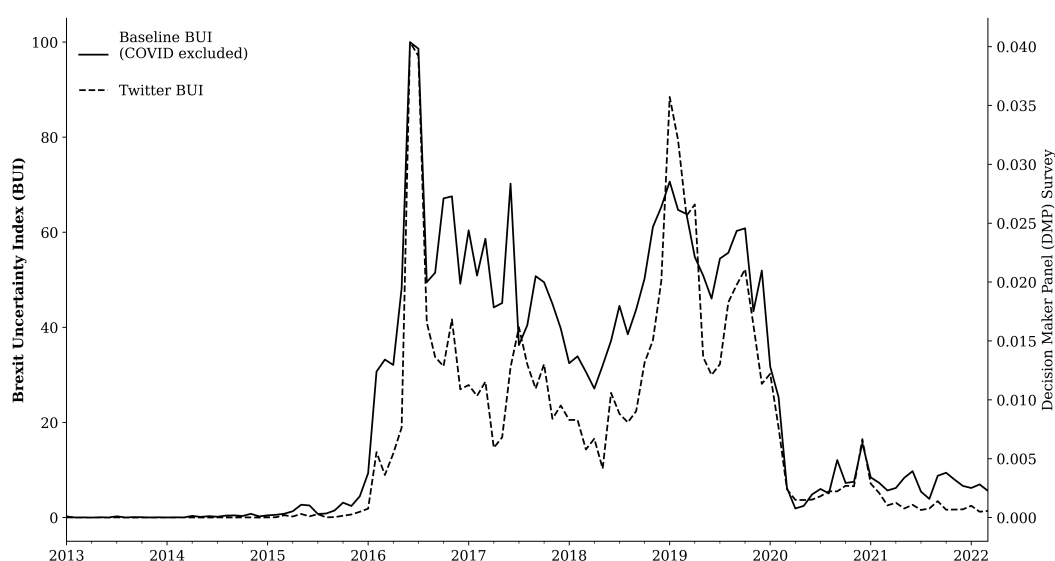


**Figure 1.6:** Our EPU Index and the BBD-EPU Index. The figure shows our EPU index constructed with our BUI measure and the BBD-EPU index constructed by [Baker et al. \(2016\)](#). Both EPU indices are based on the same eleven UK newspapers listed in Section 2.1.

### 1.6.3 A New BUI Using the X (Twitter) API

Although informative, our previous validation exercises do not address doubts over the reliability and accuracy of news data per se. For example, journalistic preferences for certain types of news story may bias the results. As a further verification exercise we measure Brexit uncertainty based on tweets on X, formerly known as Twitter and hereafter referred to as Twitter in this paper. We construct a Twitter-based aggregate level BUI based on the fre-

quency of tweets that discuss issues related to Brexit uncertainty. We count the number of tweets that contains “Brexit,” and “uncertain” or “uncertainty,” and “UK.” We then scale the raw counts by the total number of tweets containing “today” and “UK” over the same period. Retweets are not included in the analysis. The data are collected through the Twitter API for academic research. Figure 1.7 compares the Twitter based and news-based BUIs and shows that they follow similar trends and have a correlation of 0.89 with the the standard BUI and 0.91 with the lower bound (excluding COVID) BUI.



**Figure 1.7:** Our BUIs and the X (twitter) BUI. The figure shows our aggregate BUI and lower bound aggregate BUI from Figure 1.3, and the twitter BUI that we constructed based on twitter data.

### 1.6.4 Political Leanings in Newspaper Coverage of Brexit Uncertainty

A concern one might have is that our indices are unduly influenced by the political leaning of the newspapers in our sample. The Conservative Party, typically considered to be on the

center-right on the political spectrum, was in power for the entire period of our analysis. Our measure, then, may be skewed if left (right) leaning newspapers overstate (downplay) the extent of Brexit uncertainty. To this end, we examine how the political leanings of newspapers affect our BUI. The results show that the BUIs generated from left-wing and right-wing newspapers are highly correlated with a correlation coefficient of 0.98.<sup>23</sup>

### 1.6.5 BUI and Stock Market Volatility

As a final validation exercise and a way to illustrate a simple use case for our indices we investigate whether there is any correlation between movements in our BUIs and stock market volatility. Hence, we plot BUI against the FTSE 100 Volatility Index (VFTSE), representing the implied volatility on the FTSE 100, available until June 2019.<sup>24</sup> Comparing our different BUIs against the VFTSE we find that the aggregate BUI and the currency BUI are the two that most closely match the VFTSE. As shown in Figure 1.8, our aggregate BUI and VFTSE follow similar trends after the Brexit referendum (a correlation of 0.60). Since the referendum, our currency BUI and VFTSE also follow similar trends with a correlation of 0.58. However, after 2018 the VFTSE was influenced by the surge in US stock market volatility and other international influences, e.g., the US-China trade war in late 2018. Reassuringly, our BUI does not react to such events.<sup>25</sup>

Results for the FTSE 100 realized volatility (which is available for the whole time span of our index) are fairly similar.<sup>26</sup> In this case, we select the pre-2020 period for comparison

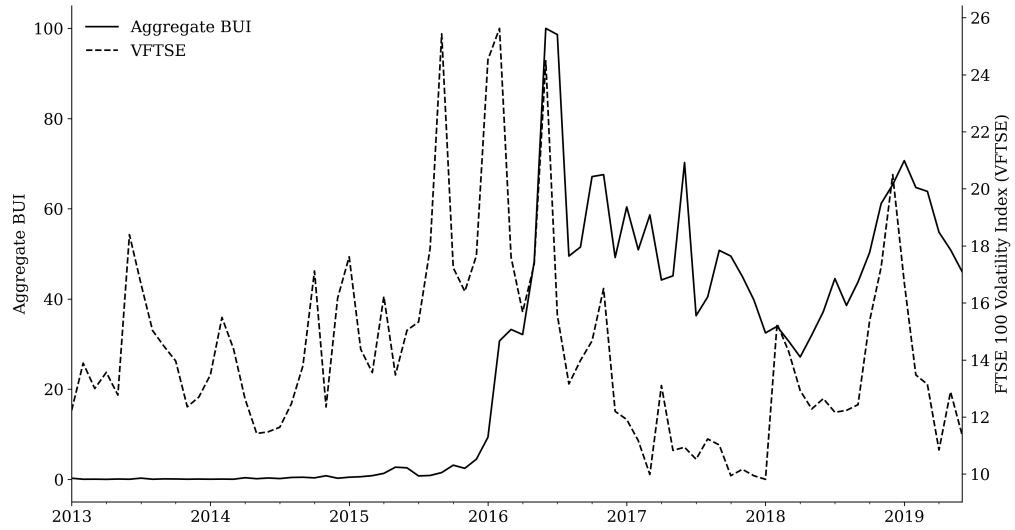
---

<sup>23</sup>Based on a classification from <https://www.oxford-royale.com/articles/a-guide-to-british-newspapers/> we classify left leaning newspapers as *The Guardian* and *The Mirror* and right leaning newspapers as *The Daily Telegraph*, *The Times*, *The Daily Mail*, *The Mail on Sunday* and *The Daily Express*. Figures showing the index values by political leaning are available from the authors upon request.

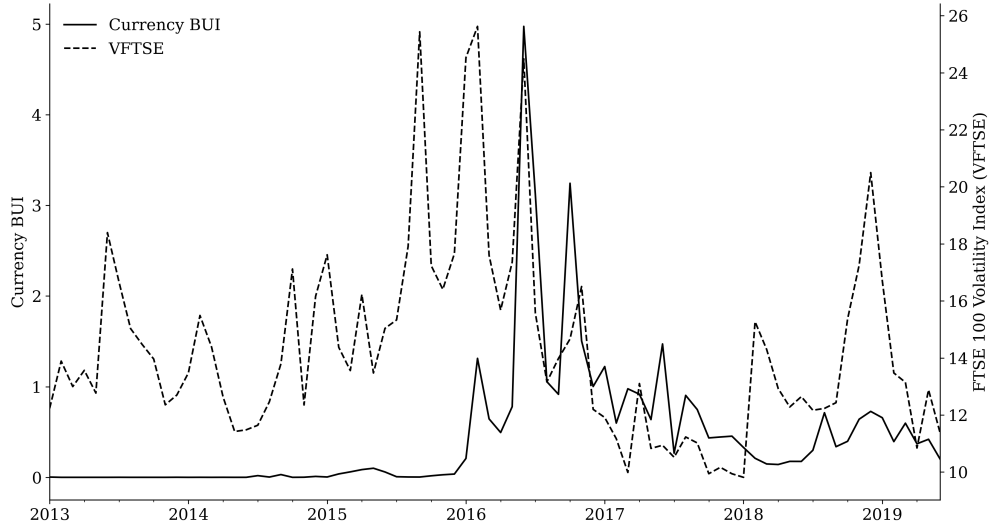
<sup>24</sup>Datasource: Bloomberg.

<sup>25</sup>The VIX, which is the implied volatility of the US stock market spiked in early 2018.

<sup>26</sup>Datasource: “Oxford-Man Institute’s realized library” (Heber et al. 2009).



(a) Aggregate BUI Compared to FTSE 100 Volatility Index (VFTSE) from Jan 2013 to Dec 2019



(b) Currency BUI Compared to FTSE 100 Volatility Index (VFTSE) from Jan 2013 to June 2019.

**Figure 1.8:** Aggregate BUI, Currency BUI, and FTSE 100 Volatility Index (VFTSE)

because after 2020 the COVID shock had a significant impact on the stock market that overshadowed Brexit uncertainty effects and is a period when stock market volatility and Brexit uncertainty are not significantly correlated. Although not shown, the results show again that post referendum, the currency and aggregate BUIs are correlated to the FTSE 100 realized volatility (correlations of 0.68 and 0.54, respectively).

## 1.7 Application: Public Sentiment and BUIs

In this section, we demonstrate a simple application of our indices by examining how public sentiment responds differently to different Brexit-related topics and their associated uncertainties. In particular, we construct monthly public sentiment scores using Twitter data and relate these scores to our topic-specific BUIs. Sentiment plays an important role in human decision making, including investment and financial decisions (Dolan 2002, Nofsinger 2005), which in turn can influence stock market performance (Bollen et al. 2011, Ranco et al. 2015). The connection between Brexit uncertainty and public sentiment is therefore worth exploring.

We focus on all Twitter posts that mention “Brexit” and were posted in the UK between January 2016 and April 2022. We remove retweets and duplicated posts to avoid double counting. Our main sample contains a total of 1,680,308 tweets. We also consider a subset of tweets which mention “Brexit” as a hashtag (i.e., with a specific Brexit label), giving us 362,629 tweets in total.

We then construct a Brexit sentiment score for each tweet using a Bidirectional Encoder Representations from Transformers (BERT) model, a deep neural network for language understanding, developed by researchers at Google (Devlin et al. 2018).<sup>27</sup> Specifically, we

---

<sup>27</sup>The BERT model has achieved state-of-the-art accuracy in previous sentiment analysis tasks, e.g., Sun

use a BERT model, trained by around 60,000 English tweets labeled with positive, negative or neutral, to predict the sentiment of each tweet in our sample.<sup>28</sup> Put simply, inputting a tweet, the model outputs the predicted probabilities that the tweet falls into the positive, negative, and neutral categories. The probabilities that the sentiment of a tweet falls into one of these three categories sums to 1. We compute the sentiment score of a tweet by subtracting the probability of being predicted as negative from that of being predicted as positive, i.e.,  $Sentiment\_score = p_{(positive)} - p_{(negative)}$ . Sentiment scores range from -1 (most negative) to 1 (most positive). To illustrate, we include four example tweets and their sentiment scores. More example tweets of different Brexit topics and their sentiment scores are shown in the appendix A.6. It is important to note that sentiment analysis captures the ‘tone’ of the entire tweet and not necessarily the opinion that the person sending the tweet has on the topic. For example, in relation to Brexit a tweet may express a negative view on the government in relation to Brexit but also mention immigration or Northern Ireland so the tweet will have a negative sentiment score although the poster is not necessarily putting forward an opinion on whether they are pro- or anti-immigration. Examples of positive and negative tweets include the following:

Positive tweets:

*“An incredible achievement from @BorisJohnson and Her Majesty’s Government. Great news! Also a clear example to the British people of the benefits of being an independent nation and not forced to be tied into things against our national interest. Thank goodness for Brexit!”* (Sentiment score: 0.987).

---

et al. (2019), Xu et al. (2019), Gao et al. (2019). Using the older sentiment analysis tool, namely the Valence Aware Dictionary for Sentiment Reasoner (VADER), provides us with a similar results with a correlation of 0.73.

<sup>28</sup>In our application of BERT, we use the BERT model implemented by the Python package *pySentimiento*. See Pérez et al. (2021), and also Nakov et al. (2019), Aguilar et al. (2020), Nguyen et al. (2020) for details.

*“#Brexit gives excellent opportunity for UK to be a great nation @BorisJohnson We can build a great NHS and a great nation”* (Sentiment score: 0.991).

Negative tweets:

*“@KateHoeyMP @BBCNews Where’s the Leave Plan that was made before referendum, before Article 50 or before today? I’m so sad that you still keep spouting the same mantras. #brexit is a disaster!”* (Sentiment score: -0.971).

*“#Brexit is a scandal full stop. Built on lies, delusion and hatred. #StopBrexit-SaveBritain #StopBrexitSaveDemocracy”* (Sentiment score: -0.978).

Figure 1.9 presents the monthly average Brexit sentiment scores for all Brexit-related tweets and for the sub-sample of those that contains the term “Brexit” in a hashtag, and the evolution in the monthly volumes of total Brexit-related tweets from the UK. Figure 1.9 shows consistent negative sentiment towards Brexit among posters on Twitter. The Brexit sentiment series are characterized by an overall downward trend, indicating that public sentiment towards Brexit has become increasingly negative over time albeit with positive reaction at the time of the UK’s formal withdrawal from the EU in January 2020.<sup>29</sup>

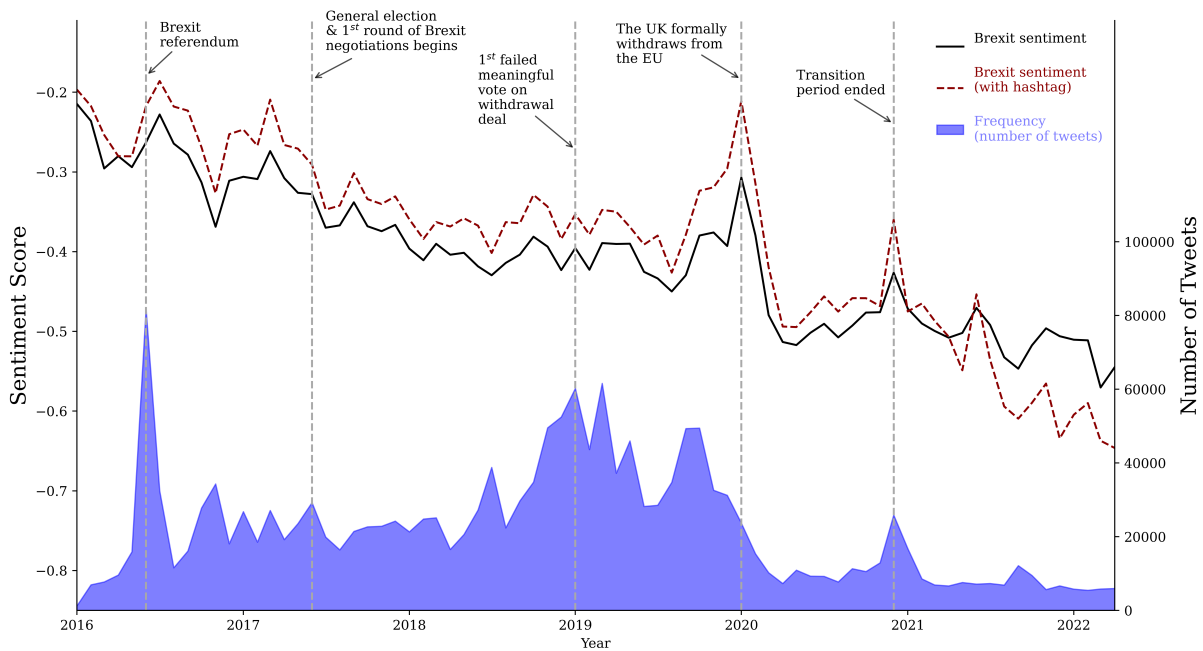
Next, we examine how public sentiment reacts to our BUIs by running the following regression:

$$sentiment_t = \alpha_0 + \alpha_1 BUI_t + \alpha_2 freq_t + \epsilon_t, \quad (1.1)$$

where  $sentiment_t$  is the Brexit sentiment score in month  $t$ ,  $BUI_t$  is a topic-specific BUI, and  $freq_t$  is the log of total number of tweets in a given month, as a time-varying control which allows us to control for total news coverage, and other potential concerns such as the degree

---

<sup>29</sup>It is worth mentioning that the existence of sarcasm, that is, when people express their negative sentiments using positive tones, may contribute to an upward bias of the sentiment score (Rajadesingan et al. 2015, Joshi et al. 2017).



**Figure 1.9:** Brexit sentiment, Brexit sentiment (with hashtag), and the number of Brexit-related tweets in each month. The Brexit sentiment score shows the monthly average sentiment score of all tweets containing “Brexit.” The Brexit sentiment (with hashtag) shows the monthly average sentiment score of all tweets with a hashtag that includes “Brexit” in it. The frequency shows the monthly counts of tweets that mention “Brexit” in each month.

of sarcasm. We focus on contemporaneous effects as Twitter users tend to react rapidly to world events and news (for instance, in Figure 1.9 we can see the highest total number of tweets recorded was precisely when the referendum happened in June 2016). Results are reported in Table 3.2.<sup>30</sup>

Columns (1) - (9) of Table 3.2 report results of specification (1.1) for different topic-specific BUIs. In Column (1), we can see that public sentiment reacts positively to macroeconomy-

<sup>30</sup>Results using lagged (past month) BUIs are qualitatively robust and available upon request. Including a year trend absorbs all our effects except for the results for supply chains and food. Our results should therefore be interpreted with caution, as this may imply that both BUIs and sentiment scores can be explained by some other observable annual trend.

**Table 1.3: Brexit Uncertainty and Brexit Sentiment**

Dep var:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
sentiment	Macro- economy	Trade policy	NI	Supply chain	Energy& climate	Fishing	Immi- gration	Employment	Food
BUI	0.067*** (0.009)	0.061*** (0.009)	-0.029*** (0.010)	-0.033*** (0.009)	0.005 (0.010)	-0.015 (0.009)	0.051*** (0.009)	0.053*** (0.011)	-0.016 (0.010)
lnfreq	-0.014 (0.012)	-0.007 (0.012)	0.055*** (0.013)	0.042*** (0.011)	0.037*** (0.013)	0.039*** (0.012)	0.008 (0.012)	-0.007 (0.014)	0.048*** (0.014)
Constant	-0.265** (0.114)	-0.336*** (0.116)	-0.944*** (0.128)	-0.813*** (0.112)	-0.766*** (0.125)	-0.783*** (0.120)	-0.486*** (0.115)	-0.342** (0.140)	-0.872*** (0.133)
Obs.	76	76	76	76	76	76	76	76	76
R-squared	0.503	0.455	0.210	0.263	0.120	0.145	0.385	0.329	0.145

Notes: The control variable, freq, is the log of total number of tweets in a given month. All dependent variables (BUIs) are standardised. Standard errors are reported in parentheses. \*\*\* indicates significance at the one percent level, \*\* indicates significance at the five percent level and \* indicates significance at the ten percent level. Results using all tweets with a Brexit hashtag are qualitatively robust.

related BUI. A one standard deviation increase in macroeconomy-related BUI is associated with a 0.067 increase in the sentiment score. The coefficient for trade policy-related BUI is also statistically significant and positive. An interpretation is that the positive sentiment is capturing a general feeling of optimism about the UK's future macroeconomic performance and independent trade policy during the Brexit negotiation process.

Similarly, Columns (7) and (8) suggest that immigration and employment-related BUIs are significantly and positively associated with Brexit sentiment. These two topics are closely related. While one might expect a negative reaction to the loss of skilled migrants (and loss of future skilled migrants), the possibility that this will open up more opportunities for locals may have triggered a positive reaction on Twitter. The issue is further complicated by tweets that welcome more immigration from outside of the EU or indeed, those that believe that Brexit does not go far enough to control immigration and therefore react negatively.

What is clear is that a large number of tweets express a negative view on the government but the tweet also mentions Brexit which further complicates interpretation. Hence, sentiment scores should be interpreted with caution (see Appendix D for example tweets to illustrate the different perspectives on immigration). In future research we hope to investigate the reasons for the positive immigration sentiment score by looking at the political leanings of the posters and how that impacts their tweet sentiment.

In contrast, negative and significant coefficients are reported in Columns (3) and (4) indicating that the public had a negative view on Northern Ireland issues (including the NI Protocol) and the future of supply chains. For NI, it is likely that both pro-Brexiters and remainers would see NI as a cause for concern that either undermines the Brexit process or was a reason why Brexit was a bad idea in the first place. There is no significant response to other BUIs such as energy & climate, fishing, and food. The results suggest that Brexit uncertainties matter for public sentiment towards Brexit, and that the direction and degree of impact varies across topics. Future research is needed to make a distinction between sentiment related to the topic (e.g. immigration) and that related to the government. For example, an individual tweet may be anti-immigration and pro-government or anti-government and pro-immigration both of which may be recorded as having a negative sentiment score.

## 1.8 Conclusions

The uncertainties triggered by the withdrawal of the UK from the EU are large and protracted. From the pre-referendum era to the current post-Brexit period the risks associated with the withdrawal have generated various uncertainties including, but not limited to, supply chains, employment, immigration, and the macroeconomy more generally. A major concern is that these uncertainties affect the economic environment in the UK, for example,

investment, trade, and employment. However, quantifying the degree of Brexit uncertainty is a challenge.

In this paper we adapt two ML techniques from computational linguistics to develop a novel news-based measure of aggregate and topic specific Brexit uncertainty. Our methodology allows us to consider all news pertaining to Brexit uncertainty and decompose the news by topic to obtain the time-series share of aggregate and topic-level uncertainty. We are also able to exclude the impact of the COVID pandemic to give us a lower-bound BUI index that is more closely aligned with the attitudes of businesses during the pandemic period. In this way, we can quickly and cost effectively measure the evolution of this uncertainty and its component topics over time and in close to real-time. An online tool enables us to extend the analysis presented in this paper quickly and efficiently so that policymakers can be kept up-to-date with how uncertainty continues to evolve.

Validation exercises demonstrate that our indices appear to be reasonable proxies for Brexit uncertainties. We find that our aggregate index is strongly correlated with both the time-series proportion of firms reporting Brexit as the largest source of uncertainty in the DMP survey and the original BBD-EPU index developed by [Baker et al. \(2016\)](#). Moreover, our UK EPU index constructed using our ML approaches shows a clear co-movements with the BBD-EPU index. Finally, our currency-related BUI and aggregate BUI follow fairly similar trends to indices of UK stock market volatility between the Brexit referendum and the COVID outbreak.

Our measure of Brexit uncertainty opens up a number of possible channels for future research. For example, we briefly show how our BUIs can be used in sentiment analysis to show how the general public react to different aspects of the Brexit process. Specifically, the ability to measure Brexit uncertainty should allow researchers to quantify the impact of Brexit across a range of different dimensions. Our approach may also be valuable for policy

makers who want to see how a particular event or policy affects different types of Brexit uncertainty or economic policy uncertainty more generally.

# A Appendix for Chapter 1

## A.1 Word sets

Table A.1 shows the words we use to filter out articles related to Brexit uncertainty from the mass of news. An article containing “Brexit” and at least one word from the “uncertainty” and “UK” sets is taken as Brexit uncertainty related articles. All words or word stems associated with “uncertainty” and “UK” are selected based on the outputs of the Word2Vec model.

**Table A.1:** Word Sets

<b>Brexit</b>	Brexit
<b>Uncertainty</b>	uncertain*; instab*; unstabl*; risk*; unpredict*; volatile*; unclear*; worry*; fear*; tension*; anxiet*; anxious; nervous*; jitter*; unsettl*; precar*; unknow*; indecis*; angst*
<b>UK</b>	UK (United Kingdom); British; Britain

## A.2 Why We Choose the LDA Model for Topic Modeling

In natural language processing, topic modeling is a statistical approach used to identify the underlying “topics” within a set of documents. The LDA model is one of the most mature and widely used topic model. In this section, we discuss why we apply the LDA model rather than other topic models.

We chose the LDA model over other techniques, such as semi-supervised models or Dynamic Topic Modeling (DTM), due to LDA’s robustness, simplicity, and effectiveness in managing large datasets without requiring prior knowledge. While semi-supervised models like GuidedLDA allow users to input “seed” words to guide topic formation, they presuppose knowledge of the topics and could bias the discovery process. Our goal is to automate the detection of Brexit-related topics.

In addition, Dynamic Topic Modeling (DTM), which tracks the evolution of topics over time, adds unnecessary complexity, as Brexit uncertainty-related topics are relatively stable. We also considered transformer-based models like BERTopic, which excel with short texts, such as social media posts or news headlines. However, BERTopic may not perform as well with longer texts, such as newspaper articles, due to the inherent characteristics of transformers, and it typically assigns only one topic per article. This single-topic assignment does not reflect the multifaceted nature of newspaper articles, which often discuss multiple topics simultaneously.

## A.3 LDA: Model and estimation

We discard the notation defined in the previous section. Suppose the corpus contains  $M$  documents in total, where  $N_m$  denotes the total number of words in document  $m$ . Let  $z_{m,n}$

be the topic for the  $n$ -th word in document  $m$ ,  $w_{m,n}$  be the  $n$ -th word in document  $m$ ,  $K$  be the number of latent topics, and  $V$  be the set of all unique words  $t$  in the corpus. The objects of the model are twofold: (a) to estimate the mixture component for each topic,  $\Phi = \{\varphi_k\}_{k=1}^K$  ( $K \times V$  matrix), in the form of probability distributions over  $V$  words; (b) to estimate the topic mixture proportion for each document,  $\Theta = \{\theta_m\}_{m=1}^M$  ( $M \times K$  matrix), in the form of probability distributions over  $K$  topics.  $\alpha$  and  $\beta$  are the pre-defined hyperparameters, which determine the prior weight of each topic in a document and prior weight of each word in a topic. In this section, we briefly describe the LDA model and the estimation method. For more detailed information and explanation, see [Griffiths & Steyvers \(2004\)](#) and [Heinrich \(2005\)](#).

### A.3.1 Model

As we mentioned in the previous section, the LDA model is a generative model: LDA generates all the observable words  $w_{m,n}$ , which in turn generate the whole corpus. With the notation we defined above, the probability that the word  $w_{m,n}$  is the term  $t$  can be expressed as:

$$p(w_{m,n} = t | \theta_m, \Phi; \alpha, \beta) = \sum_{k=1}^K p(w_{m,n} = t | \varphi_k) p(z_{m,n=k} | \theta_m)$$

Then, the generation process of the corpus  $W = \{w_m\}_{m=1}^M$  are:

$$p(W | \Theta, \Phi; \alpha, \beta) = \prod_{m=1}^M p(w_m | \theta_m, \Phi; \alpha, \beta) = \prod_{m=1}^M \prod_{n=1}^{N_m} p(w_{m,n} | \theta_m, \Phi; \alpha, \beta)$$

### A.3.2 Estimation

We apply the Gibbs sampling algorithm introduced by [Griffiths & Steyvers \(2004\)](#) to estimate the LDA model. Both target distributions  $\Phi$  and  $\Theta$  can be interpreted with the observed words  $w_{m,n}$  and their corresponding topics  $z_{m,n}$ . The goal of the inference could therefore be the distribution:

$$p(Z|W; \alpha, \beta) = \frac{p(Z|W; \alpha, \beta)}{p(W; \alpha, \beta)}$$

Gibbs sampling algorithm uses the full conditional  $p(z_i|Z_{-i}, W; \alpha, \beta)$  to simulate this distribution. To generate the full conditional, we first draft joint distribution, that is:

$$p(W, Z|\alpha, \beta) = p(W|Z, \beta)p(Z|\alpha)$$

The transition from the left to the right side of the equation relies on conditional independence. These two components can be processed separately, and the joint distribution can be written as:

$$p(Z, W|\alpha, \beta) = \prod_{z=1}^K \frac{\Delta(N_z + \beta)}{\Delta(\beta)} \prod_{m=1}^M \frac{\Delta(N_m + \alpha)}{\Delta(\alpha)}$$

where  $N_z = \{n_z^{(t)}\}_{t=1}^V$ , and  $N_m = \{n_m^{(k)}\}_{k=1}^K$ . Then, employing chain rule and letting word index be  $i = (m, n)$ , the full conditional distribution can be derived as:

$$\begin{aligned}
p(z_i = k | Z_{-i}, W) &= \frac{p(W, Z)}{p(W, Z_{-i})} = \frac{p(W|Z)}{p(W_{-i}|Z_{-i})p(w_i)} \cdot \frac{p(Z)}{p(Z_{-i})} \\
&\propto \frac{\Delta(N_z + \beta)}{\Delta(N_{z,-i} + \beta)} \cdot \frac{\Delta(N_m + \alpha)}{\Delta(N_{m,-i} + \alpha)} \\
&= \frac{\Gamma(n_k^{(t)} + \beta)\Gamma(\sum_{t=1}^V n_{k,-i}^{(t)} + \beta)}{\Gamma(n_{k,-i}^{(t)} + \beta)\Gamma(\sum_{t=1}^V n_k^{(t)} + \beta)} \cdot \frac{\Gamma(n_m^{(k)} + \alpha)\Gamma(\sum_{k=1}^K n_{m,-i}^{(k)} + \alpha)}{\Gamma(n_{m,-i}^{(k)} + \alpha)\Gamma(\sum_{k=1}^K n_m^{(k)} + \alpha)} \\
&= \frac{n_{k,-i}^{(t)} + \beta}{\sum_{t=1}^V n_{k,-i}^{(t)} + \beta} \cdot \frac{n_{m,-i}^{(k)} + \alpha}{[\sum_{k=1}^K n_m^{(k)} + \alpha] - 1} \\
&\propto \frac{n_{k,-i}^{(t)} + \beta}{\sum_{t=1}^V n_{k,-i}^{(t)} + \beta} (n_{m,-i}^{(k)} + \alpha)
\end{aligned}$$

Finally, the two target distributions can be estimated as:

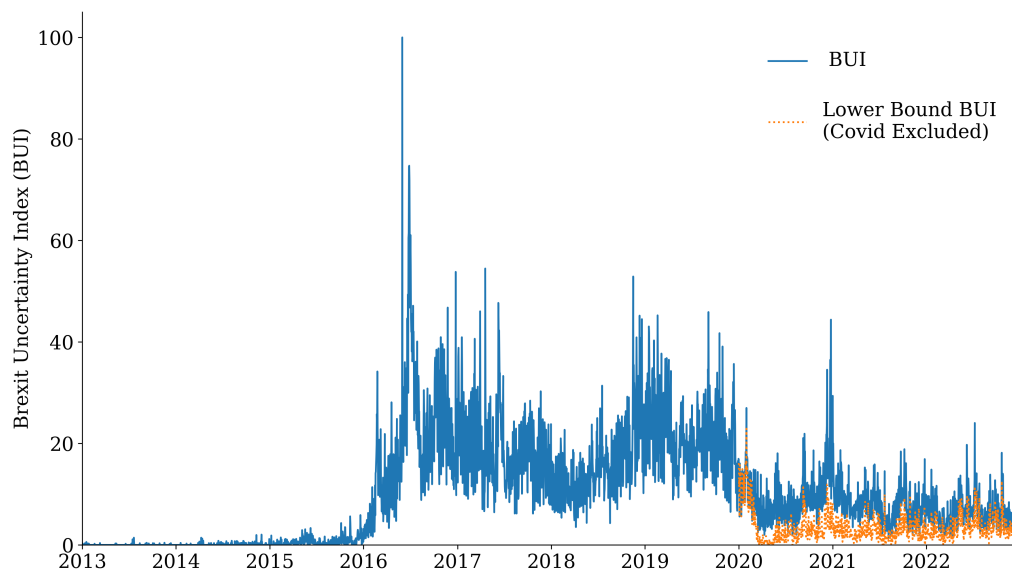
$$\hat{\theta}_{m,k} = \frac{n_m^{(k)} + \alpha}{\sum_{k=1}^K n_m^{(k)} + \alpha}$$

$$\hat{\varphi}_{k,t} = \frac{n_k^{(t)} + \beta}{\sum_{t=1}^V n_k^{(t)} + \beta}$$

where  $n_m^{(k)}$  denotes the number of times that topic  $k$  has been observed with a word in document  $m$ , and  $n_k^{(t)}$  refers to the number of times that term  $t$  has been observed with topic  $k$ .

## A.4 Daily Brexit Uncertainty Index

Figure A.1 shows the daily BUI and lower bound BUI (COVID excluded). The index is normalized to a maximum value of 100, as in our monthly BUI.



**Figure A.1:** Daily BUI and Lower Bound BUI (COVID Excluded).

## A.5 Correlation of Topics in Articles

Table [A.2](#) shows which topics appear together more frequently in the same article by calculating the correlations between topics in the topic distribution of the articles.

## A.6 Example Tweets of Brexit Topics and their Sentiment Scores

Below we list three sets of example tweets with extremely positive, extremely negative, and neutral sentiment scores, respectively. The brackets that follow each tweet show the topic(s) being discussed and the related sentiment score.

### Positive tweets:

*"@MartynBowler The 'Brexit' aim was to leave the EU That is now done. Therefore #brexit is already a success. We achieved our goal. Now our wages are increasing, our economy's growing, we continue to make trade deals around the globe. Could not be happier. I wish we could do it again!"* (Macroeconomy, trade policy, employment; Sentiment score: 0.991).

*"Many congratulations to @trussliz and @tradegovuk for agreeing a huge post-Brexit Free Trade Agreement with Japan. It will cover 99% of British exports to Japan and boost trade by over £15bn. The UK's new trade policy is really starting to take shape. Onwards and upwards!"* (Trade policy; Sentiment score: 0.991).

*"God dam best move ever. I'm a working class manual labor man. Before brexit I was competing against cheap labor from several eu country's. After brexit I got job security, a wage increase and more job opportunities.. Brexit is so so good for the working classes..."* (Employment; Sentiment score: 0.988).

*"@danielrhamilton @LauraLundahl Luckilyly Brexit means we can welcome more immigrants from outside the European Union. Great news!"* (Immigration; Sentiment score: 0.991).

Negative tweets:

*"Politics and Parliament are killing the UK economy! The working class are simply suffering from all this turmoil while all the upper class politicians wine and argue over matters! £££ droppings and the public suffer! #brexit"* (Macroeconomy; Sentiment score: -0.982).

*"Devastating for many many people . More food banks . More misery. UK retailers warn of 40% tariffs on food in no-deal Brexit"* (Trade policy, food; Sentiment score: -0.978).

*"Foster still pushing the economic catastrophe of no deal #Brexit on Northern Ireland. A deluded, reckless sociopath hell bent on getting a hard border at any cost. The mask has truly slipped."* (Northern Ireland; Sentiment score: -0.980).

*"I mean this "no deal will be great for Britain" stuff is indefensible IMO. It would be a catastrophe, killing complex supply chain businesses at a stroke, but also a catastrophe that causes huge volatility in the markets, and volatility = profit."* (Supply chain; Sentiment score: -0.967).

*"Tory denial continues. Brexit is unworkable fiasco. Every sector is under staffed, supply chains broken and no clue how to implement Brexit. Meanwhile staff shortages are killing people #JohnsonLiedPeopleDied #BrexitIsNotWorking"* (Supply chain, employment; Sentiment score: -0.980).

*"This rotten Tory government has failed. It has failed on the economy, on the climate crisis, on investment for the future, on public services and on Brexit."* (Energy & Climate; Sentiment score: -0.981).

*"@GMB The problem with immigration is #Brexit The government, particularly the lying #CrimeMinister, told everyone that #Brexit would solve immigration. Those with half a brain knew it would make it worse. Guess what ... it is worse! #GMB #ToryLiars"* (Immigration; Sentiment score: -0.981).

*"This is a flat lie. The food industry especially is massively damaged. As for fishing. I'm astounded at the bare faced mendacity Brexit apologists continue to apply; do you imagine we're all stupid?"* (Food, fishing; Sentiment score: -0.978).

*"Warnings of food shortages and price hikes from the food industry if there is no deal. How much more do we need before we end this whole sorry business? So tired of #brexit"* (Food; Sentiment score: -0.969).

#### Neutral tweets:

*"@Brexit @michelbarnier Northern Ireland will claim back it's fishing waters, 8% of Eire's economy. Tariffs to cross UK to EU carrying goods"* (Northern Ireland, fishing; Sentiment score: 0.008).

*"Personally feel that any Brexit supporter who now needs immigrant workers should be required to hire from the third world."* (Immigration; Sentiment score: -0.005).

*"So, currently you can buy and bring back unlimited cheaper (tax included) booze, and after Brexit you will be limited to 1 litre tax free. How is this beneficial?"* (Trade policy; Sentiment score: 0.003).

*"#EUreferendum #Remain need to offer vision- exciting future of tech innovation & fighting climate change, #Brexit would leave us stuck in past"* (Energy & climate; Sentiment score: 0.006).

*"If #Brexit gets delayed till end June, gives plenty of time for manufacturing businesses to secure new premises, sign leases and leave U.K. "* (Manufacturing; Sentiment score: 0.002).

**Table A.2:** Correlation of Topics in Articles

<b>Topic</b>	<b>Associated topics in the same article</b>
Government Spending & Budget	Tax (0.106)
Tax	Government Spending & Budget (0.106), Housing Price (0.029)
Currency	N/A
Trade Policy	Food (0.058), Supply Chain (0.022)
Housing Price	Tax (0.029)
Northern Ireland	Supply Chain (0.021)
Supply Chain	Food (0.050), Employment (0.027), Trade Policy (0.022)
Energy & Climate	Food (0.047), Manufacturing (0.020)
Immigration	Employment (0.079)
Employment	Immigration (0.079), Supply Chain (0.027)
Food Industry	Trade Policy (0.058), Supply Chain (0.050), Energy & Climate (0.047)
Manufacturing	Energy & Climate (0.020)
Scotland	N/A

Notes: Table A.2 shows the top three other associated topics with a correlation of 0.02 or higher for each topic. Listing less than three topics indicates that no more than three associated topics meet the criteria; N/A indicates that no associated topic meets the criteria. Correlation coefficients are in parentheses.

## Chapter Two

# A Picture is Worth a Thousand Words: An Application of Machine Learning and Computer Vision Image Processing Applied to the Brexit Referendum

## Abstract

In this paper, we introduce an automated method to quantify visual media bias in images from online news articles during the 2016 Brexit referendum. Using computer vision and machine learning techniques, we analyze a dataset of 64,089 images from seven leading UK newspapers, distinguishing between front-page and content-page images, as well as between tabloids and broadsheets. Our findings reveal significant partisan visual media bias in tabloid newspapers. Specifically, pro-Leave tabloids were more likely to depict pro-Leave politicians with positive facial expressions and in favorable visual contexts, while portraying pro-Remain politicians more negatively. This bias was especially pronounced in front-page images and among a few key politicians, with the exception of Boris Johnson, who was consistently depicted in a negative light across newspapers, without reflecting partisan bias. As the Brexit referendum approached, media became increasingly cautious in displaying clear partisan biases, resulting in a reduction of such bias leading up to the referendum. Following the referendum, the bias diminished. In contrast, no comparable bias was detected in broadsheet newspapers. Our study contributes to the literature by offering one of the first automated analyses of visual partisan bias, introducing a scalable approach for examining both facial expressions and contextual sentiments, and providing the first empirical evidence of visual media bias in UK newspapers during the Brexit referendum.

