



DEPARTMENT
OF POLITICAL AND
SOCIAL SCIENCES

The Rise of Central Bank Talk: Essays in Central Bank Communication and Independence

Lauren Leek

Thesis submitted for assessment with a view to
obtaining the degree of Doctor of Political and Social Sciences
of the European University Institute

Florence, 7 October, 2025.

**European University Institute
Department of Political and Social Sciences**

**The Rise of Central Bank Talk
Essays in Central Bank Communication and
Independence**

Lauren Leek

Thesis submitted for assessment with a view to
obtaining the degree of Doctor of Political and Social Sciences
of the European University Institute

Examining Board

Prof. Simon Hix, Institution (EUI Supervisor)
Prof. Waltraud Schelkle, Institution (EUI Co-supervisor)
Prof. Kenneth Benoit, Singapore Management University
Prof. Ana Carolina Garriga, University of Essex

© Leek, 2025

No part of this thesis may be copied, reproduced or transmitted without prior
permission of the author

Researcher declaration to accompany the submission of written work
Department of Political and Social Sciences - Doctoral Programme

I Lauren Leek certify that I am the author of the work The Rise of Central Bank Talk: Essays in Central Bank Communication and Independence I have presented for examination for the Ph.D. at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the Code of Ethics in Academic Research issued by the European University Institute (IUE 332/2/10 (CA 297)).

The copyright of this work rests with its author. Quotation from it is permitted, provided that full acknowledgement is made. This work may not be reproduced without my prior written consent. This authorisation does not, to the best of my knowledge, infringe the rights of any third party.

I declare that this work consists of 90,141 words.

Statement of inclusion of previous work:

I confirm that chapter 2 was jointly co-authored with Mr Simeon Bischl and I was responsible for conceptualisation, theory, writing, data curation, empirical strategy and final edits while the co-author contributed to data curation, empirical strategy and final edits. I also confirm that chapter 2 draws upon an earlier article published:

Leek, L., & Bischl, S. (2025). How central bank independence shapes monetary policy communication: A large language model application. European Journal of Political Economy, 87, 102668. <https://doi.org/10.1016/j.ejpoleco.2025.102668>

I confirm that chapter 4 was jointly co-authored with Mr Simeon Bischl and Dr. Maximilian Freier and I was responsible for 60% of the work. This paper is based on work as a PhD trainee in DG Economics at the European Central Bank conducted in May to November 2023. The views expressed are those of the authors and do not necessarily reflect those of the ECB.

Signature and date:

Lauren Leek

25/07/2025

Abstract

Both central bank independence and central bank communication have drastically expanded globally in recent decades. This thesis addresses a puzzling paradox: how do central banks reconcile their independent status with extensive public communication that goes beyond their narrow mandates? I argue that central bank communication is not merely a neutral tool for information transmission; it is dynamically shaped by and employed to address perceived challenges to their independence. This perspective conceptualises independence as an ongoing process, requiring continuous responsiveness to economic and societal realities beyond its formal institutional design. Central banks, therefore, consider more than just their statutory independence; they actively seek public support and macro-economic coordination to effectively fulfil their mandates.

To examine the relationship between central bank independence and communication, I leverage Large Language Models to systematically uncover latent insights from an extensive dataset of speeches from 119 central banks worldwide (1997-2024). This significantly broadens the empirical scope beyond traditional studies focusing mainly on central banks in high-income democracies. The publicly available dataset, which includes novel metadata, is accessible at centralbanktalk.eu.

The dissertation comprises three empirical chapters. Chapter 2 examines how central bank independence shapes communication in response to monetary and financial pressures. Chapter 3 explores how independent central banks - particularly within the multi-level framework of the Eurosystem - respond in communication to broader societal issues and regionally diverse political demands. Chapter 4 introduces a novel textual measure of central bank pressure on monetary policy, introducing a distinction between coordination and dominance in macroeconomic policy linkages. Together, these chapters offer new insights into the evolving political economy of central banking and the role of communication in sustaining an adaptive form of central bank independence.

Thesis accompaniments

This dissertation is accompanied by the following resources, which provide additional context, access and visualisation of the data and code for replication:

- **Thesis website:** <https://www.centralbanktalk.eu>
 - Offers search functions and raw links for all the speeches in the database.
 - Offers insights into and various visualisation options for the speeches data and novel meta-data used throughout this thesis.
 - Contains pre-prints, replication codes, docker containers, presentations and extra materials for all papers of this thesis.
- **Replication codes of all papers:** https://github.com/laurencleek/thesis_replication
 - Completely modularised and commented code which allows for direct replication from the raw data to the final thesis.
 - Check the README for instructions how to run the replication code on your local device and/or GPU.

For further technical details regarding the data and replication, installation guides and additional documentation, please consult the resources above.

Acknowledgements

What a journey! A PhD is never a solo performance. Mine has featured an extraordinary cast spread across three countries and four institutions. If I've forgotten to mention someone, please know the fault lies with my memory, not with my gratitude.

First and foremost, both of my supervisors, Simon Hix and Waltraud Schelkle deserve a massive thank you for all their support, both within and outside the academic premises.

Simon, you have been an exceptional supervisor. Your enthusiasm, encouragement, flow of ideas and optimism never ran short. You (and your family) offered endless support and advice - not just academically, but far beyond. I couldn't have completed this journey without your guidance, belief, and numerous pep-talks. I consider myself lucky to have had you as a supervisor.

Waltraud, my gratitude to you stretches back long before this PhD. You believed in me as (a very young) LSE MSc student, encouraged me to apply for doctoral studies, and later joined my supervision team. You also made it possible for me to be at the ECB. This PhD most certainly would never have happened without you and for the continuation of your support throughout.

I am also grateful to my committee members, Kenneth Benoit and Ana Carolina Garriga, for their thoughtful and constructive feedback.

As I started with, this (almost 4 year) journey spanned multiple countries and institutions, each contributing something invaluable - shaped by those I met along the way.

Florence was both its starting and end point. In the early days, I was fortunate to benefit from the support of scholars such as Erik Jones, Philipp Genschel, Sebastian Diessner, Jens van t'Klooster, Bjorn Bremer, and Manuela Moschella. In particular, a big thanks to Erik for his central banking enthusiasm! I am also thankful for the many invitations in the early days of my PhD to be part of workshops: the UCL 'How people think about the economy', the Sciences Po 'Central Bank Independence in the 21st Century', the MPIfG 'Politics of Central Banking' and the 'LLMs in Political Science' in Barcelona. Besides conferences, for feedback on Chapter 2, I would like to thank Davide Romelli, for Chapter 3 our 'LLMs in Political Economy group', for Chapter 4, the AI in Economics group at the ECB, and for feedback on all chapters, I am thankful to the participants of the many good sessions of Simon's and Waltraud's colloquium.

Throughout this time in Florence, my 'EUI crew' kept me grounded: Jordy, Maxine, Leon, Jessica, Sven, and Milan - thank you for the dinners, city trips, drinks, distractions, cycles, coffees, visits in different countries, and constant friendship. I am certain that this PhD would have been a lot less fun without you. To my great PhD cohort: thank you for being in this with me (you are slightly too many to mention). I'm also grateful to my co-authors: Lorenzo, Simeon, Jordy, and Philipp and those with whom work is still in progress. Also special mention goes to Johanna, Robert, Cathal, Robin, John, Mirko, Luuk, Emil, Martin, Fabian, Magdalena and many more who during the first or last year have been great support. On a different note, sitting still is definitely not my talent, so this PhD is powered by the countless runs; cycling deep into the Tuscan hills; five-a-side matches at Villa Schifanoia (shout-out to Unathletic Bilbo, PSG, and under 'slight' pressure I mention Emil and the Pancakes) - thanks to all who joined me!