JEL Codes: L82, D91, D83.

**Keywords:** Visual media bias; visual analysis; machine learning.

## 2.1 Introduction

Visual content has become increasingly prevalent in the current digital media landscape and plays a crucial role in political communication (Verser & Wicks 2006, Grabe & Bucy 2009). Compared to verbal information, visuals tend to be more memorable and more impactful in shaping opinions (Sullivan & Masters 1988, Graber 1990, 1996) as they can engage users' heuristic processing systems (Schmuck & Matthes 2017). Attributes such as facial expressions can significantly shape our impressions of politicians and affect voting behaviors (Tiedens 2001, Todorov et al. 2005, Mutz 2007). To attract specific audiences, media outlets increasingly disclose their political affiliations (Groeling 2013), thereby introducing bias. Media bias can be conveyed through visual content (Barrett & Barrington 2005). However, due to the complexity of image analysis and the high costs associated with manually coding large volumes of images, existing research on media bias has focused almost exclusively on text-based bias (e.g., Groseclose & Milyo 2005, Gentzkow & Shapiro 2010, Groeling 2013, Martin & Yurukoglu 2017).

The purpose of this paper is to use computer vision and machine learning techniques to quantify visual media bias embedded in images from online news articles. More specifically, we look explicitly at seven leading media outlets in the UK during the 2016 Brexit referendum. Our analysis is based on 64,089 images collected from these UK newspaper websites from 21 February to 31 August 2016, which covers the period before and after the June 23 referendum date. We aim to answer the question of whether newspapers portray politicians who align with their position in a more favorable light, in terms of more favorable facial expressions and in a more positive context, to see if images were used to influence the voting during the referendum. Media bias can be defined as the systematic and significant distortion of reality by the media in its portrayal of reality (Groeling 2013). This broad concept encompasses various aspects of bias, such as the portrayals of race, gender, geography, or social class. In this paper, we narrow the focus to a specific form, that is, partisan bias,

which involves persistent and clear favoritism towards one political group or ideology over its opposition in news coverage (Waldman & Devitt 1998, Groeling 2013, Peng 2018).

Newspapers in the UK are generally divided into two groups: Broadsheets, known as the ‘quality press’ for their serious, intellectual content and in-depth political reporting, and tabloids, known as the ‘popular press,’ which focus on more attention-grabbing content, often highlighting celebrity news and human interest stories. Tabloids frequently use sensational language and selective reporting, which can shape public perception in a biased way (Chadwick et al. 2018). Given this context, we expect tabloids to be more likely to exhibit media bias.

In this study, we distinguish between front-page and content-page images. A front-page image is the primary picture accompanying a news article, displayed in the preview on news websites and when shared on social media, allowing readers to see it even if they do not view the full article. On the other hand, content-page images are less prominent and are used to support detailed stories within the newspaper. A number of previous studies focused exclusively on front-page images and found notable nonverbal media bias during the 2016 and 2020 US elections (Boxell 2021, Caprini 2023). Peng (2018) included both kinds of images, but did not distinguish between them. To better understand visual media bias, we separately analyze front-page and content-page images, acknowledging that, while we consider both, front-page images are often more strategically used for propaganda purposes and therefore have a greater potential to convey bias.

This paper makes three distinct contributions. First, this paper introduces one of the first automated analyses of visual partisan bias. Most existing studies of nonverbal bias utilize manual coding to analyze visual information (Waldman & Devitt 1998, Barrett & Barrington 2005, Hehman et al. 2012). Peng (2018) is the first study to use automated computer vision tools, followed by Boxell (2021), both focusing solely on analyzing the

faces of politicians. In contrast, our approach considers both the politician’s face and their contextual images using automated algorithms. Furthermore, unlike [Caprini \(2023\)](#), who analyze contextual information in images utilizing textual tagging of the factors in images such as objects and colors, we extend the framework by incorporating sentiment analysis to evaluate the emotional tone conveyed in contextual images.

Second, previous research relies heavily on third-party application programming interfaces (APIs) for visual content analysis ([Peng 2018](#), [Boxell 2021](#), [Caprini 2023](#)). To the best of our knowledge, existing commercially available APIs do not provide the function to automatically categorize image sentiment. To address this gap, we develop a deep learning model based on a Vision Transformer (ViT) architecture to measure contextual image sentiment. ViTs are well-suited for image analysis as they can effectively capture patterns and contextual relationships in visual data ([Dosovitskiy et al. 2020](#)).

Third, this paper presents the first evidence of visual media bias in UK media. It is widely documented that British newspapers exhibit strong political leanings ([Brandenburg 2006](#)). For example, following the 2016 Brexit referendum, researchers started to investigate the UK press coverage of the Brexit vote, examining a range of factors including the main arguments, the tone of news articles, and specific topics of coverage ([Levy et al. 2016](#), [Deacon, Wring, Harmer, Downey & Stanyer 2016](#), [Aftab 2018](#)). However, prior studies focus exclusively on textual contents of media. This paper contributes to the literature on the UK media and the Brexit referendum by summarizing possible bias embedded in the visual content of UK newspapers.

Our methodological approach is to use computer vision and machine learning techniques to recognize and locate politicians from images, to perform emotion analysis on the politician’s facial expressions, and to perform sentiment analysis on the contextual settings. For example, to measure media bias in facial expressions of politicians, we rely on the Ama-

zon Rekognition API, a cloud-based visual information analysis tool provided by Amazon, to detect faces from images, recognize politicians, and analyze the emotions of their facial expressions. Using the API, we extract 19,288 faces of 27 politicians and classify each face into eight emotion categories: ‘happy,’ ‘calm,’ ‘surprised,’ ‘confused,’ ‘disgusted,’ ‘angry,’ ‘fear,’ and ‘sad.’

To measure media bias embedded in contextual images, we leverage a pre-trained Vision Transformer (ViT) model (Dosovitskiy et al. 2020). Although the model is not explicitly designed for sentiment analysis, it possesses extensive domain knowledge in image recognition. To adapt this model for image sentiment classification, we employ transfer learning. This process involves fine-tuning the pre-trained model with new image datasets that are explicitly labeled by sentiment. We modify the model’s architecture by replacing its original classification head with a new one containing only two target sentiment categories, positive and negative. The model evaluates various elements of the image, such as objects, colors, gestures, and facial expressions, to make predictions. The trained model achieves 90.3% accuracy on the test dataset. Next, we use the model to make predictions on the sentiments conveyed by the contextual images of the politicians. The contextual images are constructed by blurring the politician’s face to exclude the facial information.

Our main results can be summarized as follows. We identify clear evidence of partisan visual media bias in how politicians were visually depicted and the contextual settings in which they were portrayed, particularly in tabloid newspapers. In tabloids, when covering pro-Leave politicians, pro-Leave newspapers were less likely to display them with negative emotions (e.g., angry and confused) and more likely to display them with positive emotions, compared to pro-Remain newspapers. The bias is especially pronounced in front-page images and was concentrated among a select group of key politicians, with the notable exception of Boris Johnson, who was consistently portrayed negatively across both newspaper types. Moreover, the visual media bias was mainly reflected in the coverage of pro-Leave politicians,

with the portrayal of pro-Remain politicians consistently having similar degrees of favorability in the two positions of the newspapers over the voting cycle. Immediately after the vote, this bias disappeared. In broadsheets, however, no such bias was detected.

Our results for contextual images were similar. In tabloids, when covering pro-Leave politicians, pro-Leave newspapers were statistically significantly more inclined to place them in a more positive context, compared to pro-Remain newspapers. This bias was observed only in the front-page images, with no similar bias found in the content-page images. In broadsheets, again, no such bias was found. In general, these results suggest that partisan tabloid media conveyed bias during the 2016 Brexit vote by using relatively more negative emotions to depict opposing politicians and by placing these politicians in relatively negative image contexts.

The remainder of the paper proceeds as follows. Section 2 introduces background. Section 3 describes the relevant literature. Section 4 introduces data and computer vision methods for measuring media bias in facial expressions. Section 5 discusses the media bias in facial expressions. Section 6 discusses visual media bias from contextual images. Section 7 concludes.

## 2.2 Background

The Brexit referendum held on June 23, 2016, narrowly decided that the UK should leave the European Union (EU), with 51.9% voting to leave and 48.1% voting to remain. Since joining the European Economic Community (EEC) in 1973, concerns have emerged over issues such as loss of national sovereignty, increasing EU regulations, and limited control over immigration. Many argued that EU membership restricted the ability of the UK to independently manage main areas such as trade, immigration, and domestic policy.

The financial crisis in 2008 and the subsequent Eurozone crisis exacerbated these concerns, even though the UK is not part of the Eurozone. Many believe that EU economic policies are exacerbating economic issues across Europe. In addition, the expansion of the EU in the 2000s, which allowed Eastern European citizens to migrate freely to the UK, has sparked debates about immigration, competition for jobs, and pressure on public services.

In response to electoral pressure from the UK Independence Party (UKIP), led by Nigel Farage, which advocated for leaving the EU and thus appealed to many of the Eurosceptic voters, Prime Minister David Cameron, who supported remaining in the EU, announced in January 2013 that he would hold a referendum on EU membership if the Conservative Party won the 2015 general election. The strategy of David Cameron relied on a favorable renegotiation of the terms of British membership within the EU, followed by a public vote that he expected pro-Remain to win.

The referendum was officially scheduled after the Conservative win in the 2015 general election. Two main campaign groups emerged: (1) Vote Leave advocated for the UK's withdrawal from the EU, and (2) Britain Stronger in Europe supported remaining in the EU. The campaign highlighted issues of national sovereignty, immigration, and the economic costs and benefits of being a member of the EU, with Leave supporters emphasizing the need to reclaim control over borders and laws, while Remain supporters warned of the economic risks associated with leaving the EU's single market.

On February 21, 2016, Boris Johnson, then Mayor of London, announced his support for the Leave campaign, marking a pivotal moment that brought a high-profile figure to the pro-Leave side of the Brexit debate. Prior to this, Johnson's stance on Brexit had been ambiguous. Many saw his decision to support Brexit as a strategic move for his future leadership role within the Conservative Party. His endorsement provided a substantial boost to the Leave campaign, giving it both credibility and momentum, which it had previously

lacked due to the absence of prominent leadership.

The official campaign for the Brexit referendum was launched on April 15, 2016, after the Electoral Commission designated Vote Leave and Britain Stronger in Europe as the official representatives of the Leave and Remain campaigns, respectively. The referendum debate quickly intensified, with sharp contrasts emerging between the two sides. The Leave campaign, encapsulated by the slogan “Take Back Control,” centered its arguments on restoring UK sovereignty, controlling immigration, and reducing the influence of EU regulations. In contrast, the Remain campaign emphasized the economic risks of leaving the EU, particularly highlighting the uncertainty surrounding future trade agreements, financial stability, and jobs. This period saw increasingly polarizing rhetoric, as both sides aimed to sway public opinion in a vote that would have profound and lasting consequences for the UK.

On June 23, 2016, the Leave campaign won. This result led to significant political consequences, starting with the resignation of David Cameron on June 24, 2016. Although Boris Johnson was a prominent figure in the Leave campaign, he declined to run for the leadership of the Conservative Party shortly thereafter. Instead, Theresa May, a pro-Remainer, became the new leader of the Conservative Party and on July 13, 2016, took office as Prime Minister. Table 2.1 documents the main events during the Brexit vote, starting with Prime Minister David Cameron’s initial announcement of support for the Brexit referendum in January 2013.

### **2.2.1 UK Newspapers: Broadsheets and Tabloids**

UK newspapers are typically categorized into two groups: (1) Broadsheets, which are known for their serious and intellectual content and are collectively referred to as the ‘quality press,’

**Table 2.1:** Brexit Vote Main Events

23 January 2013	Prime Minister David Cameron announces his backing for the EU referendum
9 October 2015	The “Vote Leave” campaign group is launched
12 October 2015	The “Britain Stronger in Europe” campaign group is launched
4 January 2016	Prime Minister David Cameron confirms that his ministers would be allowed to support and campaign for both the Leave and Remain sides
17-18 February 2016	David Cameron gets a deal at the EU summit, granting the UK special status within the EU, including reforms on migration and economic governance
20 February 2016	Prime Minister David Cameron declares the referendum date and that he will campaign for Britain to remain
20 February 2016	Cabinet splits: Six cabinet ministers announce their support for Vote Leave and will campaign for Brexit
21 February 2016	Boris Johnson declares his intention to campaign for Brexit
13 April 2016	The Electoral Commission designates “Vote Leave” (advocating for leaving the EU) and “Britain Stronger in Europe” (advocating for remaining in the EU) as the official campaign groups for the “Leave” and “Remain”
15 April 2016	The official campaign period starts.
23 June 2016	Brexit Referendum data. A turnout of 72.2 percent of registered voters voted, of which 51.9 percent voted Leave and 48.1 percent voted Remain. The UK will leave the EU
24 June 2016	Prime Minister David Cameron resigns
30 June 2016	Boris Johnson announces he will not run for Conservative leader
11 July 2016	Theresa May becomes Conservative leader
13 July 2016	Theresa May serves as the new Prime Minister

Notes: This table presents the timeline of main events during the Brexit vote.

and (2) tabloids, also known as the ‘popular press.’ Broadsheets primarily cover in-depth political reporting and international news, whereas tabloids often emphasize celebrity news and human interest stories.

Broadsheets are widely regarded as the cornerstone of traditional journalism, emphasizing nuanced reporting on major issues. They typically feature comprehensive articles on politics, economics, culture, and global affairs written by experienced journalists and subject matter experts. Newspapers such as *The Financial Times*, *The Guardian*, and *The Daily Telegraph* are known for their investigative journalism and analytical articles that provide insight into current events. The larger format of the broadsheets, with longer articles and detailed reporting, appeals to readers who value in-depth, well-researched information.

On the other hand, tabloids appeal to audiences by offering content that is more accessible and visually appealing. Newspapers such as *The Sun*, *The Daily Mirror*, *The Daily Mail*, and *The Express* are known for their eye-catching headlines, vivid images, and strong focus on entertainment news. The front pages usually feature celebrity gossip, sports, and human interest stories, catering to readers who enjoy light, easy-to-read news. The smaller format of tabloids allows for casual reading, and their informal, straightforward writing style resonates with a wider audience.

## 2.3 Related Literature

In this section, we begin by describing computer vision techniques and the studies that have used them in economics and the social sciences. We then present the literature on visual media bias and research that addresses media bias in the UK.

### 2.3.1 Image as Data with Computer Vision Techniques

Although both text and image data have been used in economics and social science research, images are easier to understand as there is no language barrier (Jurgenson 2019). However, there are only a few studies that utilize images as data compared to those that utilize textual data. One reason for the lack of previous research is the challenges associated with effectively retrieving and analyzing images and also because of the lack of easy-to-use tools (Ghermandi & Sinclair 2019). Recently, thanks to the rapid development of machine learning and computer vision technologies, automated visual content analysis is becoming more accessible and is replacing time-consuming and labor-intensive manual coding (see, for example, Ashkezari-Toussi et al. (2019) and Koylu et al. (2019)). In order to better understand the image data, we show in Table 2.2 how it compares to the text data.

**Table 2.2:** Comparisons between Text and Image Data

<b>Text</b>	<b>Image</b>
Linear: a sequence of words	Spatial: a matrix of pixels
Low ambiguity at the word level	High ambiguity at all levels
Compact size; easy to transfer and store	Larger size
Fixed vocabulary	Undefined vocabulary
Language-dependent	Universal applicability
Logical in nature	Evokes emotions

**Note:** This table presents the comparisons between textual and image data.

Computer vision is a field of artificial intelligence that uses machine learning algorithms to understand images and perform a range of tasks. For example, image classification is the computer vision task that involves automatically assigning category labels to images, such as categorizing images as containing happy, sad, or neutral emotions (Gorin et al. 2023), as well as face recognition (Boxell 2021). In object detection tasks, techniques can help iden-

tify and localize specific objects in an image, such as detecting a face in a photograph (Peng 2018). In image captioning tasks, a textual description of the contents of the image can be automatically generated.

From a technical perspective, computer vision techniques rely heavily on Convolutional Neural Networks (CNNs), a deep learning architecture that has become the most powerful algorithm for computer vision since Krizhevsky et al. (2012). More recent techniques include the Vision Transformer (ViT) (Dosovitskiy et al. 2020) and Swin Transformer (Liu et al. 2021), deep neural networks based on the Transformer architecture. The Transformer was first proposed by Vaswani et al. (2017), and its best-known applications are a series of generative large language models such as ChatGPT and Gemini. So far, the choice between CNNs and Transformer-based models depends on the specific task at hand, as no single model demonstrates a universal advantage across all visual tasks. For example, Transformer-based models may be preferable when handling large-scale datasets; on the other hand, CNNs tend to be more efficient and perform better on tasks with smaller datasets due to their lower computational requirements.

Image data that can be used in economics and social science studies includes, but is not limited to, images from news, social media, and books (e.g., Obaid & Pukthuanthong 2022, Adukia et al. 2023), satellite data (Donaldson & Storeygard 2016), maps (De Bellefon et al. 2021), images from videos (Hu & Ma 2021), and paintings (Gorin et al. 2023). For example, Adukia et al. (2023) examine the evolution of gender and race of characters in children’s books over the last 100 years and found that women and Black individuals were underrepresented, but the representation gradually increased over time. Obaid & Pukthuanthong (2022) use a computer vision model to classify the sentiment of news photos and propose that news photo sentiment can predict stock market prices and trading volume, after controlling for the sentiment of the text. Orellana et al. (2024) use computer vision techniques to analyze the faces depicted in advertisement images and find significant differences between them

and a wide range of demographic characteristics, as evidenced by a lower body mass index, especially for women. [Gorin et al. \(2023\)](#) innovatively measure the emotions contained in paintings and use the emotions to indicate historical economic activity, finding that painting emotion scores can predict historical trade openness, GDP per capita, and democracy. Using demographic data, including profile images of Trump and Clinton followers on X (Twitter) during the 2016 US election, [Wang et al. \(2016\)](#) found that Trump supporters were more polarized in terms of their social impact, while Clinton supporters had more racial diversity.

[Haim & Jungblut \(2021\)](#) investigate visual political communication during the 2019 European Parliamentary elections by examining candidate imagery on social media and in news outlets across all 28 EU member states. They find that while candidates' self-presentations on social media platforms tend to include more smiling, news outlets portrayed candidates with less smiling and used more varied camera angles, including bird's-eye and close-up shots. This suggests that news media tend to present politicians more neutrally or negatively, in contrast to the more favorable self-depictions seen on social media. [Glaeser et al. \(2018\)](#) studied images of houses obtained from Google Street View and argued that improvements in the appearance of houses significantly increase housing prices. [Naik et al. \(2016\)](#) also utilized Google Street View images and found that perceived safety from the street views was positively associated with city population density and household income. In urban economics, [De Bellefon et al. \(2021\)](#) use maps of buildings to segment urban areas. [Hu & Ma \(2021\)](#) further analyze videos of startups seeking funding from investors on three dimensions: visual, audio, and verbal, and find that positive presentations increase the chances of receiving investment.

### 2.3.2 Visual Media Bias

Previous research has tended to rely on manual annotation to identify visual media bias. This can be organized in two ways. The first is the so-called checklist method, a relatively objective annotation, which involves labeling based on some features that are pre-defined by the designers, such as facial expressions, gestures, and contextual information (e.g., other objects and people in the same image), and structural characteristics including angles at which the photos are taken and the color of the photos (Moriarty & Popovich 1991, Waldman & Devitt 1998, Verser & Wicks 2006, Grabe & Bucy 2009).

Specifically, Waldman & Devitt (1998) focused on the 1996 US presidential campaign and analyzed newspaper images of the two presidential candidates, Clinton and Dole, across five dimensions: candidates' expression, activity, their interaction with other people and objects, contextual information of the image, and camera angle. They found clear evidence of bias, and the degree of bias varied between newspapers. Verser & Wicks (2006) labeled presidential candidate images during the 2000 US election on 28 attributes including pose, interaction with people (including approachability and friendliness), posture, distance from others, image condition (including color), and camera angle. They found that each candidate carefully selected campaign photos to portray a positive appearance. In general, this method provides insights into how media bias manifests in particular visual representations. However, as Barrett & Barrington (2005) points out, the effects of various visual attributes on image favorability can be uneven. It is also uncertain if these research-selected features actually affect the perceptions of the readers (Grabe & Bucy 2009).

Other studies have taken relatively subjective approaches, namely, directly rating the positivity of images. For example, Barrett & Barrington (2005) collected hundreds of photos of US political candidates, and each image was rated by three coders on its favorability. High favorability usually implies that the candidates have positive facial expressions or have

active supporters in the background. Candidates in low favorability photos usually have confused or negative facial expressions and awkward postures. They found that in newspapers, candidates who usually aligned with the political leanings of the newspaper had more favorable photos than candidates of the opposite stance. [Hehman et al. \(2012\)](#) coded photos of political figures in terms of warmth and competence and found that the press tends to portray politically aligned politicians more positively on these two dimensions. However, the drawbacks of manual annotation are obvious. It includes limited scalability due to the time-consuming and labor-intensive nature of the task and susceptibility to human bias, which can affect the objectivity of the annotations.

[Peng \(2018\)](#) addressed this concern by leveraging recently developed computer vision techniques to automatically analyze how visual attributes in media coverage can influence audience perceptions and reveal underlying partisan biases. Specifically, they used the Face++ API to analyze the politicians' facial features, including facial expressions, face size, and skin condition. The study found that media across political ideologies characterized the 2016 US presidential candidates variously in terms of a variety of facial features. Similarly, [Boxell \(2021\)](#) used computer vision techniques, specifically Microsoft's Face API, to investigate nonverbal media bias during the 2016 election by analyzing nearly one million images from news websites. The study found that news websites displayed more images of politicians who aligned with their stances and often portrayed them with more positive facial expressions than opposing politicians. [Ash et al. \(2021\)](#) analyzed over two million articles from prominent US news outlets with machine learning algorithms, revealing disparities in how gender and ethnic groups are depicted. They showed that men and White individuals are disproportionately represented, while women and Hispanics are less frequently portrayed. These studies focused exclusively on facial features to examine visual media bias, implying that media outlets of different political stances use different visual languages.

[Caprini \(2023\)](#) further analyzed information on contents in news leading images, in-

cluding faces and contextual messages. They used Microsoft APIs to obtain text tags corresponding to faces and general pictures (e.g., objects and colors). In their study, observers were asked to guess the political stance of an image based on these tags. The probability of correctly guessing the political stance of an image was significantly higher than random chance, which implies that media outlets of different political stances use distinct visual languages. Our paper contributes to this literature by extending the framework for analyzing visual media bias from a sentiment analysis perspective. Specifically, we encode perceptual attributes inherent in images, i.e., the sentiments that humans intuitively perceive as conveyed by visual content.

### 2.3.3 Media Bias in the UK

Studies on bias in British media outlets have mainly focused on textual analysis. Using news coverage during the 2005 general election campaign, [Brandenburg \(2006\)](#) found from a quantitative study of news-related text that British newspapers were clearly partisan, with the Labour Party being the biggest beneficiary of media bias. [Ward \(2017\)](#) analyzed news coverage during the 2010 election and found that Black, Asian, and Minority Ethnic (BAME) female candidates were reported more negatively compared to their White counterparts.

The positions taken by UK newspapers with regard to the Brexit referendum varied greatly between newspapers. [Levy et al. \(2016\)](#) looked at how the mainstream press in the UK covered the EU referendum and showed that Brexit-related news coverage primarily centered on politicians and campaign spokespeople. They showed that the Daily Mirror had the highest percentage of pro-Remain articles, followed by The Guardian and the Financial Times, while the Daily Express had the highest percentage of pro-Leave articles, followed by the Daily Mail, The Sun, and the Daily Telegraph.

Deacon, Downey, Harmer, Stanyer & Wring (2016) summarized that news coverage in the UK is heavily polarized, with pro-Leave newspapers highly concentrating on pro-Leave arguments and vice versa. News coverage mainly favored the Leave campaign, with a 6:4 ratio of pro-Leave to pro-Remain coverage, a ratio that widened further to 8:2 after circulation was taken into account. Simpson & Startin (2023) focused exclusively on the British tabloids and argued that there is far more coverage of pro-Leave than pro-Remain, and in particular, there is a great deal of space allocated to anti-immigrant issues.

## 2.4 Data and Methodology

### 2.4.1 News Data

Our empirical analysis uses politician images from four pro-Leave UK newspapers, *The Daily Mail*, *The Express*, *The Sun*, and *The Telegraph*, and three pro-Remain UK newspapers, *The Mirror*, *The Guardian*, and *Financial Times*.<sup>1</sup> The positions and partisanship scores of these newspapers during the referendum come from Levy et al. (2016). We collected news articles and associated images published between 21 February 2016 and 31 August 2016 that included the terms “Referendum” and/or “Brexit”.<sup>2</sup>

A major challenge in data collection is that most existing online news databases, such

---

<sup>1</sup>We do not include *The Times*, given its ambiguous position. While it carried more pro-Leave news articles, according to Levy et al. (2016), it ultimately endorsed Remain about a week before the EU referendum.

<sup>2</sup>Our sample period starts on 21<sup>st</sup> February 2016, the day Boris Johnson announced his support for Brexit. This came one day after Prime Minister David Cameron officially set the date for the referendum and declared that he would campaign for Remain. Moreover, in Boxell (2021), a roughly 3:1 ratio was used for the period before and after the US election day in his study on visual media bias. We therefore select August 31 as the end of our sample period to maintain a similar pre- and post-referendum time ratio for our analysis.

as Nexis and Factiva, include only textual content and do not incorporate news images. To address this, we built web scrapers for each news website. We do not utilize some news aggregation services, for example, Google News, as they only include a limited selection of news articles from each newspaper. More specifically, we first obtained links to the news articles of interest. For *Financial Times* and *The Daily Mail*, we searched their websites directly using relevant keywords and recorded links to news articles that include “Referendum” and/or “Brexit”. Since keyword search results on the remaining news websites do not extend back to 2016, we first identified the titles of the qualifying news articles in the Nexis news database using the keyword searches. We then conducted Google searches using a combination of the newspaper name and the news title to find the corresponding news links. For example, we searched for ‘Express’ along with the title of a specific *Express* news article. We then recorded the links on the search results page that contain ‘www.express.co.uk.’

Following the links to the news articles, we collected information about these articles, including the time of publication and the links to the images. We then downloaded the images using the collected image links. The result was that we collected a total sample of 64,089 images.

## 2.4.2 Measuring Facial Expression Emotions

Our facial analysis consists of three tasks: detecting faces from images, recognizing faces (e.g., identifying whether the face belongs to Boris Johnson), and analyzing facial emotions. The Amazon Rekognition API, a paid service provided by Amazon, can complete all three tasks by leveraging computer vision and deep learning algorithms. These algorithms are trained by Amazon using millions of manually labeled data. We opted for the API over other open-source algorithms to allow for easy replication of the results (Peng 2018).

The first step is to upload a set of manually labeled politician images to the Amazon Web Services (AWS) Simple Storage Service (S3). Our dataset intentionally includes diverse images of each politician, capturing variations in facial angles, lighting conditions, and image resolutions. Image diversity ensures that the model can effectively recognize these politicians in various real-world environments. An 80% confidence threshold is used in the face recognition task; that is, the system must be more than 80% confident that the detected face matches a known politician before we consider the face matching to be valid.

The Amazon Rekognition system analyzes each face across eight dimensions of emotion, including ‘happy,’ ‘calm,’ ‘surprised,’ ‘confused,’ ‘disgusted,’ ‘angry,’ ‘fear,’ and ‘sad.’ Previous studies have used Microsoft’s Face API for facial emotion detection (Boxell 2021, Caprini 2023). However, Microsoft retired the facial emotion recognition feature in 2022 due to concerns that it could lead to stereotyping, discrimination, or unfairness.<sup>3</sup> Microsoft’s solution was similar in that it also included eight emotion categories; the difference was that Microsoft did not include ‘calm’ and ‘confused’ emotions but instead included ‘contempt’ and ‘neutral.’

Another key difference between Microsoft’s and Amazon’s emotion recognition APIs is that Microsoft generates emotion scores that sum to approximately 100 (Boxell 2021), indicating that while a face can show a mix of emotions, these emotions are mutually exclusive to some extent. In contrast, Amazon treats emotions as non-mutually exclusive, allowing a face to exhibit multiple strong emotions simultaneously, with scores for different emotions not constrained to sum to 100. From a machine learning perspective, the Microsoft API functions as a classifier with eight emotion outputs, whereas the Amazon API operates as multiple classifiers, each dedicated to a specific emotion category.

Figure 2.1 provides examples of the outputs from the Amazon Rekognition API. For

---

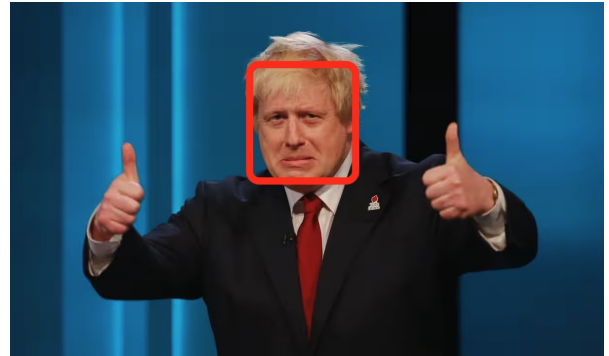
<sup>3</sup>For more details on Microsoft’s decision to discontinue its facial emotion recognition capabilities, see <https://azure.microsoft.com/en-us/blog/responsible-ai-investments-and-safeguards-for-facial-recognition/>.

**Figure 2.1:** Example Outputs of the Amazon Rekognition API



(a) Face recognition: David Cameron

- *Happy*: 100.0
- *Surprised*: 0.0
- *Angry*: 0.0
- *Confused*: 0.0
- *Disgusted*: 0.0
- *Fear*: 0.0
- *Calm*: 0.0
- *Sad*: 0.0



(b) Face recognition: Boris Johnson

- *Sad*: 69.4
- *Confused*: 54.4
- *Disgusted*: 1.5
- *Angry*: 0.9
- *Calm*: 0.1
- *Fear*: 0.1
- *Surprised*: 0.0
- *Happy*: 0.0

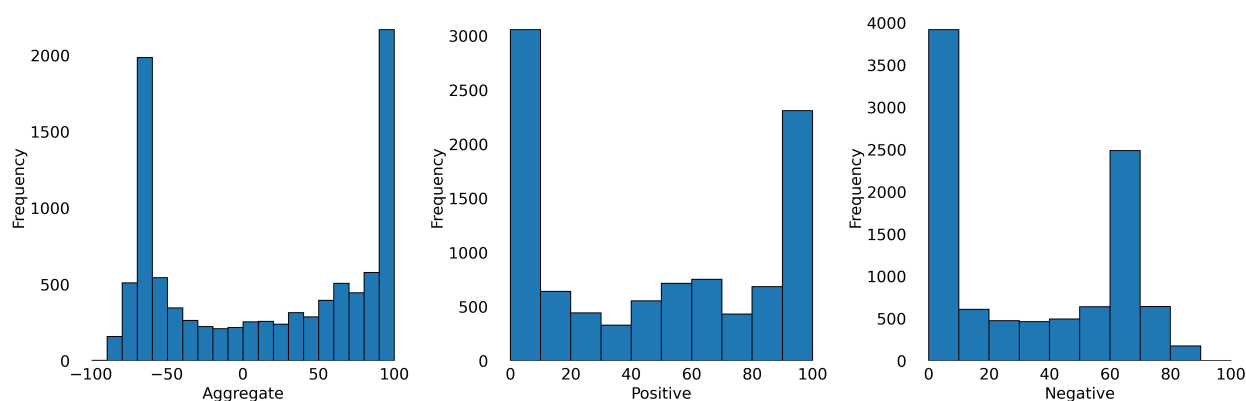
Notes: This figure presents example outputs from the Amazon Rekognition API including the detected faces (marked with red bounding boxes), the results of face recognition, such as David Cameron (Figure a) and Boris Johnson (Figure b), and the confidence scores to each emotion category.

each image, the system first detects faces and locates them, as indicated by the red bounding boxes in the figure. It then identifies who they are, e.g., David Cameron for the face in Figure 2.1 (a) and Boris Johnson for the face in Figure 2.1 (b), and finally analyzes the emotions, providing a confidence score for each emotion category.

To build a single emotion score for each facial expression, we first compute the positive emotion score, which is the sum of the scores of the positive emotions ('happy' and 'calm'),

and the negative emotion score, which is the sum of the negative emotions (‘surprised,’ ‘confused,’ ‘disgusted,’ ‘angry,’ ‘fear,’ and ‘sad’), both normalized to a range of 0-100, with 100 representing very positive and very negative, respectively. We then define the aggregate emotion score as the difference between the positive emotion score and the negative emotion score, ranging from -100 (most negative) to 100 (most positive). Figure 2.2 displays the distribution of aggregate, positive, and negative emotion scores for the pre-referendum period.

**Figure 2.2:** Distribution of Emotion Scores: Pre-Referendum Period



### 2.4.3 Politician Sample

Our politician sample includes all the cabinet ministers at the time of the referendum, 24 of whom supported Remain and 6 supported Leave.<sup>4</sup> We also include Nigel Farage and Jeremy Corbyn, and exclude all politicians who appear less than 30 times. This results in a final sample featuring 19,288 faces of 27 politicians, 7 of whom are pro-Leave.<sup>5</sup> The decision to focus the politician sample on cabinet ministers along with Nigel Farage and Jeremy Corbyn is driven by their central roles in the Brexit referendum and their prominence in both media

<sup>4</sup>The list of politicians and their stances can be found at <https://www.bbc.co.uk/news/uk-politics-eu-referendum-35616946>.

<sup>5</sup>See Tables B.1 - B.4 in the Appendix for the list of politicians and some summary statistics.

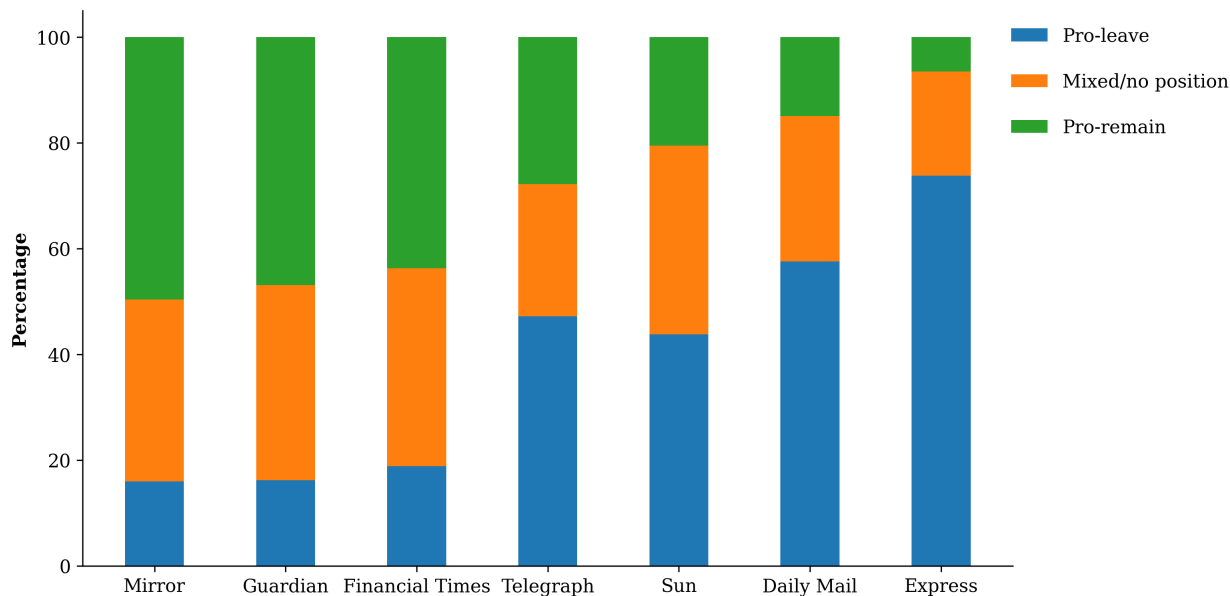
coverage and public discourse. Cabinet ministers, as key government figures, were responsible for shaping and executing national policy, making them the focal points of the campaign. Nigel Farage, as the leader of UKIP, was the face of the pro-Leave movement, while Jeremy Corbyn, as leader of the opposition, played a crucial role in shaping the Remain narrative. Including these figures ensures that the study captures the most influential voices during the referendum, who were central to shaping both the policy debate and the media's portrayal of the Brexit issue.

#### 2.4.4 Newspaper Partisanship on Brexit Vote

To measure newspaper partisanship, we rely on [Levy et al. \(2016\)](#), who analyzed news coverage of UK newspapers over a four-month Brexit campaign period and reported the proportions of articles in each newspaper that were pro-Leave, pro-Remain, and neither, respectively. Their sample consisted of a total of 2,378 news articles for two days a week (Tuesdays and Saturdays) over the campaign period, from Prime Minister David Cameron's post-summit Cabinet meeting on 20 February to the Brexit referendum held on 23 June. Categorizing articles was achieved by manual labeling. [Figure 2.3](#) shows that the most pro-Leave newspaper is *The Express*, followed by *The Daily Mail*, *The Sun*, and *The Telegraph*, while *The Mirror* is the most pro-Remain newspaper, with *The Guardian* and *Financial Times* next. Notably, we can also see that pro-Leave newspapers are working harder on their campaigns, with the most pro-Leave newspaper, *The Express*, having 73% Leave articles and only 6.5% Remain articles, a difference of 66.5%, while the most pro-Remain newspaper, *The Mirror*, has 49.6% Remain articles and 16% Leave articles, a difference of merely 33.6%. According to [Levy et al. \(2016\)](#), 41% of the total media coverage during the Brexit campaign was pro-Leave, compared to 27% pro-Remain, suggesting a media environment that was more inclined to support the Leave campaign.

We then compute newspaper partisanship scores by subtracting the proportion of pro-Remain articles from the proportion of pro-Leave articles and then normalizing the result to be between 0 (most pro-Remain) to 100 (most pro-Leave).<sup>6</sup>

**Figure 2.3:** Newspaper Partisanship from [Levy et al. \(2016\)](#)



Notes: This figure plots the the shares of pro-Leave, pro-Remain, and mixed/no position articles across newspapers.

## 2.5 Visual Media Bias in Politician Facial Expression Emotions

Following [Boxell \(2021\)](#), we analyze visual media bias in politician facial expressions in two ways: examining who they choose to cover and how they report on them. In the first instance, we restrict our sample to tabloids only.

<sup>6</sup>Table [B.5](#) reports the partisanship score, count of images, and count of detected politician faces for each newspaper.

To explore visual media bias in the choice of who to cover, Figure 2.4 Panel A reports the proportion of pro-Remain and pro-Leave tabloid newspapers displaying a particular politician or group of politicians during the pre-referendum period. Overall, pro-Leave politicians received more visual coverage in pro-Leave newspapers than in pro-Remain newspapers (46.1% vs. 41.9%). In more detail, Cameron, as then Prime Minister and a central figure in the Remain camp, had more coverage than any other political figure. The pro-Leave newspapers gave Cameron more coverage but less to other pro-Remain politicians compared to pro-Remain newspapers. On the other hand, the bulk of the coverage of the Leave camp centered on Johnson and Farage. Johnson received slightly more coverage in the pro-Leave newspapers compared to pro-Remain newspapers (20.4% vs. 18.3%). Farage was the opposite, with a larger share of images in pro-Remain newspapers. Overall, these findings indicate that there is no clear and consistent bias in the selection of politicians to cover.