Frankfurt pushed me to engage with the real world. My time at the European Central

Bank in the Fiscal Policy Division (DG Economics) had a lasting impact. Max, thank you for bringing me in (and extending my stay) - which required some creative rewriting of the rules. Siria (and the strategic communications group), thank you for all your insights. Simeon, thank you for being my go-to for methods chats, later flatmate when you began your PhD in Florence when I entered my final year, and (so far) double co-author. Working with you at the ECB and for a long-time after definitely improved this thesis. I'm also grateful to, in particular, Benoit, and also to Tom, Fabio, and Julia. A special shout-out to Anna for dragging me to just about every central banking conference.

London is where the idea of this PhD first took root. Returning to LSE and teaching/working at the DPE at King's College helped shape my work in important ways. Thank you to Jonathan Hopkin for hosting me at LSE for an extended year, to Chris Anderson, to my European Institute office mates and to a selected bunch of Government Department PhD students - for the many great conversations. I'm also grateful to the Methodology department (Dan de Kadet). Thank you also to Christel Koop, Daniele Girardi, and Julian Limberg at KCL. Another special shout-out to Irene and Agni - the ECB London girls - for being such good company. I also cannot forget to mention the many (non-academic) friends out there still keeping me (sort of) sane in London! The fact that I'm writing this having once again moved back to London is, I suppose, a good sign.

On the home front: friends and those who became more than friends - your visits and support wherever I was meant the world. To my mum, for your endless support and well-meaning critiques (*Jallah!*). To my dad, for sparking my political-economic interests and curiosities very early on. To my brother, Andrew - thank you for always being there. And to my Oma and Opa: Opa, you told me that your only regret was not being there to see me defend my PhD. The day you passed was the day I wrote this: *Ik denk aan je.*

List of Tables

Table	Page
1.1 Topic frequency by CBI	7
1.2 Overview of the data, empirical methods and text-as-data methods . .	29
2.1 Examples of sentence classifications	53
2.2 Effect heterogeneity by country characteristics	67
2.3 2SLS estimation of the impact of CBI on monetary and financial dominance	69
2.4 Aggregated event-study estimates of supervision changes	77
4.1 Classification examples.	133
4.2 Classification Share (in Percent)	145
4.3 Validation metrics of final model.	145
4.4 Validation metrics for GPT 3.5-0301, 3.5-0125, 4-0125.	147
4.5 Few shot learning and gemini	148
4.6 Distribution of topics	154
B1.1 Speech metadata	185
B2.2 Classification examples	187
B3.3 Dataset coverage	193
B3.4 Descriptive Statistics of Main Variables	196
B4.5 Single coefficient estimates	204
B4.6 Effect of CBI by audience	210
B5.7 Hyper-parameter tuning options	220

LIST OF TABLES

B5.8	Final model parameters	223
B6.9	Validation metrics	224
B7.10	Lagged dependant variable definition	230
B7.11	Instrumental variable regressions with interaction term	232
C1.1	Overview of the number of speeches cumulative over time	247
C1.2	Overview of the number of speeches per national central bank	248
C1.3	Variables names, data source and timespan	249
C1.4	Summary Statistics	249
C1.5	Country heterogeneity in terms of variables	253
C2.6	Topics 1 to 15: most important terms with c-TF-IDF scores	262
C2.7	Topics 16 to 30: most important terms with c-TF-IDF scores	263
C2.8	Topics 31 to 45: most important terms with c-TF-IDF scores	264
C5.9	Impact of NCB Topic \times Sintra interaction on ECB responsiveness	294
D2.1	List of central banks, time periods and number of speeches	308
D2.2	Variables in the database	314
D3.3	Level 1 classification examples	320
D3.4	Level 2 classification examples	321
D3.5	Level 3 classification examples	322
D3.6	Coder overlap matrices	324
D3.7	Coder reliability scores	324
D4.8	Validation metrics of system messages.	327
D4.9	Performance metrics of different prompt configurations	327

List of Figures

Figure		Page
1.1	World average CBI.	2
1.2	Map of Central Bank Independence scores in 2022.	3
1.3	Top 10 topics in central bank speeches over time	6
1.4	The analytical structure of the thesis	10
1.5	Speeches dataset frequency	20
1.6	Speaker position of all speeches	21
1.7	Audiences over time	21
1.8	Text-as-data methods overview	24
2.1	Main relationship between independence and communication	41
2.2	Relationship between monetary policy communication and CBI	47
2.3	Workflow: data and measurement used to create our database	50
2.4	Distribution of central bank speeches according to country group	52
2.5	LLM classification categories breakdown	55
2.6	Indicators of dominance	57
2.7	Dataset coverage and number of independence changes	58
2.8	Distribution of independence changes	62
2.9	Event studies: dynamic effects on monetary and financial dominance	63
2.10	Event study estimates of the impact of CBI on policy pressures	71
2.11	Relationship of inflation and financial pressures with dominance	73
2.12	Effect of CBI on financial dominance	74

LIST OF FIGURES

2.13	Alternative estimation samples	75
3.1	Text-as-data workflow steps	95
3.2	Smoothed topic proportions over time	98
3.3	Empirical strategy overview	99
3.4	First movers on emerging topics	106
3.5	Markov Transition Matrix of central banks conditional on the topic . .	107
3.6	Coefficient plot using three-month lags	109
3.7	Interaction grid of three-way interactions	111
4.1	Geographical spread of the number of speeches available per central bank.	129
4.2	Three stage coding scheme to identify dominance and coordination. . .	131
4.3	Variation in accuracy with different temperature settings on level 3 classification.	138
4.4	Relationship between sentence count, accuracy and token usage. . . .	140
4.5	Confusion matrix	144
4.6	Share of sentences classified with most frequent label by agreement of human coders.	150
4.7	Level 1 classification results of the topic model.	153
4.8	Level 2 classification results of the descriptive/normative classification.	155
4.9	Level 3 classification results of policy coordination and dominance. . .	156
4.10	Inflation and policy linkage indicators	158
4.11	Evolution of policy linkage indicators in political regimes over time . .	160
4.12	Dominance and monetary-fiscal coordination during crisis and non-crisis times	162
B4.1	Controlling for linear trends and macroeconomic controls	198
B4.2	Event study with continuous treatments	200

LIST OF FIGURES

B4.3	Alternative event study estimators	202
B4.4	Effect of CBI sub-indicator changes on dominances	207
B4.5	Robustness sample construction	208
B4.6	The effect of CBI on audiences	209
B4.7	Dataset coverage and number of independence changes	211
B4.8	Correlation of CBI scores	212
B4.9	Comparison of CBI events	213
B4.10	Event study estimates with different CBI datasets	215
B4.11	Treatment randomisation placebo (aggregated)	216
B4.12	Treatment randomisation placebo (dynamic effects)	218
B5.13	Hyper-parameter influence on validation metrics	222
B6.14	Confusion Matrices of LLM classifiers	225
B6.15	Monetary and fiscal dominance against a model based measure	226
B6.16	Financial pressure and financial dominance	227
C1.1	Frequency of speeches per central bank and over time	248
C1.2	Google search interest for Germany	250
C1.3	Google search interest for France	251
C1.4	Trust in EU institutions	252
C1.5	FT salience	253
C2.6	Wordcloud of the topics 1-22	257
C2.7	Wordcloud of the topics topics 23-44	258
C2.8	Static representation of the 2D intertopic distance map	259
C2.9	Topic regrouping map	260
C2.10	UMAP intertopic distance map	261
C3.11	First-mover graph crisis management	266
C3.12	First-mover graph climate	266

LIST OF FIGURES

C3.13	Matrix window robustness: banking supervision	268
C3.14	Matrix window robustness: monetary policy	269
C3.15	Matrix window robustness: climate	270
C3.16	Matrix window robustness: crisis management	271
C3.17	Matrix window robustness: financial markets	272
C3.18	Matrix window robustness: international economics	273
C3.19	Matrix window robustness: national economy	274
C3.20	Matrix window robustness: payment systems	275
C3.21	Matrix window robustness: digital finance	276
C3.22	Matrix window robustness: economic indicators	277
C4.23	Coefficient plot pre and post-crisis	279
C4.24	Coefficient plot with leads	280
C4.25	Coefficient plot half-year lags	281
C4.26	Margin plots: monetary policy and economic indicators	283
C4.27	Margin plots: climate and crisis management	284
C4.28	Margin plots: financial markets and banking supervision	285
C4.29	Margin plots: national economics and payment systems	286
C4.30	Margin plots: international economics and digital finance	287
C4.31	Interaction grid of half yearly three-way interactions	288
C5.32	Markov transition matrix between central banks during the eurocrisis years	290
C5.33	Coalition interaction grid	292
C6.34	Markov Transition Matrix of topic transitions	297
C6.35	Topic transition matrix time window robustness	299
C6.36	Intra-topic distance monetary policy	301
C6.37	Intra-topic distance climate	301

LIST OF FIGURES

C6.38	Permutations of runs	304
D3.1	Three-level coding classification overview.	318
D4.2	Relationship between sentence count, accuracy and token usage	328
D4.3	Variation in accuracy with different temperature settings	328
D5.4	Level 2 classification by country group.	329
D5.5	Scatter chart of correlation of bond spreads with coordination and dominances in advanced and non-advanced countries.	330
D5.6	High and low polarization countries over time.	331
D5.7	Correlation of GDP per capita (PPP) with policy-linkages by economic development	332
D5.8	Means of advanced and emerging and developing economies over time. .	333

Table of Contents

List of Tables	viii
List of Figures	x
1 Introduction: The Rise of Independence and Communication	1
1.1 The paradox of modern central banking	1
1.2 My argument and contributions	8
1.2.1 Paper contributions	12
1.2.2 Underlying rationales	15
1.3 Measuring communication	18
1.3.1 Data: speeches and meta-data	18
1.3.2 Quantifying central bank speeches	22
1.3.2.1 General overview	22
1.3.2.2 LLMs: challenges and pitfalls	24
1.3.3 General empirical strategy	28
1.3.4 Measurement and data contributions	29
1.4 Thesis outlook	31
2 How Central Bank Independence Shapes Monetary Policy Communication: A Large Language Model Application	33
2.1 Introduction	34
2.2 A theory of central bank independence and communication	39

TABLE OF CONTENTS

2.2.1	Observable expectations	44
2.3	Data and measurement	49
2.4	Empirical analysis	59
2.4.1	Staggered difference-in-differences	59
2.4.2	Results	62
2.4.3	Instrumental variable approach	66
2.4.4	Mechanisms	70
2.4.5	Alternative explanations	72
2.4.6	Robustness checks	77
2.5	Conclusion	79
3	Who Sets the Agenda? The Role of National Central Banks in the Eurosystem	83
3.1	Introduction	84
3.2	Central bank agenda-setting and issue-responsiveness	87
3.3	The Eurosystem set-up	90
3.4	Data and methodology	94
3.4.1	Speeches-as-data and topic modeling	94
3.4.2	Empirical strategy	98
3.5	Results	105
3.5.1	Sequence analysis	105
3.5.2	Cross-sectional timeseries	108
3.5.3	Heterogeneous effects	111
3.6	Alternative explanations	113
3.7	Conclusion	115
4	Introducing Textual Measures of Central Bank policy linkages Using GPT-3.5	117

TABLE OF CONTENTS

4.1	Introduction	118
4.2	Monetary, fiscal and financial policy linkages in central bank communication	120
4.3	Constructing and validating our indices	125
4.3.1	Choice of GPT-3.5	126
4.3.2	Corpus and pre-processing	128
4.3.3	Identifying policy linkages	130
4.3.4	Validation and prompt engineering	133
4.3.4.1	Manual validation	134
4.3.4.2	Prompt engineering steps	135
4.3.4.3	Validation scores	142
4.3.4.4	Improvements and uncertainty measures	145
4.3.5	From raw data to constructing indices	149
4.4	Descriptive variation of the indices	151
4.5	Conclusion	162
5	Conclusion: More Than Just Talk?	165
5.1	Summary	165
5.2	Limitations and further research	166
5.3	Broader implications	172
5.4	Final remarks	175
Appendices		177
Appendix Introduction		178
A1	Extra information	178
A1.1	Figure 1.3	178
A1.2	Table 1.1	179

TABLE OF CONTENTS

A1.3	Figure 1.4	179
Appendix Chapter 2		181
B1	BIS dataset processing	181
B1.1	Speeches corpus pre-processing	181
B1.2	Metadata extraction	182
B2	Codebook	185
B3	Extended dataset description	187
B3.1	BIS Speeches database	187
B3.2	CBI data	187
B3.3	Supplementary datasets for subgroup & mechanism analyses	193
B4	Event study robustness checks	196
B4.1	Relaxing the parallel trends assumption	196
B4.2	Accounting for treatment intensity and multiple treatments	199
B4.3	Heterogeneity robust estimators	199
B4.4	Dynamic treatment effect aggregation	202
B4.5	Sample and treatment indicator variations	204
B4.6	Effect of CBI on audiences	207
B4.7	Alternative CBI dataset	210
B4.8	Randomisation tests	215
B5	LLM finetuning	219
B6	LLM validation	223
B6.1	Policy pressures	223
B6.2	External validation	225
B7	Instrumental variable design	228
B7.1	Choice of lagged dependent variable	228
B7.2	Heterogeneity estimates using the IV approach	230

TABLE OF CONTENTS

B8	Prompts	232
B8.1	Policy pressure channel	233
B8.2	Metadata extraction	235
B8.3	Audience classification	241
B8.4	Geographic location	244
Appendix Chapter 3		247
C1	Data	247
C1.1	Speeches-as-data	247
C1.2	Additional data	248
C1.2.1	Google trends data	250
C1.2.2	Eurobarometer data	251
C1.2.3	Media salience	252
C1.2.4	Ranking countries	253
C2	Topic modeling	254
C2.1	Speeches corpus pre-processing	254
C2.2	BERTopic model steps and choices made	255
C2.3	Topic creation	256
C3	Sequence analysis robustness checks	265
C3.1	First-mover graphs	265
C3.2	Markov windows robustness checks	267
C4	Timeseries analysis robustness checks	278
C4.1	Two-way interactions	278
C4.2	Three-way interaction margin plots	282
C4.3	Three-way interaction grid robustness tests	288
C5	Heterogeneous effects	289
C5.1	Salience: Eurocrisis	289

TABLE OF CONTENTS

C5.2 Coalitions	290
C5.3 Sintra effects	292
C6 Alternative explanations	295
C6.1 Between topic transitions	295
C6.2 Within-topic responsiveness	300
C6.3 Permutation-based validation of transition patterns	302
Appendix Chapter 4	305
D1 Replication files on Github	305
D1.1 Instructions to run codes	305
D1.2 Included files	306
D1.3 Replication of Charts	307
D1.4 Prompts	308
D2 Database Construction	308
D2.1 Pre-processing steps	313
D2.2 Variables in the Database	314
D3 Validation	318
D3.1 Graphical coding overview	318
D3.2 Ambiguous sentence coding guidelines	319
D3.3 Classification Examples	320
D3.4 Coder reliability scores	323
D3.5 Validation metrics	324
D4 Additional prompt engineering results	326
D5 Additional variation in indices	329
D6 Prompt variations	334
D6.1 Minimal Prompt	334
D6.2 Extended Prompt	334

TABLE OF CONTENTS

Bibliography	339
--------------	-----

Chapter 1

Introduction: The Rise of Independence and Communication

“Look back with nostalgia at your time of independence, it was nice while it lasted” (*Charles Goodhart quoted in Marsh, 2021*).