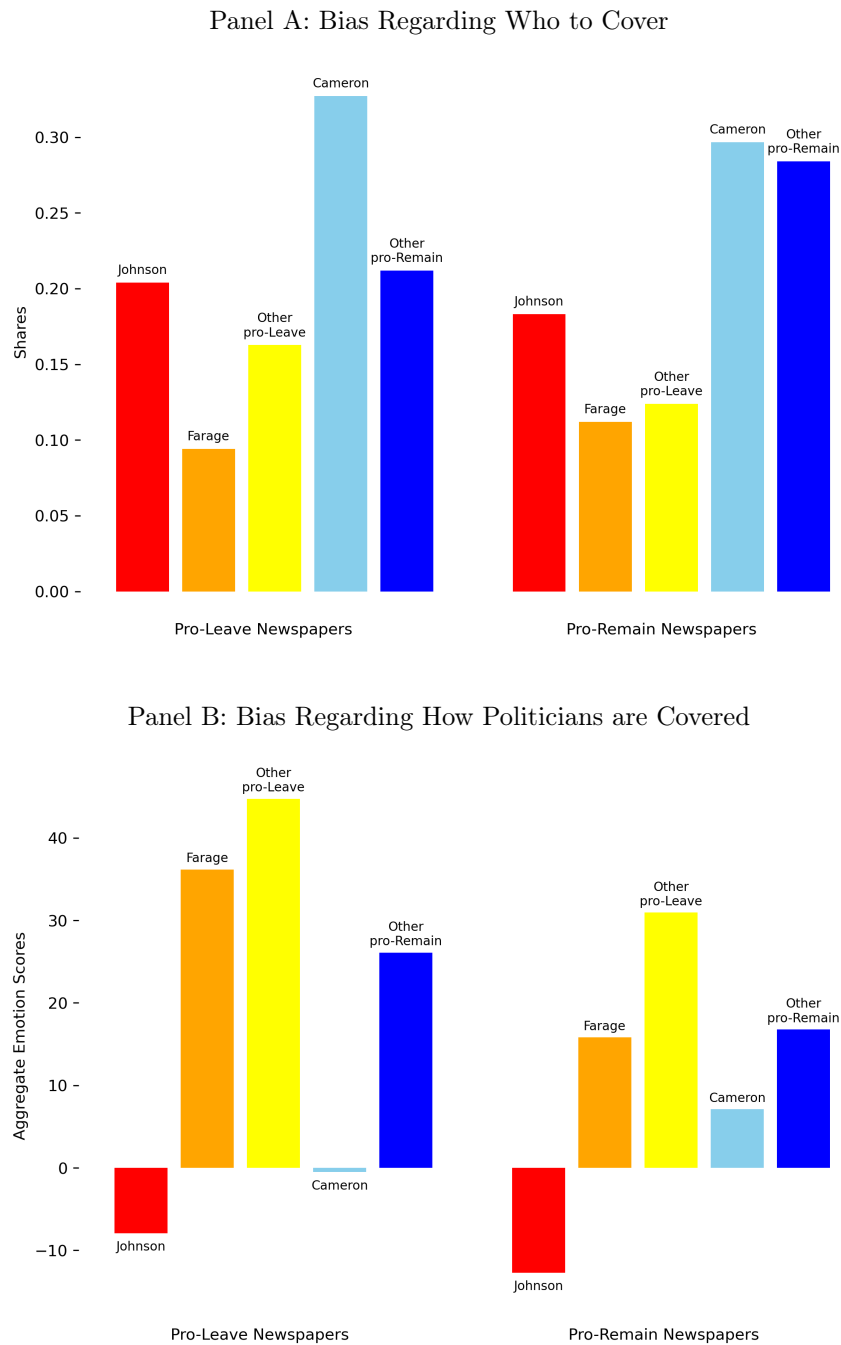
To explore visual media bias in how politicians are covered, Figure 2.4 Panel B reports the average aggregate emotion scores for specific politicians or groups of politicians across pro-Leave and pro-Remain newspaper websites.<sup>7</sup> We can see that Johnson, Farage, and other pro-Leave politicians were given relatively more favorable coverage in pro-Leave newspapers than in pro-Remain newspapers. In contrast, Cameron was portrayed in a relatively unfavorable light in the pro-Leave newspapers. Despite the relative favorability in pro-Leave newspapers, Johnson appears to have negative facial expressions in both types of newspapers. Something similar was found by Boxell (2021) in US newspapers, where Trump always looks more negative than other politicians on websites of various political leanings. In broadsheet newspapers, however, we do not observe clear media bias in both who to report and how they are reported.<sup>8</sup>

---

<sup>7</sup>Figure B.1 shows the average scores across emotion categories for Johnson, Farage, Cameron, and other pro-Remain and pro-Leave politicians. Figure B.2 shows that even though the Amazon API does not consider emotions to be mutually exclusive, for the most part, each face is still dominated by individual emotions, as evidenced by the correlations between emotions that are largely negative.

<sup>8</sup>See Figure B.3 for broadsheets.

**Figure 2.4:** Visual Media Bias During the Brexit Vote



Notes: Panel A presents the proportions of individual politicians or groups of politicians appearing in images from pro-Leave and pro-Remain newspaper websites. These proportions are calculated relative to the total number of detected politicians' faces in pro-Leave or pro-Remain newspapers. Panel B shows the average aggregate emotion scores for individual politicians or groups of politicians across Pro-Leave and Pro-Remain newspaper websites. Pro-Leave newspapers are defined as those that publish more articles supporting leaving the EU, while pro-Remain newspapers publish more articles in favor of staying in the EU, based on [Levy et al. \(2016\)](#).

## 2.5.1 Empirical Evidence

To formally examine visual media bias in the coverage of politicians, we follow the empirical strategy of [Boxell \(2021\)](#) and estimate the following equation using ordinary least squares (OLS):

$$\begin{aligned} \text{FacialEmotion}_{ijt} = & \gamma_1 + \beta_1(\text{NewspaperPartisanship}_j \times \text{PoliticianStance}_i) \\ & + \text{Politician FE}_i + \text{Newspaper FE}_j + \epsilon_{ijt}. \end{aligned} \tag{2.1}$$

where  $\text{FacialEmotion}_{ijt}$  denotes the facial emotion score for instance  $t$  of the politician  $i$  that is published in newspaper  $j$ ,  $\text{NewspaperPartisanship}_j$  is the partisanship score of the newspaper  $j$ ,  $\text{PoliticianStance}_i$  is a dummy, which is equal to 1 if politician  $i$  is in the Leave campaign, and 0 otherwise, and  $\text{PoliticianFE}_i$  denotes politician fixed effects, that controls for time-invariant politician characteristics, for example, certain politicians may consistently show particular emotions across all newspapers.<sup>9</sup> Newspaper  $\text{FE}_j$  represents newspaper fixed effects that controls for time-invariant unobserved newspaper-specific characteristics, for example, some newspapers tend to portray all politicians in a certain light. The parameter  $\epsilon_{it}$  is the error term.

Table [2.3](#) presents the regression results for tabloid newspapers. Columns (1) - (10)

---

<sup>9</sup>We code  $\text{PoliticianStance}_i$  as a dummy variable, where 1 indicates a Leave politician and 0 indicates Remain. A concern was raised that with this 0/1 coding, the interaction term  $\text{NewspaperPartisanship}_j \times \text{PoliticianStance}_i$  is primarily identified from the relatively small number of Leave politicians. To address this, we also estimate the model using  $\text{PoliticianStance}$  coded as -1 and +1, which balances the contributions of both Leave and Remain politicians and isolates the general effect. The results of this alternative specification are presented in Table [B.6](#), where the interaction coefficient under 0/1 coding is found to be twice the magnitude of the coefficient under -1/1 coding. Importantly, the substantive conclusions remain consistent across both coding schemes. We retain the 0/1 coding in the main analysis for its straightforward interpretation and clarity.

show the results of Specification 2.1 for each of the eight emotions, the sum of negative emotions (Column (9)), and the aggregate emotion score (Column (10)). Panels A to C examine images published in the run-up to the Brexit referendum. Panel A examines all images, Panel B focuses exclusively on front-page images, and Panel C focuses exclusively on content-page images.<sup>10</sup> Panel D examines all images published in the post-referendum period. Standard errors (in parentheses) are clustered by newspaper. Media bias in how politicians are depicted exists when pro-Leave and pro-Remain press portray the same politicians differently.

In Table 2.3 Panel A, Column (10), the significant coefficient for the aggregate emotion score implies that a one unit increase in the newspaper partisanship score, for example, an increase from 50 to 51 (more pro-Leave), is associated with around 19.28 points relative increase in the aggregate emotion score of pro-Leave politicians compared to pro-Remain politicians. Within the categories of emotions (Columns (1) to (8)), we find that pro-Leave newspapers make relatively more use of the Calm emotion and less use of the Angry and Confused emotions to portray pro-Leave politicians. Column (9) refers to the sum of negative emotions (Columns (3) - (8)) with the coefficients being negative and statistically significant at the conventional level.

Panels B and C of Table 2.3 report the results that distinguish between front-page images and content-page images. Focusing on Column (10), we find qualitatively similar but quantitatively different results across the two panels. Specifically, we observe a considerably greater degree of bias in the images featured on the front page, although both sets of images show significant slanting. Columns (3) - (8) suggest that pro-Leave politicians in the front-page images are more rarely depicted in pro-Leave newspapers as angry, fear, sad, and confused. Panel D of Table 2.3 reports results in the post-referendum sample. We find that in all regressions, only the coefficient of fear is significantly negative, and the coefficient is

---

<sup>10</sup>The slight deviation between the sum of observations for front-page images and content-page images and all images is due to the removal of singleton observations.

**Table 2.3:** Newspaper Partisanship and Politician Emotions

		Dependent Variable: Emotion Score $\times 100$									
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)		
Happy (P)	Calm (P)	Disgusted (N)	Angry (N)	Fear (N)	Sad (N)	Confused (N)	Surprised (N)	Negative	AggregateEmo.		
<b>Panel A: All images</b>											
Partisanship x	-0.567	0.379	-2.735**	-0.301	-0.590	-7.821**	-2.229	-8.595***	19.28***		
PoliticianStance	(1.676)	(0.450)	(0.562)	(0.168)	(1.809)	(1.356)	(1.426)	(0.333)	(0.862)		
Observations	7946	7946	7946	7946	7946	7946	7946	7946	7946		
<b>Panel B: Front page images</b>											
Partisanship x	0.796	2.165	-3.493***	-2.199***	-2.897**	-9.258***	-0.518	-10.47***	24.54***		
PoliticianStance	(1.907)	(0.831)	(0.474)	(0.264)	(0.688)	(1.417)	(0.557)	(0.639)	(1.443)		
Observations	2026	2026	2026	2026	2026	2026	2026	2026	2026		
<b>Panel C: Content page images</b>											
Partisanship x	-1.913	-0.266	-2.334	0.330	0.0626	-5.963**	-2.973	-7.203***	15.60***		
PoliticianStance	(1.139)	(0.996)	(0.828)	(0.206)	(2.184)	(1.268)	(2.027)	(0.553)	(1.513)		
Observations	5917	5917	5917	5917	5917	5917	5917	5917	5917		
<b>Panel D: All images, post-referendum period</b>											
Partisanship x	-2.894	0.373	-0.157	-0.239**	0.888	2.110	2.618	3.616	-9.398		
PoliticianStance	(2.771)	(0.221)	(1.533)	(0.0672)	(1.682)	(1.805)	(2.416)	(2.460)	(5.753)		
Observations	7580	7580	7580	7580	7580	7580	7580	7580	7580		
Clusters	4	4	4	4	4	4	4	4	4		
Newspaper F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y		
Politician F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y		

Notes: The table presents estimates of  $\beta_1$  from Eq. 2.1, where the dependent variable is the emotion score derived from the face of the politician and fixed effects for newspapers and politicians are included. Panel B focuses exclusively on the front page images. Panel C focuses exclusively on the content page images. 'NewspaperPartisanship  $\times$  PoliticianStance' is the interaction between the newspaper partisanship scores with an indicator for whether the politician is Pro-leave. The column headers specify the emotion score used as the dependent variable in each regression. The (P) and (N) labels after each emotion in the column headers indicate whether the emotion is classified as positive (P) or negative (N). 'Negative' indicates the sum of disgusted, angry, fear, sad, confused, and surprised. 'AggregateEmo.' indicates the aggregate emotion score that is the sum of happy and calm minus negative. Standard errors clustered by newspaper are in parentheses. \*\*\* indicates significance at the one percent level, \*\* indicates significance at the five percent level.

fairly small. This suggests that the media bias disappeared and the partisan newspapers stopped their propaganda in the aftermath of the Brexit vote.

We further explore the bias in the visual depiction of six major politicians in the two camps during the vote, namely Boris Johnson, Nigel Farage, and Michael Gove, who were in the Leave camp, and David Cameron, George Osborne, and Jeremy Corbyn, who were in the Remain camp. To do so, we run six separate regressions as specified in Equation 2.1, each focusing on one of the six major politicians paired against all politicians from the opposing camp. To show the results more intuitively, in each regression equation we set  $\text{PoliticianStance}_i$  to 1 for these major politicians and 0 for politicians in the opposite camp.

Table 2.4 reports the results. Column (10) of Table 2.4 shows that Farage and Gove, both pro-Leave politicians, have relatively higher aggregate emotion scores as newspaper partisanship scores increase, indicating that newspapers with stronger pro-Leave stances portray them more favorably. In contrast, Cameron, Osborne, and Corbyn, who were pro-Remain politicians, display different trends, with their aggregate emotion scores decreasing as newspapers become more pro-Leave. Among them, the bias showing Farage is the most pronounced: a one-unit increase in newspaper partisanship score (more pro-Leave) is associated with an approximate 35.66-point relative increase in the aggregate emotion score of Farage compared to pro-Remain politicians, followed by Gove, with a 26.89-point increase.

On the other hand, the bias portraying Cameron is the largest in the Remain camp, with a 20.42-point relative decrease in his aggregate emotion score as the newspaper partisanship increased by one unit. However, Panel A Column (10) suggests that newspapers with different stances do not show a statistically significant bias when it comes to covering Boris Johnson, although Johnson is less likely to exhibit anger (Column (4)) and surprise (Column (8)) emotions in pro-Leave newspapers.<sup>11</sup>

---

<sup>11</sup>Table B.7 in the Appendix presents three sets of results comparing Cameron vs. Johnson, Cameron vs. Farage, and Cameron vs. both Johnson and Farage. The results are consistent, showing no clear bias

**Table 2.4:** Newspaper Partisanship and Politician Emotions: Major Politicians

	Dependent Variable: Emotion Score $\times 100$								(9)	(10)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)		
	Happy (P)	Calm (P)	Disgusted (N)	Angry (N)	Fear (N)	Sad (N)	Confused (N)	Surprised (N)	Negative	AggregateEmo.
<b>Panel A: Johnson and all pro-Remain politicians</b>										
Partisanship x	0.345	4.305	0.188	-4.002**	0.164	-0.0690	1.429	-2.182**	-2.891	7.539
PoliticianStance	(1.217)	(3.821)	(0.283)	(1.006)	(0.0854)	(2.015)	(1.783)	0.428	(1.893)	(4.688)
Observations	5937	5937	5937	5937	5937	5937	5937	5937	5937	5937
<b>Panel B: Farage and all pro-Remain politicians</b>										
Partisanship x	-3.580	23.18***	0.202	-0.437	-1.544**	-4.211	-14.94***	-3.929	-16.07***	35.66***
PoliticianStance	(1.392)	(1.314)	(0.209)	(2.550)	(0.377)	(1.363)	(1.355)	2.745	(1.663)	(4.363)
Observations	5123	5123	5123	5123	5123	5123	5123	5123	5123	5123
<b>Panel C: Gove and all pro-Remain politicians</b>										
Partisanship x	-4.322	18.14***	1.043	-4.770	-0.00909	0.500	-15.93**	-1.063	-13.08***	26.89***
PoliticianStance	(4.797)	(2.822)	(1.207)	(4.934)	(0.322)	(2.716)	(2.915)	2.401	(1.233)	(3.296)
Observations	4940	4940	4940	4940	4940	4940	4940	4940	4940	4940
<b>Panel D: Cameron and all pro-Leave politicians</b>										
Partisanship x	-0.441	-10.86**	-0.496	-0.0919	0.394	-1.380	13.35***	2.337	9.127***	-20.42***
PoliticianStance	(1.165)	(1.903)	(0.404)	(1.589)	(0.172)	(2.275)	(1.461)	2.213	(0.764)	(2.229)
Observations	6152	6152	6152	6152	6152	6152	6152	6152	6152	6152
<b>Panel E: Osborne and all pro-Leave politicians</b>										
Partisanship x	-1.203	-5.593	-1.796**	7.432***	0.483	3.560***	-2.297**	3.076**	6.760**	-13.55**
PoliticianStance	(5.532)	(3.636)	(0.440)	(1.199)	(0.217)	(0.424)	(0.542)	0.896	(1.406)	(3.682)
Observations	4300	4300	4300	4300	4300	4300	4300	4300	4300	4300
<b>Panel F: Corbyn and all pro-Leave politicians</b>										
Partisanship x	1.276	-12.18**	0.518***	7.510***	0.516	4.675	-1.624	2.118	8.864***	-19.76***
PoliticianStance	(2.896)	(2.969)	(0.0463)	(0.669)	(0.203)	(3.420)	(3.303)	1.246	(0.475)	(0.776)
Observations	4124	4124	4124	4124	4124	4124	4124	4124	4124	4124
Clusters	4	4	4	4	4	4	4	4	4	4
Newspaper F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Politician F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The table reports the estimates of  $\beta_1$  from Eq.2.1, where the dependent variable is the emotion score derived from the face of the politician and fixed effects for newspapers and politicians are included. Panels A-F focus exclusively on a particular politician and the opposing group of politicians. The column headers specify the emotion score used as the dependent variable in each regression. The (P) and (N) labels after each emotion in the column headers indicate whether the emotion is classified as positive (P) or negative (N). 'Negative' indicates the sum of disgusted, angry, fear, sad, confused, and surprised. 'AggregateEmo.' indicates the aggregate emotion score that is the sum of happy and calm minus negative. Standard errors clustered by newspaper are in parentheses. \*\*\* indicates significance at the one percent level, \*\* indicates significance at the five percent level.

To investigate whether our results are driven by a specific major politician, we conduct a series of regressions, as outlined in Equation 2.1, on samples that exclude each of the six major politicians, respectively. Columns (9) and (10) of Panels A-F in Table 2.5 show that all coefficients on the negative emotion and aggregate emotion scores are statistically significant at the 1 percent level. This suggests that excluding any single major politician does not distort our findings. Next, we examine the sample that excludes all six major politicians to identify whether the visual bias is driven by these key figures in the Brexit campaign. Panel G shows that no significant bias remains in how other politicians are portrayed (Column (10)). Combined with previous results in Table 2.4, our findings suggest that the media bias is primarily concentrated in the portrayal of a few major politicians, particularly Farage, Gove, Cameron, and Corbyn, rather than being distributed across all politicians.

Taken together, our empirical findings provide strong support for the existence of visual media bias in how politicians are covered during the Brexit referendum by depicting the opposing politician with more unfavorable emotions, especially in front-page images. The major politicians of both Remain and Leave camps have a slanted portrayal. The one exception is that Boris Johnson was depicted without bias across newspapers and consistently in a negative light.<sup>12</sup>

---

between Cameron and Johnson, but a significant bias in how Farage is depicted compared to Cameron.

<sup>12</sup>For broadsheets, Table B.12 reports the results of Specification 2.1 across eight emotion categories (Columns (1) - (8)), along with the sum of negative emotions in Column (9) and the aggregate emotion score in Column (10). In Columns (9) and (10), none of the coefficients for the negative emotion score or the aggregate emotion score are statistically significant, suggesting that the broadsheets do not exhibit a clear bias in their visual portrayal of politicians. In addition, Table B.13 reports the results for the sample that includes all newspapers.

**Table 2.5:** Newspaper Partisanship and Politician Emotions: Excluding Major Politicians

	Dependent Variable: Emotion Score $\times 100$								(9)	(10)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)		
	Happy (P)	Calm (P)	Disgusted (N)	Angry (N)	Fear (N)	Sad (N)	Confused (N)	Surprised (N)		
<b>Panel A: All politicians except Johnson</b>										
Partisanship x	-1.094	16.07***	0.519	-1.811*	-0.638*	-0.949	-14.48***	-2.245	-12.67***	27.64***
PoliticianStance	(2.822)	(1.077)	(0.591)	(0.729)	(0.242)	(1.720)	(1.144)	(2.599)	(1.628)	(4.041)
Observations	6357	6357	6357	6357	6357	6357	6357	6357	6357	6357
<b>Panel B: All politicians except Farage</b>										
Partisanship x	0.564	6.890*	0.436	-3.584*	0.159	0.796	-5.266*	-1.645	-5.884***	13.33***
PoliticianStance	(1.837)	(2.488)	(0.530)	(1.133)	(0.118)	(2.008)	(1.677)	(1.000)	(0.564)	(1.512)
Observations	7171	7171	7171	7171	7171	7171	7171	7171	7171	7171
<b>Panel C: All politicians except Gove</b>										
Partisanship x	0.165	9.928**	0.245	-2.326*	-0.354*	-0.791	-6.268**	-2.478	-7.739***	17.83***
PoliticianStance	(2.456)	(1.968)	(0.313)	(0.890)	(0.143)	(1.649)	(1.508)	(1.281)	(0.365)	(0.995)
Observations	7354	7354	7354	7354	7354	7354	7354	7354	7354	7354
<b>Panel D: All politicians except Cameron</b>										
Partisanship x	-1.822	11.90***	0.246	-6.511**	-0.190	-3.227*	-0.811	-1.873**	-7.993***	18.07***
PoliticianStance	(2.573)	(0.715)	(0.498)	(1.244)	(0.157)	(1.244)	(1.251)	(0.414)	(0.977)	(2.874)
Observations	5392	5392	5392	5392	5392	5392	5392	5392	5392	5392
<b>Panel E: All politicians except Osborne</b>										
Partisanship x	-0.964	12.42***	0.100	-1.781*	-0.267	-0.0121	-9.852**	-2.030	-8.948***	20.40***
PoliticianStance	(1.042)	(1.060)	(0.446)	(0.636)	(0.155)	(2.090)	(1.739)	(1.537)	(0.113)	(0.538)
Observations	7244	7244	7244	7244	7244	7244	7244	7244	7244	7244
<b>Panel F: All politicians except Corbyn</b>										
Partisanship x	-0.518	11.16**	0.557	-1.871*	-0.267	0.127	-9.560***	-2.189	-8.534***	19.17***
PoliticianStance	(2.185)	(2.044)	(0.499)	(0.745)	(0.163)	(1.560)	(1.097)	(1.674)	(0.342)	(0.881)
Observations	7420	7420	7420	7420	7420	7420	7420	7420	7420	7420
<b>Panel G: All politicians except Johnson, Farage, Gove, Cameron, Osborne, and Corbyn</b>										
Partisanship x	-1.991	11.97**	-0.909	-1.648	1.277***	2.771**	-12.57**	2.273	-5.691	15.67
PoliticianStance	(10.32)	(2.810)	(1.223)	(5.014)	(0.0880)	(0.769)	(2.926)	(1.348)	(6.824)	(15.59)
Observations	1208	1208	1208	1208	1208	1208	1208	1208	1208	1208
Clusters	4	4	4	4	4	4	4	4	4	4
Newspaper F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Politician F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The table reports the estimates of  $\beta_1$  from Eq.2.1, where the dependent variable is the emotion score derived from the face of the politician and fixed effects for newspapers and politicians are included. Panels A-F show results excluding a specific major politician in the Brexit campaign. Panel G show results excluding six major politicians in the Brexit campaign, namely, Jhnsn, Farage, Gove, Cameron, Osborne, and Corbyn. The column headers specify the emotion score used as the dependent variable in each regression. The (P) and (N) labels after each emotion in the column headers indicate whether the emotion is classified as positive (P) or negative (N). ‘Negative’ indicates the sum of disgusted, angry, fear, sad, confused, and surprised. ‘AggregateEmo.’ indicates the aggregate emotion score that is the sum of happy and calm minus negative. Standard errors clustered by newspaper are in parentheses. \*\*\* indicates significance at the one percent level, \*\* indicates significance at the five percent level.

## 2.5.2 Robustness

We consider three sets of robustness tests: (1) excluding each emotion respectively to account for potential misclassification in categorizing emotions as positive or negative, (2) using an alternative partisanship measure, and (3) using an alternative sample starting from 15 April, the official start of the campaign period. More specifically, following [Boxell \(2021\)](#), we categorize ‘surprise’ as a negative emotion. However, as [Caprini \(2023\)](#) suggests, the algorithm might misinterpret an open mouth as an indication of surprise, particularly when politicians are delivering speeches. We therefore perform a robustness check that excludes surprise emotions (see [Table B.8](#)), and the results remain robust. To ensure robustness, we also performed separate checks by excluding each emotion individually from the analysis. The results consistently aligned with our main findings (see [Table B.9](#)).

Another concern may be the reliance on [Levy et al. \(2016\)](#) to measure newspaper partisanship, as their study is based exclusively on print newspapers and includes only articles published on Tuesdays and Saturdays. We therefore implement an alternative partisanship measure, coding newspapers as either 0 for pro-Remain or 1 for pro-Leave, and the results remain robust (see [Table B.10](#)). Lastly, we perform an additional robustness check by restricting the sample period to start from April 15, when the official campaign period began, and find consistent results (see [Table B.11](#)).

## 2.5.3 Trends in Visual Media Bias during the Brexit Vote

The next stage is to investigate the dynamics in visual media bias during the referendum. [Figure 2.5](#) shows trends in the average aggregate emotion score of visual coverage of pro-Leave and pro-Remain politicians across pro-Leave and pro-Remain newspapers for each calendar week. Panel A indicates that when reporting on pro-Leave politicians, pro-Leave newspapers

nearly always display more positive emotions compared to pro-Remain newspapers before the referendum. The differences in aggregate emotion scores between pro-Leave and pro-Remain newspapers have a clear dynamic inconsistency over the pre-referendum period.

Panel A also reveals a narrowing gap in the emotional portrayals of pro-Leave politicians between pro-Leave and pro-Remain newspapers as the referendum date approached, indicating a moderation of visual media bias over time. This pattern aligns with [Levy et al. \(2016\)](#)'s analysis of textual coverage, which shows that the percentage of pro-Leave articles peaked in the early stages of the campaign period but gradually declined as the referendum date approached, while the share of mixed or no-position articles increased significantly (as shown in [Figure B.4](#)). In contrast, the percentage of pro-Remain articles remained relatively unchanged throughout the campaign. This suggests that media, especially pro-Leave outlets, became more cautious in showing clear leaning as the vote drew near in both visual and textual content.

Panel B of [Figure 2.5](#) shows that the two groups of newspapers do not exhibit significant variation in aggregate emotion scores in their visual coverage of pro-Remain politicians. These results suggest that visual media bias is predominantly reflected in the coverage of pro-Leave politicians.<sup>13</sup> In both Panels A and B, the aggregate emotion scores converge towards zero as the referendum date approaches, indicating a shift towards more neutral portrayals of both pro-Leave and pro-Remain politicians in both pro-Leave and pro-Remain newspapers. This trend reconfirms our earlier conclusion that newspapers tend to avoid showing bias as the referendum approaches.<sup>14</sup>

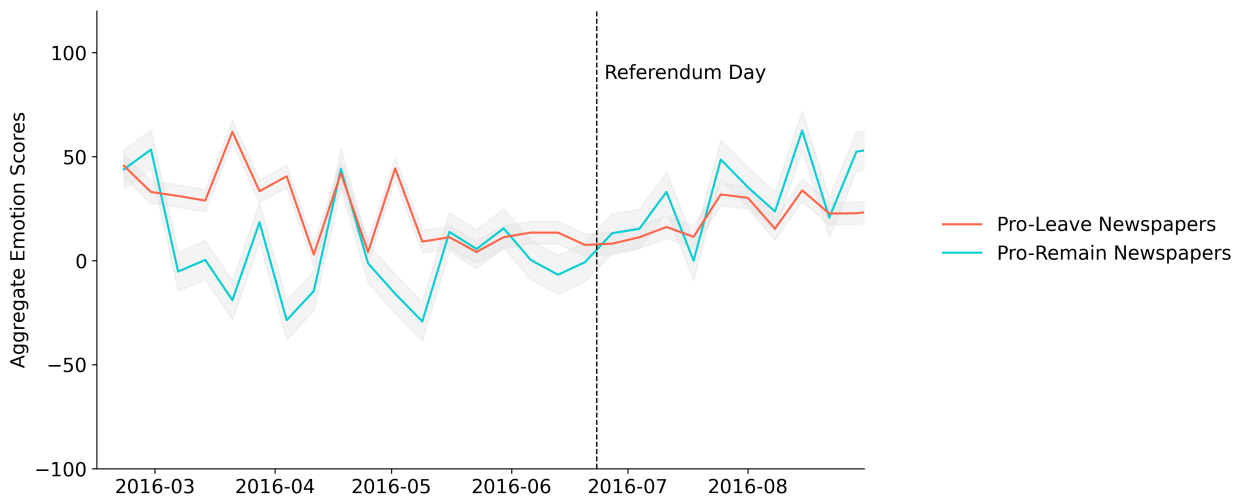
---

<sup>13</sup>See [Table B.14](#) for regression results.

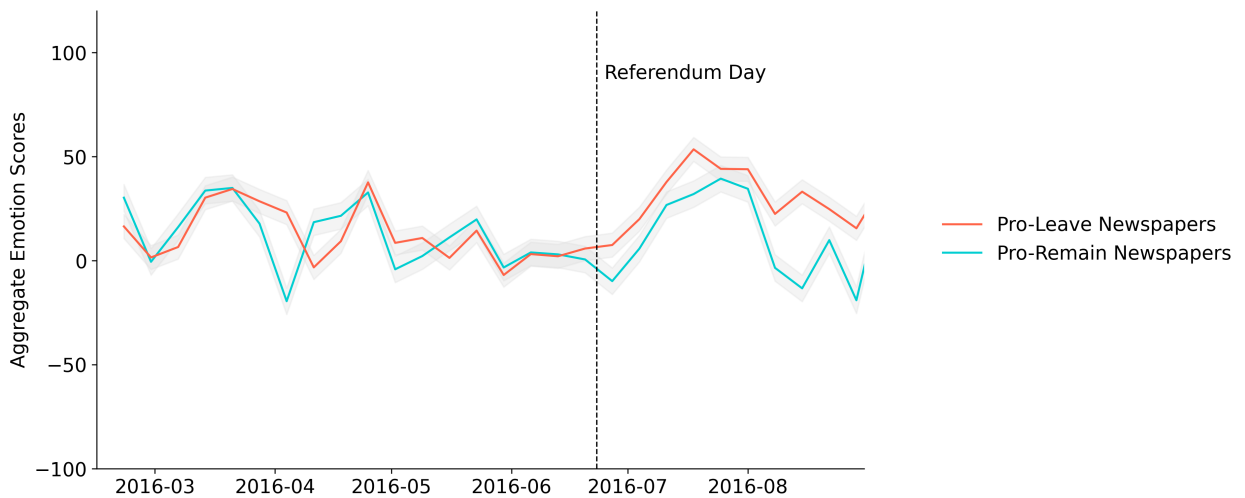
<sup>14</sup>See [Figures B.5](#) and [B.6](#) in the Appendix for the trend of each emotion category.

**Figure 2.5:** Dynamics of Visual Media Bias During the Brexit Vote

Panel A: Average Aggregate Emotion Scores Towards Pro-Leave Politicians



Panel B: Average Aggregate Emotion Scores Towards Pro-Remain Politicians



Notes: The figure shows the aggregate emotion scores towards pro-Leave and pro-Remain politicians, across the set of pro-Leave and pro-Remain newspapers respectively for each week. Pro-Leave newspapers are those that publish more articles supporting leaving the EU, while Pro-Remain newspapers are those that publish more articles in favor of staying in the EU, based on [Levy et al. \(2016\)](#). The shaded grey areas represent 95% confidence intervals.

## 2.6 Bias in Contextual Images

### 2.6.1 Measuring Contextual Image Sentiment

In this section, we extend our analysis to examine contextual images, which involves creating contextual images of politicians. We first pick out all the images that contain politicians and ensure that the politicians are of the same position. That is, for example, we exclude all images that feature both David Cameron (pro-Remain) and Boris Johnson (pro-Leave) due to their differing campaigns, and retain images that either feature a single politician or multiple politicians within the same campaign, such as Boris Johnson and Michael Gove (both pro-Leave). Subsequently, we blur each politician’s face to create a contextual image that obscures the facial expression while maintaining the general color distribution of the original image. For images that include multiple politicians, we blur each politician’s face individually and generate a contextual image for each politician.

To measure image sentiments, we utilize a pre-trained Vision Transformer (ViT) model, a transformer-based architecture designed for image recognition tasks and developed by researchers at Google (Dosovitskiy et al. 2020). Specifically, we use the ViT-B/16 configuration, which is a base-sized model that processes images by dividing them into  $16 \times 16$  pixel patches (Dosovitskiy et al. 2020). ViT has demonstrated exceptional performance in various computer vision tasks, including image classification, image segmentation, and autonomous driving (Han et al. 2022, Khan et al. 2022). The training process of the ViT model involves two stages. First, the ViT model is trained on the ImageNet-21k dataset, one of the most widely used datasets in the field of computer vision due to its comprehensive coverage of diverse object categories. The ImageNet-21k dataset includes around 14 million images across 21,843 classes, at a resolution of  $224 \times 224$  pixels. Second, this model was further fine-tuned on the ImageNet 2012 dataset, comprising around 1 million images categorized

into 1,000 classes, also at a resolution of  $224 \times 224$  pixels.

By leveraging the pre-trained ViT model, we can mitigate the substantial requirement for manually labeled data to train deep neural networks. Specifically, using transfer learning, a machine learning technique that adapts a model trained on one task for use on a related task, we can utilize the domain knowledge the ViT model has acquired from the ImageNet-21k and ImageNet 2012 datasets. This approach allows us to fine-tune the pre-trained model with a smaller dataset, significantly reducing the need for extensive manual labeling while still achieving a high level of accuracy and performance in our image classification tasks.

Technically speaking, we begin by loading the pre-trained model from the PyTorch library and modifying its classification head. The original classification head, designed for the 1,000 categories in the ImageNet 2012 dataset, was replaced with a new layer tailored specifically for our task, featuring two categories corresponding to positive and negative sentiments.

To fine-tune the pre-trained model, we utilized two datasets with sentiment-labeled images. The first dataset, constructed by [You et al. \(2015\)](#), consists of 1,269 images sourced from Twitter. Each image was labeled as either “positive” or “negative” by five online annotators from Amazon Mechanical Turk (AMT). Although these images originate from social media rather than news articles, [Obaid & Pukthuanthong \(2022\)](#) demonstrated that the dataset can reliably inform sentiment classification in news-related images. The second dataset, provided by [Hassan et al. \(2022\)](#), comprises 3,679 disaster-related images. These images were labeled by five online annotators from Microworkers, an online crowdsourcing platform. For each image, five annotators were asked, “What emotion is evoked after seeing this picture?” using a Likert scale from 1 to 9, where 1 indicates “very negative,” 5 is “neutral,” and 9 represents “very positive.” For our study, we categorized responses from 1 to 4 as negative sentiment and responses from 6 to 9 as positive sentiment, excluding images

with neutral responses (5) from the analysis. To ensure the reliability of the datasets for fine-tuning, we included only those images where all five annotators agreed on the sentiment labels. This filtering process reduced the first dataset to 882 images and the second dataset to 764 images, resulting in a combined total of 1,646 images across both datasets.

We use 80% of the dataset as the training sample and the remaining 20% as the testing sample. The fine-tuning process involves 100 iterations, during which we adjust the pre-trained ViT model’s parameters for 100 training cycles. Additionally, we set the learning rate to 0.001 and the batch size to 64. Accuracy is a common metric used to evaluate the performance of a model:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}},$$

where correct prediction means that the true sentiment agrees with the predicted sentiment, and the total number of predictions indicates the total number of images in the testing dataset. The fine-tuned ViT model gives us an accuracy score of 90.3%.<sup>15</sup>

Next, we resize our contextual images to  $224 \times 224$  pixels, ensuring consistency with the ImageNet dataset on which the Vision Transformer (ViT) model was pre-trained. We then use the fine-tuned model to predict the sentiment of our contextual images. The predicted image sentiment score for each image ranges from 0 (most negative) to 1 (most positive). To illustrate, we show examples of contextual images in Figure 2.6.

A main advantage of this method for categorizing images into sentiments is its inherent objectivity and reproducibility. This approach allows any researcher to replicate the process. In contrast, using human classification for sentiment analysis is not only more costly but also

---

<sup>15</sup>We also experimented with EfficientNet-V2-S, ResNet-50, and Inception V3 models, and the results indicated that, for our specific task, the ViT achieved the highest accuracy.

Figure 2.6: Examples of Positive and Negative Contextual Images

Panel A: Examples of Negative Images



Panel B: Examples of Positive Images



Notes: This figure presents a selection of contextual images used in our sentiment analysis study. Panel A includes examples of images that were classified as negative, while Panel B includes examples of images that were classified as positive. The images have been processed by blurring the faces of politicians to obscure their facial expressions, thus focusing the sentiment classification on the broader visual context. The fine-tuned ViT model was used to categorize the sentiment of these images.

susceptible to subjective judgments and biases.

## 2.6.2 Empirical Evidence

To examine partisan bias in the contextual images, we estimate the following regression via OLS:

$$\begin{aligned} \text{ImageSenti}_{ijt} = & \gamma_2 + \beta_2(\text{NewspaperPartisanship} \times \text{PoliticianStance}) \\ & + \text{Politician FE}_i + \text{Newspaper FE}_j + \epsilon_{ijt}. \end{aligned} \tag{2.2}$$

where  $\text{ImageSenti}_{ijt}$  denotes the contextual image sentiment for instance  $t$  of politician  $i$  that is published in newspaper  $j$ . The coefficient of interest  $\beta_2$  reflects the extent to which pro-Leave newspapers place pro-Leave politicians in a more favourable context compared to pro-Remain politicians. The coefficient should be positive if the newspaper places politicians of the same position in a more favourable context.

Table 2.6 reports the results based on Equation 2.2. We can see that newspaper partisanship is significantly correlated with the contextual image sentiment of politicians (Column (1)). In Column (2), the analysis is restricted to front-page images. The coefficient remains statistically significant at the 1 percent level and suggests that a one unit increase in the newspaper partisanship score (more pro-Leave) is associated with a 0.087 point increase in the contextual sentiment of pro-Leave politicians in the front-page news images, compared to pro-Remain politicians. Column (3) focuses on content-page images, where the coefficient of interest is not statistically significant, suggesting that there is no evidence of newspapers placing opposing politicians in a more negative contextual image among content-page images. Column (4) shows results for all images in the post-referendum period. The coefficient of interest is not statistically significant, indicating that after the referendum, newspaper

partisanship does not significantly influence the contextual image sentiment of pro-Leave versus pro-Remain politicians.

To conclude, our empirical results provide evidence that visual media bias also appears in the contextual images of politicians. Newspapers conveyed media bias by placing politicians of the same position in more positive background images, but this bias only appeared in front-page images. Moreover, the slant disappeared after the referendum. We did not find significant media bias in the contextual images from broadsheets.<sup>16</sup>

**Table 2.6:** Newspaper Partisanship and Politician Contextual Image Sentiment

	Dependent Variable: Contextual Image Sentiment $\times 100$			
	(1)	(2)	(3)	(4)
	All Images	Front Page Images	Content Page Images	All Images (Post)
Partisanship x	0.024***	0.087***	-0.010	-0.005
PoliticianStance	(0.002)	(0.012)	(0.017)	(0.011)
Observations	7396	1748	5645	6854
Clusters	4	4	4	4
Newspaper F.E.	Y	Y	Y	Y
Politician F.E.	Y	Y	Y	Y

Notes: The table presents estimates of  $\beta_2$  from Eq.2.2, where the dependent variable is the contextual image sentiment score of politician and fixed effects for newspapers and politicians are included. ‘NewspaperPartisanship  $\times$  PoliticianStance’ is the interaction between the newspaper partisanship scores with an indicator for whether the politician is Pro-leave. The column headers specify the image sample that used in each regression. Column (1)-(3) use the pre-Referendum sample. Column (4) use the sample that includes all images in post-Referendum period. Standard errors clustered by newspaper are in parentheses. \*\*\* indicates significance at the one percent level, \*\* indicates significance at the five percent level.

<sup>16</sup>See Table B.15 in the Appendix for the results of Broadsheets.

## 2.7 Conclusion

In this paper, we adapt computer vision and machine learning techniques to develop a novel measure of visual media bias. We use images from online news articles published during the 2016 Brexit vote and show the existence of visual media bias, particularly in tabloids. The results suggest that tabloid newspapers exhibited bias in the run-up to the Brexit vote by using relatively more negative emotions to depict politicians with opposing stances and by placing these politicians in more negative contexts in their images.

In the first part of our analysis, we examine facial expressions, finding that pro-Leave tabloids systematically showed pro-Leave politicians with more positive emotions like happiness and calmness. In contrast, pro-Remain politicians were often shown with more negative emotions, including anger, sadness, and confusion. This emotional disparity was especially pronounced in front-page images. Notably, this bias was mainly concentrated around a few key political figures. By contrast, no such bias was observed in broadsheet newspapers with no significant difference in the visual emotional portrayal of politicians from opposing camps.

In the second part of the study, we focus on the broader visual context of images. To isolate the contextual background from the politicians' facial expressions, we blurred the politicians' faces and applied sentiment analysis to the remaining elements of the images. This analysis revealed that the pro-Leave tabloids also placed pro-Leave politicians in more positive visual settings, while placed pro-Remain politicians in more negative contexts. This form of bias, however, was only present in front-page images and was notably absent after the referendum.

From a methodological perspective, we show how newspaper images can be used to develop and evaluate new measures of interest to researchers. In this context, it is worth

emphasizing that in recent years, visual news content has become increasingly popular in newspapers all over the world. The vast amount of image data, coupled with the rapid development of computer vision technology, offers great possibilities to utilize images to deepen our understanding of a wide range of economic, political, and historical developments through systematic empirical research. One limitation of our method is that we did not explore the source of the sentiments embedded in the contextual images, e.g., whether they are derived from objects, colors, or other human expressions. Despite this limitation, our approach achieves an automated and scalable analysis of visual contents using computer vision techniques.

These findings contribute to the broader literature on media bias by extending the focus beyond text to include visual content. As visual media continues to play a dominant role in how information is consumed, the techniques developed in this paper offer valuable insights into the subtle ways bias can manifest. These insights are critical in understanding how visual framing, especially during pivotal political events like the Brexit referendum, can shape public perceptions and ultimately influence voting behavior.

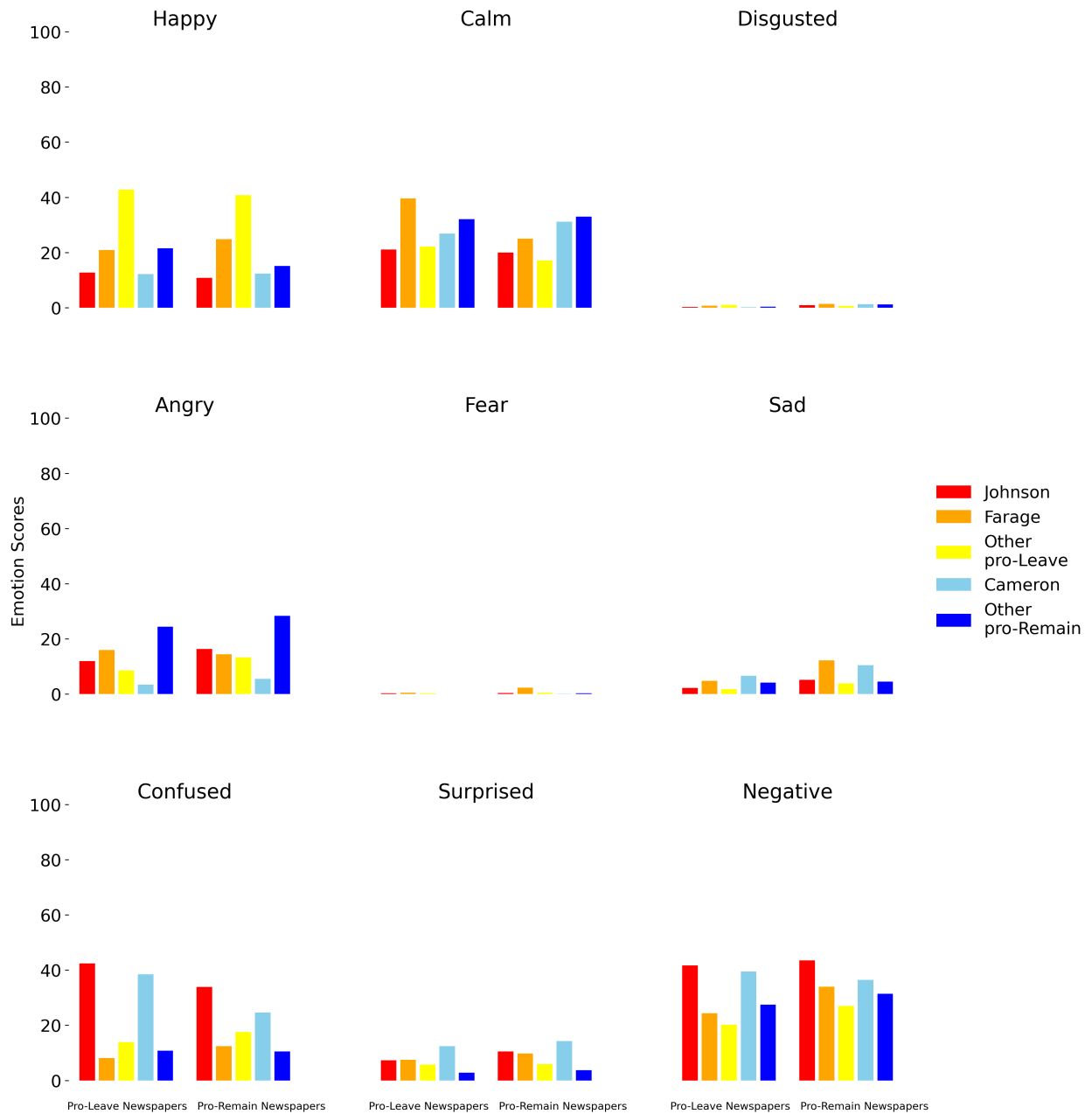
## B Appendix for Chapter 2

This appendix provides additional figures and tables to support the main analysis of the paper. Figure [B.1](#) displays the bias in how politicians were visually portrayed across emotion categories in tabloids. Figure [B.2](#) shows the correlation between different emotions. Figure [B.3](#) compares the portrayal of politicians in broadsheets. Figure [B.4](#) plots the percentages of pro-Leave, pro-Remain, and mixed/no position articles over time, based on the analysis by [Levy et al. \(2016\)](#). Figures [B.5](#) and [B.6](#) present the weekly dynamics of each emotion category for pro-Leave and pro-Remain politicians, respectively.

Table [B.1](#) lists the politicians included in the sample along with their official roles during the referendum period. Table [B.2](#) reports the number of times each politician's face appeared in tabloid and broadsheet images, both before and after the referendum. Tables [B.3](#) and [B.4](#) present the average emotion scores for each politician in tabloids and broadsheets, respectively. Table [B.5](#) reports the type of newspaper (broadsheet or tabloid), partisanship score, the total count of images, and the count of detected politician faces for each newspaper. Table [B.7](#) presents three sets of results comparing Cameron vs. Johnson, Cameron vs. Farage, and Cameron vs. both Johnson and Farage. Tables [B.8](#) to [B.11](#) provide results for robustness checks as discussed in Section [2.5.2](#). Tables [B.12](#) and [B.13](#) report the results for broadsheets and all newspapers, respectively. Table [B.14](#) shows the results across different groups of politicians. Finally, Table [B.15](#) presents the analysis of contextual image

sentiment in broadsheet newspapers.

**Figure B.1:** Bias regarding How Politicians are Covered (Tabloids)

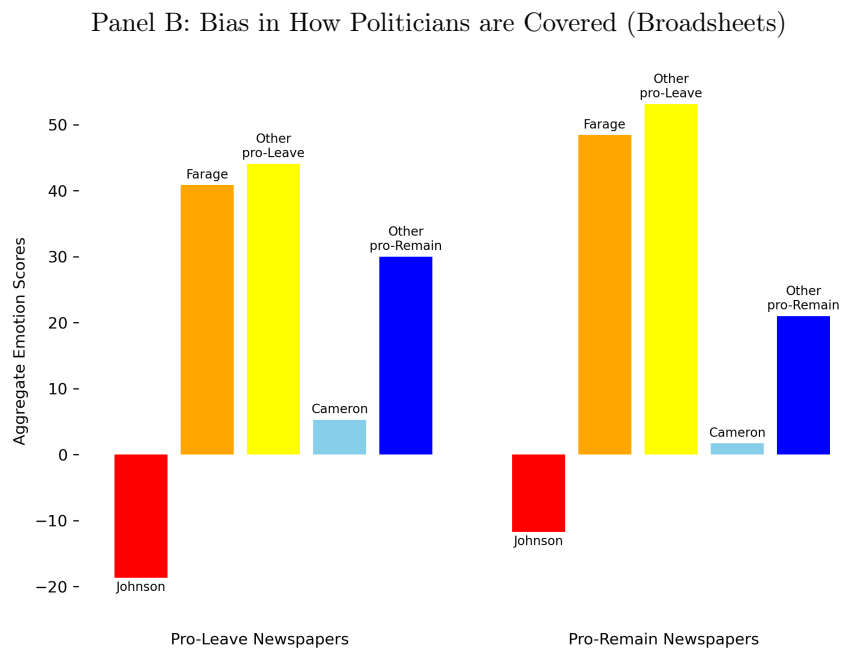
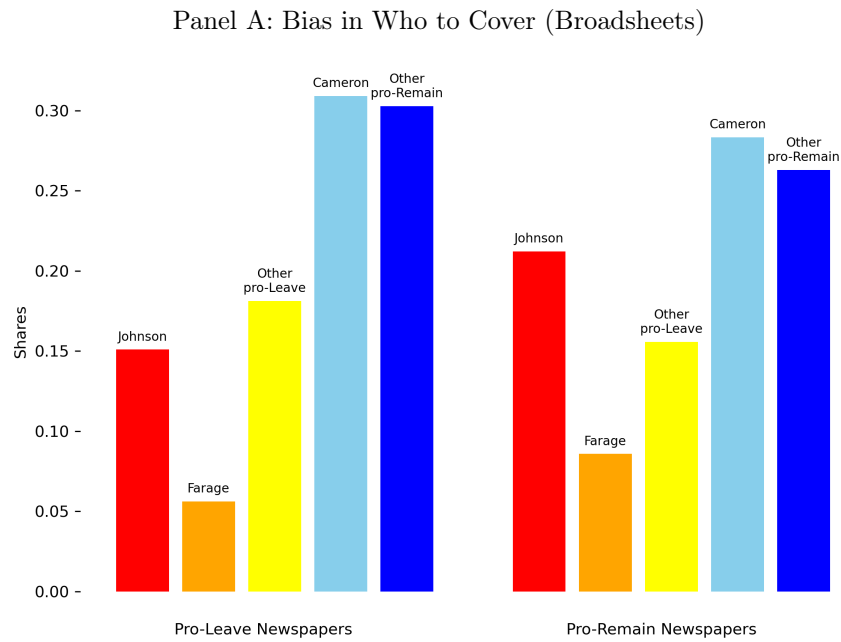


Notes: The figure plots the average scores across emotion categories for specific politicians or groups appearing across pro-Leave and pro-Remain newspaper website images. Pro-Leave newspapers are those that publish more articles supporting leaving the EU, while Pro-Remain newspapers are those that publish more articles in favor of staying in the EU, based on [Levy et al. \(2016\)](#).

Figure B.2: Correlation between Emotions for Amazon API

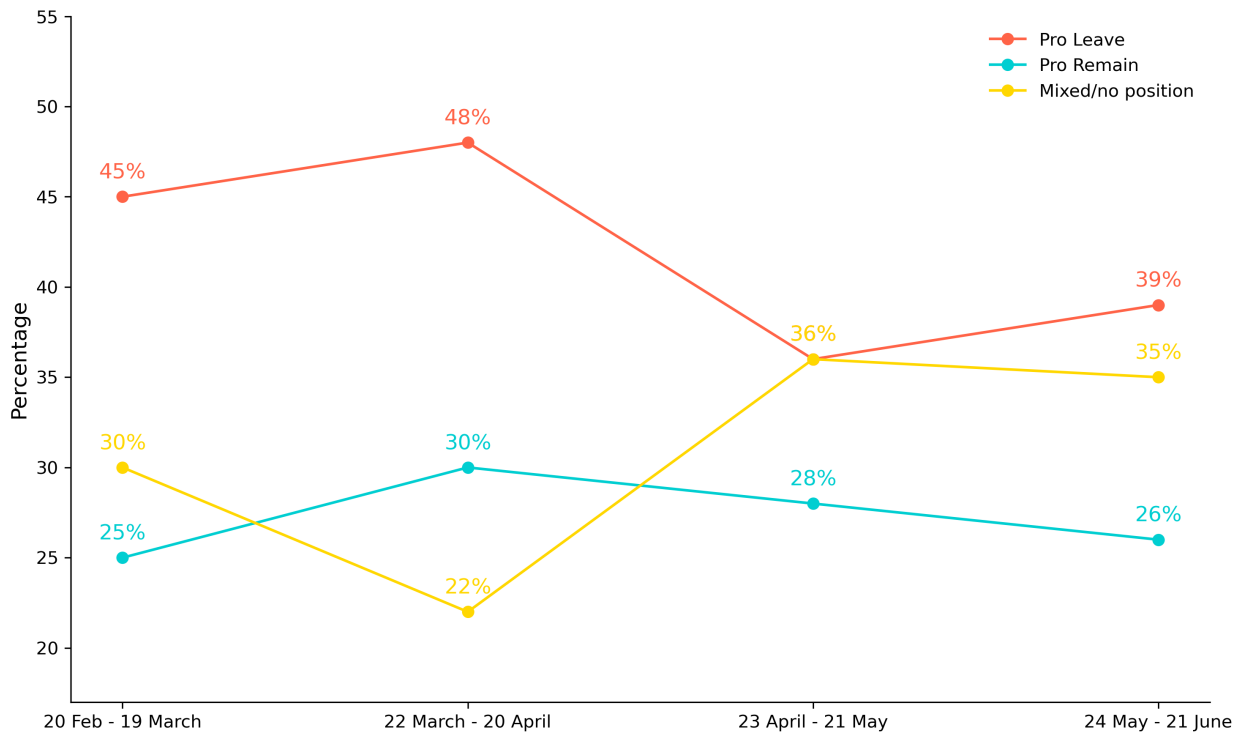


**Figure B.3:** Visual Media Bias During the 2016 Brexit Vote (Broadsheets)



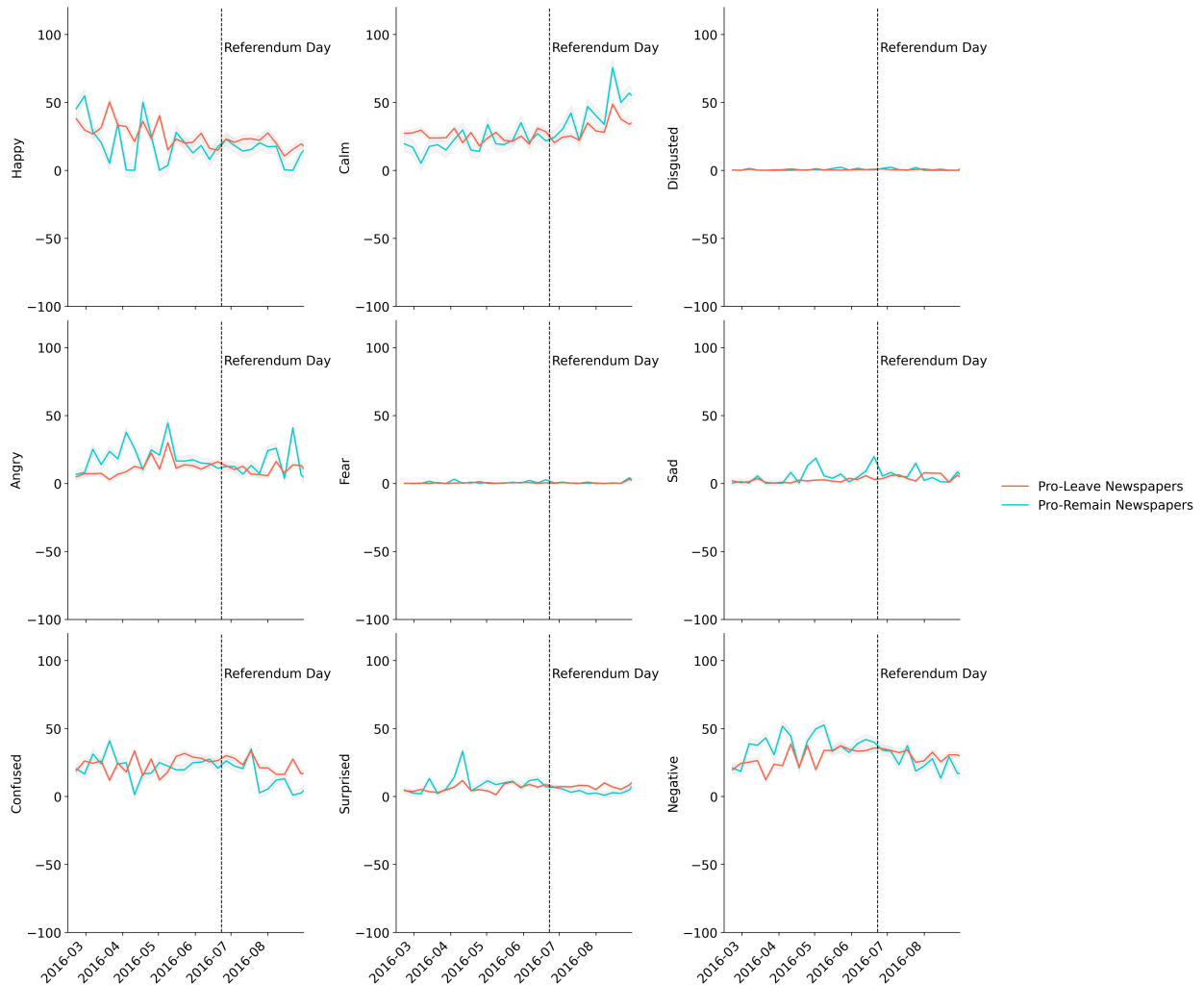
Notes: Panel A plots the shares of specific politicians or groups of politicians appearing across pro-Leave and pro-Remain newspaper website images. The shares are calculated relative to the total number of detected politicians' faces. Panel B plots the average aggregate emotion scores for specific politicians or groups of politicians across Pro-Leave and Pro-Remain newspaper websites. Pro-Leave newspapers are those that publish more articles supporting leaving the EU, while Pro-Remain newspapers are those that publish more articles in favor of staying in the EU, based on [Levy et al. \(2016\)](#).

**Figure B.4:** Position of the Brexit Referendum Articles Over Time from [Levy et al. \(2016\)](#)



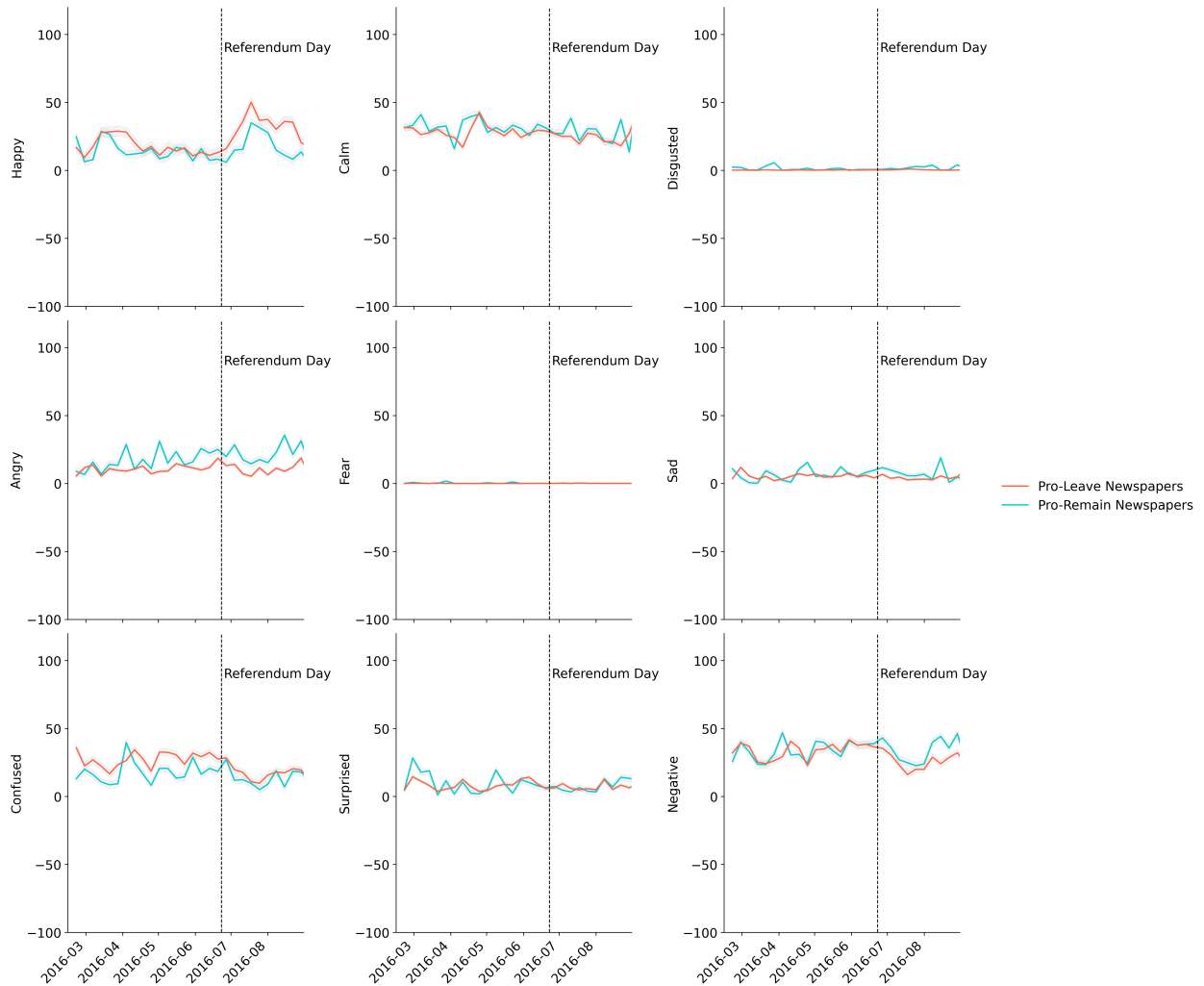
Notes: This figure plots the percentages of pro-Leave, pro-Remain, and mixed/no position articles over four time periods during the Brexit campaign period: 20 Feb - 19 March, 22 March - 20 April, 23 April - 21 May, and 24 May - 21 June. Data is based on the analysis by [Levy et al. \(2016\)](#).

**Figure B.5:** Dynamics of Visual Media Bias During the Brexit Vote - Pro-Leave Politicians



Notes: The figure shows the emotions used to portray pro-Leave politicians, across the set of pro-Leave and pro-Remain newspapers respectively for each week. Pro-Leave newspapers are those that publish more articles supporting leaving the EU, while Pro-Remain newspapers are those that publish more articles in favor of staying in the EU, based on [Levy et al. \(2016\)](#). The shaded grey areas represent 95% confidence intervals.

**Figure B.6:** Dynamics of Visual Media Bias During the Brexit Vote - Pro-Remain Politicians



Notes: The figure shows the emotions used to portray pro-Remain politicians, across the set of pro-Leave and pro-Remain newspapers respectively for each week. Pro-Leave newspapers are those that publish more articles supporting leaving the EU, while Pro-Remain newspapers are those that publish more articles in favor of staying in the EU, based on [Levy et al. \(2016\)](#). The shaded grey areas represent 95% confidence intervals.

**Table B.1:** Politicians and Their Roles During the Referendum Period (21 February – 31 August 2016)

<b>Politician</b>	<b>Roles</b>
<b>Pro-Leave Politicians</b>	
Boris Johnson	Mayor of London (until May 2016); Prominent Leave campaigner afterwards
Nigel Farage (UKIP)	Leader of the United Kingdom Independence Party (UKIP)
Michael Gove	Secretary of State for Justice
Chris Grayling	Leader of the House of Commons
John Whittingdale	Secretary of State for Culture, Media and Sport (until July 2016)
Theresa Villiers	Secretary of State for Northern Ireland
Priti Patel	Minister of State for Employment (until July 2016); Secretary of State for International Development (from July 2016)
<b>Pro-Remain Politicians</b>	
David Cameron	Prime Minister of the United Kingdom (until July 2016)
Jeremy Corbyn (Labour)	Leader of the Opposition (Labour Party)
George Osborne	Chancellor of the Exchequer (until July 2016)
Theresa May	Home Secretary (until July 2016); Prime Minister (from July 2016)
Philip Hammond	Foreign Secretary (until July 2016); Chancellor of the Exchequer (from July 2016)
Sajid Javid	Secretary of State for Business, Innovation and Skills (until July 2016); Secretary of State for Communities and Local Government (from July 2016)
Alun Cairns	Secretary of State for Wales
Justine Greening	Secretary of State for International Development (until July 2016); Secretary of State for Education (from July 2016)
Jeremy Hunt	Secretary of State for Health
Greg Clark	Minister for Universities, Science and Cities (until July 2016); Secretary of State for Business, Energy and Industrial Strategy (from July 2016)
Patrick McLoughlin	Secretary of State for Transport (until July 2016); Conservative Party Chairman (from July 2016)
Elizabeth Truss	Secretary of State for Environment, Food and Rural Affairs (until July 2016); Lord Chancellor and Secretary of State for Justice (from July 2016)
Oliver Letwin	Minister for Government Policy (until July 2016); Chancellor of the Duchy of Lancaster (until July 2016)
Nicky Morgan	Secretary of State for Education (until July 2016)
David Mundell	Secretary of State for Scotland
Michael Fallon	Secretary of State for Defence
Amber Rudd	Secretary of State for Energy and Climate Change (until July 2016); Home Secretary (from July 2016)
Greg Hands	Chief Secretary to the Treasury (until July 2016); Minister of State for Trade and Investment (from July 2016)
Mark Harper	Chief Whip of the House of Commons (until July 2016)
Anna Soubry	Minister for Small Business, Industry and Enterprise (until July 2016)

Notes: The table lists politicians and their official roles during the referendum period, specifically between 21 February 2016 and 31 August 2016. If a politician held different roles during this period, both positions are listed with the date of transition noted. All politicians are members of the Conservative Party unless otherwise indicated by their party affiliation in parentheses, e.g., Labour for the Labour Party or UKIP for the United Kingdom Independence Party.

**Table B.2:** Number of Politician Faces

Politician	Tabloids		Broadsheets		Total
	Pre-Referendum	Post-Referendum	Pre-Referendum	Post-Referendum	
<b>Pro-Leave Politicians</b>					
Boris Johnson	1589	1081	352	245	3267
Nigel Farage	775	914	137	128	1954
Michael Gove	592	544	159	122	1417
Chris Grayling	178	90	45	24	337
John Whittingdale	110	23	36	4	173
Theresa Villiers	92	22	28	11	153
Priti Patel	262	64	67	17	410
<b>Pro-Remain Politicians</b>					
David Cameron	2554	1143	587	249	4533
Jeremy Corbyn	526	831	163	218	1738
George Osborne	702	351	223	91	1367
Theresa May	170	1604	32	438	2244
Philip Hammond	62	227	13	63	365
Sajid Javid	50	72	16	19	157
Alun Cairns	10	17	1	7	35
Justine Greening	15	86	3	24	128
Jeremy Hunt	35	77	15	20	147
Greg Clark	3	39	1	13	56
Patrick McLoughlin	5	29	6	3	43
Elizabeth Truss	46	44	24	12	126
Oliver Letwin	8	25	0	3	36
Nicky Morgan	40	56	17	9	122
David Mundell	5	28	3	15	51
Michael Fallon	15	38	6	7	66
Amber Rudd	47	108	21	38	214
Greg Hands	23	5	8	1	37
Mark Harper	26	12	10	3	51
Anna Soubry	6	50	0	5	61
Total	7946	7580	1973	1789	19288

Notes: The table reports the number of faces of politicians appearing in Tabloids and Broadsheets before and after the Brexit referendum. The total row is the sum across all politicians.

**Table B.3:** Average Emotion by Politician, Pre-Referendum, Tabloids

Politician	Happ.	Calm	Disgusted	Angry	Fear	Sad	Confused	Surprised	Pos.	Neg.	AggregateEmo.
<b>Pro-Leave Politicians</b>											
Boris Johnson	12.40	20.87	0.39	12.75	0.33	2.75	40.91	7.88	33.25	42.02	-8.77
Nigel Farage	21.74	36.41	0.87	15.60	0.87	6.46	9.08	8.00	58.13	26.43	31.70
Michael Gove	19.86	24.82	1.37	17.19	0.53	3.90	24.35	7.59	44.66	35.50	9.16
Chris Grayling	29.61	41.64	0.04	1.41	0.01	1.17	11.70	2.33	71.21	10.77	60.45
John Whittingdale	72.09	16.31	0.11	0.28	0.02	0.79	6.76	1.70	88.35	6.25	82.11
Theresa Villiers	93.20	3.04	0.05	0.09	0.11	0.00	0.17	2.17	96.20	1.67	94.52
Priti Patel	72.19	8.47	1.12	3.57	0.22	0.05	2.33	6.97	80.63	9.21	71.41
Total	24.73	24.40	0.67	12.16	0.44	3.33	24.99	7.18	49.11	31.53	17.58
<b>Pro-Remain Politicians</b>											
David Cameron	12.23	27.58	0.39	3.77	0.06	7.28	36.04	12.75	39.79	38.97	0.82
Jeremy Corbyn	1.75	16.76	0.26	52.76	0.02	4.60	20.95	4.01	18.50	53.40	-34.89
George Osborne	32.41	37.20	0.65	17.75	0.01	0.99	2.29	0.59	69.58	14.40	55.18
Philip Hammond	13.84	63.83	0.30	6.48	1.36	2.50	4.89	1.39	77.63	10.94	66.69
Theresa May	20.98	29.15	1.68	12.30	0.06	7.92	18.82	7.69	50.11	31.33	18.78
Sajid Javid	18.73	48.62	0.14	4.23	0.14	1.63	16.95	9.63	67.32	21.16	46.17
Alun Cairns	26.70	38.42	0.08	0.58	0.12	18.37	16.38	0.39	65.09	23.22	41.87
Justine Greening	11.51	64.34	0.11	2.73	0.28	5.37	8.25	1.14	75.82	11.55	64.27
Jeremy Hunt	40.46	26.51	0.11	10.12	0.98	0.74	7.68	4.91	66.94	15.87	51.07
Greg Clark	33.35	30.72	0.04	30.50	0.12	0.05	1.90	0.23	64.03	21.22	42.81
Patrick McLoughlin	66.75	8.16	0.64	2.33	0.00	0.49	13.76	0.32	74.88	11.34	63.54
Elizabeth Truss	22.81	58.48	0.84	1.13	0.23	3.70	2.86	3.28	81.25	7.79	73.46
Oliver Letwin	36.86	23.10	0.24	0.48	0.00	25.05	0.28	10.52	59.93	23.64	36.29
Nicky Morgan	33.28	34.13	0.55	16.29	0.02	0.86	6.71	5.63	67.37	19.42	47.95
David Mundell	17.66	32.46	0.04	6.42	0.29	8.83	9.62	19.89	50.09	29.15	20.93
Michael Fallon	10.59	49.84	0.06	13.00	0.10	6.83	15.24	0.47	60.40	23.09	37.31
Amber Rudd	21.23	33.92	0.85	21.79	0.24	14.33	4.76	0.17	55.12	27.24	27.88
Greg Hands	12.44	45.80	0.03	1.47	0.62	22.81	19.99	0.20	58.21	29.17	29.04
Mark Harper	4.28	56.48	0.04	2.95	1.26	31.89	4.54	0.11	60.73	26.37	34.36
Anna Soubry	35.83	5.88	5.42	2.62	0.04	0.01	15.51	33.60	41.69	36.98	4.71
Total	15.43	29.54	0.47	12.68	0.09	6.02	25.58	8.73	44.95	34.62	10.32

Notes: The table reports the average emotion scores across politicians in Tabloids during the pre-referendum period.

**Table B.4:** Average Emotion by Politician, Pre-Referendum, Broadsheets

Politician	Happ.	Calm	Disgusted	Angry	Fear	Sad	Confused	Surprised	Pos.	Neg.	AggregateEmo.
<b>Pro-Leave Politicians</b>											
Boris Johnson	10.82	19.09	0.21	10.14	0.41	3.51	45.22	9.88	29.90	44.84	-14.95
Nigel Farage	24.10	41.68	1.12	13.42	0.50	4.40	6.53	6.02	65.75	20.68	45.07
Michael Gove	21.50	30.07	1.17	14.00	0.85	2.38	16.41	12.71	51.55	30.72	20.83
Chris Grayling	26.59	46.36	0.01	1.18	0.01	1.33	8.95	2.67	72.91	9.15	63.76
John Whittingdale	72.16	11.61	0.00	1.36	0.07	0.25	8.30	4.30	83.73	9.23	74.50
Theresa Villiers	95.11	1.49	0.00	0.20	0.00	0.01	0.02	0.11	96.55	0.23	96.32
Priti Patel	69.89	9.25	0.76	7.68	0.19	0.16	2.79	7.81	79.10	12.53	66.57
Total	26.30	24.73	0.56	10.02	0.44	2.79	24.65	8.65	51.00	30.45	20.55
<b>Pro-Remain Politicians</b>											
David Cameron	15.73	25.43	0.53	4.21	0.03	5.39	35.07	12.58	41.14	37.37	3.76
Jeremy Corbyn	3.09	14.38	0.31	50.80	0.00	3.71	25.47	3.65	17.46	54.26	-36.81
George Osborne	23.13	50.69	0.10	13.77	0.04	0.60	2.56	0.83	73.79	11.57	62.22
Philip Hammond	22.53	40.27	0.01	3.63	0.07	0.13	17.95	3.14	62.78	16.11	46.67
Theresa May	5.58	41.50	1.89	9.38	0.01	4.15	19.04	22.23	47.06	36.65	10.41
Sajid Javid	40.13	27.71	0.05	1.52	0.11	1.52	20.11	7.79	67.81	20.11	47.71
Alun Cairns	0.05	83.25	0.27	0.58	0.00	9.41	0.04	0.00	83.26	6.65	76.61
Justine Greening	0.01	87.93	0.01	0.30	1.24	4.14	0.77	1.83	87.90	5.35	82.55
Jeremy Hunt	32.84	28.42	0.04	7.91	0.09	0.21	3.09	23.04	61.23	22.22	39.01
Greg Clark	93.16	0.00	0.02	0.05	0.02	0.00	0.42	0.37	93.12	0.56	92.56
Patrick McLoughlin	17.18	48.16	0.37	13.99	0.24	2.06	10.83	2.26	65.31	19.23	46.08
Elizabeth Truss	13.53	70.15	0.28	0.38	0.21	2.42	2.54	4.22	83.64	6.50	77.14
Nicky Morgan	35.33	33.45	0.17	9.01	2.18	0.85	7.80	16.24	68.74	23.43	45.31
David Mundell	27.04	29.52	0.29	10.39	0.02	27.08	3.08	0.32	56.53	26.62	29.91
Michael Fallon	0.02	36.33	0.08	9.93	0.01	0.14	54.33	0.85	36.33	42.23	-5.90
Amber Rudd	8.31	48.86	2.15	13.97	3.67	13.24	2.27	0.32	57.14	23.03	34.12
Greg Hands	12.53	36.43	4.99	3.08	0.99	11.37	24.35	0.67	48.93	29.38	19.55
Mark Harper	0.62	60.75	0.06	0.74	0.94	33.35	2.78	0.34	61.34	24.70	36.63
Total	15.65	31.69	0.47	13.04	0.15	4.32	23.84	8.53	47.32	32.54	14.78

Notes: The table reports the average emotion scores across politicians in Broadsheets during the pre-referendum period.

**Table B.5:** Type, Partisanship and Counts by Newspapers: Whole Period

Newspaper	Type	Partisanship Score	Count of Images	Count of Politician Faces
Mirror	Tabloid	0.00	8865	2852
Guardian	Broadsheet	2.87	7649	1079
Financial Times	Broadsheet	8.72	5017	569
Telegraph	Broadsheet	52.53	7716	2114
Sun	Tabloid	56.39	6179	2609
Daily Mail	Tabloid	75.62	14179	5407
Express	Tabloid	100.00	14484	4658
		Total	64089	19288

Notes: The table reports type, partisanship score, count of images, and count of detected politician faces for each newspaper.

**Table B.6:** Newspaper Partisanship and Politician Emotions

	Dependent Variable: Emotion Score $\times 100$									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Happy (P)	Calm (P)	Disgusted (N)	Angry (N)	Fear (N)	Sad (N)	Confused (N)	Surprised (N)	Negative	AggregateEmo.
<b>Panel A: All images</b>										
Partisanship x	-0.284	5.630***	0.190	-1.368**	-0.151	-0.295	-3.911**	-1.115	-4.298***	9.640***
PoliticianStance	(0.838)	(0.699)	(0.225)	(0.281)	(0.084)	(0.295)	(0.678)	(0.713)	(0.167)	(0.431)
Observations	7946	7946	7946	7946	7946	7946	7946	7946	7946	7946
<b>Panel B: Front page images</b>										
Partisanship x	0.398	6.640**	1.083	-1.747***	-1.100***	-1.448**	-4.629***	-0.259	-5.235***	12.270***
PoliticianStance	(0.953)	(1.224)	(0.416)	(0.237)	(0.132)	(0.344)	(0.708)	(0.279)	(0.320)	(0.722)
Observations	2026	2026	2026	2026	2026	2026	2026	2026	2026	2026
<b>Panel C: Content page images</b>										
Partisanship x	-0.957	5.155***	-0.133	-1.167	0.165	0.031	-2.982**	-1.487	-3.602***	7.800***
PoliticianStance	(0.570)	(0.498)	(0.171)	(0.414)	(0.103)	(1.092)	(0.634)	(1.014)	(0.277)	(0.756)
Observations	5917	5917	5917	5917	5917	5917	5917	5917	5917	5917
<b>Panel D: All images, post-referendum period</b>										
Partisanship x	-1.447	-1.446	0.187	-0.079	-0.120**	0.444	1.055	1.309	1.808	-4.699
PoliticianStance	(1.385)	(0.738)	(0.111)	(0.767)	(0.034)	(0.841)	(0.903)	(1.208)	(1.230)	(2.877)
Observations	7580	7580	7580	7580	7580	7580	7580	7580	7580	7580
Clusters	4	4	4	4	4	4	4	4	4	4
Newspaper F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Politician F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The table presents estimates of  $\beta_1$  from Eq.2.1, where the dependent variable is the emotion score derived from the face of the politician and fixed effects for newspapers and politicians are included. *PoliticianStance* is coded as -1 for Remain and +1 for Leave to address identification concerns associated with the interaction term being driven primarily by the small number of Leave politicians in a 0/1 coding scheme. Consequently, the interaction coefficient in this -1/1 specification is half the magnitude of what it would be under a 0/1 coding scheme. We also provide results using 0/1 coding in Table X, which confirm the robustness of our findings. Panel B focuses exclusively on the front page images. Panel C focuses exclusively on the content page images. ‘NewspaperPartisanship  $\times$  PoliticianStance’ is the interaction between the newspaper partisanship scores and *PoliticianStance*. The column headers specify the emotion score used as the dependent variable in each regression. The (P) and (N) labels after each emotion in the column headers indicate whether the emotion is classified as positive (P) or negative (N). ‘Negative’ indicates the sum of disgusted, angry, fear, sad, confused, and surprised. ‘AggregateEmo.’ indicates the aggregate emotion score that is the sum of happy and calm minus negative. Standard errors clustered by newspaper are in parentheses. \*\*\* indicates significance at the one percent level, \*\* indicates significance at the five percent level.