1.1 The paradox of modern central banking

In the last few decades, modern democracies have extensively delegated governmental functions to non-majoritarian institutions run by neutral experts – technocrats (Thatcher and Sweet, 2010; Thatcher et al., 2023). These institutions have been constrained by narrow mandates, have only limited instruments at their disposal, are not directly elected and are subject to potential override by elected politicians. Through delegation, public authority can be exercised without direct involvement of elected politicians. By varying the amount of authority delegation and its constraints, lawmakers can protect their (broader) political priorities (Epstein and O’Halloran, 1999; Huber and Shipan, 2002). Independent institutions have been established in various realms, including in the regulatory, supervisory, judicial, executive and monetary realms. One of the most independent among these non-majoritarian institutions are central banks. Specifically, the functional or technical aim of Central bank independence (CBI) was to shield monetary policymakers from political manipulations by “tying one’s hands” through delegation and institutionalised commitments to a narrow policy remit, typically an inflation target

(Barro and Gordon, 1983b; Giavazzi and Pagano, 1988).¹

Although CBI goes back to the early 20th century (Eichengreen, 2008), it was not until the late 1980s and early 1990s that support for the idea of CBI became widespread (McNamara, 2002). This support spread globally, establishing an ‘ideal’ type for both developing and advanced economy central banks, democracies and, curiously, also autocracies (e.g., see Romelli, 2022; Aklin et al., 2022; Garriga, 2025a) and was promoted by international organisations such as the OECD and the IMF. The spread of CBI nowadays continues with formal institutional set-ups and legal mandates reminiscent of these times. Figure 1.1 shows the changes over time of CBI world-wide from 1970 to 2022 of various sub-indicators and an aggregated score. Figure 1.2 shows current independence. Both figures use the data of Romelli (2022) based on the codification of central banks statutes.

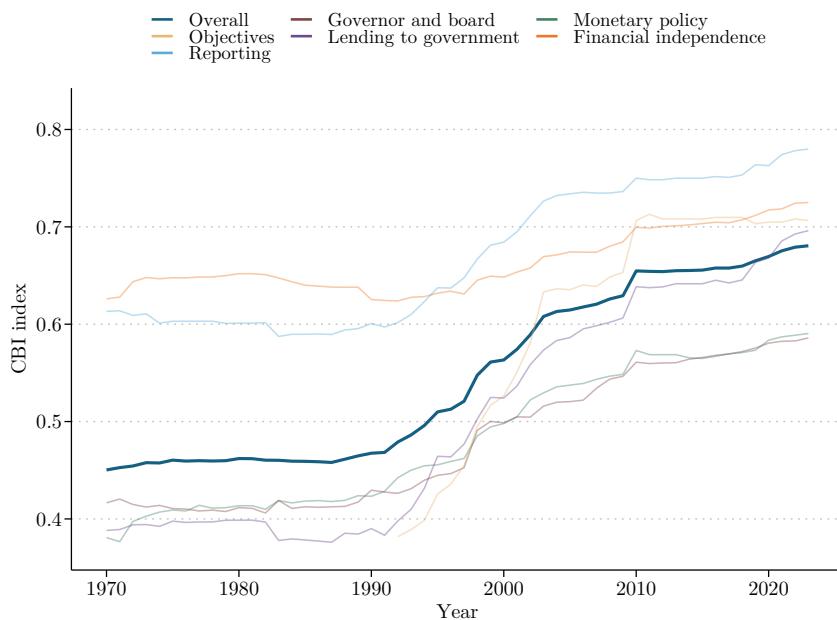


Figure 1.1: World average CBI.

Note: The graph shows the average CBI scores of Romelli (2022) from 1970 to 2022. The dark blue line shows the aggregated CBI score and the lighter lines represent the sub-indicators of the aggregated score.

¹This purely functional aim of CBI, was not the only one, as was already recognised by, *inter alia*, McNamara (2002) who argued that delegating has important legitimising and symbolic proprieties.

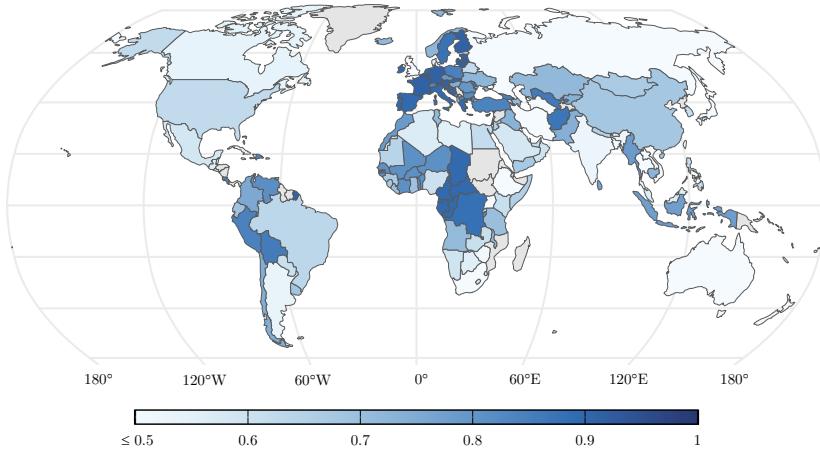


Figure 1.2: Map of Central Bank Independence scores in 2022.

Note: The colour scale indicates the level of central bank independence. The countries shown are the countries which are included in the independence dataset of Romelli (2022). Countries that are missing are shown in grey. This graph shows that CBI is not limited to advanced democratic countries but, rather, is a global phenomenon and implemented regardless of the political system and economic development level.

In parallel with the rise of CBI, another trend emerged. Namely, with increased independence and inflation targeting, the role of central banks to provide information about the economy and monetary policy has also become part of its core functions, leading to the rise of central bank communication. In just a few decades central banks went from:

“If I seem unduly clear to you, you probably misunderstood what I said.”

(Alan Greenspan as Chairman of the Fed, 1987 as quoted in the Guardian on November 4, 2005)

“The effects of monetary policy depend critically on the public getting the message about what policy will do months or years in the future.” *(Janet Yellen as Chairwoman of the Fed, 2013, in a speech at the Society of American Business Editors and Writers’ 50th Anniversary Conference)*

In other words, central bank communication has gone from being something ‘not done’ to a core tool of central banks over the past three decades. Before the 1990s, central banks were known to be secretive. Montagu Norman of the Bank of England after WWI once famously said: ‘Never excuse, never explain’. Exemplary of this is also the title of William Greider’s bestselling book in 1987 which focused on the Federal Reserve: ‘Secrets of the Temple’. He and central bankers of the time saw central banking as something incomprehensible to the general public. This all changed in the 1990s - a shift Blinder (2004) referred to as the ‘quiet revolution’ and described it as a remarkable change in the consensus of central bank communication over the past three decades. More specifically, Haldane et al. (2020) describe the paradigm shift toward greater transparency as the first of two waves of the quiet revolution. The shift towards transparency in central bank communication occurred as Casiraghi and Perez (2022) argue, especially with the adoption of inflation targeting. When an inflation target is set, central banks can clearly communicate this, which helps the formation of expectations. In 1989, the Reserve Bank of New Zealand pioneered this.