**Table B.7:** Newspaper Partisanship and Politician Emotions: Tabloids

	Dependent Variable: Emotion Score $\times 100$								(9)	(10)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)		
	Happy (P)	Calm (P)	Disgusted (N)	Angry (N)	Fear (N)	Sad (N)	Confused (N)	Surprised (N)		
<b>Panel A: Cameron and Johnson images</b>										
Partisanship x	1.418	3.822	0.317	-1.237	0.0627	1.804	-4.071	-2.111	-3.385	8.622
PoliticianStance	(0.619)	(4.318)	(0.228)	(2.259)	(0.0843)	(2.489)	(1.979)	(0.697)	(2.599)	(6.821)
Observations	4143	4143	4143	4143	4143	4143	4143	4143	4143	4143
<b>Panel B: Cameron and Farage images</b>										
Partisanship x	-2.710	22.93***	0.349	2.332	-1.640**	-2.249	-20.44***	-4.036	-16.61***	36.81***
PoliticianStance	(0.893)	(1.085)	(0.164)	(1.836)	(0.384)	(1.704)	(1.547)	(3.467)	(1.365)	(3.341)
Observations	3329	3329	3329	3329	3329	3329	3329	3329	3329	3329
<b>Panel C: Cameron, Johnson and Farage images</b>										
Partisanship x	-0.230	11.33**	0.327	0.145	-0.597**	0.210	-10.40***	-2.906	-8.549***	19.65***
PoliticianStance	(0.273)	(2.334)	(0.205)	(1.217)	(0.182)	(2.165)	(1.704)	(1.735)	(1.209)	(3.283)
Observations	4918	4918	4918	4918	4918	4918	4918	4918	4918	4918
Clusters	4	4	4	4	4	4	4	4	4	4
Newspaper F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Politician F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The table reports the results from OLS regression, where the dependent variable is the emotion score derived from the face of the politician and fixed effects for newspapers and politicians are included. Panel A-C focuses exclusively on two particular politicians from different camps. The column headers specify the emotion score used as the dependent variable in each regression. 'Negative' indicates the sum of disgusted, angry, fear, sad, confused, and surprised. 'AggregateEmo.' indicates the aggregate emotion score that is the sum of happy and calm minus negative. Standard errors clustered by newspaper are in parentheses. \*\*\* indicates significance at the one percent level, \*\* indicates significance at the five percent level.

**Table B.8:** Newspaper Partisanship and Politician Emotions: Excluding Surprised Emotion, Tabloids, Pre-Referendum

	Tabloids		Broadsheets	
	(1) Negative	(2) AggregateEmo.	(3) Negative	(4) AggregateEmo.
<b>Panel A: All images</b>				
Partisanship x	-8.063***	18.75***	9.097	-22.12
PoliticianStance	(0.858)	(0.876)	(6.329)	(17.98)
Observations	7946	7946	1971	1971
<b>Panel B: Front page images</b>				
Partisanship x	-11.42***	25.49***	13.06	-30.85
PoliticianStance	(0.740)	(1.452)	(5.979)	(16.51)
Observations	2026	2026	1300	1300
<b>Panel C: Content page images</b>				
Partisanship x	-5.985**	14.38**	-7.273**	15.11*
PoliticianStance	(1.755)	(2.531)	(0.562)	(1.699)
Observations	5917	5917	670	670
<b>Panel D: All images, post-referendum period</b>				
Partisanship x	2.167	-7.950	8.584	-27.08
PoliticianStance	(1.997)	(5.071)	(5.472)	(11.83)
Observations	7580	7580	1788	1788
Clusters	4	4	3	3

Notes: The table reports regression results where the dependent variables are either the negative emotion score or the aggregate emotion score, with the surprised emotion excluded. Fixed effects for newspapers and politicians are included in all regressions. Columns (1)-(2) report results for tabloids, Columns (3)-(4) report results for broadsheets. \*\*\* indicates significance at the one percent level, \*\* indicates significance at the five percent level.

**Table B.9:** Newspaper Partisanship and Politician Emotions: Excluding Each Emotion Individually, Tabloids, Pre-Referendum

	Dependent Variable: Aggregate Emotion Score $\times 100$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	$\sim$ Happy	$\sim$ Calm	$\sim$ Disgusted	$\sim$ Angry	$\sim$ Fear	$\sim$ Sad	$\sim$ Confused
<b>Panel A: All images</b>							
Partisanship x	19.85***	8.03**	19.53***	17.56***	19.10***	20.02***	14.22***
PoliticianStance	(1.291)	(1.915)	(0.838)	(0.623)	(0.898)	(1.494)	(1.659)
<b>Panel B: Front page images</b>							
Partisanship x	25.08***	12.60***	27.69***	23.32***	24.27***	24.14***	19.28***
PoliticianStance	(3.085)	(1.514)	(2.100)	(1.550)	(1.406)	(1.059)	(1.384)
<b>Panel C: Content page images</b>							
Partisanship x	17.52***	5.290**	15.43***	14.14***	15.83***	16.63***	11.74**
PoliticianStance	(1.218)	(1.572)	(1.641)	(0.982)	(1.430)	(0.883)	(2.239)
Newspaper F.E.	Y	Y	Y	Y	Y	Y	Y
Politician F.E.	Y	Y	Y	Y	Y	Y	Y

Notes: The table reports the regression results where the aggregate emotion score (dependent variable) is calculated by excluding one emotion at a time. Each column corresponds to a regression where the specified emotion is excluded. For example, " $\sim$ Happy" means the aggregate emotion score is calculated excluding the "Happy" emotion. The regression controls for both newspaper and politician fixed effects. The regression controls for both newspaper and politician fixed effects. \*\*\* indicates significance at the one percent level, \*\* indicates significance at the five percent level.

**Table B.10:** Newspaper Stance (0 for Pro-Remain and 1 for Pro-Leave) and Politician Emotions: Tabloids

	Dependent Variable: Emotion Score									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Happy (P)	Calm (P)	Disgusted (N)	Angry (N)	Fear (N)	Sad (N)	Confused (N)	Surprised (N)	Negative	AggregateEmo.
<b>Panel A: All images</b>										
NewspaperStance x	-1.846	10.48***	0.596**	-2.359**	-0.379***	-1.438	-6.485**	-0.704	-6.962***	15.59***
PoliticianStance	(0.641)	(0.674)	(0.181)	(0.698)	(0.0585)	(1.303)	(1.964)	(0.832)	(0.951)	(2.194)
Observations	7946	7946	7946	7946	7946	7946	7946	7946	7946	7946
<b>Panel B: Front page images</b>										
NewspaperStance x	-0.916	13.73***	2.497***	-3.099**	-2.113***	-3.023**	-8.842**	-0.315	-9.628***	22.43***
PoliticianStance	(0.867)	(0.146)	(0.318)	(0.600)	(0.175)	(0.553)	(1.574)	(0.596)	(0.588)	(1.360)
Observations	2026	2026	2026	2026	2026	2026	2026	2026	2026	2026
<b>Panel C: Content page images</b>										
NewspaperStance x	-2.353	9.287***	-0.0148	-2.001	0.147	-0.930	-5.277**	-0.952	-5.835***	12.77***
PoliticianStance	(0.930)	(0.809)	(0.169)	(0.799)	(0.132)	(1.499)	(1.427)	(1.105)	(0.845)	(1.997)
Observations	5917	5917	5917	5917	5917	5917	5917	5917	5917	5917
Clusters	4	4	4	4	4	4	4	4	4	4
Newspaper F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Politician F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The table reports the results from OLS regression, where the dependent variable is the emotion score derived from the face of the politician and fixed effects for newspapers and politicians are included. Panel A-C focuses exclusively on tabloid newspapers. The column headers specify the emotion score used as the dependent variable in each regression. The (P) and (N) labels after each emotion in the column headers indicate whether the emotion is classified as positive (P) or negative (N). 'Negative' indicates the sum of disgusted, angry, fear, sad, confused, and surprised. 'AggregateEmo.' indicates the aggregate emotion score that is the sum of happy and calm minus negative. Standard errors clustered by newspaper are in parentheses. \*\*\* indicates significance at the one percent level, \*\* indicates significance at the five percent level.

**Table B.11:** Newspaper Partisanship and Politician Emotions: Tabloids, April 15 - Referendum date

	Dependent Variable: Emotion Score $\times 100$								(9)	(10)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)		
	Happy (P)	Calm (P)	Disgusted (N)	Angry (N)	Fear (N)	Sad (N)	Confused (N)	Surprised (N)		
<b>Panel A: All images</b>										
Partisanship x	1.677	7.948***	-0.258	0.142	-0.252	-1.025	-7.219	-3.427**	-7.782***	17.40***
PoliticianStance	(1.391)	(1.272)	(0.443)	(0.647)	(0.296)	(1.591)	(2.337)	(0.818)	(0.782)	(1.154)
Observations	5452	5452	5452	5452	5452	5452	5452	5452	5452	5452
<b>Panel B: Front page images</b>										
Partisanship x	0.869	13.67**	0.551	-2.679**	-2.783**	-4.261	-7.636**	0.653	-10.44**	24.98**
PoliticianStance	(1.570)	(3.990)	(0.572)	(0.598)	(0.501)	(1.397)	(1.656)	(0.783)	(1.840)	(4.488)
Observations	1402	1402	1402	1402	1402	1402	1402	1402	1402	1402
<b>Panel C: Content page images</b>										
Partisanship x	1.052	5.060**	-0.583	1.589	0.566	0.130	-5.575	-4.727**	-5.559***	11.67***
PoliticianStance	(1.234)	(1.465)	(0.426)	(0.730)	(0.318)	(1.669)	(2.198)	(1.047)	(0.819)	(0.910)
Observations	4046	4046	4046	4046	4046	4046	4046	4046	4046	4046
Clusters	4	4	4	4	4	4	4	4	4	4
Newspaper F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Politician F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The table reports the results from OLS regression, where the dependent variable is the emotion score derived from the face of the politician and fixed effects for newspapers and politicians are included. Panel A-C focuses on the period between April 15 and the Referendum date, using tabloid newspapers. The (P) and (N) labels after each emotion in the column headers indicate whether the emotion is classified as positive (P) or negative (N). ‘Negative’ indicates the sum of disgusted, angry, fear, sad, confused, and surprised. ‘AggregateEmo.’ indicates the aggregate emotion score that is the sum of happy and calm minus negative. Standard errors clustered by newspaper are in parentheses. \*\*\* indicates significance at the one percent level, \*\* indicates significance at the five percent level.

**Table B.12:** Newspaper Partisanship and Politician Emotions: Broadsheets

	Dependent Variable: Emotion Score $\times 100$								(9)	(10)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)		
	Happy (P)	Calm (P)	Disgusted (N)	Angry (N)	Fear (N)	Sad (N)	Confused (N)	Surprised (N)	Negative	AggregateEmo.
<b>Panel A: All images</b>										
Partisanship x	-10.85	-2.179	1.165	0.669	-0.737	1.423**	9.945	2.745	9.833	-22.85
PoliticianStance	(10.34)	(1.436)	(1.065)	(0.700)	(0.534)	(0.156)	(8.791)	(2.588)	(7.275)	(18.93)
Observations	1971	1971	1971	1971	1971	1971	1971	1971	1971	1971
<b>Panel B: Front page images</b>										
Partisanship x	-5.759	-12.04	0.269	5.141	-0.613	4.420***	7.829	3.779	13.46	-31.25
PoliticianStance	(7.413)	(3.145)	(0.657)	(1.315)	(0.472)	(0.251)	(9.041)	(4.201)	(7.759)	(18.29)
Observations	1300	1300	1300	1300	1300	1300	1300	1300	1300	1300
<b>Panel C: Content page images</b>										
Partisanship x	-2.887	10.73**	6.141	-11.43***	-2.093**	-3.034**	0.444	8.443	-0.986	8.827
PoliticianStance	(0.759)	(0.379)	(0.602)	(0.163)	(0.112)	(0.0601)	(1.586)	(1.071)	(0.195)	(0.943)
Observations	670	670	670	670	670	670	670	670	670	670
<b>Panel D: All images, post-referendum period</b>										
Partisanship x	3.885	-22.39	0.131	1.560	0.662	-6.832***	16.24	4.020**	10.20	-28.70
PoliticianStance	(8.649)	(14.89)	(0.269)	(1.782)	(0.505)	(0.199)	(6.052)	(0.582)	(5.074)	(11.43)
Observations	1788	1788	1788	1788	1788	1788	1788	1788	1788	1788
Clusters	3	3	3	3	3	3	3	3	3	3
Newspaper F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Politician F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The table reports the results from OLS regression, where the dependent variable is the emotion score derived from the face of the politician and fixed effects for newspapers and politicians are included. Panel A-D focuses exclusively on the specified politician and image type. The (P) and (N) labels after each emotion in the column headers indicate whether the emotion is classified as positive (P) or negative (N). 'Negative' indicates the sum of disgusted, angry, fear, sad, confused, and surprised. 'AggregateEmo.' indicates the aggregate emotion score that is the sum of happy and calm minus negative. Standard errors clustered by newspaper are in parentheses. \*\*\* indicates significance at the one percent level, \*\* indicates significance at the five percent level.

**Table B.13:** Newspaper Partisanship and Politician Emotions: All Newspapers

	Dependent Variable: Emotion Score $\times 100$								(9)	(10)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)		
	Happy (P)	Calm (P)	Disgusted (N)	Angry (N)	Fear (N)	Sad (N)	Confused (N)	Surprised (N)		
<b>Panel A: All images</b>										
Partisanship x	-1.287	8.397**	0.460	-1.573	-0.270	-0.809	-5.155	-1.922	-5.991**	13.10
PoliticianStance	(1.861)	(2.945)	(0.398)	(1.012)	(0.188)	(1.555)	(3.040)	(1.057)	(2.246)	(5.353)
Observations	9919	9919	9919	9919	9919	9919	9919	9919	9919	9919
<b>Panel B: Front page images</b>										
Partisanship x	-1.218	6.441	1.103	-0.148	-1.059	-1.268	-4.054	-1.273	-4.329	9.550
PoliticianStance	(2.656)	(3.827)	(0.822)	(1.671)	(0.674)	(1.007)	(3.890)	(1.514)	(3.780)	(9.037)
Observations	3330	3330	3330	3330	3330	3330	3330	3330	3330	3330
<b>Panel C: Content page images</b>										
Partisanship x	-0.692	9.339***	0.0978	-2.518**	0.218	-0.0707	-6.210***	-2.525	-7.116***	15.76***
PoliticianStance	(2.313)	(2.165)	(0.472)	(0.880)	(0.219)	(2.041)	(1.033)	(1.671)	(0.630)	(1.561)
Observations	6588	6588	6588	6588	6588	6588	6588	6588	6588	6588
<b>Panel D: All images, post-referendum period</b>										
Partisanship x	-2.812	-3.472	0.258	-0.361	-0.367	0.451	3.411	2.852	4.037	-10.32
PoliticianStance	(2.510)	(3.282)	(0.216)	(1.375)	(0.160)	(1.355)	(2.004)	(2.225)	(2.035)	(4.862)
Observations	9369	9369	9369	9369	9369	9369	9369	9369	9369	9369
Clusters	7	7	7	7	7	7	7	7	7	7
Newspaper F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Politician F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The table reports the results from OLS regression, where the dependent variable is the emotion score derived from the face of the politician and fixed effects for newspapers and politicians are included. Panel A-D focuses on different groups of images. The (P) and (N) labels after each emotion in the column headers indicate whether the emotion is classified as positive (P) or negative (N). ‘Negative’ indicates the sum of disgusted, angry, fear, sad, confused, and surprised. ‘AggregateEmo.’ indicates the aggregate emotion score that is the sum of happy and calm minus negative. Standard errors clustered by newspaper are in parentheses. \*\*\* indicates significance at the one percent level, \*\* indicates significance at the five percent level.

**Table B.14:** Newspaper Partisanship and Politician Emotions by Politicians

	Pro-Leave Politicians		Pro-Remain Politicians	
	(1)	(2)	(3)	(4)
	Negative	AggregateEmo.	Negative	AggregateEmo.
Partisanship Score	-7.569***	16.87***	2.015	-4.572
	(1.130)	(2.843)	(2.506)	(5.733)
Clusters	4	4	4	4
Observations	3598	3598	4348	4348

Notes: The table reports the results, where the dependent variable is the emotion scores. ‘Negative’ indicates the sum of disgusted, angry, fear, sad, and confused. ‘AggregateEmo.’ indicates the aggregate emotion score that is the sum of happy and calm minus negative. Standard errors clustered by newspaper are in parentheses. \*\*\* indicates significance at the one percent level, \*\* indicates significance at the five percent level.

**Table B.15:** Newspaper Partisanship and Politician Contextual Image Sentiment: Broadsheets

	Dependent Variable: Contextual Image Sentiment			
	(1)	(2)	(3)	(4)
	All Images	Front Page Images	Content Page Images	All Images (Post)
Partisanship x	0.037	0.081	0.055	0.130
PoliticianStance	(0.081)	(0.116)	(0.030)	(0.045)
Observations	1742	1145	595	1571
Clusters	3	3	3	3
Newspaper F.E.	Y	Y	Y	Y
Politician F.E.	Y	Y	Y	Y

Notes: The table presents estimates of  $\beta_2$  from Eq. 2.2, where the dependent variable is the contextual image sentiment score of politician and fixed effects for newspapers and politicians are included. ‘NewspaperPartisanship  $\times$  PoliticianStance’ is the interaction between the newspaper partisanship scores with an indicator for whether the politician is Pro-leave. The column headers specify the image sample that used in each regression. Column (1)-(3) use the pre-Referendum sample. Column (4) use the sample that includes all images in post-Referendum period. Standard errors clustered by newspaper are in parentheses. \*\*\* indicates significance at the one percent level, \*\* indicates significance at the five percent level.

## Chapter Three

### Modelling the Effects of Brexit

### Uncertainty on the UK Economy

# Abstract

In this paper, we present new empirical evidence on the impact of Brexit-induced uncertainty on the UK economy. We employ a proxy structural vector autoregressive (proxy-SVAR) model and introduce a novel high-frequency instrument constructed from daily data to identify Brexit uncertainty shocks. Our findings reveal that Brexit uncertainty shocks exerted a temporary boost in economic activity, including increases in hours worked, GDP, and trade openness in goods towards the EU. These results align with theories suggesting that anticipation of future adverse conditions can prompt firms in the tradable sector to increase production and net exports in the short run. These results are robust across various identification strategies and alternative instruments for recovering Brexit uncertainty shocks.

JEL Codes: E37, C32, C36.

**Keywords:** Brexit uncertainty shocks, Proxy-SVAR, high-frequency instrument.

## 3.1 Introduction

More than fifty percent of the UK voters supported leaving the European Union (EU) leading to the UK's formal exit from the EU in January 2020. The period between 2016 and 2020 was marked by significant and sustained uncertainty for the UK economy, as evidenced by notable spikes in the UK economic policy uncertainty index with the most pronounced spike occurring during the referendum period. Moreover, following the referendum result, according to [Bloom et al. \(2019\)](#), over a third of firms identified Brexit as a primary source of uncertainty, a figure that rose to around fifty percent by April 2020.

Uncertainty plays a crucial role in shaping economic outcomes by affecting the decisions of firms, consumers, and policymakers. A substantial body of literature has emphasized the impact of uncertainty on the economy. For example, [Bloom \(2009\)](#) suggests that uncertainty causes firms to delay investment and hiring, leading to a reduction in economic activity and a slowdown in overall growth. By contrast, [Segal et al. \(2015\)](#) categorizes macroeconomic uncertainty into “good” and “bad” components, finding that “bad” uncertainty slows economic growth and reduces asset prices, while “good” uncertainty can boost economic activity, including consumption, output, and investment. Given the significant impact of uncertainty, it is essential to understand the role of Brexit-induced uncertainty on the UK economy.

In this paper, we estimate a proxy structural vector autoregressive model (proxy-SVAR) to examine the dynamic effects of Brexit uncertainty shocks on a series of macroeconomic aggregates, such as GDP, trade openness, and hours worked, for the UK economy from 2016 to 2022. We rely on the text-based Brexit Uncertainty Index (BUI) developed by [Chung et al. \(2022\)](#), as detailed in Chapter 1 of this thesis, to construct the external instruments for the identification of structural uncertainty shocks. The BUI reflects the frequency of news coverage regarding Brexit uncertainty.

Early empirical studies on uncertainty primarily rely on vector autoregressive (VAR) models with recursive (Cholesky) identification strategies to examine the economic impact of uncertainty (Bloom 2009, Bachmann et al. 2013, Caggiano et al. 2014, Jurado et al. 2015, Baker et al. 2016, Scotti 2016). These studies typically assume that uncertainty is exogenous, implying that uncertainty does not respond instantly to changes in economic variables, while economic variables have an instantaneous reaction to uncertainty. Subsequent studies, however, challenge the exogenous nature of uncertainty (Ludvigson et al. 2021, Carriero et al. 2018) and suggest that uncertainty may exhibit contemporaneous co-movement with real economic activity and react endogenously to macroeconomic shocks.

To address possible endogeneity concerns, this paper applies the proxy-SVAR model proposed by Stock & Watson (2012) and Mertens & Ravn (2013) to identify a structural shock using an external instrument. A pertinent study is Baker et al. (2024), in which they employ an external instrument constructed from a panel of indicators for natural disasters, terrorist attacks, coups, and revolutions to analyze the effects of uncertainty. Further, recognizing that financial markets can react rapidly to uncertainty shocks, low-frequency estimates, for example, using monthly data, may miss such rapid responses. To this end, we follow the approach of Alessandri et al. (2023) and identify structural shocks from a high-frequency VAR.

Specifically, we take a two-step approach. First, we estimate a daily VAR, including the daily BUI and a series of economic variables to recover the daily Brexit uncertainty shock series. Then, the shock series is aggregated to a monthly frequency by averaging the daily shock series. We use the resulting monthly shock series as an external instrument to identify the uncertainty shock in a proxy-SVAR. Two tests verify that the instrument is valid, specifically that our instrument is associated with structural uncertainty shocks and not significantly correlated with other contemporaneous shocks.

Some studies have documented the impact of Brexit. For example, [Dhingra et al. \(2017\)](#) examine the impact of Brexit on UK household welfare as measured by real consumption per capita in the medium to long term, focusing on trade and financial transfers, and argue that Brexit will lead to a loss of welfare of 1.3-2.7%. [Broadbent et al. \(2023\)](#) consider Brexit as an expected negative shock to productivity growth in the tradable sector and show that, in the short run, the Brexit vote outcome temporarily boosts the tradable sector and exports as firms accelerate production to take advantage of current conditions before the expected drop in productivity materializes. Notably, Brexit increased uncertainty (second moment) and may have lowered expectations for the future economy (first moment). [Bloom et al. \(2019\)](#) suggest that the metrics of the first and second moments are highly correlated, making it challenging to distinguish the effects of first and second moments. We thus, following [Bloom et al. \(2019\)](#), interpret our estimated effects as a combination of first and second moment effects.

Other studies have directly demonstrated the negative economic impacts of Brexit uncertainty, such as on investment and productivity ([Bloom et al. 2019](#)), employment ([Javorcik et al. 2020](#)), the EU-UK bilateral trade in goods ([Crowley et al. 2018](#), [Douch et al. 2020](#), [Graziano et al. 2021](#)), and trade in services ([Ahmad et al. 2023](#)). However, [Steinberg \(2019\)](#) model the costs of Brexit uncertainty and find that they are small, with a consumption-equivalent welfare cost of around 0.2%, compared with the long-term impacts of Brexit, which range from 0.4% to 1.2%, suggesting that the short-term macroeconomic impact of uncertainty on the UK economy is very limited.<sup>1</sup>

The main contribution of this paper is that we employ a proxy-SVAR and introduce

---

<sup>1</sup>Typically, the literature on economic uncertainty highlights a negative impact on the economy, as shown by, for example, [Bloom \(2014\)](#), [Baker et al. \(2016\)](#), [Baker et al. \(2024\)](#), and [Fernández-Villaverde et al. \(2015\)](#). Other studies, including [Gilchrist & Williams \(2005\)](#), [Jones et al. \(2005\)](#), [Kung & Schmid \(2015\)](#), and [Malkhozov \(2014\)](#), investigate alternative economic mechanisms that could lead to a positive relationship between uncertainty and investment, and thus economic growth.

a new instrument to identify Brexit uncertainty shocks, using the continuous Brexit Uncertainty measure provided by [Chung et al. \(2022\)](#), to investigate the dynamic causal effect of Brexit-induced uncertainty. This approach builds on the methodology used in prior studies, such as [Piffer & Podstawski \(2018\)](#) for identifying general uncertainty, [Ha et al. \(2022\)](#) for identifying geopolitical uncertainty, and [Miescu \(2023\)](#) for identifying uncertainties in emerging economies.

To briefly summarize our results, Brexit uncertainty shocks had a short-term positive impact on the real economy, leading to growth in GDP, hours worked, and trade openness in goods towards the EU, with these effects disappearing within around six months after the shock. This finding is consistent with [Broadbent et al. \(2023\)](#)'s research, which suggests that the referendum outcome temporarily stimulated the tradable sector, leading to a short-term expansion in economic activity before broader economic adjustments took effect. The results remain consistent across various identification strategies, including shock identification through a recursive scheme, employing monthly data in a proxy-SVAR, utilizing an informative prior in a Bayesian SVAR framework and shock identification through a sign-restricted daily VAR. Sensitivity checks suggest that our results are robust to the use of other proxies for Brexit uncertainty. Using the Brexit uncertainty series from the Decision Makers Panel to recover the Brexit shock generates macroeconomic responses that align with our results in direction.

The remainder of the paper proceeds as follow. Section 2 introduces the related literature. Section 3 describes our empirical methodology, i.e., the proxy-SVAR model. Section 4 introduces the data. Section 5 presents how we construct external instruments for Brexit uncertainty shocks. Section 6 depicts the results. Section 7 shows the sensitivity check. The last section concludes.

## 3.2 Literature Review

This paper relates to three main strands of the literature. The first strand of the literature documents the effects of Brexit itself. The second strand of the literature explicitly looks at the impacts of Brexit uncertainty. The third strand of the literature studies the effects of uncertainty shocks using VAR type models.

### 3.2.1 Brexit Effects

Several studies predicted the potential impact of Brexit before the EU referendum. For example, [Kierzenkowski et al. \(2016\)](#) argue that Brexit can be viewed as a tax on GDP, resulting in a sustained negative impact on the UK economy. [Ebell & Warren \(2016\)](#) identified three main channels through which Brexit would affect the UK economy: (1) increased trade barriers with EU countries, (2) a reduction in foreign direct investment (FDI), and (3) a decrease in the UK's financial contributions to the EU. [Dhingra et al. \(2016\)](#) suggested that a decline in trade as a result of Brexit might cost the UK economy more than the savings from reducing its contribution to the EU budget, leading to a decrease in per capita income.

More recently, [Broadbent et al. \(2023\)](#) employed a small open economy model with tradable and non-tradable sectors to analyze the impacts of the Brexit vote, taking the referendum outcome served as a signal for future declines in productivity growth within the tradable sector. This anticipation of lower future productivity led to a temporary boom in the tradable sector, as firms and investors sought to capitalize on current opportunities before the expected downturn. Meanwhile, the non-tradable sector experienced stagnation, as resources were diverted away from it in response to changing expectations. The reallocation of resources was primarily driven by shifts in relative prices: as the price of non-tradable goods fell relative to tradable goods, the profitability of producing and investing in the

tradable sector increased, prompting a reallocation of labor and capital. In the long run, however, UK GDP will be 3.6% lower than in the no-Brexit counterfactual scenario.

Some studies have documented the realized effects of Brexit. [Du et al. \(2023\)](#), using a synthetic difference-in-differences approach, showed that Brexit persistently and significantly decreased exports of goods. [Bailey et al. \(2023\)](#) found that new trade barriers particularly impacted small firms within manufacturing supply chains. Using synthetic control methods they reveal that up to March 2019, the Brexit referendum led to a 17% increase in UK investment deals into EU countries, driven by UK firms establishing European subsidiaries to maintain access to EU markets. In contrast, the number of outward investment deals from EU countries into the UK declined. From a labor market perspective, [Di Iasio & Wahba \(2023\)](#) using a Difference-in-Differences approach showed a decrease in net migration flows from EU countries to the UK post-referendum.

### 3.2.2 Brexit Uncertainty Effects

The influential research by [Bloom \(2009\)](#) sparked significant interest in understanding the relationships between uncertainty and different macroeconomic variables and explored the underlying transmission mechanisms. There is a small existing literature that has looked at how uncertainty related to Brexit has impacted certain sectors. [Graziano et al. \(2021\)](#) utilize product-level bilateral monthly trade data from Eurostat and suggest that Brexit uncertainty reduced UK-EU bilateral trade flows and trade participation. Using transaction-level data from Her Majesty's Revenue and Customs (HMRC), [Crowley et al. \(2018\)](#) investigate the role of trade agreement renegotiation uncertainty induced by Brexit and find that British exporters were deterred from entering EU markets and accelerated exit as British exporters started to target non-EU markets. They also argue that the renegotiation deterred British firms from introducing new products into the EU and that the effects were larger for products

facing high potential tariff and non-trade barriers. Using the same dataset, [Douch et al. \(2020\)](#) suggest that the negative effects also exist in terms of intensive margins with declining bilateral trade values. They consider exporter size heterogeneity and add that small and medium enterprises (SMEs), with relatively high exposure to Brexit uncertainty due to slow production adjustment, were more affected by uncertainty than larger businesses.

Using the Brexit referendum as a quasi-natural experiment, [Fernandes & Winters \(2021\)](#) find that exchange rate and uncertainty shocks make Portuguese exporters reduce the volume and price of their exports to the UK. [McGrattan & Waddle \(2020\)](#) showed that exit from the EU decreased FDI into the UK by about 22%. They also evidenced, using data from the Office for National Statistics (ONS), that Brexit uncertainty lowered trade in services. [Javorcik et al. \(2020\)](#) investigate the impact of potential trade barriers induced by Brexit on the UK labor market and find a notable reduction in online job postings, especially for higher-skilled positions, in regions with higher exposure to potential barriers on professional services exports to the EU.

From a macro economic perspective, [Faccini & Palombo \(2019\)](#) suggested before the Brexit deal was reached that the long-run effects of Brexit were around 8.3% and 4.8% decrease in GDP for hard and soft Brexit, respectively. Moreover, in the short run, firms would be driven by wait-and-see effects with the expectation of a worsening trade circumstances. [Faccini & Palombo \(2021\)](#) suggest that when uncertainty is about the future rather than the present, and is expected to last for a long time, the UK economy will decline slowly rather than in a V-shape. However, [Steinberg \(2019\)](#) study the macroeconomic impact of Brexit-induced TPU and find that the impacts are rather limited.

As part of Brexit, the UK has to renegotiate preferential trade agreements (PTAs) with not only the EU but also non-EU countries as a result of no longer being subject to EU common trade policy, thus, the Brexit-related uncertainty may have had a global

effect. [Hassan et al. \(2024\)](#) examined firms headquartered in 71 countries and splits firms into potential Brexit beneficiaries and losers through sentiment analysis of their quarterly earnings calls. They argue that, in general, international firms with Brexit risk exposure tend to reduce trade, hiring, and investment. [Graziano et al. \(2020\)](#) also found that Brexit uncertainty has led to a reduction in trade flows between the UK and non-European countries with which the UK had PTAs prior to 2016. [Makrychoriti & Spyrou \(2023\)](#) proposed a new measure of Brexit uncertainty by constructing a binary variable using the occurrence of 17 Brexit uncertainty related events, and then converting the binary variable into a continuous latent variable as a measure of Brexit uncertainty by using the Qual VAR developed by [Dueker \(2005\)](#). They suggested that Brexit uncertainty hurts the UK economy but boosts the economies of some EU countries especially France, Spain and Italy by estimating a structural Factor-Augmented VAR.

### 3.2.3 VAR Models and Uncertainty

[Sims \(1980\)](#) first introduced the Structural VAR model, a multivariate time-series model that describes the dynamic structure of the variables. VAR models are now widely used to study the causal effects of structural economic shocks, including the effects of uncertainty shocks. The recursive identification methods in VAR are widely employed to identify uncertainty shocks. For example, [Bachmann et al. \(2013\)](#), using a recursively identified VAR, suggest that while shocks in business uncertainty can initially lead to declines in economic activity, their high-frequency impact is minimal, indicating that uncertainty may be more a consequence than a cause of economic downturns. [Caggiano et al. \(2014\)](#) study the impact of uncertainty shocks on unemployment dynamics during recessions using a non-linear VAR and find significant asymmetric effects, with uncertainty shocks having a more severe influence on unemployment during recessions. [Jurado et al. \(2015\)](#) estimate a VAR and argue that macroeconomic uncertainty shocks explain a larger degree of fluctuations in production

and hours worked than stock market volatility shocks. [Scotti \(2016\)](#) examine the recursive impulse responses of key macroeconomic variables and suggest that uncertainty measures related to financial conditions have a stronger effect on real-activity variables. Other examples include [Bekaert et al. \(2013\)](#), [Bloom \(2014\)](#), and [Gilchrist et al. \(2014\)](#).

Yet, the restrictions when using a recursive identification strategy are somewhat problematic as they do not capture the contemporaneous response of uncertainty to other economic shocks ([Baker & Bloom 2013](#)). The proxy-SVAR model can mitigate these concerns by identifying shocks with an external instrument. The proxy-SVAR model uses an exogenous series with information outside the VAR system to recover structural shocks.

The proxy-SVAR model has become popular in analyzing the impact of shocks on macroeconomic indicators ([Klein & Linnemann 2021](#), [Miescu & Rossi 2021](#), [Känzig 2021](#), [Miranda-Agrippino et al. 2022](#), [Miranda-Agrippino & Ricco 2023](#)). Of those studies that look specifically at uncertainty, [Piffer & Podstawski \(2018\)](#) identify uncertainty shocks in a proxy-SVAR model that utilizes variations in the price of gold around selected events as an external instrument to identify uncertainty shocks and find that uncertainty shocks identified in the proxy-SVAR have a larger and quicker macroeconomic impact than those found in a recursive setting.

[Carriero et al. \(2015\)](#), using Monte Carlo simulation experiments, show that impulse responses from recursive VARs including uncertainty indicators as endogenous variables are downward biased when the uncertainty measures do not capture the true uncertainty perfectly. They suggest that using the indicators as exogenous instruments for uncertainty shocks in a proxy-SVAR offers a potential solution to this issue. Another application of proxy-SVAR in uncertainty studies is [Miescu \(2023\)](#), who uses global uncertainty shocks as instruments for country-level uncertainty shocks to explore, for the first time, the impact on emerging economies and show that the dampening effect of uncertainty shocks is much

stronger in emerging market economies compared to developed economies.

Another concern is that, due to limited data availability on macroeconomic variables, the estimates in previous studies are usually made at the quarterly or monthly level, which can miss the near-instant impact of shocks that can happen, especially in capital markets. As a result, [Alessandri et al. \(2023\)](#) develop a novel method that utilizes variables at high frequency to examine the responses of lower frequency variables. Specifically, they identify shocks from a daily VAR, average them to a monthly frequency, and then use the monthly series as instruments in a monthly VAR model.

All of the previously mentioned literature demonstrates a dampening effect of uncertainty on the real economy. Some literature, however, raises the possibility that certain forms of uncertainty may actually increase economic activity. For example, [Segal et al. \(2015\)](#) investigate the effects of macroeconomic uncertainty by distinguishing between “good” and “bad” uncertainties. Their findings reveal that good uncertainty can lead to heightened economic activity, while bad uncertainty can lead to economic slowdowns and falling asset prices. In other research, [Kraft et al. \(2018\)](#) suggest that although overall market volatility may not benefit equity prices, the specific uncertainty associated with individual firms (idiosyncratic volatility) could enhance equity prices, especially for companies with substantial growth prospects. Additionally, [Pástor & Veronesi \(2006\)](#) propose that heightened uncertainty about the future profitability of technology firms drove both the high prices and volatility in Nasdaq stocks.

### 3.3 Empirical Framework

In this section, we provide details on our empirical methods. Specifically, we describe the proxy-SVAR model that identifies structural shocks using external instruments to explore

the impact of Brexit uncertainty on a series of macroeconomic indicators.

A  $p^{\text{th}}$ -order reduced-form VAR model can be written as:

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + \mu_t, \quad (3.1)$$

where  $y_t$  is a vector of  $k$  endogenous variables,  $A_j (j = 1, \dots, p)$  is the matrix of coefficients,  $\mu_t$  is a vector of  $k$  reduced form shocks,  $c$  is a vector of  $k$  constant terms. Covariance matrix of the reduced form shocks are  $\Sigma = E(\mu_t \mu_t')$ . We can easily compute  $A_j$  and  $\mu_t$  by equation-by-equation OLS regression, and thus obtain  $\Sigma$ . The corresponding structural model is a linear transformation of the reduced-form VAR:

$$B^{-1} y_t = B^{-1} c + B^{-1} A_1 y_{t-1} + B^{-1} A_2 y_{t-2} + \dots + B^{-1} A_p y_{t-p} + \epsilon_t, \quad (3.2)$$

where  $B$  is the impact matrix,  $\epsilon_t$  is a vector of  $k$  structural shocks.  $\mu_t$  are assumed to be linearly related to the structural shocks  $\epsilon_t$ , as expressed by:

$$\mu_t = B \epsilon_t. \quad (3.3)$$

Here, unlike the reduced-form, the endogenous variables are contemporaneously correlated, and therefore the coefficient estimates cannot be obtained by simple OLS. In this model setup, the dynamics of the system are driven by the  $k$  structural shocks corresponding to the  $k$  elements in  $\epsilon_t$ . Structural shocks  $\epsilon_t$  are uncorrelated across equations. Our objective is to identify the responses of all the variables in the model to a single structural shock, i.e., an estimate of the specific column of matrix  $B$  that corresponds to the target structural shock of interest. For convenience, we set the target shock to be the first shock,  $\epsilon_{1,t}$ , and

thus, the relevant column of  $B$  is the first column,  $B_1$ .