In line with this, the literature widely recognises effective central bank communication as a critical component of their toolkit (Blinder et al., 2008). The standard argument being that communication serves to shape market expectations and enhance predictability, thereby aiding macroeconomic objectives (Woodford, 2005). Scholars have extensively investigated the macroeconomic and financial effects of central bank announcements, generally finding both short- and long-term impacts on financial markets by ‘creating news’ and ‘reducing noise’ (e.g., Brand et al., 2010; Jarociński and Karadi, 2020; Geraats, 2002). This became particularly evident during the Global Financial Crisis, where communication, notably forward guidance, became an active policy tool to anchor expectations after conventional monetary policy space had been exhausted (Casiraghi and Perez, 2022).

Taking the usage of central bank communication as a monetary policy tool seriously,

we can see a puzzling trend. Not only do central banks with various levels of independence, both in advanced and developing economies, democratic and autocratic, engage in a wide range of topics, massively increasing the use of communication for monetary policy over time (Woodford, 2005; Blinder et al., 2008; Hansson, 2021; Dincer and Eichengreen, 2018), this communication has also gone far beyond communication narrowly concerning primary mandates (e.g., see Hansson, 2021). In other words, independent central banks with narrow mandates seem to use communications far beyond monetary policy.

Figure 1.3 confirms this trend by previewing the topics discussed by all central banks from 1997 to 2024. Over time, we can see that topics such as inflation, interest rates and monetary policy are stable and high. However, we can also see the prominence of other topics relating to financial markets (financial stability, financial inclusion, risk management) and fiscal policy. Table 1.1 presents the distribution of topics across speeches given by central banks with high versus low levels of independence and the ratio of how frequent a certain topic appears in speeches of high CBI compared to low CBI countries. The results show that, curiously, central banks in high CBI countries are significantly more likely to address topics related to a broad range of topics outside a typical central bank mandate such as structural reforms, fiscal consolidation and climate change. By contrast, central banks in low CBI countries focus more on short-term economic concerns such as exchange rates, unemployment, and economic growth.

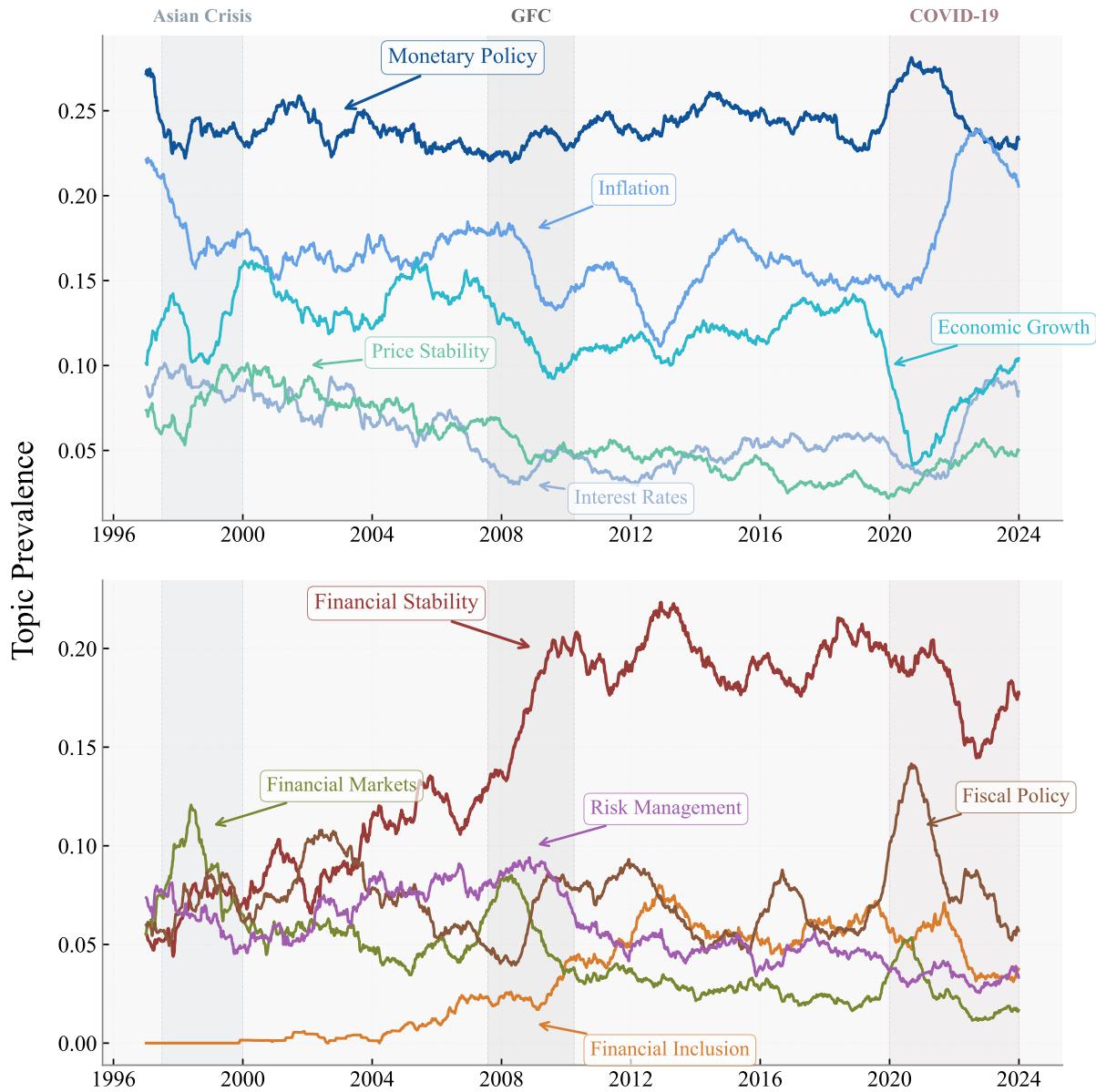


Figure 1.3: Top 10 topics in central bank speeches over time

Note: I use all speeches available of 119 central banks from 1997 to 2024 (see centralbanktalk.eu for the entire database and meta-data). See Appendix A1 the technical details of how this topic model is run.

Table 1.1: Topic frequency by CBI

Topic	Share of speeches containing topic		Ratio of topic frequency
	Low CBI	High CBI	
Topics predominantly covered by high CBI			
Structural Reforms	2.27%	14.23%	6.27
Financial Integration	0.90%	3.41%	3.78
Fiscal Consolidation	1.16%	3.82%	3.30
Price Stability	6.21%	15.96%	2.57
Climate Change	2.07%	5.05%	2.44
Basel III	1.49%	3.54%	2.37
Macroprudential Policy	1.56%	3.69%	2.37
Banking Supervision	1.32%	3.01%	2.27
Banks	1.09%	2.04%	1.88
Capital Requirements	1.61%	2.92%	1.81
Topics predominantly covered by low CBI			
Exchange Rate	6.03%	1.92%	0.32
Global Economy	4.00%	2.18%	0.55
Financial System	3.70%	2.12%	0.57
Financial Sector	2.73%	1.57%	0.58
Risk Management	12.18%	7.26%	0.60
Unemployment	3.44%	2.21%	0.64
Productivity	4.63%	3.07%	0.66
Inflation Expectations	3.27%	2.37%	0.72
Economic Growth	23.10%	18.16%	0.79
Communication	2.13%	1.68%	0.79

Note: This table uses the same speeches database as Figure 1.3 and lists topics predominantly addressed by either high or low CBI countries. The ratio of topic frequency indicates how much more (or less) frequently a topic appears in speeches from high CBI countries compared to low CBI countries. A value above 1 means the topic is more common in high CBI countries, while a value below 1 means it is more common in low CBI countries. See Appendix A1 for the technical details.

Taken together, these developments show the parallel rise of central bank independence and central bank communication, whereby we especially observe that the most independent central banks were not intended to seek outside support, yet seem to use communication to this end. This leads us to the central question in this thesis: how do central banks reconcile their independent status with extensive public communication,

moving beyond their narrow mandates?

In what follows, I present the argument of this thesis reflecting how communication is shaped by and shapes independence. This is followed by a discussion of the contributions I make to the literature. Next, in Section 1.3, I describe the central bank communication database I use in this thesis and how innovations within the measurement realm have enabled the systematic measurement of central bank communication. This is followed by an overview of the data and measurement contributions. In the last section, Section 1.4, I provide an outlook of the thesis.

1.2 My argument and contributions

Crucial to understanding this puzzling trend and answering the central question of this thesis is how I define central bank independence. The central banking literature focuses mainly on statutory dimensions of central bank independence understood as independence from government. A distinction is often made between goal independence (a central bank can set its own target) and instrument or functional independence (a central bank can choose its own instruments to achieve a target that is set by the government) (Fischer, 1995). In addition, personal independence is often mentioned as referring to the nomination and dismissal of the Governor and of the members of the central bank's decision-making bodies pertaining to the political authorities (as was first identified by Parkin and Bade, 1978) which includes the term of office, professional qualifications, political affiliations and collegiality. Moreover, a central bank cannot credibly operate in an independent way without proper financial means. The concept of financial independence is whether any third party is able to exercise either direct or indirect influence, not only over central bank tasks, but also over its ability (in both operational (e.g., staffing) and financial (e.g., resource adequacy) terms) to fulfil its

mandate (Stella, 2005). These various components of statutory independence are often used to create CBI indices, first by Cukierman et al. (1992) and later by many more scholars including Garriga (2025b) and Romelli (2022). Some authors use *de facto* rather than *de jure* measures of independence, such as Dreher et al. (2010) who examine the probability that a central bank governor is replaced before the end of their legal term in office. These authors already assume a more dynamic version of independence.

However, throughout this thesis, rather than treating central bank independence as a static institutional feature or using governor turnover rate to proxy it, I argue that CBI is a dynamic construct that requires ongoing efforts to address changing economic and societal realities. Central banks care not only about their statutory independence, but also about securing public support and coordinating their policies. Therefore I argue that independence allows for, and may even require, a degree of responsiveness to outside pressures, i.e., offering credible commitment and being responsive are not mutually exclusive. This is also in line with the literature on other technocratic agencies for which it has been found that they are responsive to public contention, at both the national and supranational level (e.g., see Koop and Lodge, 2020; Koop et al., 2022; Busuioc and Lodge, 2016). The more general point is that central banks are embedded within their broader political economy environment (e.g., see Qanas and Sawyer, 2024) and their monetary policy has linkages with other macroeconomic policies (e.g., see Smets, 2014). This dynamic view of central bank independence fundamentally challenges ‘institutional autonomy purists’ who argue that external pressures and perceptions thereof should not influence a central bank, as it could compromise independence (e.g., see Cukierman et al., 1992).

When one takes this dynamic definition one can identify various = challenges to the independence of central banks which legalistic definitions using statutory or turnover rates of governors overlook. It then becomes clear that the CBI delegation logic was

developed under peculiar socio-historical conditions which formed the ideal conditions for CBI to thrive, i.e., relative macroeconomic stability and broad political support. There was limited reflection regarding further consequences which only became clear over time, especially recently. Namely, CBI has been subject to various economic, political and societal pressures not fully foreseen in the original independence delegation formulation. My main argument being that communication is used to aid in addressing these perceived threats. The nature and strength of these threats is shaped by the level of central bank independence and central bank communication is both shaped by and used to address these threats. I refer to these ‘threats’ to independence very broadly (as will be identified in the individual chapters: financial and monetary pressures - Chapter 2, changing societal expectations - Chapter 3, and fiscal misalignment - Chapter 4. Figure 1.4 summarises the analytical structure of the thesis as a whole.

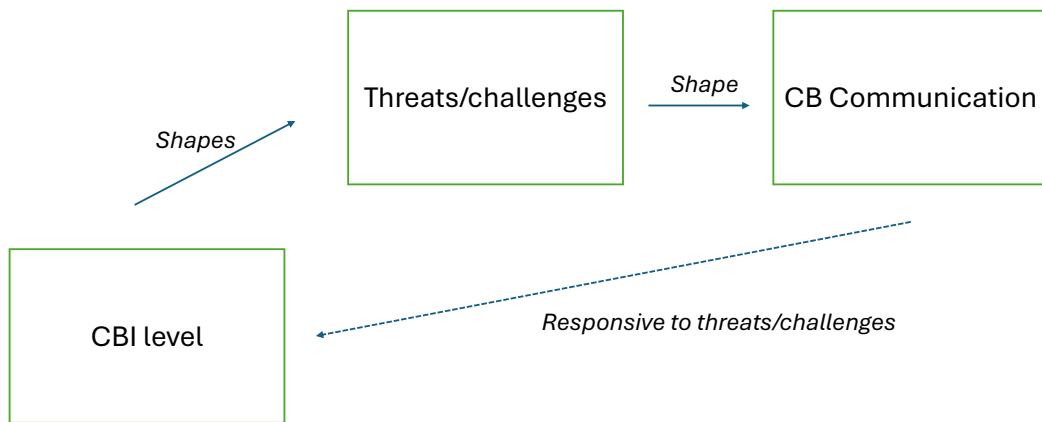


Figure 1.4: The analytical structure of the thesis

These unanticipated financial, political and societal pressures are not limited to the usual case studies in the literature. Namely, the focus has thus far mainly been

on explaining high income democratic country central banks and have mainly focused on the ‘usual suspects’, i.e., the European Central Bank, the Bank of England, the Federal Reserve Bank and the Bank of Japan. These pressures also affect emerging, developing, and autocratic countries - often to an equal or greater extent. Or on the contrary, certain financial pressures that seemed to be only a problem for emerging economies are impacting high income countries now. In this thesis I go beyond these high-income countries and propose a more general argument across advanced, emerging, and developing economies, as well as both democracies and autocracies.

This extension challenges some of the theoretical foundations of central bank independence theory. Namely, the logic of delegating to non-majoritarian institutions such as central banks stems from an inherent problem in democracies: there is a requirement of regular elections as a powerful constraint on the arbitrary use of powers in the short-term while various problems faced by societies require long-term solutions (Majone, 1996). Modern democracies have found a solution to this problem as a means to gain credibility for long-term policymaking: extensive delegation of governmental functions to non-majoritarian institutions with neutral experts - technocrats (Thatcher and Sweet, 2010; Thatcher et al., 2023). The logic of central bank independence also has its foundations in these democratic theories. In the context of monetary policy, economists have long discussed this dilemma as a rules versus discretion debate: whether governments should tailor policies to current economic circumstances (discretionary policy) or conduct policy according to pre-announced rules (Kydland and Prescott, 1977). However, subsequent studies by Garriga (2010) and Aklin et al. (2022) find that CBI is also important for autocratic countries which depend, at least to some degree, on international trade and international debt markets and, therefore, need credible backing of their currency to settle international payments and equilibrate the balance of payments. I complement this literature by showing that delegation has not isolated advanced economy central banks

from similar pressures as emerging markets.

In my empirical papers, I confirm differences in communication depending on the political regime. Although communication is equally important in autocracies and democracies, more transparency (especially in monetary policy) is conveyed in central bank communication in autocracies; this supports the rationale of Broz (2002), the transparency of monetary commitments and the transparency of political systems are substitutes. Second, there is a difference in macroeconomic pressures between developing and developed economy countries. Low-income countries, arguably, often already have high levels of financial pressures regardless of the level of independence (e.g., see Evdokimova et al., 2023) which, in turn, affects their communication strategies.