Identifying the values of at least one column of  $B$  requires additional identification restrictions. Following [Stock & Watson \(2012\)](#), [Mertens & Ravn \(2013\)](#), and [Stock & Watson \(2018\)](#), we identify the target shock using an external instrument in a proxy-SVAR model. The employed instruments must exhibit correlation with the target shock while remaining uncorrelated with other shocks, thereby capturing some external variation present in the target shock. That is, the following two conditions should be satisfied:

$$E(z_t \epsilon_{1,t}) = a \neq 0 \quad (\text{Relevance condition}), \quad (3.4)$$

$$E(z_t \epsilon_{j,t}) = 0 \text{ for } j \neq 1 \quad (\text{Exogeneity condition}), \quad (3.5)$$

where  $z_t$  denotes the external instrument (or “proxy”) for the target shock  $\epsilon_{1,t}$ , and  $a$  is an unknown scalar. The first condition conveys that the external instruments are correlated with the target shock. The second condition conveys that  $z_t$  is uncorrelated with other contemporaneous shocks. In our case, these two conditions imply that the instrument we construct is correlated with contemporaneous Brexit uncertainty shocks but not with other shocks. Consequently, we have:

$$E(z_t \mu_t) = BE(z_t \epsilon_t) = B_1 a, \quad (3.6)$$

Here, it is  $B_1$  rather than  $B$  because the elements other than  $B_1$  are not correlated with  $z_t$ . Suppose  $B_1(1)$  is the first element of  $B_1$  and  $B_1(k)$  is the  $k$ -th element of  $B_1$  for  $k \neq 1$ , indicating the extent to which the target shock affects the first and the  $k$ -th variable, respectively, we can have:

$$\frac{B_1(k)}{B_1(1)} = \frac{B_1(k)a}{B_1(1)a} = \frac{E(z_t\mu_{t,k})}{E(z_t\mu_{t,1})} = \gamma_k, \quad (3.7)$$

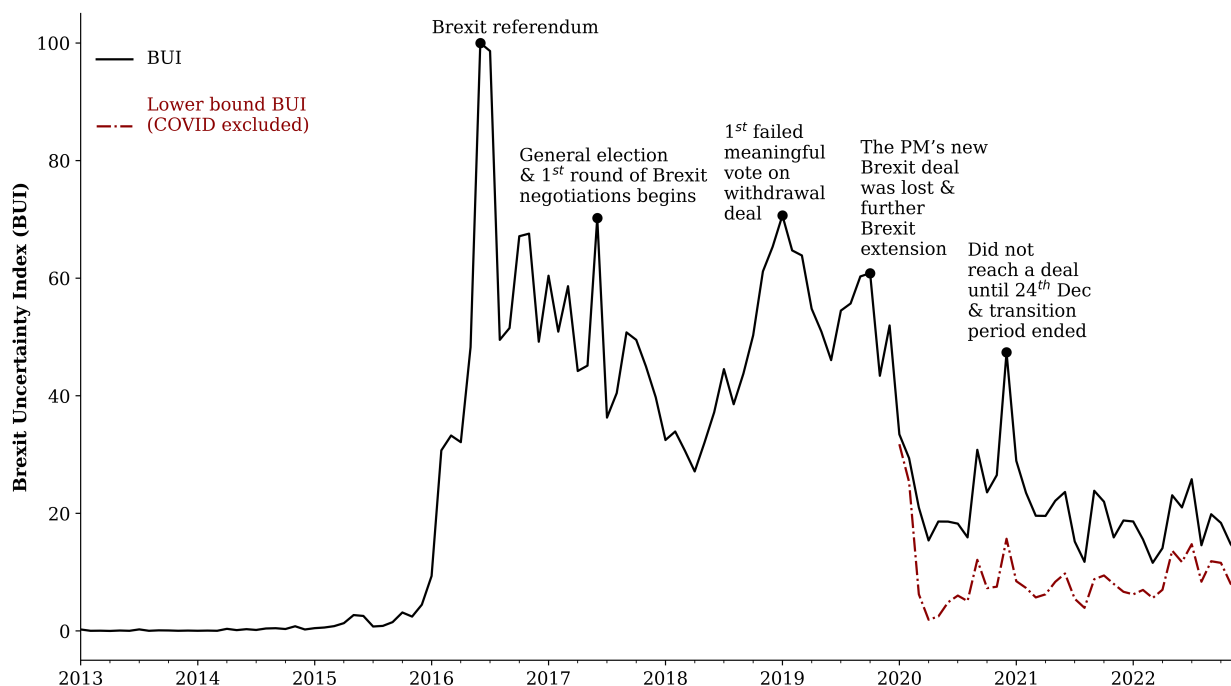
where  $z_t$  and  $\mu_t$  are observable, and thus, we can compute the ratio vector  $\gamma$ . This equation highlights the relative impact of the structural shock on the  $k$ -th variable compared to the first variable  $y_{1,t}$ . To determine the absolute effects of a one standard deviation uncertainty shock, we need to further solve for  $B_1(1)$  with the covariance restrictions of the reduced-form model, i.e.,  $\Sigma = E(\mu_t\mu_t')$ . Based on equation 3.3, the covariance matrix of  $\mu_t$  can be rewritten as  $E(\mu_t\mu_t') = E(B\epsilon_t\epsilon_t'B') = E(BB') = \Sigma$ . Appendix C.1 provides further details on the procedure to solve for  $B_1(1)$ . Once  $B_1(1)$  is determined, the ratio vector  $\gamma$  allows us to compute all the other elements of  $B_1$ . To reasonably set the sign, we always include the Brexit Uncertainty Index (BUI) as the first variable in subsequent VAR analyses and set the sign to be positive, as a positive Brexit uncertainty shock should be associated with an increase in the BUI.

## 3.4 Data

### 3.4.1 Brexit Uncertainty Measure

In this paper, we use the Brexit Uncertainty Index (BUI) developed by [Chung et al. \(2022\)](#) as our measure of Brexit uncertainty. [Chung et al. \(2022\)](#) relies on 11 major UK newspapers and uses the frequency of news articles containing the term ‘‘Brexit,’’ at least one term denoting uncertainty, and at least one term denoting the country, UK, to construct the aggregate BUI. Figure 3.1 shows the monthly index. As depicted in Figure 3.1, they further provide a lower bound BUI that controls for the exacerbating effect of the COVID-19 pandemic on Brexit uncertainty by excluding all COVID-related articles. This paper adopts the lower

bound BUI to avoid the effects of the COVID-19 pandemic distorting our results. Hereafter, the BUI we refer to is the lower bound BUI. See Chapter 1 of the thesis for more details about the Brexit uncertainty measurement.



**Figure 3.1:** BUI and lower bound BUI (COVID excluded). BUI is a scaled monthly index based on news articles that include “Brexit” and one or more words in the “uncertainty” word set, and one or more terms in the “UK” term set listed in Appendix A.1. The index is normalized to a maximum value of 100. Lower bound BUI stands for scaled monthly amount of news articles including the identical triplet of terms as BUI but excluding all articles containing one or more words suggesting “COVID” that are issued after January 2020.

### 3.4.2 Daily Data for Instrument Construction

Our daily estimation covers the period 2015D233 to 2022D166, which corresponds to the period December 2015 to August 2022. Notably, our daily sample includes only business days, as financial markets are closed on weekends and public holidays. The daily VAR consists of, in

this order, 21 lags of the daily BUI, FTSE 100, the FTSE Implied Volatility Index (IVI), the nominal effective exchange rate, and credit spreads, that is, the investment grade corporate bond yield relative to the gilt yield.<sup>2</sup> The credit spread is derived by subtracting the S&P UK Investment Grade Corporate Bond Index from the S&P UK Gilt Index. All variables are in levels. The FTSE 100 and FTSE IVI are obtained via Bloomberg. The nominal effective exchange rate is provided by the Bank for International Settlements (BIS). The two variables used to calculate the credit spread, the S&P UK Investment Grade Corporate Bond Index and the S&P UK Gilt Index, are also obtained via Bloomberg.

### 3.4.3 Data for Monthly VAR Estimation

Our estimation covers the period 2015M12 to 2022M8 as the BUI is close to 0 prior to December 2015. The 21 lags in the daily VAR analysis generate the average-to-month shock starting in 2016M1. We, hence, use monthly UK data from 2016M1 to 2022M8. Our baseline specification comprises four lags of the five endogenous variables: the BUI, hours worked per week, the log of real GDP, trade openness (TO) as the sum of exports and imports divided by GDP, and the log of the broad real effective exchange rate (REER) indices covering 64 economies. Hours series, GDP, and trade openness are adjusted for seasonal variations. The hours series is in per-capita terms using the total working-age population. GDP is presented in real per-capita terms.

The selection of variables in our baseline specification is based on capturing the main macroeconomic indicators and variables that are anticipated to respond most directly to the Brexit-related uncertainty shocks. Specifically, hours worked captures labor market reactions and GDP reflects overall economic activity. Trade openness is a key indicator, as it represents a primary channel through which Brexit impacts the UK economy, given the

---

<sup>2</sup>The 21 lags represent 21 consecutive business days, which roughly correspond to one calendar month.

potential disruptions in the trade sector. Lastly, the real effective exchange rate provides valuable insights into the external economic environment, reflecting how fluctuations in the currency respond to the uncertainties.

In more detail, Table 3.1 summarizes the variables used in our baseline monthly VAR model, along with the additional variables utilized to produce the supplementary results in Section 3.6.2. We perform some treatments on the variables before feeding them into the VAR. In particular, we convert some variables into per capita terms by dividing the original series by the number of people of working age (16-64 years old). Annual data on the working-age population come from the Office for National Statistics (ONS) and are based on mid-year estimates (e.g., each year's estimate represents the population as of June 30). To obtain monthly estimates from these annual data points, we use linear interpolation, which assumes a constant rate of change between each pair of annual estimates. We also seasonally adjust some variables. While some variables obtained from data sources are already seasonally adjusted, for others that require seasonal adjustment, we use the X-13 seasonal adjustment method.

Some variables are not available directly and require calculation. For example, we calculate hours worked per week using a series of data from ONS, specifically,  $\text{Hours} = (\text{average actual weekly hours of work for full-time workers} \times \text{the number of persons in full-time employment}) + (\text{average actual weekly hours of work for part-time workers} \times \text{the number of persons in part-time employment})$ . This total is then divided by the working-age population to obtain per-capita hours worked. Trade openness (TO) is obtained by dividing the sum of total imports and total exports by GDP, with all series sourced from ONS. The GDP is provided as a monthly index. We scale the obtained TO series so that the TO in January 2016 is set to 100. Subgroups of trade openness are calculated in a similar way to the TO, for example, openness in goods towards the EU is measured by the ratio of total imports and exports of goods with the EU to GDP, scaled such that January 2016 equals

**Table 3.1:** Variables in Monthly VAR

Variable	Source	Seasonal Adjustment	Per capita series	Log
Brexit uncertainty index (BUI)	<a href="#">Chung et al. (2022)</a>			
GDP and Sectoral GDP	ONS	Y	Y	Y
Hours worked	Calculated based on variables from ONS	Y	Y	
Trade openness (TO)	Calculated based on variables from ONS	Y		
Real effective exchange rate (REER)	BIS			Y
Unemployment rate	ONS	Y		
Real wage	ONS	Y		Y
All subgroups of trade openness	Calculated based on variables from ONS	Y		

Notes: This table reports the variables in our VAR systems. ONS stands for Office for National Statistics. BIS stands for Bank for International Settlements. CIPS stands for Chartered Institute of Procurement & Supply.

### 3.5 Instrument Construction

Following [Alessandri et al. \(2023\)](#), we construct the instrument to identify Brexit uncertainty shocks using a two-step process: (1) estimate a high-frequency VAR to identify high-frequency uncertainty structural shocks through a Cholesky decomposition; (2) aggregate the high-frequency structural shocks to a lower frequency to obtain structural shocks at a lower frequency. The resulting low-frequency shocks are then used as instruments to identify low-frequency Brexit uncertainty shocks.

Specifically, we first estimate a daily VAR under recursive identification. Appendix [C.2](#) describes how we obtain a recursive identification. The VAR consists of, in this order, 21 lags of the daily BUI, FTSE 100, the FTSE Implied Volatility Index (IVI), the nominal exchange rate, and credit spreads, that is, the investment-grade corporate bond yield relative to the gilt yield as discussed in Section [3.4.2](#). All variables are in levels. The estimated daily

uncertainty shocks are then aggregated to the monthly level by averaging. We use the obtained monthly series as the proxy for Brexit uncertainty shocks.

A concern is that there may not be sufficient information for the high-frequency VAR to properly recover the target shock due to the lack of macroeconomic variables that are only available in low-frequency systems. According to [Forni & Gambetti \(2014\)](#), the VAR should contain enough information about the economy, meaning that no state variables of the economy Granger cause the estimated shock. Following [Forni & Gambetti \(2014\)](#), we conduct a test to check if the aggregated low-frequency shock is orthogonal to the state variables of the economy. As in [Miranda-Agrippino & Ricco \(2023\)](#), we extract the first six principal components from large macro data sets as state variables and conduct a test as in [Eq. 3.8](#):

$$z_t = c + \beta_0 + \beta_1 z_{t-1} + \sum_{j=1}^6 \beta_{pc_j} pc_{j,t-1} + u_t, \quad (3.8)$$

where the  $z_t$  is the aggregated monthly shock and the proxy in monthly VAR, the  $pc_{j,t-1}$  denotes the one lag of principal components, and  $c$  is a constant. The test result shows that state variables do not Granger cause the estimated structural shock, which is not rejected, suggesting that our VAR system contains sufficient information to identify the shock. The invertibility condition is also required to hold for a valid impulse response analysis in the low-frequency proxy-SVAR. As we are interested in a single shock, we only need the invertibility of the target shock to hold ([Miranda-Agrippino & Ricco 2023](#)).

We find that the estimated shock does not Granger cause the residuals of the Brexit uncertainty equation in the monthly VAR. We, therefore, conduct the standard estimation process for the proxy-SVAR model using the estimated shock as an external instrument, following [Noh \(2017\)](#) and [Alessandri et al. \(2023\)](#). The estimation of both the low and

high-frequency SVAR models is done using the Bear toolbox, constructed by [Dieppe et al. \(2016\)](#).

### 3.5.1 Evaluation of the Instrument

As discussed in Section 3.3, a valid instrument must satisfy two conditions: (1) the instrument must carry information about the target shock (relevance condition), and (2) the instrument must not be associated with other contemporaneous shocks (exogeneity condition). In this section, we demonstrate that our estimated shock fulfills both conditions.

*Relevance condition: Test on the strength of the instrument.*

The external instrument is weak if  $a$  in Eq. (3.4) is close to zero. A common method for testing the strength of instruments is the first-stage F-test, with the null hypothesis that instruments are uncorrelated with the target structural shock  $\epsilon_{1,t}$ , i.e.,  $a = 0$ . A difficulty, however, is that the target shock  $\epsilon_{1,t}$  is unobservable. Following [Gertler & Karadi \(2015\)](#), we instead test the relationship between the instrument and the estimated residuals from the first equation of the reduced-form model,  $\mu_{t,1}$ , given that  $\mu_{t,1}$  is a linear transformation of  $\epsilon_{1,t}$ .<sup>3</sup> The computed F-statistics in our setting all exceed the rule-of-thumb value of 10, suggesting that our instrument is strong.<sup>4</sup>

*Exogeneity Condition*

The exogeneity condition is violated if, for example, the instrument is correlated with other contemporaneous shocks that would affect the macro variables in our system. We, thus, verify the exogeneity condition by exploring the correlation between our instrument and

---

<sup>3</sup>See also [Piffer & Podstawski \(2018\)](#), [Klein & Linnemann \(2021\)](#) and [Miescu \(2023\)](#).

<sup>4</sup>For example, the F statistic in the following baseline specification is 26.48.

other structural shocks. Specifically, we build on [Piffer & Podstawski \(2018\)](#) and estimate the following regression:

$$z_t = c + \gamma_h x_{ht} + u_{ht}, \tag{3.9}$$

where the vector  $x_{ht}$  stands for other contemporaneous shocks.

In particular, we consider credit risk, economic policy uncertainty, and monetary policy shocks. The exogeneity condition holds if we accept the null hypothesis that the instrumental variable and other structural shocks are uncorrelated, i.e.,  $\gamma = 0$ . Specifically, we use the UK five-year sovereign credit default swap (CDS) in sterling as a proxy for credit risk, reflecting market pricing of credit risk. We use the UK Economic Policy Uncertainty Index provided by [Baker et al. \(2016\)](#) to capture economic policy uncertainty shocks, and a monetary policy shock series from [Willems et al. \(2022\)](#) to capture monetary policy shocks.<sup>5</sup> The monetary policy shock series is somewhat shorter due to data availability. To comply with the principle of like-for-like comparison, we use the level form of these shock series.

The regression results are displayed in [Table 3.2](#). We report the estimated  $\gamma$  and  $p$ -values. Our instrument is not significantly associated with credit risk, uncertainty, and monetary policy shocks. In all cases, we do not reject the null hypothesis at the 1% significance level, with all  $p$ -values larger than 0.10. This means that our estimated external

---

<sup>5</sup>The UK five-year sovereign credit default swap (CDS) contract on the UK government's debt is a financial agreement designed to protect the buyer against the possibility of the UK government defaulting on its debt within a 5-year period. Essentially, the CDS acts as insurance for the buyer. By paying premiums, the buyer hedges against the risk of losing their investment if the UK government defaults. The data source is Thomson Reuters. We collect the daily UK CDS data denominated in US Dollars (USD). We then calculate the monthly average of the CDS series and use the monthly average exchange rate of the US dollar to sterling to obtain the monthly CDS series priced in sterling.

instruments satisfy the exogeneity condition.

**Table 3.2:** Validating the Exogenous Condition of the Instrument

Shock	Format	$\gamma$	p-value	Sample Period
Credit Risk	Raw	4.40	0.29	2016M1 - 2022M8
Uncertainty	Raw	67.52	0.16	2016M1 - 2022M8
Monetary Policy	Raw	0.01	0.70	2016M1 - 2019M12

Notes: This table reports the results of regressing the instrument for the uncertainty shock ( $z_t$ ) on other contemporaneous shock ( $x_{ht}$ ) (see Eq. 3.9). Other contemporaneous shocks are listed in the rows of the table.

## 3.6 Results

In this section, we present the empirical results. We start with the baseline specification by estimating a proxy-SVAR model that consists of the BUI, GDP, hours worked, trade openness, and real effective exchange rate. We compute the impulse response functions (IRFs) of the endogenous variables to a one standard deviation Brexit uncertainty shock based on 10,000 moving block bootstrap replications, over a horizon of up to 36 months. We then present additional results by incorporating a wider range of macroeconomic variables. We also show the IRFs of the variables in the baseline specification under alternative identification strategies, including a recursive scheme, a proxy-SVAR that identifies shocks using low-frequency data, a Bayesian SVAR with prior knowledge from theoretical models and a proxy-SVAR that identifies shocks from a sign-restricted daily VAR.

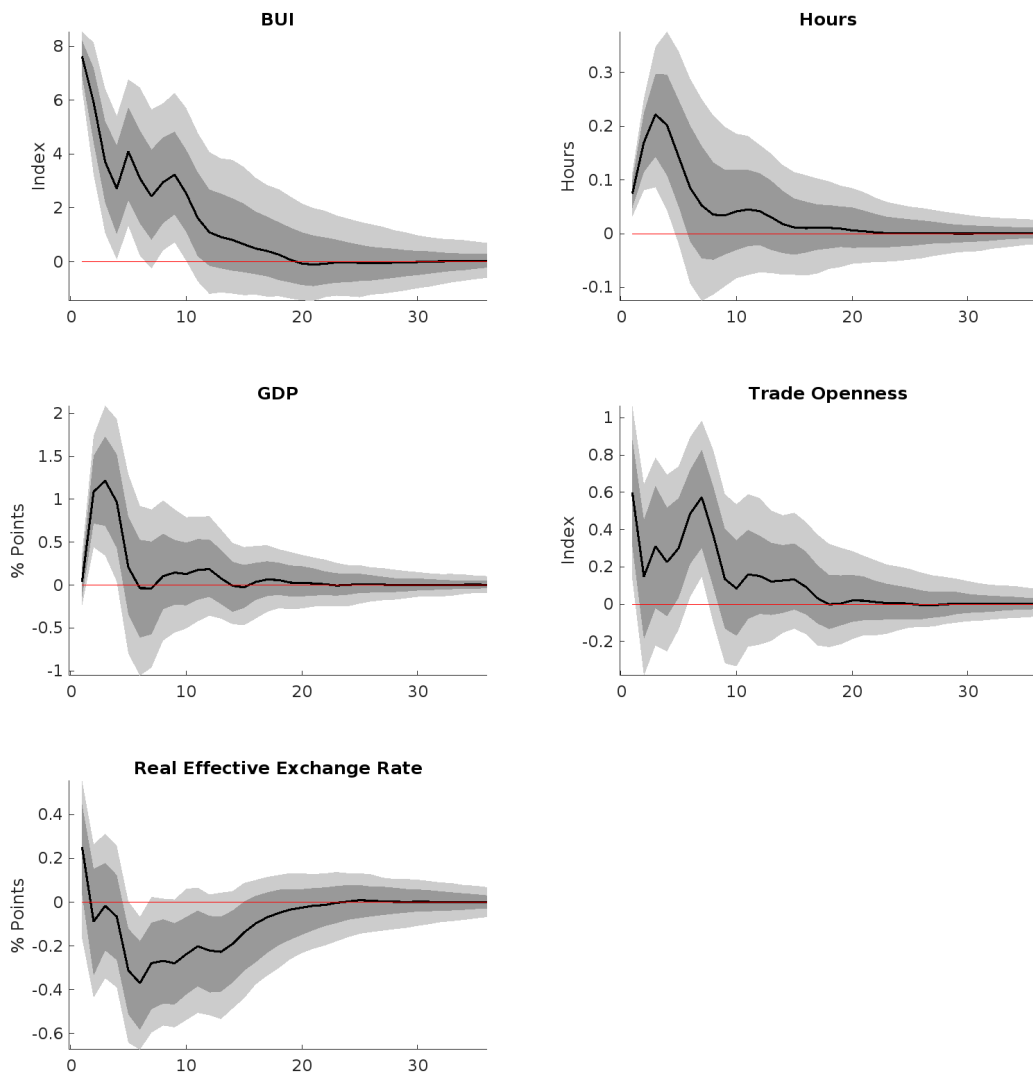
### 3.6.1 Baseline Results

Figure 3.2 shows the IRFs of the endogenous variables in our baseline model, with the shaded areas denoting 68% and 90% confidence intervals.<sup>6</sup> One immediate observation is that Brexit uncertainty shocks increase GDP, hours worked, and trade openness. In particular, hours worked increases on impact, and the response continues to increase, peaking at the third month with an increase of 0.2 hours before fading. This can be explained by two factors. First, workers may choose to work more to secure their income during periods of uncertainty. This is consistent with [Parker et al. \(2005\)](#), who find that income uncertainty leads self-employed individuals to work longer hours as a form of self-insurance. Second, firms are likely to increase hours worked as a precautionary measure to maintain production and capitalize on short-term opportunities. [Broadbent et al. \(2023\)](#) demonstrate that the Brexit vote creates a temporary advantage for the tradable sector as firms anticipate a future decline in productivity, leading to an increase in the relative price of tradable goods, prompting an increase in output, and thus potentially an increase in hours worked. GDP does not respond immediately to the Brexit uncertainty shock but begins to rise from the second month onward in a hump-shaped fashion, peaking at around 1%. This increase likely reflects the cumulative impact of the earlier rise in hours worked, which gradually boosts production and economic output.

The trade openness responses tend to manifest in the impact period and in the seventh month following the shock, which we discuss in further detail in Section 3.6.2. Second, the Brexit uncertainty shock has little effect on the real effective exchange rate, with the effect only being statistically significant in the sixth month after the shock occurs, and the effect is negative. Notably, the impact of the Brexit uncertainty shock is relatively short-lived, with responses peaking early on and all fading away very quickly within eight months following the shock.

---

<sup>6</sup>All the IRFs in this paper are generated in the Bear toolbox ([Dieppe et al. 2016](#)).



**Figure 3.2:** Impulse response to a one-standard-deviation Brexit uncertainty shock. The solid lines and shaded areas exhibit the mean and the 68% (dark) and 90% (light) confidence intervals.

To investigate the extent to which the uncertainty shock drives variation in our main variables, Table 3.3 presents the forecast error variance decomposition (FEVD) of the uncertainty shock at various horizons. The uncertainty shock from the high-frequency structural VAR model explains around 18% of the forecast error variance of GDP, around 23% of the

**Table 3.3:** Forecast Error Variance Decomposition

Horizon	BUI	GDP	Working Hours Per Capita	Trade Openness	Real effective Exchange Rate
h = 6	0.68	0.18	0.23	0.17	0.12
h = 12	0.64	0.18	0.23	0.20	0.22
h = 24	0.62	0.18	0.24	0.21	0.24
h = 36	0.62	0.18	0.24	0.21	0.24

Notes: The forecast error variance decompositions at horizons 6, 12, 24 and 36 months, based on 1,000 bootstrap replications. All listed values are significantly different from zero at the 90 % level.

forecast error variance of total hours worked, and around 20% of the forecast error variance of trade openness. The Brexit uncertainty shock accounts for only about 12% of fluctuations in the exchange rate in the short run, whereas this share doubles in the long run.

### 3.6.2 Additional Results

In this section, we display the response of additional variables of interest to the Brexit uncertainty shock. To generate additional IRFs, we re-estimate the models that replace the real effective exchange rate in our previous baseline model with the additional variables. We replace the real effective exchange rate because it produces the strongest instrumental variable, that is, the largest F-value, which we discussed in Section 3.5.1. Figure 3.3 presents the responses. The impulse responses of the other initial variables in the system exhibit analogous patterns to those depicted in Figure 3.2, and are thus not reiterated here.

First, we explore the propagation of Brexit uncertainty shocks on employment in more detail (panel (a)). Figure 3.3 reveals that the effect of the Brexit uncertainty shock on unemployment rates is largely insignificant, except for a small significant positive effect in the

immediate impact period. Real wages, on the other hand, show an initial positive response followed by a reversed response. This pattern can be attributed to the initial surge in labor demand driven by firms expanding production, which temporarily elevates wages. However, as the labor supply continues to increase, the excess supply exerts downward pressure on wages, leading to the subsequent decline and eventual stabilization.

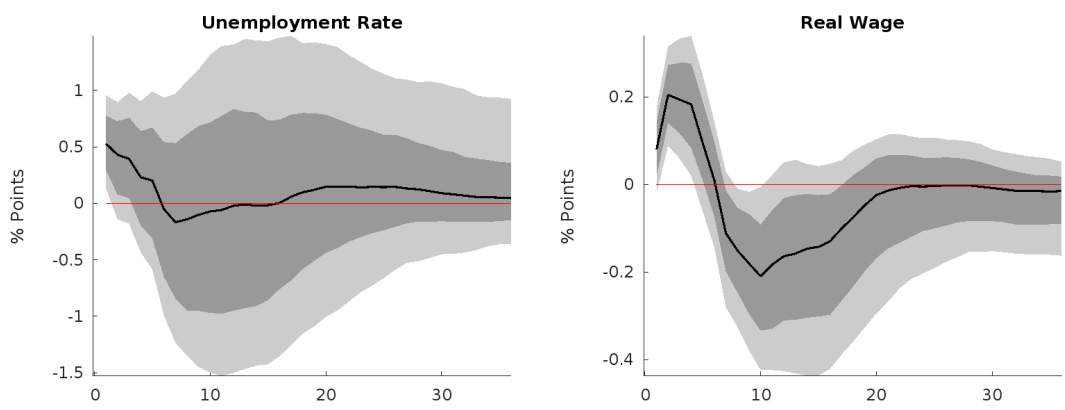
Next, we look at the dynamics of four distinct measures of trade openness (panel (b)). The IRFs indicate that the rise in trade openness following the Brexit uncertainty shock is largely driven by a significant and immediate increase in goods trade, while trade openness in services remains unaffected. Moreover, trade openness towards the EU increases sharply after the shock, whereas trade openness towards non-EU countries shows no significant response, suggesting that the impact of the shock is primarily concentrated on trade in goods with the EU.<sup>7</sup>

### 3.6.3 Alternative Identification Strategies

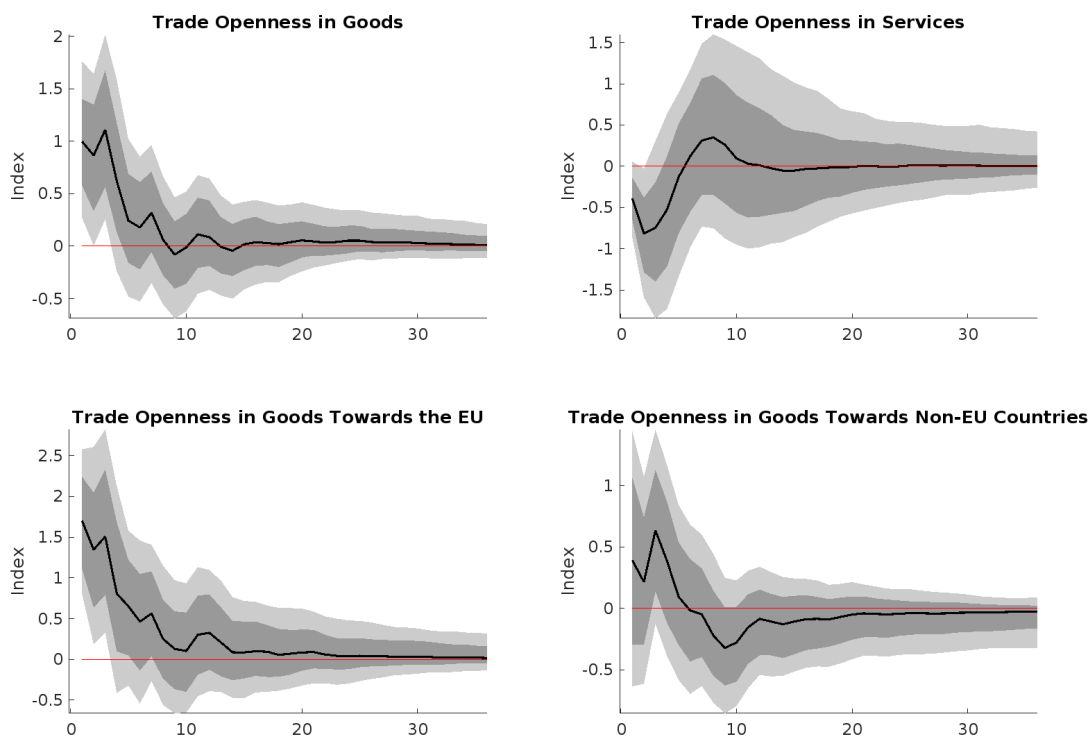
We explore how varying shock identification strategies affect the responses of the macroeconomic variables. We compare our baseline results with the dynamics generated from a VAR model with the prevalent recursive identification scheme, a proxy-SVAR that uses the monthly BUI index to construct the external instruments, and a Bayesian VAR. We also consider an external instrument generated from a daily VAR estimated with sign restrictions.

---

<sup>7</sup>We include the response plots of GDP at various sectors in Appendix C.3.



(a) Employment



(b) Trade

**Figure 3.3:** Impulse response to a one-standard-deviation Brexit uncertainty shock. The solid lines and shaded areas exhibit the mean and the 68% (dark) and 90% (light) confidence intervals.

### 3.6.3.1 SVAR with with a recursive identification scheme

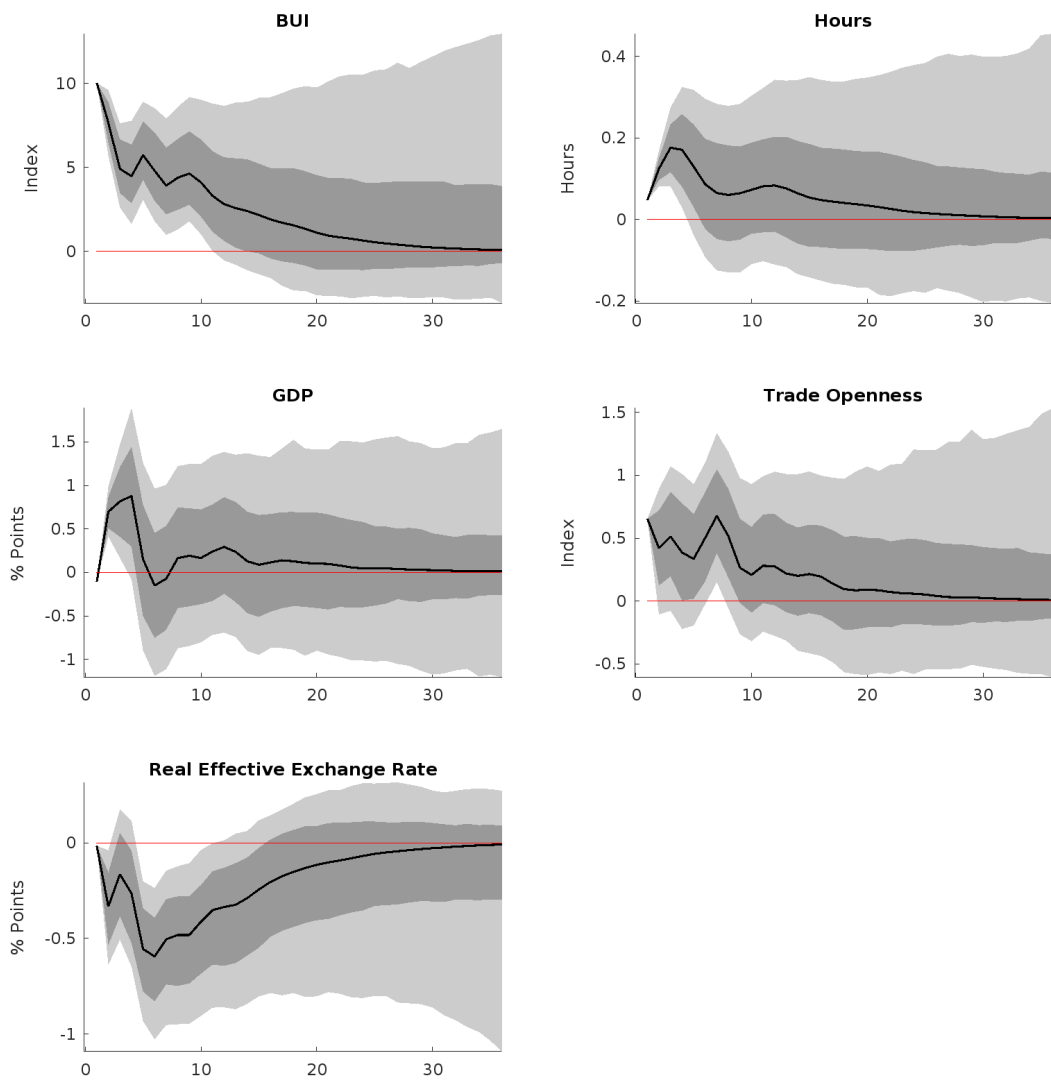
We fit a monthly VAR that includes a constant and four lags of (in this order) the BUI index, the log of the REER, the TO, the TWH, and the log of GDP, and use a Cholesky decomposition to identify the uncertainty shock. In this identification method, restrictions on the contemporaneous impact of shocks are imposed according to the order of the variables in the VAR. This implies that the first variable, BUI, only contemporaneously responds to its own shocks and its response to all other shocks is constrained to zero on impact; the second variable, REER, responds to the Brexit uncertainty shock and to its own shocks on impact, and so on. Figure 3.4 shows the dynamics.<sup>8</sup> Overall, the impulse responses of recursively identified uncertainty shocks closely resemble the trends in the responses in our baseline proxy-SVAR with a high-frequency instrument. One exception is the more persistent negative impact of the Brexit uncertainty shock on the real effective exchange rate under the recursive setting.

### 3.6.3.2 Proxy-SVAR that Identifies Structure Shocks with Low-frequency Data

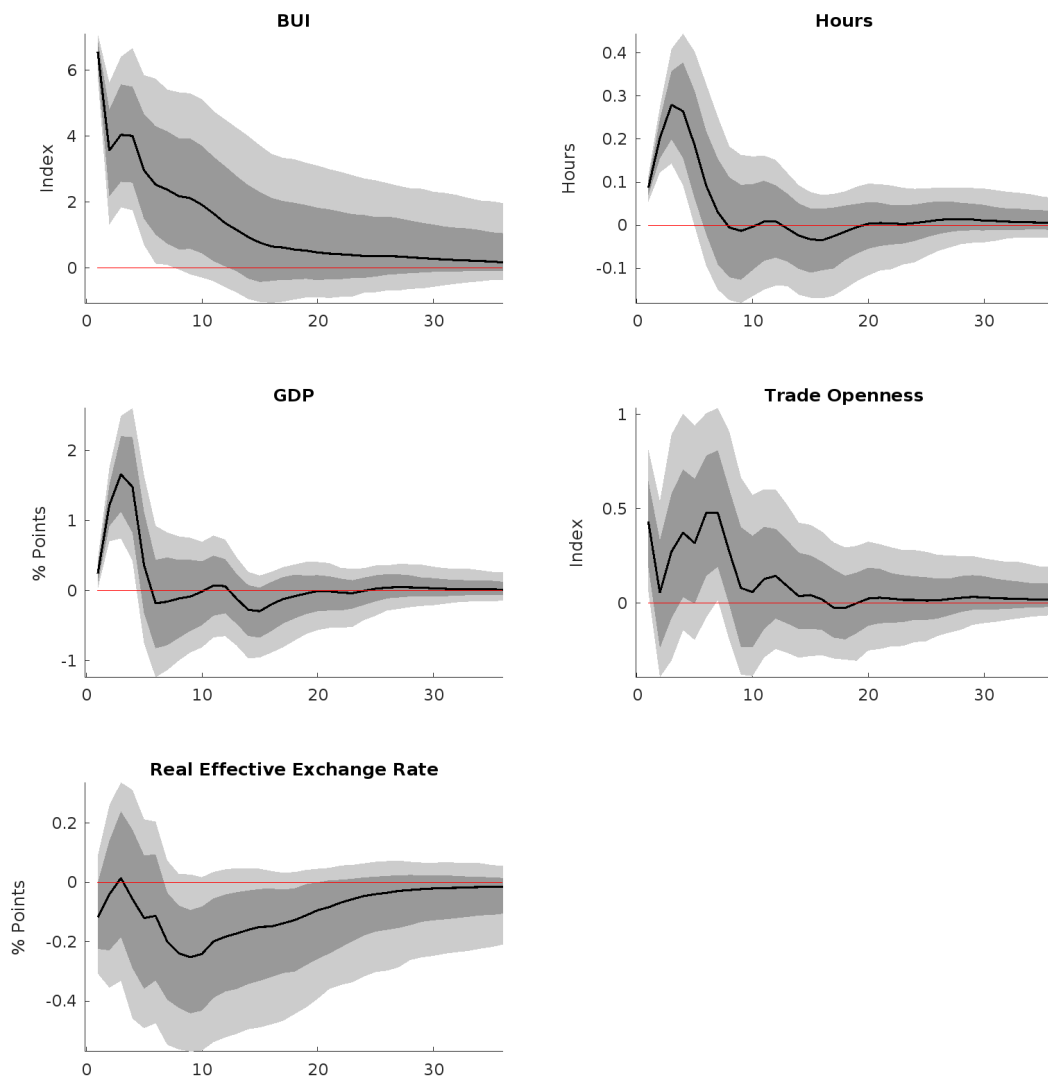
To compare the baseline results with the responses to the shock that is identified from low-frequency data, we estimate a proxy-SVAR using the monthly BUI index to construct the external instrument for the Brexit uncertainty shock. Appendix C.4 provides more details on its construction. Figure 3.5 displays the impulse response functions. In general, shocks identified using high-frequency surprises as external instruments also produce similar macro-variable responses as the shocks identified in the monthly proxy-SVAR, and this similarity is even more pronounced than with the recursive SVAR.

---

<sup>8</sup>The variable order in Figure 3.4 corresponds to Figure 3.2 and does not represent the order for the Cholesky identification.



**Figure 3.4:** Identifying the shocks with recursive scheme. Impulse response to a one-standard-deviation Brexit uncertainty shock under recursive identification scheme. The solid lines and shaded areas exhibit the mean and the 68% (dark) and 90% (light) confidence intervals.



**Figure 3.5:** Identifying the shocks with proxy-SVAR. Impulse response to a Brexit uncertainty shock under recursive identification scheme. The solid lines and shaded areas exhibit the mean and the 68% (dark) and 90% (light) confidence intervals.

### 3.6.3.3 Bayesian SVAR

One challenge in using VAR models for our analysis is the relatively short length of the dataset (80 monthly periods) compared to the large number of parameters that need to be estimated. Bayesian methods address this over-parameterization issue by incorporating prior information about the model parameters. Unlike traditional VAR models, which treat parameters as fixed values, Bayesian VAR (BVAR) models treat parameters as random variables. BVAR estimates posterior probability distributions for the parameters by integrating observed data with prior knowledge from theoretical models and other macroeconomic phenomena.

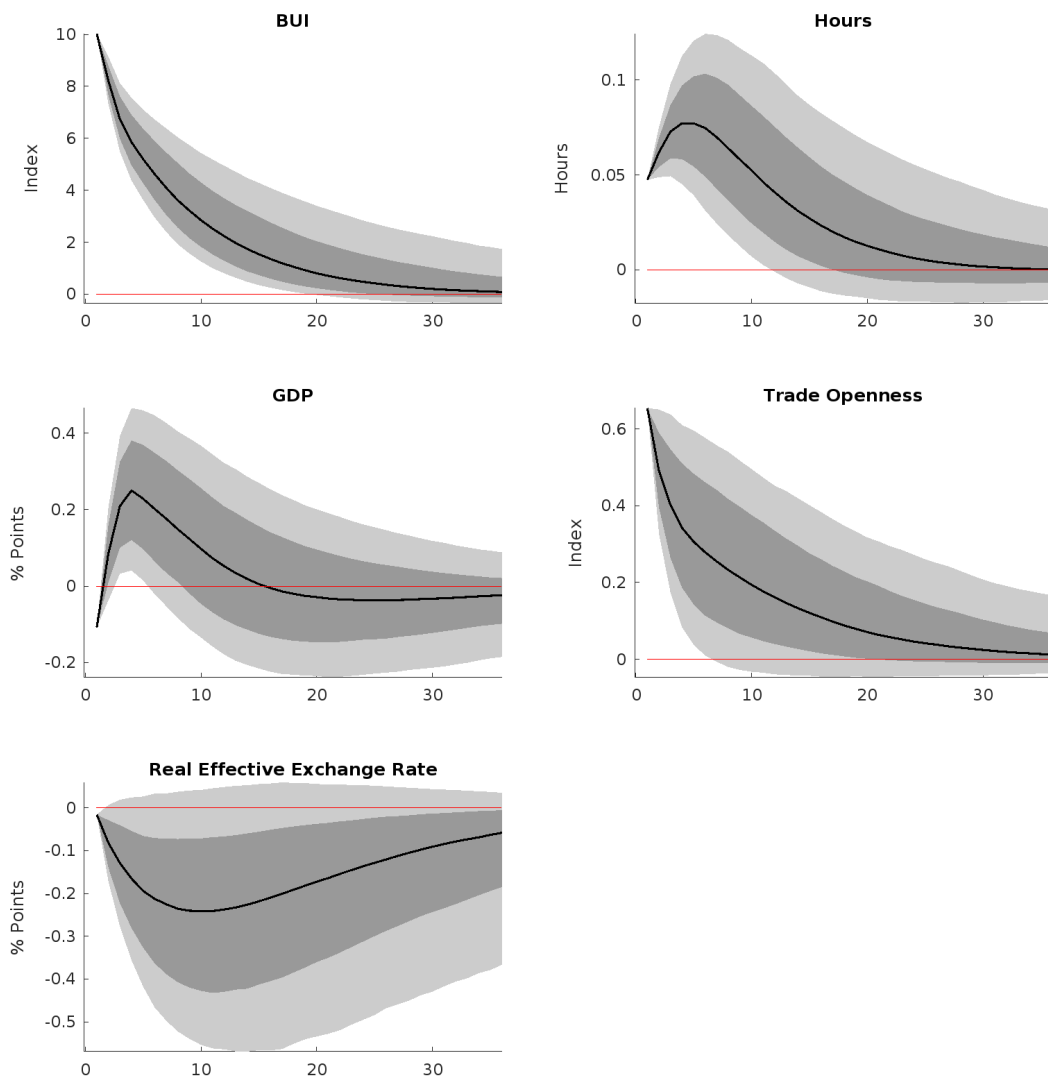
In our empirical analysis, we use a Minnesota prior as proposed by [Litterman \(1986\)](#) and adopt the default hyperparameters from the BEAR toolbox.<sup>9</sup> Our focus is on examining whether the direction of the impact of Brexit uncertainty on macroeconomic variables changes using a Bayesian approach. To simplify identification, we employ a recursive identification scheme. [Figure 3.6](#) presents the resulting impulse response functions (IRFs). We find that the Brexit uncertainty shock still exhibits positive effects on GDP, hours worked, and trade openness under the Bayesian approach, albeit to a different extent and duration than in the baseline setting.

### 3.6.3.4 Sign-Restricted Daily VAR Estimation

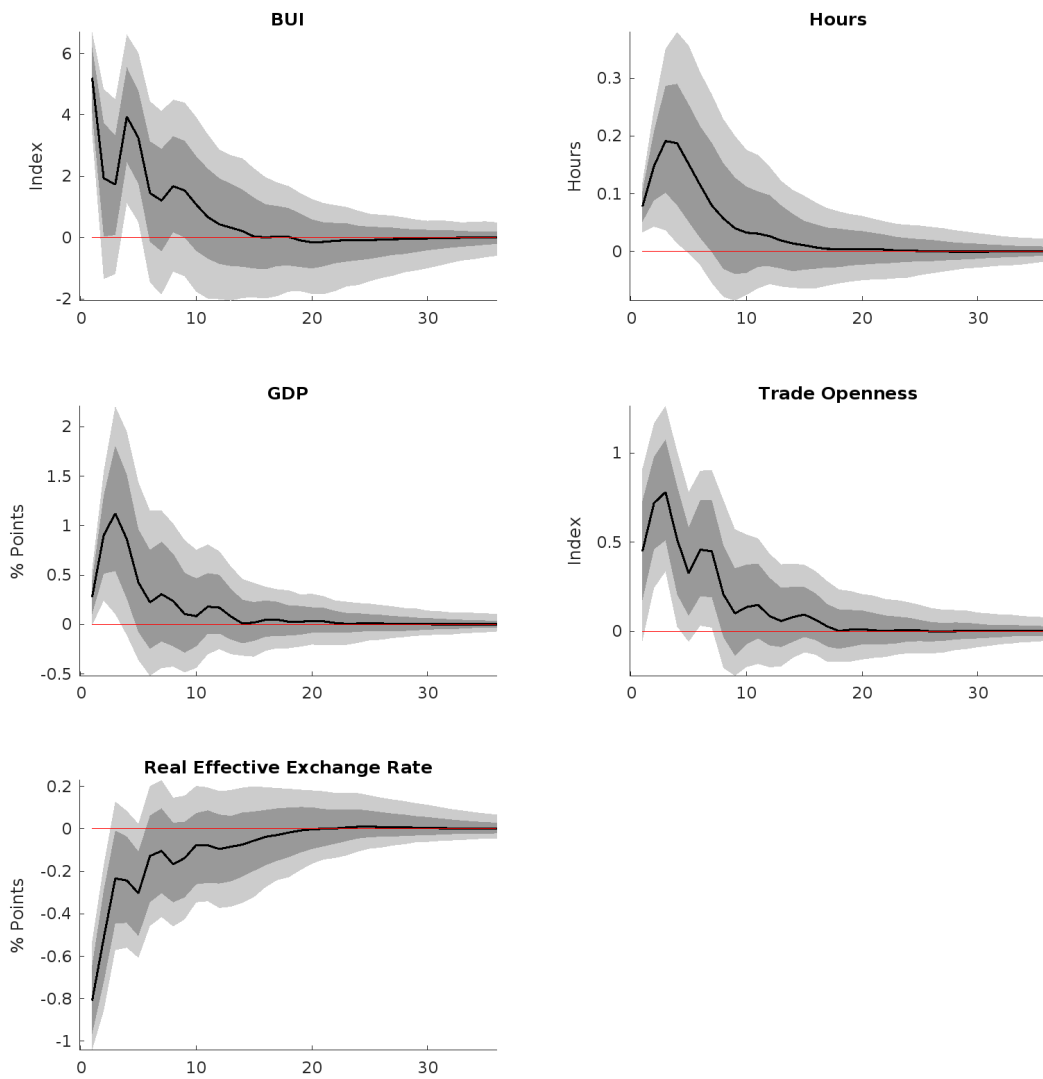
We examine the robustness of our daily VAR identification by replacing the baseline recursive approach with sign restrictions. Specifically, we impose that a Brexit uncertainty shock raises the Brexit Uncertainty Index (BUI) and depreciates the nominal effective exchange rate on

---

<sup>9</sup>See [Dieppe et al. \(2016\)](#) for a detailed explanation of Bayesian estimation and drawing from the correct posterior distribution in the SVAR model.



**Figure 3.6:** Identifying the shocks with Bayesian VAR using recursive identification scheme. Impulse response to a Brexit uncertainty shock under recursive identification scheme. The solid lines and shaded areas exhibit the mean and the 68% (dark) and 90% (light) confidence intervals.



**Figure 3.7:** Impulse responses to a Brexit uncertainty shock identified via sign restrictions in the daily VAR. The solid lines and shaded areas exhibit the mean and the 68% (dark) and 90% (light) confidence intervals.

impact. We leave the responses of the FTSE100, the FTSE Implied Volatility Index, and the credit spread unrestricted, allowing them to adjust freely to the shock.

After recovering the implied daily shock series, we aggregate it to the monthly level by averaging (as before). We then use this monthly measure as an external instrument in our proxy-SVAR to identify structural uncertainty shocks. Figure 3.7 shows that the Brexit uncertainty shock continues to produce positive effects on GDP, hours worked, and trade openness, though the magnitude and duration differ somewhat from the baseline. This close alignment with our original findings reinforces confidence in the validity and robustness of our results.

### 3.7 Sensitivity Analysis: Alternative Instrument

Here, we provide a comparison of the above baseline results with the responses to uncertainty shocks that are identified using the Brexit uncertainty series from the Decision Makers Panel (DMP). We identify Brexit uncertainty shocks using time-varying changes in the proportion of managers who view Brexit as the largest source of uncertainty.<sup>10</sup> Since the DMP series has only monthly data starting in 2016m9, we identify the shock in the proxy SAVR, i.e. in a monthly frequency model, with the data period from 2016m9. Figure 3.8 shows the dynamics of variables under the DMP uncertainty shock. The responses of GDP and hours worked are similar to those in the baseline model, showing a sharp hump-like movement. However, the responses of trade openness and REER are slightly different, with neither of their responses being significant. This difference should be attributed to the difference in sample periods. In particular, the model here excludes the referendum period, when Brexit uncertainty shocks are highly significant. We, therefore, believe that the baseline results are

---

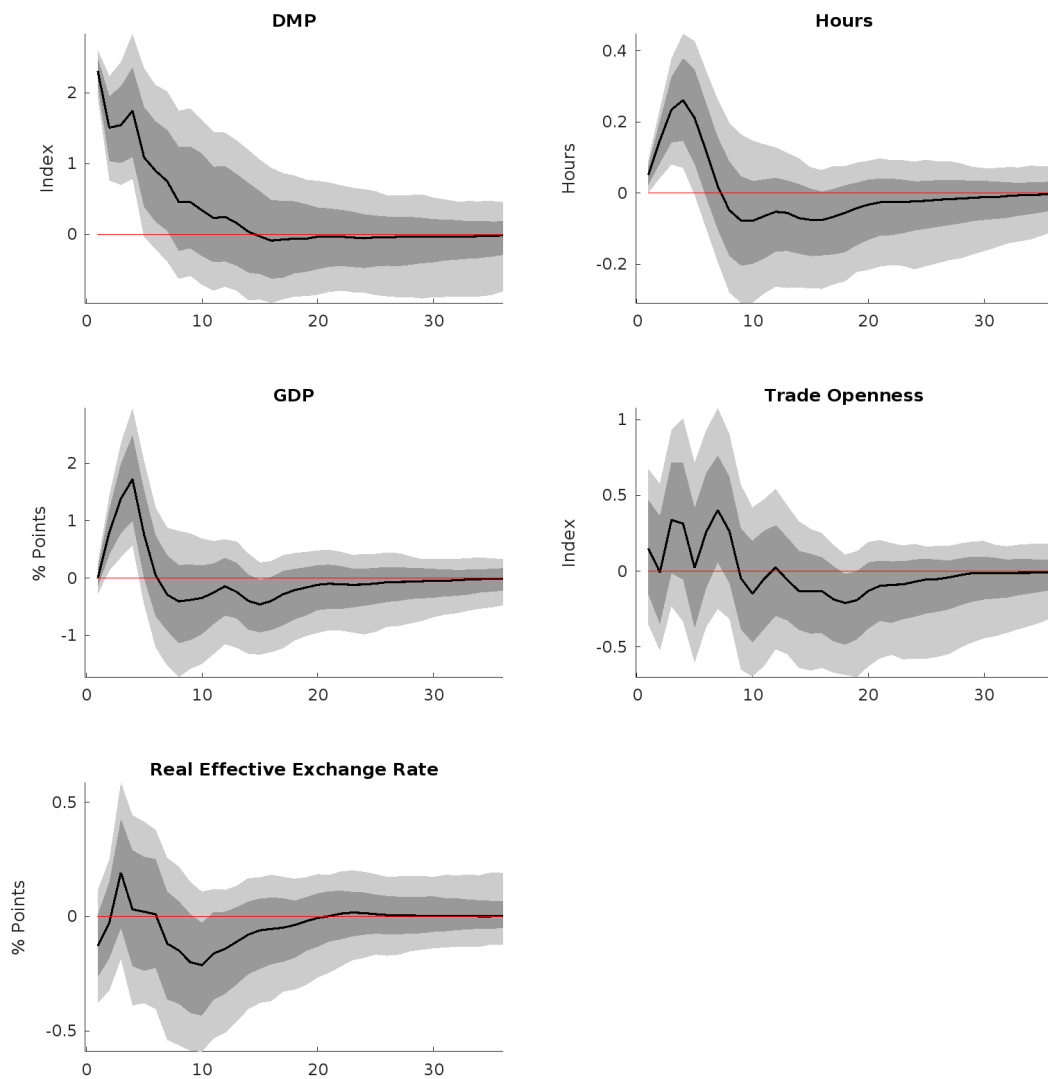
<sup>10</sup>Identifying shocks using the changes in the proportion of managers who see Brexit as one of the top three sources of uncertainty generates similar IRFs.

robust to using alternative external instruments to recover uncertainty shocks.

### 3.8 Conclusion

How did Brexit uncertainty shocks affect the UK economy? In this article, we investigate this question using a proxy-SVAR model and introduce a novel high-frequency instrument variable based on the Brexit uncertainty index developed by [Chung et al. \(2022\)](#) to identify Brexit uncertainty shocks. The process involves two steps: first, we recover uncertainty shocks using daily data on the Brexit uncertainty index, stock index, stock market implied volatility, nominal exchange rates, and credit spreads in a recursive SVAR. Second, we aggregate the resulting shocks to monthly frequencies and apply the aggregated shock as an exogenous instrument in a monthly proxy-SVAR model of the UK economy. This identification approach enables us to capture endogenous economic reactions, including financial market responses, within each month that may be overlooked by lower-frequency analyses. Our test results suggest that this high-frequency instrument provides sufficient information to proxy for Brexit uncertainty shocks and is exogenous to other macroeconomic shocks.

Our empirical analyses show that Brexit uncertainty shocks have a short-term, significant, and positive impact on the UK economy, boosting GDP, hours worked, and trade openness. Moreover, by examining various measures of trade openness, we find that Brexit uncertainty shocks predominantly impact goods trade with the EU, with trade openness in goods toward the EU responding immediately to the shock. In contrast, trade openness in goods towards non-EU countries and in services shows no significant response. All effects fade away typically within six months. This short-term positive response of the economy can be attributed to a temporary boost in the tradable sector following the Brexit vote, as suggested by [Broadbent et al. \(2023\)](#).



**Figure 3.8:** Using DMP series to construct instrument. The solid lines and shaded areas exhibit the mean and the 68% (dark) and 90% (light) confidence intervals.

While this paper examines the macroeconomic implications of aggregate Brexit uncertainty for the UK economy, future research should focus on the effects of topic-specific Brexit uncertainty. For example, the impact of trade policy-related uncertainty on UK trade openness, trade volumes, and trade diversion merits further exploration.

# Appendix for Chapter 3

## C.1 Identification of the Proxy-SVAR

Following the discussion in the main text, this section provides additional details on the procedure to recover  $B_1(1)$  using the covariance restrictions from of reduced-form model. We focus on the method of using a single instrument to identify a single structural shock, as outlined in [Stock & Watson \(2012\)](#). While [Mertens & Ravn \(2013\)](#) employed two instruments in their SVAR analysis to identify two distinct structural shocks, this paper does not cover the multi-instrument case. In a VAR with  $K + 1$  variables, the impact matrix  $B$  can be written as:

$$B = \begin{pmatrix} B_1(1) & B_1(K)' \\ B_K(1) & B_K(K) \end{pmatrix} \quad (3.10)$$

In this context,  $B_1(1)$  represents the scalar value in the  $(1, 1)$  position of the impact matrix  $B$  indicating the impact of the target structural shock,  $\epsilon_{1,t}$ , on the first endogenous variable  $y_{1,t}$ . The vector  $B_1(K)'$  is a  $1 \times K$  row vector that captures the effect of the same target shock on the remaining  $K$  endogenous variables in the model, excluding the first one. The column vector  $B_K(1)$  represents the responses of the first variable to the  $K$  remaining structural shocks. Finally,  $B_K(K)$  is a  $K \times K$  matrix that measure the relationships among the remaining variables and corresponding structural shocks. To recover  $B_1(1)$ , we start by

considering the relationship between the reduced-form residuals  $\mu_t$  and the structural shocks  $\epsilon_t$  given by:

$$\mu_t = B\epsilon_t. \quad (3.11)$$

In a structural VAR model, the structural shocks  $\epsilon_t$  are typically assumed to be orthogonal and standardized, meaning that their covariance matrix is the identity matrix  $I$ , that is,  $E(\epsilon_t\epsilon_t') = I$ . Thus, we have:

$$\Sigma_\mu = E(\mu_t\mu_t') = E(B\epsilon_t\epsilon_t'B') = BB' \quad (3.12)$$

We can write the matrix form of  $\Sigma_\mu$  as:

$$\Sigma_\mu = \begin{pmatrix} \sigma_{11} & \sigma'_{1K} \\ \sigma_{K1} & \Sigma_{KK} \end{pmatrix} \quad (3.13)$$

By combining the matrix form of  $B$  from Equation 3.10 with the matrix form of  $\Sigma_\mu$  from Equation 3.13 and the relationship given in Equation 3.12, we can derive the following set of equations:

$$B_1(1)^2 + B_1(K)'B_1(K) = \sigma_{11}, \quad (3.14)$$

$$B_1(1)B_K(1) + B_K(K)B_1(K) = \sigma_{K1}, \quad (3.15)$$

$$B_K(1)B_K(1)' + B_K(K)B_K(K)' = \Sigma_{KK}. \quad (3.16)$$

As we discussed in the main text,  $\gamma$  is the ratio vector that captures the relationship between  $B_1(1)$  and  $B_K(1)$ , specifically representing the impact of the target shock on the remaining variables relative to the first variable. By substituting  $B_K(1) = \gamma B_1(1)$  into our equations, we obtain:

$$B_1(1)^2 + B_1(K)'B_1(K) = \sigma_{11}, \quad (3.17)$$

$$B_1(1)^2\gamma + B_K(K)B_1(K) = \sigma_{K1}, \quad (3.18)$$

$$B_1(1)^2\gamma\gamma' + B_K(K)B_K(K)' = \Sigma_{KK}. \quad (3.19)$$

From Equation 3.17, we can write  $B_1(1)$  as:

$$B_1(1) = \pm\sqrt{\sigma_{11} - B_1(K)'B_1(K)}. \quad (3.20)$$

To solve for  $B_1(1)$ , we need to obtain  $B_1(K)'B_1(K)$ . Subtracting Equation 3.17 times  $\gamma$  from Equation 3.18 gives:

$$\begin{aligned} B_K(K)B_1(K) - B_1(K)'B_1(K)\gamma &= \sigma_{K1} - \sigma_{11}\gamma \\ \Rightarrow (B_K(K) - \gamma B_1(K)')B_1(K) &= \sigma_{K1} - \sigma_{11}\gamma \\ \Rightarrow B_1(K) &= (B_K(K) - \gamma B_1(K)')^{-1}(\sigma_{K1} - \sigma_{11}\gamma). \end{aligned} \quad (3.21)$$

Thus,

$$B_1(K)'B_1(K) = (\sigma_{K1} - \sigma_{11}\gamma)'(B_K(K) - \gamma B_1(K)')^{-1}(B_K(K) - \gamma B_1(K)')^{-1}(\sigma_{K1} - \sigma_{11}\gamma) \quad (3.22)$$

with

$$\begin{aligned} \Gamma &= (B_K(K) - \gamma B_1(K)')(B_K(K) - \gamma B_1(K)')' \\ &= B_K(K)B_K(K)' - B_K(K)B_1(K)\gamma' - \gamma B_1(K)'B_K(K)' + \gamma B_1(K)'B_1(K)\gamma'. \end{aligned} \quad (3.23)$$

By subtracting Equation 3.18 (multiplied by  $\gamma'$ ) from Equation 3.19, we obtain:

$$\begin{aligned} B_K(K)B_K(K)' - B_K(K)B_1(K)\gamma' &= \Sigma_{KK} - \sigma_{K1}\gamma' \\ \Rightarrow B_K(K)B_1(K)\gamma' &= B_K(K)B_K(K)' - (\Sigma_{KK} - \sigma_{K1}\gamma') \end{aligned}$$

By substituting this expression and its transpose into  $\Gamma$ , we obtain:

$$\Gamma = -B_K(K)'B_K(K) + \gamma B_1(K)'B_1(K)\gamma' + (\Sigma_{KK} - \sigma_{K1}\gamma') + (\Sigma_{KK} - \sigma_{K1}\gamma')' \quad (3.24)$$

To eliminate the term  $B_K(K)'B_K(K)$ , we start by taking Equation 3.19 and subtracting Equation 3.17, after multiplying the latter by  $\gamma$  and then post-multiplying by its transpose. This operation yields:

$$B_K(K)B_K(K)' - \gamma B_1(K)'B_1(K)\gamma' = \Sigma_{KK} - \sigma_{11}\gamma\gamma' \quad (3.25)$$

Taking the transpose of this equation provides:

$$B_K(K)'B_K(K) - \gamma'B_1(K)'B_1(K)\gamma = \Sigma_{KK} - \sigma_{11}\gamma\gamma' \quad (3.26)$$

Since most terms are symmetric, we substitute this result back into  $\Gamma$  to obtain:

$$\Gamma = -(\Sigma_{KK} - \sigma_{11}\gamma\gamma') + (\Sigma_{KK} - \sigma_{K1}\gamma') + (\Sigma_{KK} - \sigma_{K1}\gamma')' \quad (3.27)$$

Simplifying further, we have:

$$\Gamma = \Sigma_{KK} + \sigma_{11}\gamma\gamma' - \sigma_{K1}\gamma' - \gamma\sigma'_{K1} \quad (3.28)$$

With the  $\Gamma$ , we can solve for  $B_1(1)$ .

## C.2 Recursive Identification

To simplify, we assume a three-variable VAR model. The impact matrix  $B$  is:

$$B = \begin{bmatrix} B_1(1) & B_2(1) & B_3(1) \\ B_1(2) & B_2(2) & B_3(2) \\ B_1(3) & B_2(3) & B_3(3) \end{bmatrix}$$

In order to pin down the impact matrix  $B$ , additional restrictions need to be imposed. The recursive identification strategy imposes a contemporaneous exclusion restriction. This

strategy assumes that the first variable only contemporaneously responds to its own shocks, the second variable contemporaneously responds to both its own shocks and the shocks of the first variable, and so on. Therefore, we have  $B_2(1) = B_3(1) = B_3(2) = 0$ , and:

$$B = \begin{bmatrix} B_1(1) & 0 & 0 \\ B_1(2) & B_2(2) & 0 \\ B_1(3) & B_2(3) & B_3(3) \end{bmatrix}$$

We have:

$$BB' = \begin{bmatrix} B_1(1) & 0 & 0 \\ B_1(2) & B_2(2) & 0 \\ B_1(3) & B_2(3) & B_3(3) \end{bmatrix} \times \begin{bmatrix} B_1(1) & B_1(2) & B_1(3) \\ 0 & B_2(2) & B_2(3) \\ 0 & 0 & B_3(3) \end{bmatrix}$$

$$= \begin{bmatrix} B_1(1)^2 & B_1(1) \cdot B_1(2) & B_1(1) \cdot B_1(3) \\ B_1(1) \cdot B_1(2) & B_1(2) \cdot B_2(2) & B_1(2) \cdot B_2(3) \\ B_1(1) \cdot B_1(3) & B_1(3) \cdot B_2(3) & B_1(3) \cdot B_3(3) \end{bmatrix} = \sum_{\mu} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

So we have:

$$B_1(1)^2 = \sigma_{11}$$

$$B_1(1) \cdot B_1(2) = \sigma_{12}$$

$$B_1(1) \cdot B_1(3) = \sigma_{13}$$

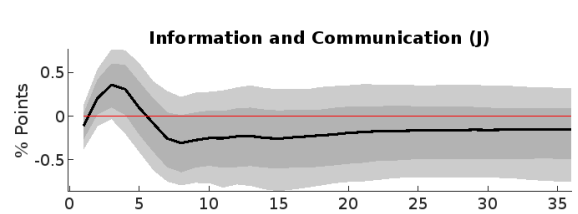
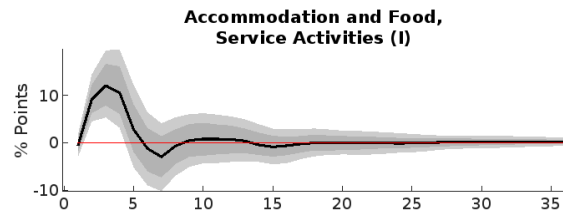
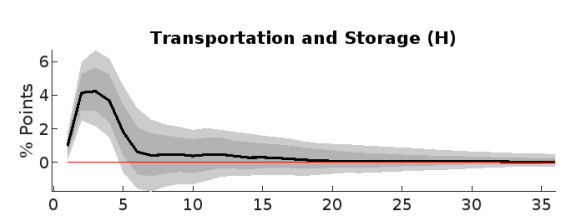
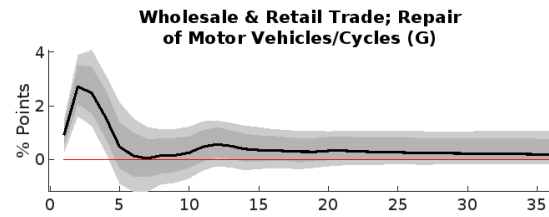
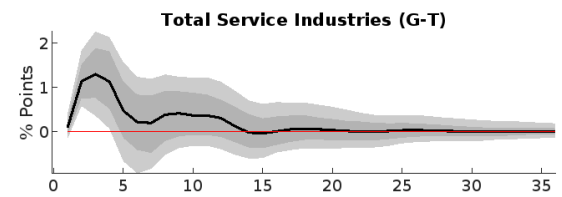
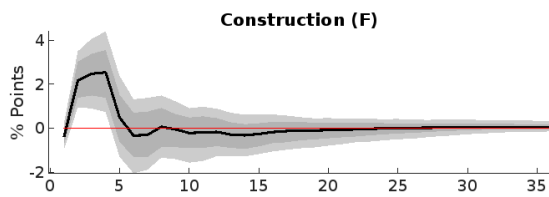
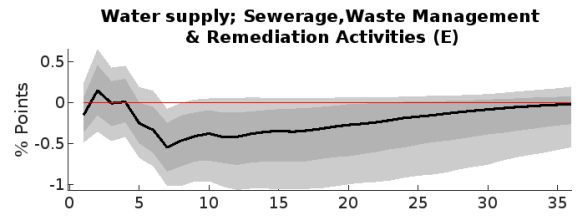
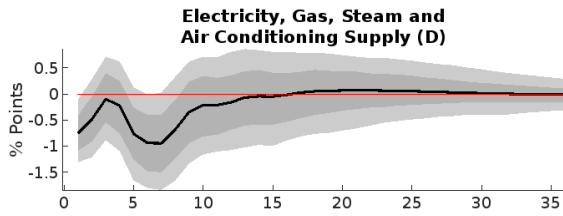
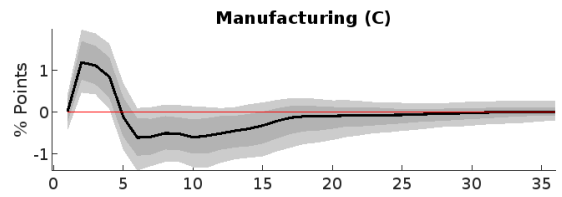
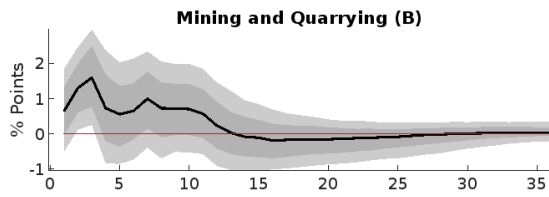
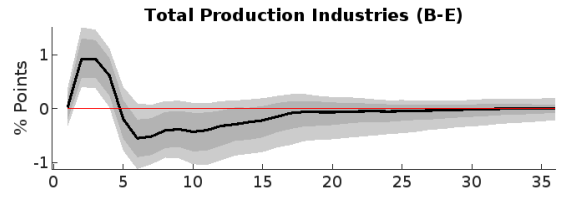
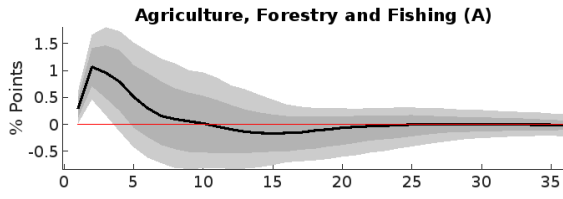
$$B_1(2) \cdot B_2(2) = \sigma_{22}$$

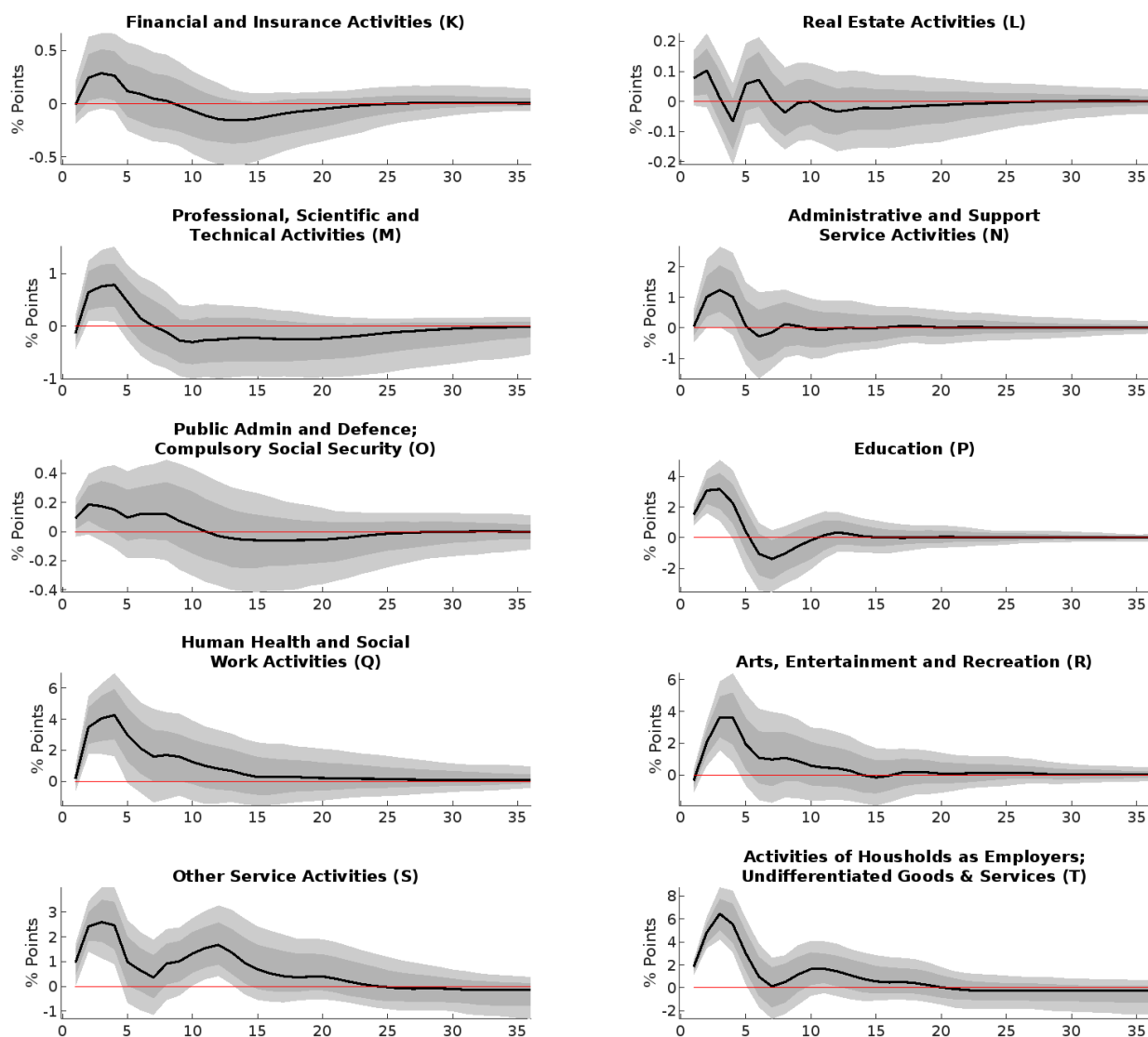
$$B_1(2) \cdot B_2(3) = \sigma_{23}$$

$$B_1(3) \cdot B_3(3) = \sigma_{33}$$

Solving these equations, we can identify  $B$ . We can thereafter compute the structural shocks  $\epsilon_t = \mu_t \cdot B^{-1}$ .

### C.3 Sectoral GDP





**Figure C.1:** Impulse response to a one-standard-deviation Brexit uncertainty shock. The solid lines and shaded areas exhibit the mean and the 68% (dark) and 90% (light) confidence intervals.

## C.4 Proxy SVAR using Monthly Index

We construct an external instrument, controlling for macroeconomic beliefs that develops before Brexit uncertainty takes hold, and its own lags. Specifically, our BUIs contains messages about Brexit uncertainty shocks. The indices, however, may be driven by prior macroeconomic expectations, and also its own past evolution. Therefore, in defining a pure and contemporaneous Brexit uncertainty shock, it is necessary to take into accounts these two sets of variables.

We aim to eliminate endogenous changes in our indices due to expectations of future economic performance that arise from other instantaneous shocks, and to its own past trends. Doing so increases the chances that the IV is associated solely with the current uncertainty shocks, which is critical for accurate identification.

To be exact, following [Miranda-Agrippino et al. \(2022\)](#), we take the residual  $\eta_t$  from the below regression as our instrument:

$$bui_t = c + \sum_{i=1}^9 \beta_i bui_{t-i} + \sum_{j=1}^5 \delta_j E_{t-j}(m_T) + \eta_t \quad (3.29)$$

where  $bui_t$  is the monthly growth rate of our BUI, i.e.  $bui_t = (LnBUI_t - LnBUI_{t-1}) \times 100$ .  $E_{t-j}(m_T)$  refers to the average of the forecast values of the macroeconomic variables for the present year  $T$  that are made in the three periods from  $t - j$  to  $t - j - 2$ . The vector  $m_T$  encompasses forecasts of year-on-year percentage changes in GDP ( $y_T$ ), CPI ( $\pi_T$ ), as well as government consumption ( $c_T$ ). The forecast values are taken from the forecasts for the UK economy database compiled by HM Treasury. For each month, the database provides the average of forecasts of year-on-year GDP and CPI growth received in the three months prior, e.g. in September 2018 it publishes the average of the GDP growth forecasts for the

period July to September 2018. In our setup, we use this series to measure forecasts of GDP.

The database also provides time series of independent forecasts from 26 forecasters for year-on-year percentage changes in government consumption. We compute the average values across forecasters, using the average of the prior three months as the forecast series for private consumption and fixed investment.

Table C.1 shows the regression results. This table presents the regression coefficients for Eq. 3.29 along with robust standard errors in parentheses. The specifications are estimated over the period 2015M12:2022M8. The BUI does not exhibit an autocorrelation pattern, as suggested by the insignificant autoregressive coefficients. Furthermore, pre-existing beliefs, as captured by the forecast values of CPI growth rates, and government consumption, include information for Brexit uncertainty. This suggests that Brexit uncertainty may be influenced by prior news embedded in the survey forecasts.

#### C.4.1 Evaluation of the Instrument

*Relevance condition: test on the strength of the instrument*

The computed F-statistic in this setting exceeds the rule-of-thumb value of 10, which is 93.89, suggesting that our instrument is strong and satisfies the relevance condition.

*Exogeneity Condition*

We estimate the Eq. 3.9. The results are presented in Table C.2, showing the estimated  $\gamma$  and  $p$ -values. Our instrument is not significantly associated with credit risk or monetary policy shocks. In both cases, we do not reject the null hypothesis at the 1% significance level, with  $p$ -values both greater than 0.10. Moreover, our instrument is not

**Table C.1:** Constructing Instrument

Own Lags	
$bui_{t-1}$	0.17 (0.21)
$bui_{t-2}$	-0.2 (0.23)
$bui_{t-3}$	0.25 (0.21)
$bui_{t-4}$	-0.12 (0.24)
$bui_{t-5}$	0.07 (0.17)
$bui_{t-6}$	-0.17 (0.19)
$bui_{t-7}$	0.03 (0.14)
$bui_{t-8}$	-0.21 (0.15)
$bui_{t-9}$	0.14 (0.14)
Prior Beliefs	
$E_{t-1}(y_T)$	-0.19 (4.1)
$E_{t-1}(\pi_T)$	0.61 (16.1)
$E_{t-1}(c_T)$	3.2 (7.24)
$E_{t-2}(y_T)$	-2.32 (5.64)
$E_{t-2}(\pi_T)$	26.06 (24.49)
$E_{t-2}(c_T)$	-5.11 (11.61)
$E_{t-3}(y_T)$	-2.08 (4.65)
$E_{t-3}(\pi_T)$	-56.74** (26.93)
$E_{t-3}(c_T)$	11.56 (11.73)
$E_{t-4}(y_T)$	3.86 (5.18)
$E_{t-4}(\pi_T)$	32.08 (29.17)
$E_{t-4}(c_T)$	-21.97** (10.29)
$E_{t-5}(y_T)$	2.02 (3.61)
$E_{t-5}(\pi_T)$	-12.07 (16.81)
$E_{t-5}(c_T)$	13.13** (6.4)
intercept	15.25 (19.92)

Notes: Regression results of Eq. 3.29. Robust standard errors are reported in parentheses. \*\*\* indicates significance at the one percent level, \*\* indicates significance at the five percent level and \* indicates significance at the ten percent level.