Having offered an overarching framework of how communication shapes and is shaped by central bank independence across an extended geographic scope, the next subsection further develops the individual substantial contributions of the chapters. These contributions have all been uniquely identified using central bank communication and each addresses a distinct threat or challenge - shaped by the level of independence - to which communication serves as a response.

1.2.1 Paper contributions

The first contribution involves re-evaluating financial instability as a persistent challenge for central bank independence to which central banks have to respond in their communication. Chapter 2 shows that financial pressures become more prevalent following increases in independence. This directly challenges the prevailing view that financial instability concerns for central banks are primarily a post-Global Financial Crisis (GFC) phenomenon. While the original delegation logic of CBI often overlooked financial instability, assuming price stability was sufficient for financial stability, I show that financial

instability has, in fact, been a recurring problem for the most independent central banks from 1997 to now.

This finding aligns with the historical role of central banks, central banks have always had a role in financial matters: from their 17th-century origins financing governments and monetising debt, to providing financial stability and managing the gold standard in the late 19th and early 20th centuries (Bordo, 2007). Even during the 1960s and 1970s, politicians heavily influenced central banks to pursue policy goals (Jones, 2020), but the case for CBI was later solidified after the oil crisis and stagflation, based on the logic of game theory and rational expectations (Fromage et al., 2019). The focus in these years was all about measuring independence from government despite warnings from historians and practitioners (e.g., Kindleberger, 1996). While CBI subsequently became a cornerstone, financial stability was for a time considered less of a problem in the economic literature. Events such as the recent collapse of Silicon Valley Bank and the subsequent pressure on Federal Reserve Chair Jerome Powell to adjust guidance (Lee and Wessel, 2024) further underscore the persistent nature of these challenges.

The second contribution is to show how even the most independent central banks engage in communication to address broader societal threats. Chapter 3 shows that pressures faced by the most independent central bank worldwide and their governors are not confined to financial matters and their core economic functions; they increasingly encompass perceived broader societal and re-emerged political pressures to which they are responsive in their communication. I show this by unpacking certain particularities of the Eurosystem, demonstrating that the often overlooked National Central Banks (NCBs) are crucial for enabling the ‘isolated’ European Central Bank (ECB) to bring diverse issues onto its agenda and effectively respond to various member state pressures using communication. The Eurosystem has traditionally been treated in the literature as an autonomous unit since the ECB’s establishment in 1998, a process that transferred

monetary policy away from (at the time) 11 NCBs to the EU level, however, this view is incomplete. The NCBs' role was largely envisioned as limited to voting in the Governing Council and performing implementation tasks related to monetary policy, banking supervision, and economic analysis based on the subsidiarity principle. I reveal that NCBs exert greater influence than is suggested by these narrowly defined tasks; they dynamically represent their member states and are not mere implementers of ECB policy. Specifically, NCBs act as vital intermediaries that transmit national priorities to the ECB, which in turn draws on this information. The ECB demonstrates responsiveness in agenda-setting by incorporating new issues in its speeches as signals to address pressures stemming from domestic politics, indicating that NCBs potentially function as a critical transmission mechanism between national publics and the ECB. More broadly, these multi-level dynamics also speak to the broader politics of agenda-setting in supranational and international organisations.

The third contribution, developed in Chapter 4, maps how fiscal and financial pressures on central banks can be addressed in the communication through either coordination or dominance, thereby introducing two novel concepts in the context of policy-linkages with monetary policy. These concepts allow for a more nuanced understanding of how even the most independent and isolated central banks are still intricately linked to government policy and financial markets. Classic analyses have long highlighted two primary configurations in monetary-fiscal policy relations: fiscal dominance, forcing central banks to accommodate unsustainable public spending (Sargent and Wallace, 1981), and monetary dominance, forcing fiscal authorities to rein in deficit spending to preserve price stability (Sargent and Wallace, 1981). Increasingly, scholars also recognise financial dominance, forcing the central bank to prioritise extensive liquidity provision to rescue banks, with possibly inflationary consequences (Brunnermeier, 2015). In these traditional models, dominance regimes inherently assume a hierarchical strategy, where

either fiscal or financial policy compels central banks to adjust their policy stance within a competitive game framework with governments.

However, Chapter 4 finds that this relationship is neither always hierarchical nor consistently resembles a competitive game setup between three atomistic actors. Instead, evidence suggests that coordination between monetary and fiscal authorities has frequently occurred, particularly in the period following the Global Financial Crisis. Moreover, coordinated strategies have also been observed with financial markets, such as the alignment of debt management practices or the introduction of macro-prudential measures (Maddaloni et al., 2022). Recognising the flaws in the game theoretic modelling that misses coalitions between the authorities, I introduce the terms ‘monetary-fiscal coordination’ and ‘monetary-financial coordination’ to accurately describe the contemporary challenges faced by central banks. Crucially, these concepts come to light when systematically examining central bank communication.”

1.2.2 Underlying rationales

In this thesis, I do not empirically test the underlying rationales why central banks respond to certain threats, nor do I test the intentions behind governors’ communicative responses depending on their independence levels. Rather, I build on previous literature to argue this is because central banks care about their institutional reputation and public trust and communication can be used to address these perceived threats. Trust is defined as a forward-looking belief held by an audience that the central bank will act competently (Christelis et al., 2020). Reputation is a collective, backward-looking evaluation of the bank’s past performance and identity - an asset based on the bank’s past performance and identity that confers autonomy and resilience and can be lost (Busuioc and Lodge, 2016). While reputation can nurture trust (a favourable record encourages positive expectations), they are analytically distinct: trust is relational and prospective;

reputation is social and retrospective. Both trust and reputation are essential for central banks for the following reasons.

First, trust in central banks is essential for anchoring inflation expectations for price stability (Christelis et al., 2020). In line with this, Ehrmann et al. (2021) in a recent survey of former ECB policymakers, find that enhancing credibility and trust is seen as the main objective of monetary policy communication. Moreover, low trust can have important repercussions for macroeconomic variables. For instance, Bursian and Faia (2018) show that lower trust amplifies macroeconomic fluctuations and steepens the sacrifice ratio. Low trust can also have political ramifications. For instance, low trust in central banks has been linked to the rise of populism (Algan et al., 2018), which in turn could have ramifications for central bank independence (as shown by Goodhart and Lastra, 2018). Moreover, popular support can help central banks resist political interference (Berger and Kißmer, 2013; Ehrmann and Fratzscher, 2010). Second, reputation acts as a filter through which external demands are interpreted (Busuioc and Lodge, 2016), i.e., changes in the external environment are translated into organisational behaviour via the filter of bureaucratic reputation. Once reputation has been established, it is an important source of power because it can give the central bank freedom to act.

Trust and reputation can be achieved through communication. The public, or ‘Main Street,’ has become increasingly critical of central banks, as evidenced by protests (Genovese et al., 2016). Contrary to earlier assumptions, citizens do not appear to be passively accepting technocratic rule. This is compounded by a decline in public trust in central banks (Roth et al., 2022), an increase in their public salience, and politicians’ tendency to shift blame for adverse economic circumstances onto these institutions. This erosion of trust is particularly problematic, given that the currency we hold - fiat money - is a social construct without intrinsic value, reinforcing the importance of trust in anchoring inflation expectations (Christelis et al., 2020). Central banks use communication

reactively to show accountability and thereby mitigate their democratic deficit (e.g., see Puppis et al., 2014) to make up for traditional standards of input legitimacy do not apply to independent agencies, which are governed by unelected officials(Majone, 2001; Elgie, 1998; Haan and Amtenbrink, 2000; Tucker, 2018). Accordingly, central banks actively seek to legitimise their actions by offering justifications for their actions reactively. Communication can also be used to cultivate existing trust and reputation. Baker et al. (2024) show that communication can be performatively used to reinforce independence and Hayo and Neuenkirch (2015) show that communication improves perceived central bank independence of financial market participants. Moreover, as Braun (2016) argues regarding financial stability, central banks communicate on a wider range of topics to demonstrate awareness of the risks these issues pose for financial stability by watching markets, media and public audiences and demonstrating preparedness, technical capacity and knowledge in response to them.