**Table C.2:** Validating the Exogenous Condition of the Instrument

Shock	$\gamma$	p-value	Sample Period
Credit Risk	0.04	0.72	2016M9 - 2022M8
Uncertainty	0.03	0.08	2016M9 - 2022M8
Monetary Policy	3.60	0.89	2016M9 - 2019M12

Notes: This table reports the results of regressing the instrument for the uncertainty shock ( $z_t$ ) on other contemporaneous shock ( $x_{ht}$ ) (see Eq. 3.9). Other contemporaneous shocks are listed in the rows of the table.

significantly associated with economic policy uncertainty at the 5% significance level, with a  $p$ -value greater than 0.05. This means that our estimated external instrument satisfies the exogeneity condition.

# Concluding Remarks

Each of the three papers in this thesis shed light on specific aspects of Brexit. Chapter 1 focuses on the metrics of Brexit uncertainty, Chapter 2 examines the non-verbal media bias exhibited by the UK media during the Brexit referendum period, and Chapter 3 explores the macroeconomic impact of Brexit uncertainty shocks. In Chapter 1, we develop a novel method to measure Brexit-related uncertainty using machine learning to create both aggregate and topic-specific Brexit Uncertainty Indices (BUIs). These indices enable near real-time tracking of uncertainty across dimensions such as trade, immigration, and the macroeconomy. By isolating the effects of Brexit uncertainty from the COVID-19 pandemic, this approach offers added precision, providing policymakers and researchers with an effective means to quantify and monitor uncertainty over time.

Chapter 2 introduces a novel approach to measuring visual media bias during the Brexit vote, using advanced computer vision and machine learning techniques. We demonstrate that tabloid newspapers exhibit clear bias by portraying politicians with opposing views through more negative emotional expressions and less favorable visual contexts in their coverage. This bias is most evident in front-page images and is concentrated around key political figures. In contrast, broadsheets exhibit no significant bias. This study highlights the potential for leveraging computer vision for large-scale media bias research, emphasizing the growing importance of visual content in shaping political narratives.

Chapter 3 examines how Brexit uncertainty shocks affect the UK economy using a proxy-SVAR model. We use a novel high-frequency instrument, based on the daily Brexit uncertainty index developed in Chapter 1, to recover the shocks. We find that Brexit uncertainty provided a short-term boost to GDP, hours worked, and trade openness in goods towards the EU. However, these effects fade within the first six months. The temporary economic boost likely stems from increased activity in the tradable sector following the Brexit referendum.

## Limitations

This section discusses the main limitations of the research. For Chapter 1, our method for identifying uncertainty-related articles relies on a predefined term set that includes terms associated with uncertainty. While this approach is widely used in text-based uncertainty metrics, it has its limitations. For example, it does not account for phrases like “no uncertainty” or “reduced uncertainty,” which convey a reduction in uncertainty. In addition, uncertainty can be conveyed in more implicit or subtle ways that fall outside the fixed vocabulary of our term set. Phrases such as “may,” “might,” “could,” “potentially,” or “appears to” may be able to indicate uncertainty without directly using typical uncertainty-related words. These expressions, while important, are not captured by our current methodology, which could lead to an underestimation of the full range of uncertainty conveyed in the news articles analyzed.

In Chapter 2, we do not explore the specific sources of sentiment embedded in contextual images. While our method captures the overall sentiment conveyed, it does not distinguish whether the sentiment arises from particular visual elements such as objects, colors, background settings, or other human expressions. Understanding the contribution of these individual factors could provide a more detailed analysis of how visual media shapes

biases.

For Chapter 3, we do not explore the effects of Brexit uncertainty at the topic level. For example, we do not assess how uncertainty related to trade policy affects key economic outcomes such as the UK's trade openness, trade volumes, or trade diversion. Investigating these topic-specific uncertainties could provide more targeted insights into how different aspects of Brexit uncertainty influence various sectors of the economy.

## **Future Research**

To address the limitations outlined above, several directions for future research can be considered. For Chapter 1, future research could leverage advanced large language models, such as the ChatGPT API, to identify uncertainty-related articles without relying on pre-defined term sets. These models can better understand context, enabling them to discern not only whether an article discusses uncertainty but also whether that uncertainty specifically pertains to Brexit. By capturing more nuanced expressions of uncertainty, this approach could enhance the precision of our uncertainty metrics and improve the overall accuracy of media-based uncertainty analysis. For Chapter 2, future research could focus on analyzing the individual visual elements that contribute to the sentiment in contextual images. This could involve breaking down images into components such as objects, colors, and background settings to better understand the sources of emotional content. Identifying these factors would provide a more detailed view of how visual media conveys bias and shapes public perception. For Chapter 3, future research should examine the impact of topic-level Brexit uncertainty on the UK economy. For example, studies could explore how uncertainty in employment or manufacturing impacts key economic sectors.

Additionally, I am currently collaborating with the Scottish Government on a research

project that develops measures of business sentiment and uncertainty regarding key issues such as employment, immigration, and taxation for Scotland. This ongoing research aims to capture the unique challenges and uncertainties faced by Scottish businesses, providing valuable insights that could be expanded to other regions or sectors. Future research may also explore video data, which integrates text, image, and audio, offering richer insights into economic questions.

# Bibliography

- Adukia, A., Eble, A., Harrison, E., Runesha, H. B. & Szasz, T. (2023), ‘What We Teach About Race and Gender: Representation in Images and Text of Children’s Books’, *The Quarterly Journal of Economics* **138**(4), 2225–2285.
- Aftab, S. A. (2018), ‘Brexit Referendum and Media Coverage: An Appraisal’, *Journal of European Studies* **34**(1), 68–81.
- Aguilar, G., Kar, S. & Solorio, T. (2020), ‘LinCE: A Centralized Benchmark for Linguistic Code-Switching Evaluation’, *arXiv preprint arXiv:2005.04322* .
- Ahmad, S., Limão, N., Oliver, S. & Shikher, S. (2023), ‘Brexit Uncertainty and Its (Dis)Service Effects’, *American Economic Journal: Economic Policy* **15**(4), 459–485.
- Alessandri, P., Gazzani, A. & Vicondoa, A. (2023), ‘Are the effects of uncertainty shocks big or small?’, *European Economic Review* **158**, 104525.
- Algaba, A., Borms, S., Boudt, K. & Van Pelt, J. (2020), ‘The Economic Policy Uncertainty Index for Flanders, Wallonia, and Belgium’, *BFW Digitaal/RBF Numérique* **6**.
- Altig, D., Baker, S., Barrero, J. M., Bloom, N., Bunn, P., Chen, S., Davis, S. J., Leather, J., Meyer, B., Mihaylov, E. et al. (2020), ‘Economic Uncertainty Before and During the COVID-19 Pandemic’, *Journal of Public Economics* **191**, 104274.
- Arellano, C., Bai, Y. & Kehoe, P. J. (2019), ‘Financial Frictions and Fluctuations in Volatility’, *Journal of Political Economy* **127**(5), 2049–2103.

- Ash, E., Durante, R., Grebenschikova, M. & Schwarz, C. (2021), ‘Visual Representation and Stereotypes in News Media’, *CEPR Discussion Paper No. DP16624*.
- Ashkezari-Toussi, S., Kamel, M. & Sadoghi-Yazdi, H. (2019), ‘Emotional Maps Based on Social Networks Data to Analyze Cities’ Emotional Structure and Measure Their Emotional Similarity’, *Cities* **86**, 113–124.
- Azqueta-Gavaldón, A. (2017a), ‘Developing News-Based Economic Policy Uncertainty Index with Unsupervised Machine Learning’, *Economics Letters* **158**, 47–50.
- Azqueta-Gavaldon, A. (2017b), ‘Financial Investment and Economic Policy Uncertainty in the UK’, *Proceedings of the 1st International Conference on Internet of Things and Machine Learning*, 1–4.
- Azzimonti, M. (2018), ‘Partisan Conflict and Private Investment’, *Journal of Monetary Economics* **93**, 114–131.
- Bachmann, R., Elstner, S. & Sims, E. R. (2013), ‘Uncertainty and Economic Activity: Evidence from Business Survey Data’, *American Economic Journal: Macroeconomics* **5**(2), 217–249.
- Bailey, D., De Propris, L., De Ruyter, A., Hearne, D. & Ortega-Argilés, R. (2023), ‘Brexit, Trade and UK Advanced Manufacturing Sectors: A Midlands’ Perspective’, *Contemporary Social Science* **18**(2), 250–265.
- Baker, S. R. & Bloom, N. (2013), ‘Does Uncertainty Reduce Growth? Using Disasters as Natural Experiments’, *NBER Working Paper No.19475*.
- Baker, S. R., Bloom, N. & Davis, S. J. (2016), ‘Measuring Economic Policy Uncertainty’, *The Quarterly Journal of Economics* **131**(4), 1593–1636.
- Baker, S. R., Bloom, N. & Terry, S. J. (2024), ‘Using Disasters to Estimate the Impact of Uncertainty’, *Review of Economic Studies* **91**(2), 720–747.

- Bank of England (2019), 'Monetary Policy Report - November 2019: In Focus - Uncertainty and Brexit'. Available at <https://www.bankofengland.co.uk/monetary-policy-report/2019/november-2019/in-focus-uncertainty-and-brexite>.
- Barrett, A. W. & Barrington, L. W. (2005), 'Bias in Newspaper Photograph Selection', *Political Research Quarterly* **58**(4), 609–618.
- Basu, S. & Bundick, B. (2017), 'Uncertainty Shocks in a Model of Effective Demand', *Econometrica* **85**(3), 937–958.
- Bekaert, G., Hoerova, M. & Duca, M. L. (2013), 'Risk, Uncertainty, and Monetary Policy', *Journal of Monetary Economics* **60**(7), 771–788.
- Bernanke, B. S. (1983), 'Irreversibility, Uncertainty, and Cyclical Investment', *The Quarterly Journal of Economics* **98**(1), 85–106.
- Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003), 'Latent Dirichlet Allocation', *The Journal of Machine Learning Research* **3**, 993–1022.
- Bloom, N. (2009), 'The Impact of Uncertainty Shocks', *Econometrica* **77**(3), 623–685.
- Bloom, N. (2014), 'Fluctuations in Uncertainty', *Journal of Economic Perspectives* **28**(2), 153–176.
- Bloom, N., Bunn, P., Chen, S., Mizen, P., Smietanka, P. & Thwaites, G. (2019), 'The Impact of Brexit on UK Firms', *NBER Working Paper No.26218*.
- Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I. & Terry, S. J. (2018), 'Really Uncertain Business Cycles', *Econometrica* **86**(3), 1031–1065.
- Bollen, J., Mao, H. & Zeng, X. (2011), 'Twitter Mood Predicts the Stock Market', *Journal of Computational Science* **2**(1), 1–8.
- Boxell, L. (2021), 'Slanted Images: Measuring Nonverbal Media Bias During the 2016 Election', Available at SSRN 3837521.

- Brandenburg, H. (2006), ‘Party Strategy and Media Bias: A Quantitative Analysis of the 2005 UK Election Campaign’, *Journal of Elections, Public Opinion and Parties* **16**(2), 157–178.
- Broadbent, B., Di Pace, F., Drechsel, T., Harrison, R. & Tenreyro, S. (2023), ‘The Brexit Vote, Productivity Growth, and Macroeconomic Adjustments in the UK’, *Review of Economic Studies* p. rdad086.
- Burn, I., Button, P., Corella, L. F. M. & Neumark, D. (2019), ‘Older Workers Need Not Apply? Ageist Language in Job Ads and Age Discrimination in Hiring’, *NBER Working Paper No.26552*.
- Bybee, L., Kelly, B. T., Manela, A. & Xiu, D. (2020), ‘The Structure of Economic News’, *NBER Working Paper No.26648*.
- Caggiano, G., Castelnuovo, E. & Groshenny, N. (2014), ‘Uncertainty Shocks and Unemployment Dynamics in US Recessions’, *Journal of Monetary Economics* **67**, 78–92.
- Caldara, D. & Iacoviello, M. (2022), ‘Measuring Geopolitical Risk’, *American Economic Review* **112**(4), 1194–1225.
- Caldara, D., Iacoviello, M., Molligo, P., Prestipino, A. & Raffo, A. (2020), ‘The Economic Effects of Trade Policy Uncertainty’, *Journal of Monetary Economics* **109**, 38–59.
- Caprini, G. (2023), ‘Visual Bias’, *Economics Series Working Papers 1016*.
- Carriero, A., Clark, T. E. & Marcellino, M. G. (2018), ‘Endogenous Uncertainty’, *FRB of Cleveland Working Papers 1805*.
- Carriero, A., Mumtaz, H., Theodoridis, K. & Theophilopoulou, A. (2015), ‘The Impact of Uncertainty Shocks Under Measurement Error: A Proxy SVAR Approach’, *Journal of Money, Credit and Banking* **47**(6), 1223–1238.

- Chadwick, A., Vaccari, C. & O’Loughlin, B. (2018), ‘Do Tabloids Poison the Well of Social Media? Explaining Democratically Dysfunctional News Sharing’, *New Media & Society* **20**(11), 4255–4274.
- Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L. & Blei, D. M. (2009), ‘Reading Tea Leaves: How Humans Interpret Topic Models’, *Advances in Neural Information Processing Systems* pp. 288–296.
- Chung, W., Dai, D. & Elliott, R. (2022), ‘Measuring Brexit Uncertainty: A Machine Learning and Textual Analysis Approach’, *CEPR Discussion Papers 17410*.
- Crowley, M., Exton, O. & Han, L. (2018), ‘Renegotiation of Trade Agreements and Firm Exporting Decisions: Evidence from the Impact of Brexit on UK Exports’, *Society of International Economic Law (SIEL), Sixth Biennial Global Conference*.
- Davis, S. J., Hansen, S. & Seminario-Amez, C. (2020), ‘Firm-Level Risk Exposures and Stock Returns in the Wake of COVID-19’, *NBER Working Paper No.27867*.
- Davis, S. J., Liu, D. & Sheng, X. S. (2019), ‘Economic Policy Uncertainty in China Since 1949: The View from Mainland Newspapers’, *Fourth Annual IMF-Atlanta Fed Research Workshop on China’s Economy Atlanta* **19**, 1–37.
- De Bellefon, M.-P., Combes, P.-P., Duranton, G., Gobillon, L. & Gorin, C. (2021), ‘Delineating Urban Areas Using Building Density’, *Journal of Urban Economics* **125**, 103226.
- Deacon, D., Downey, J., Harmer, E., Stanyer, J. & Wring, D. (2016), ‘The Narrow Agenda: How the News Media Covered the Referendum’, *EU Referendum Analysis* pp. 34–35.
- Deacon, D., Wring, D., Harmer, E., Downey, J. & Stanyer, J. (2016), ‘Hard Evidence: Analysis Shows Extent of Press Bias Towards Brexit’, *The Conversation* **16**.
- Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2018), ‘BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding’, *arXiv Preprint arXiv:1810.04805*.

- Dhingra, S., Huang, H., Ottaviano, G., Pessoa, J. P., Sampson, T. & Van Reenen, J. (2017), ‘The Costs and Benefits of Leaving the EU: Trade Effects’, *Economic Policy* **32**(92), 651–705.
- Dhingra, S., Ottaviano, G., Sampson, T. & Van Reenen, J. (2016), ‘The Consequences of Brexit for UK Trade and Living Standards. Centre for Economic Performance’, *Center for Economic Performance* pp. 4–10.
- Di Iasio, V. & Wahba, J. (2023), ‘Expecting Brexit and UK Migration: Should I Go?’, *European Economic Review* **157**, 104484.
- Dieppe, A., van Roye, B. & Legrand, R. (2016), ‘The BEAR Toolbox’, *European Central Bank Working Paper Series* (1934).
- Dolan, R. J. (2002), ‘Emotion, Cognition, and Behavior’, *Science* **298**(5596), 1191–1194.
- Donaldson, D. & Storeygard, A. (2016), ‘The View from Above: Applications of Satellite Data in Economics’, *Journal of Economic Perspectives* **30**(4), 171–198.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. et al. (2020), ‘An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale’, *arXiv preprint arXiv:2010.11929*.
- Douch, M., Du, J. & Vanino, E. (2020), ‘Defying Gravity? Policy Uncertainty, Trade Destruction, and Diversion’, *Lloyds Banking Group Centre for Business Prosperity, Research Paper No 3*.
- Du, J., Satoglu, E. B. & Shepotylo, O. (2023), ‘How Did Brexit Affect UK Trade?’, *Contemporary Social Science* **18**(2), 266–283.
- Dueker, M. (2005), ‘Dynamic Forecasts of Qualitative Variables: A Qual VAR Model of US Recessions’, *Journal of Business & Economic Statistics* **23**(1), 96–104.

- Ebell, M. & Warren, J. (2016), ‘The Long-Term Economic Impact of Leaving the EU’, *National Institute Economic Review* **236**(1), 121–138.
- Faccini, R. & Palombo, E. (2019), ‘Policy Uncertainty in Brexit UK’, *Technical Report, Queen Mary (MIMEO)*.
- Faccini, R. & Palombo, E. (2021), ‘News Uncertainty in Brexit United Kingdom’, *American Economic Review: Insights* **3**(2), 149–164.
- Fajgelbaum, P. D., Schaal, E. & Taschereau-Dumouchel, M. (2017), ‘Uncertainty Traps’, *The Quarterly Journal of Economics* **132**(4), 1641–1692.
- Felbermayr, G., Fuest, C., Gröschl, J. K. & Stöhlker, D. (2017), ‘Economic Effects of Brexit on the European Economy’, *EconPol Policy Report* (4). EconPol Europe, ifo Institute–Leibniz Institute for Economic Research at the University of Munich.
- Fernandes, A. P. & Winters, L. A. (2021), ‘Exporters and Shocks: The Impact of the Brexit Vote Shock on Bilateral Exports to the UK’, *Journal of International Economics* **131**, 103489.
- Fernández-Villaverde, J., Guerrón-Quintana, P., Kuester, K. & Rubio-Ramírez, J. (2015), ‘Fiscal Volatility Shocks and Economic Activity’, *American Economic Review* **105**(11), 3352–3384.
- Fernández-Villaverde, J., Guerrón-Quintana, P., Rubio-Ramírez, J. F. & Uribe, M. (2011), ‘Risk Matters: The Real Effects of Volatility Shocks’, *American Economic Review* **101**(6), 2530–2561.
- Forni, M. & Gambetti, L. (2014), ‘Sufficient Information in Structural VARs’, *Journal of Monetary Economics* **66**, 124–136.
- Gandomi, A. & Haider, M. (2015), ‘Beyond the Hype: Big Data Concepts, Methods, and Analytics’, *International journal of information management* **35**(2), 137–144.

- Gao, Z., Feng, A., Song, X. & Wu, X. (2019), ‘Target-Dependent Sentiment Classification with BERT’, *IEEE Access* **7**, 154290–154299.
- Gentzkow, M., Kelly, B. & Taddy, M. (2019), ‘Text as Data’, *Journal of Economic Literature* **57**(3), 535–574.
- Gentzkow, M. & Shapiro, J. M. (2010), ‘What Drives Media Slant? Evidence from US Daily Newspapers’, *Econometrica* **78**(1), 35–71.
- Gertler, M. & Karadi, P. (2015), ‘Monetary Policy Surprises, Credit Costs, and Economic Activity’, *American Economic Journal: Macroeconomics* **7**(1), 44–76.
- Ghermandi, A. & Sinclair, M. (2019), ‘Passive Crowdsourcing of Social Media in Environmental Research: A Systematic Map’, *Global Environmental Change* **55**, 36–47.
- Ghirelli, C., Pérez, J. J. & Urtasun, A. (2019), ‘A New Economic Policy Uncertainty Index for Spain’, *Economics Letters* **182**, 64–67.
- Gilchrist, S., Sim, J. W. & Zakrajšek, E. (2014), ‘Uncertainty, Financial Frictions, and Investment Dynamics’, *NBER Working Paper No.20038*.
- Gilchrist, S. & Williams, J. C. (2005), ‘Investment, Capacity, and Uncertainty: A Putty–Clay Approach’, *Review of Economic Dynamics* **8**(1), 1–27.
- Glaeser, E. L., Kincaid, M. S. & Naik, N. (2018), ‘Computer Vision and Real Estate: Do Looks Matter and Do Incentives Determine Looks?’, *NBER Working Paper No.26218*.
- Goodwin, M. J. & Heath, O. (2016), ‘The 2016 Referendum, Brexit and the Left Behind: An Aggregate-Level Analysis of the Result’, *The Political Quarterly* **87**(3), 323–332.
- Gorin, C., Heblich, S. & Zylberberg, Y. (2023), ‘State of the Art: Economic Development Through the Lens of Paintings’, *NBER Summer Institute Development of the American Economy Conference Paper*.

- Gorodnichenko, Y., Pham, T. & Talavera, O. (2021), ‘Social Media, Sentiment and Public Opinions: Evidence from #Brexit and #USElection’, *European Economic Review* **136**, 103772.
- Grabe, M. E. & Bucy, E. P. (2009), *Image Bite Politics: News and the Visual Framing of Elections*, Vol., Oxford University Press.
- Graber, D. A. (1990), ‘Seeing is Remembering: How Visuals Contribute to Learning from Television News’, *Journal of Communication*.
- Graber, D. A. (1996), ‘Say it with Pictures’, *The Annals of the American Academy of Political and Social Science* **546**(1), 85–96.
- Graziano, A. G., Handley, K. & Limão, N. (2020), ‘Brexit Uncertainty: Trade Externalities Beyond Europe’, *AEA Papers and Proceedings* **110**, 552–556.
- Graziano, A. G., Handley, K. & Limão, N. (2021), ‘Brexit Uncertainty and Trade Disintegration’, *The Economic Journal* **131**(635), 1150–1185.
- Griffiths, T. L. & Steyvers, M. (2004), ‘Finding Scientific Topics’, *Proceedings of the National Academy of Sciences* **101**(suppl 1), 5228–5235.
- Groeling, T. (2013), ‘Media Bias by the Numbers: Challenges and Opportunities in the Empirical Study of Partisan News’, *Annual Review of Political Science* **16**(1), 129–151.
- Groseclose, T. & Milyo, J. (2005), ‘A Measure of Media Bias’, *The Quarterly Journal of Economics* **120**(4), 1191–1237.
- Ha, J., Lee, S. & So, I. (2022), ‘The Impact of Uncertainty Shocks: Evidence from Geopolitical Swings on the Korean Peninsula’, *Oxford Bulletin of Economics and Statistics* **84**(1), 21–56.

- Haim, M. & Jungblut, M. (2021), ‘Politicians’ Self-Depiction and Their News Portrayal: Evidence From 28 Countries Using Visual Computational Analysis’, *Political Communication* **38**(1-2), 55–74.
- Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y. et al. (2022), ‘A Survey on Vision Transformer’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**(1), 87–110.
- Hansen, S. & McMahon, M. (2016), ‘Shocking Language: Understanding the Macroeconomic Effects of Central Bank Communication’, *Journal of International Economics* **99**, S114–S133.
- Hansen, S., McMahon, M. & Prat, A. (2018), ‘Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach’, *The Quarterly Journal of Economics* **133**(2), 801–870.
- Hansen, S., McMahon, M. & Tong, M. (2019), ‘The Long-Run Information Effect of Central Bank Communication’, *Journal of Monetary Economics* **108**, 185–202.
- Hassan, S. Z., Ahmad, K., Hicks, S., Halvorsen, P., Al-Fuqaha, A., Conci, N. & Riegler, M. (2022), ‘Visual Sentiment Analysis from Disaster Images in Social Media’, *Sensors* **22**(10), 3628.
- Hassan, T. A., Hollander, S., Lent, L. V. & Tahoun, A. (2024), ‘The Global Impact of Brexit Uncertainty’, *The Journal of Finance* **79**(1), 413–458.
- Hassan, T. A., Hollander, S., Van Lent, L., Schwedeler, M. & Tahoun, A. (2020), ‘Firm-Level Exposure to Epidemic Diseases: COVID-19, SARS, and H1N1’, *NBER Working Paper No.26971*.
- Hassan, T. A., Hollander, S., Van Lent, L. & Tahoun, A. (2019), ‘Firm-Level Political Risk: Measurement and Effects’, *The Quarterly Journal of Economics* **134**(4), 2135–2202.

- Heber, G., Lunde, A., Shephard, N. & Sheppard, K. (2009), ‘Oxford-Man Institute’s Realized Library’. Oxford-Man Institute, University of Oxford. Library Version: 0.3.
- Helman, E., Graber, E. C., Hoffman, L. H. & Gaertner, S. L. (2012), ‘Warmth and Competence: A Content Analysis of Photographs Depicting American Presidents’, *Psychology of Popular Media Culture* **1**(1), 46.
- Heinrich, G. (2005), ‘Parameter Estimation for Text Analysis’, *Technical Report* .
- Hu, A. & Ma, S. (2021), ‘Persuading Investors: A Video-Based Study’, *NBER Working Paper No.29048*.
- Huang, Y. & Luk, P. (2020), ‘Measuring Economic Policy Uncertainty in China’, *China Economic Review* **59**, 101367.
- Husted, L., Rogers, J. & Sun, B. (2020), ‘Monetary Policy Uncertainty’, *Journal of Monetary Economics* **115**, 20–36.
- Javorcik, B., Stapleton, K., Kett, B. & O’Kane, L. (2020), ‘Unravelling Deep Integration: Local Labour Market Effects of the Brexit Vote’, *CEPR Discussion Paper 14222*.
- Jones, L. E., Manuelli, R. E., Siu, H. E. & Stacchetti, E. (2005), ‘Fluctuations in Convex Models of Endogenous Growth, I: Growth Effects’, *Review of Economic Dynamics* **8**(4), 780–804.
- Joshi, A., Bhattacharyya, P. & Carman, M. J. (2017), ‘Automatic Sarcasm Detection: A Survey’, *ACM Computing Surveys (CSUR)* **50**(5), 1–22.
- Jurado, K., Ludvigson, S. C. & Ng, S. (2015), ‘Measuring Uncertainty’, *American Economic Review* **105**(3), 1177–1216.
- Jurgenson, N. (2019), *The Social Photo: On Photography and Social Media*, Vol. , Verso Books.

- Kalamara, E., Turrell, A., Redl, C., Kapetanios, G. & Kapadia, S. (2022), ‘Making Text Count: Economic Forecasting Using Newspaper Text’, *Journal of Applied Econometrics* **37**(5), 896–919.
- Känzig, D. R. (2021), ‘The Macroeconomic Effects of Oil Supply News: Evidence from OPEC Announcements’, *American Economic Review* **111**(4), 1092–1125.
- Keith, K. A., Teichmann, C., O’Connor, B. & Meij, E. (2020), ‘Uncertainty Over Uncertainty: Investigating the Assumptions, Annotations, and Text Measurements of Economic Policy Uncertainty’, *arXiv Preprint arXiv:2010.04706*.
- Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S. & Shah, M. (2022), ‘Transformers in Vision: A Survey’, *ACM Computing Surveys (CSUR)* **54**(10s), 1–41.
- Kierzenkowski, R., Pain, N., Rusticelli, E. & Zwart, S. (2016), ‘The Economic Consequences of Brexit: A Taxing Decision’, *OECD Economic Policy Papers*.
- Klein, M. & Linnemann, L. (2021), ‘Real Exchange Rate and International Spillover Effects of US Technology Shocks’, *Journal of International Economics* **129**, 103414.
- Koylu, C., Zhao, C. & Shao, W. (2019), ‘Deep Neural Networks and Kernel Density Estimation for Detecting Human Activity Patterns from Geo-Tagged Images: A Case Study of Birdwatching on Flickr’, *ISPRS International Journal of Geo-Information* **8**(1), 45.
- Kraft, H., Schwartz, E. & Weiss, F. (2018), ‘Growth Options and Firm Valuation’, *European Financial Management* **24**(2), 209–238.
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012), ‘ImageNet Classification with Deep Convolutional Neural Networks’, *Advances in Neural Information Processing Systems* **25**.
- Kung, H. & Schmid, L. (2015), ‘Innovation, Growth, and Asset Prices’, *The Journal of Finance* **70**(3), 1001–1037.

- Larsen, V. H. (2021), ‘Components of Uncertainty’, *International Economic Review* **62**(2), 769–788.
- Larsen, V. H. & Thorsrud, L. A. (2019), ‘The Value of News for Economic Developments’, *Journal of Econometrics* **210**(1), 203–218.
- Larsen, V. H., Thorsrud, L. A. & Zhulanova, J. (2021), ‘News-Driven Inflation Expectations and Information Rigidities’, *Journal of Monetary Economics* **117**, 507–520.
- Leduc, S. & Liu, Z. (2016), ‘Uncertainty Shocks Are Aggregate Demand Shocks’, *Journal of Monetary Economics* **82**, 20–35.
- Levy, D., Aslan, B. & Bironzo, D. (2016), ‘UK Press Coverage of the EU Referendum’, *Reuters Institute for the Study of Journalism*.
- Litterman, R. B. (1986), ‘Forecasting with Bayesian Vector Autoregressions—Five Years of Experience’, *Journal of Business & Economic Statistics* **4**(1), 25–38.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. & Guo, B. (2021), ‘Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows’, *Proceedings of the IEEE/CVF International Conference on Computer Vision* pp. 10012–10022.
- Ludvigson, S. C., Ma, S. & Ng, S. (2021), ‘Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?’, *American Economic Journal: Macroeconomics* **13**(4), 369–410.
- Makrychoriti, P. & Spyrou, S. (2023), ‘To Be or Not to Be in the EU: The International Economic Effects of Brexit Uncertainty’, *The European Journal of Finance* **29**(1), 58–85.
- Malkhozov, A. (2014), ‘Asset Prices in Affine Real Business Cycle Models’, *Journal of Economic Dynamics and Control* **45**, 180–193.
- Martin, G. J. & Yurukoglu, A. (2017), ‘Bias in Cable News: Persuasion and Polarization’, *American Economic Review* **107**(9), 2565–2599.

- McGrattan, E. R. & Waddle, A. (2020), ‘The Impact of Brexit on Foreign Investment and Production’, *American Economic Journal: Macroeconomics* **12**(1), 76–103.
- Mertens, K. & Ravn, M. O. (2013), ‘The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States’, *American Economic Review* **103**(4), 1212–1247.
- Miescu, M. & Rossi, R. (2021), ‘COVID-19-Induced Shocks and Uncertainty’, *European Economic Review* **139**, 103893.
- Miescu, M. S. (2023), ‘Uncertainty Shocks in Emerging Economies: A Global to Local Approach for Identification’, *European Economic Review* **154**, 104437.
- Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013), ‘Efficient Estimation of Word Representations in Vector Space’, *arXiv Preprint arXiv:1301.3781*.
- Miranda-Agrippino, S., Hoke, S. H. & Bluwstein, K. (2022), ‘Patents, News, and Business Cycles’, *Manuscript, Bank of England*.
- Miranda-Agrippino, S. & Ricco, G. (2023), ‘Identification with External Instruments in Structural VARs’, *Journal of Monetary Economics* **135**, 1–19.
- Moriarty, S. E. & Popovich, M. N. (1991), ‘Newsmagazine Visuals and the 1988 Presidential Election’, *Journalism Quarterly* **68**(3), 371–380.
- Mutz, D. C. (2007), ‘Effects of “In-Your-Face” Television Discourse on Perceptions of a Legitimate Opposition’, *American Political Science Review* **101**(4), 621–635.
- Naik, N., Raskar, R. & Hidalgo, C. A. (2016), ‘Cities Are Physical Too: Using Computer Vision to Measure the Quality and Impact of Urban Appearance’, *American Economic Review* **106**(5), 128–132.
- Nakov, P., Ritter, A., Rosenthal, S., Sebastiani, F. & Stoyanov, V. (2019), ‘SemEval-2016 Task 4: Sentiment Analysis in Twitter’, *arXiv Preprint arXiv:1912.01973*.

- Nguyen, D. Q., Vu, T. & Nguyen, A. T. (2020), ‘BERTweet: A Pre-Trained Language Model for English Tweets’, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* pp. 9–14.
- Noailly, J., Nowzohour, L. & Van Den Heuvel, M. (2022), ‘Does Environmental Policy Uncertainty Hinder Investments Towards a Low-Carbon Economy?’, *NBER Working Paper No.30361*.
- Nofsinger, J. R. (2005), ‘Social Mood and Financial Economics’, *The Journal of Behavioral Finance* **6**(3), 144–160.
- Noh, E. (2017), ‘Impulse-Response Analysis with Proxy Variables’, *Available at SSRN 3070401*.
- Obaid, K. & Pukthuanthong, K. (2022), ‘A Picture Is Worth a Thousand Words: Measuring Investor Sentiment by Combining Machine Learning and Photos from News’, *Journal of Financial Economics* **144**(1), 273–297.
- Orellana, M., Perez, F. & Wagner, R. A. (2024), ‘Monitoring Body Image Representation in Advertisements Using Computer Vision: An Exploration of Gender Disparities’, *Available at SSRN 4748827*.
- Parker, S. C., Belghitar, Y. & Barmby, T. (2005), ‘Wage Uncertainty and the Labour Supply of Self-Employed Workers’, *The Economic Journal* **115**(502), C190–C207.
- Pástor, L. & Veronesi, P. (2006), ‘Was There a Nasdaq Bubble in the Late 1990s?’, *Journal of Financial Economics* **81**(1), 61–100.
- Peng, Y. (2018), ‘Same Candidates, Different Faces: Uncovering Media Bias in Visual Portrayals of Presidential Candidates with Computer Vision’, *Journal of Communication* **68**(5), 920–941.
- Pérez, J. M., Giudici, J. C. & Luque, F. (2021), ‘pysentimiento: A Python Toolkit for Sentiment Analysis and SocialNLP Tasks’, *arXiv Preprint arXiv:2106.09462*.

- Piffer, M. & Podstawski, M. (2018), ‘Identifying Uncertainty Shocks Using the Price of Gold’, *The Economic Journal* **128**(616), 3266–3284.
- Rajadesingan, A., Zafarani, R. & Liu, H. (2015), ‘Sarcasm Detection on Twitter: A Behavioral Modeling Approach’, *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 97–106.
- Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M. & Mozetič, I. (2015), ‘The Effects of Twitter Sentiment on Stock Price Returns’, *PLOS ONE* **10**(9), e0138441.
- Saltzman, B. & Yung, J. (2018), ‘A Machine Learning Approach to Identifying Different Types of Uncertainty’, *Economics Letters* **171**, 58–62.
- Saxegaard, E. C. A., Davis, S. J., Ito, A. & Miake, N. (2022), ‘Policy Uncertainty in Japan’, *Journal of the Japanese and International Economies* **64**, 101192.
- Schmuck, D. & Matthes, J. (2017), ‘Effects of economic and symbolic threat appeals in right-wing populist advertising on anti-immigrant attitudes: The impact of textual and visual appeals’, *Political Communication* **34**(4), 607–626.
- Scotti, C. (2016), ‘Surprise and Uncertainty Indexes: Real-Time Aggregation of Real-Activity Macro-Surprises’, *Journal of Monetary Economics* **82**, 1–19.
- Segal, G., Shaliastovich, I. & Yaron, A. (2015), ‘Good and Bad Uncertainty: Macroeconomic and Financial Market Implications’, *Journal of Financial Economics* **117**(2), 369–397.
- Shapiro, A. H., Sudhof, M. & Wilson, D. J. (2022), ‘Measuring News Sentiment’, *Journal of econometrics* **228**(2), 221–243.
- Simpson, K. & Startin, N. (2023), ‘Tabloid Tales: How the British Tabloid Press Shaped the Brexit Vote’, *JCMS: Journal of Common Market Studies* **61**(2), 302–322.
- Sims, C. A. (1980), ‘Macroeconomics and Reality’, *Econometrica: Journal of the Econometric Society* pp. 1–48.

- Steinberg, J. B. (2019), 'Brexit and the Macroeconomic Impact of Trade Policy Uncertainty', *Journal of International Economics* **117**, 175–195.
- Stock, J. H. & Watson, M. W. (2012), 'Disentangling the Channels of the 2007-2009 Recession', *NBER Working Paper No.18094*.
- Stock, J. H. & Watson, M. W. (2018), 'Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments', *The Economic Journal* **128**(610), 917–948.
- Sullivan, D. G. & Masters, R. D. (1988), '"Happy Warriors": Leaders' Facial Displays, Viewers' Emotions, and Political Support', *American Journal of Political Science* pp. 345–368.
- Sun, C., Huang, L. & Qiu, X. (2019), 'Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence', *arXiv Preprint arXiv:1903.09588*.
- Tiedens, L. Z. (2001), 'Anger and Advancement Versus Sadness and Subjugation: The Effect of Negative Emotion Expressions on Social Status Conferral', *Journal of Personality and Social Psychology* **80**(1), 86.
- Tobback, E., Naudts, H., Daelemans, W., de Fortuny, E. J. & Martens, D. (2018), 'Belgian Economic Policy Uncertainty Index: Improvement Through Text Mining', *International Journal of Forecasting* **34**(2), 355–365.
- Todorov, A., Mandisodza, A. N., Goren, A. & Hall, C. C. (2005), 'Inferences of competence from faces predict election outcomes', *Science* **308**(5728), 1623–1626.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017), 'Attention Is All You Need', *Advances in Neural Information Processing Systems* **30**.

- Verser, R. & Wicks, R. H. (2006), ‘Managing Voter Impressions: The Use of Images on Presidential Candidate Web Sites During the 2000 Campaign’, *Journal of Communication* **56**(1), 178–197.
- Wadsworth, J., Dhingra, S., Ottaviano, G. & Van Reenen, J. (2016), ‘Brexit and the Impact of Immigration on the UK’, *CEP Brexit Analysis* **5**(1), 34–53.
- Waldman, P. & Devitt, J. (1998), ‘Newspaper Photographs and the 1996 Presidential Election: The Question of Bias’, *Journalism & Mass Communication Quarterly* **75**(2), 302–311.
- Wang, Y., Li, Y. & Luo, J. (2016), ‘Deciphering the 2016 US Presidential Campaign in the Twitter Sphere: A Comparison of the Trumpists and Clintonists’, *Proceedings of the International AAAI Conference on Web and Social Media* **10**(1), 723–726.
- Ward, O. (2017), ‘Intersectionality and Press Coverage of Political Campaigns: Representations of Black, Asian, and Minority Ethnic Female Candidates at the UK 2010 General Election’, *The International Journal of Press/Politics* **22**(1), 43–66.
- Willems, T., Choi, S. & Yoo, S. Y. (2022), ‘Revisiting the Monetary Transmission Mechanism Through an Industry-Level Differential Approach’, *Available at SSRN 4037181*.
- Xu, H., Liu, B., Shu, L. & Yu, P. S. (2019), ‘BERT Post-Training for Review Reading Comprehension and Aspect-Based Sentiment Analysis’, *arXiv: 1904.02232*.
- You, Q., Luo, J., Jin, H. & Yang, J. (2015), ‘Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks’, *Proceedings of the AAAI Conference on Artificial Intelligence* **29**(1).