Following this logic, in the context of declining trust in central banks among citizens (Roth et al., 2022), increased salience of central banks, decreased output legitimacy (Jones, 2009) and politicians shifting the blame of bad economic circumstances to central banks, I argue that independent central banks have tried, through communication and responsiveness to certain issues, to fend off pressures and remain independent. This may explain why, after an initial increase in communication as a monetary policy tool, a second wave of communication followed realising that communication can be used to keep independence in place. This second wave is characterised with a greater focus on communication with the general public – central bank ‘speak’ reached Main Street. Or as Blinder et al. (2008) put it: “it may be time that we pay some attention to communication with the general public”.

In the following section, I move away from the substantive contributions and discuss how these contributions and insights have been possible due to new developments in the

applied social sciences field of Large Language Models.

1.3 Measuring communication

A core challenge of using communication to research central bank independence is to achieve a high level of interpretability while enabling systematic analysis simultaneously. The recent development of the field of Large Language Models has offered a promising avenue to address this challenge. Before delving into the measurement, the core database and the steps to clean the data are presented in Section 1.3.1. Section 1.3.2 provides a general overview of the various text-as-data methods used in this thesis to gain these systematic insights. The challenges and pitfalls associated with these methods are also discussed. Section 1.3.4 concludes this section by summarising the main measurement contributions of this thesis.

1.3.1 Data: speeches and meta-data

All three empirical chapters use the same central bank speeches dataset. Arguably, speeches can be seen as a most-likely case for central banks to use to fend off their independence. Namely, central bankers are free to determine the topic and format of speeches to target various audiences. Using speeches as data is also in line with the recent rise in using text-as-data in the central banking literature to study disagreement and preferences of governors under high secrecy and limited public information (e.g., Baerg and Lowe, 2020; Schonhardt-Bailey, 2013; Bennani and Neuenkirch, 2017; Moschella and Diodati, 2020; Ferrara, 2020). Communication via speeches is not just ‘cheap talk’ but is argued to set political directions and can lead to actual responses. For example, when central bankers talk, financial markets listen – Mario Draghi’s famous “whatever it takes” speech in 2012 at a financial conference in London is again evidence of this.

For the construction of the speeches dataset which is used throughout this thesis, I collected speeches from the Bank of International Settlements (BIS) Central Bankers' Speeches database. The BIS Central Bankers' Speeches database includes public speeches delivered by central bank officials across the globe. These speeches vary in format and are typically addressed to diverse audiences, including financial market participants, academics, policymakers, and the general public. While some speeches are technical and aimed at expert audiences (e.g., during conferences or monetary policy briefings), others are more general in tone, intended to explain policy decisions or signal institutional priorities to broader audiences. The BIS dataset starts in 1997 and is regularly updated. As of March 2024, the dataset contains 18,826 speeches. The dataset does not include all central bank speeches. However, the sample of speeches included in the database is stable over time and covers diverse geographical regions. After collecting the speeches, they are pre-processed following standard steps including deleting irrelevant textual artifacts (e.g., page numbers, URLs, page break characters, footnotes, etc.) using regular expressions. Moreover, heuristics are applied to remove non-genuine sentences such as tables and annotations. Figure 1.5 below shows the frequency and geographical distribution of the speeches database. Note that Chapter 3 uses only a selection of these speeches, namely, the speeches of the six largest Eurozone economies and the European Central Bank Executive Board (n=4659).

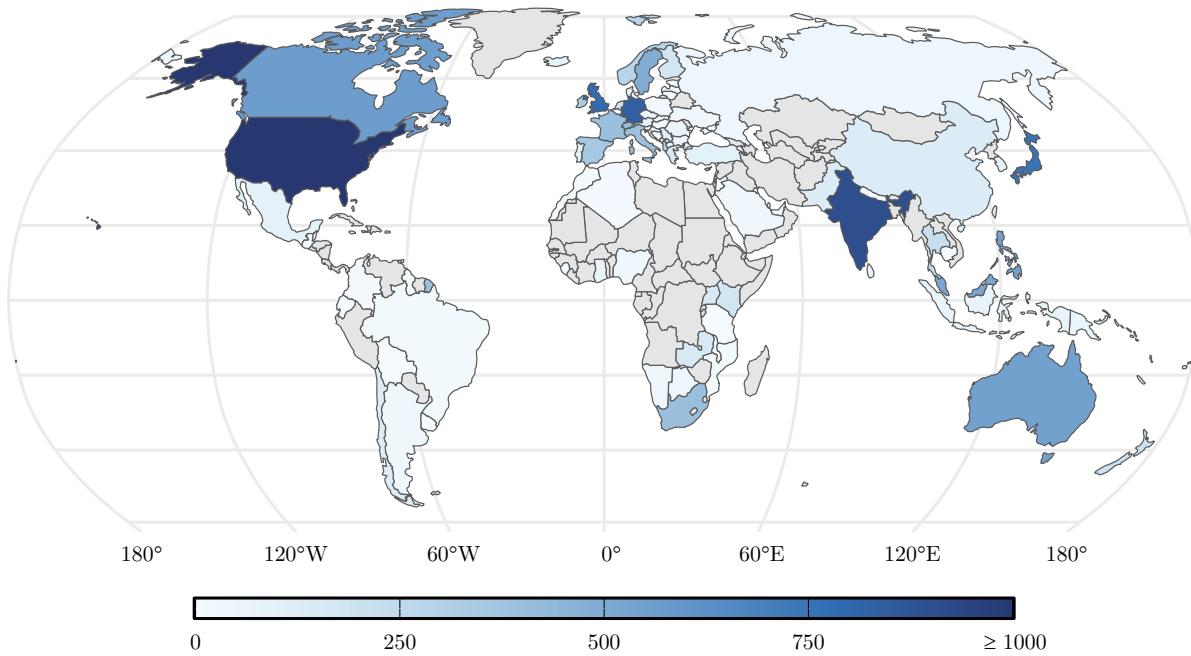


Figure 1.5: Speeches dataset frequency

Note: The colour scale indicates the number of speeches per country contained in our dataset. The figure contains all speeches used in this thesis of 119 central banks and their frequency from 1997 to 2024 retrieved from the Bank of International Settlements website. Countries that are missing are shown in grey. The central bank with the most speeches is the ECB with 2377 speeches (not shown on the map in order to be able to see which National Central Banks within the Eurosystem are included).

I also collect meta-data of the speeches to create a novel dataset. The BIS dataset contains the text of the speech, the date on which it was given, the author and a non-standardised description string which contains meta-data on the speech such as the location, the speaker and the occasion of the speech. Since the meta-data contained inside the description does not follow a standardised format and varies from speech to speech, I rely on a Large Language Model (Gemini 1.0 Pro) with a few-shot prompting approach and provide appropriate instructions and examples to correctly interpret and extract the meta-data from the description. A sample of this was manually validated. To give a preview, Figure 1.6 shows the positions of the speakers of the speeches database

and Figure 1.7 shows the audiences addressed over time.

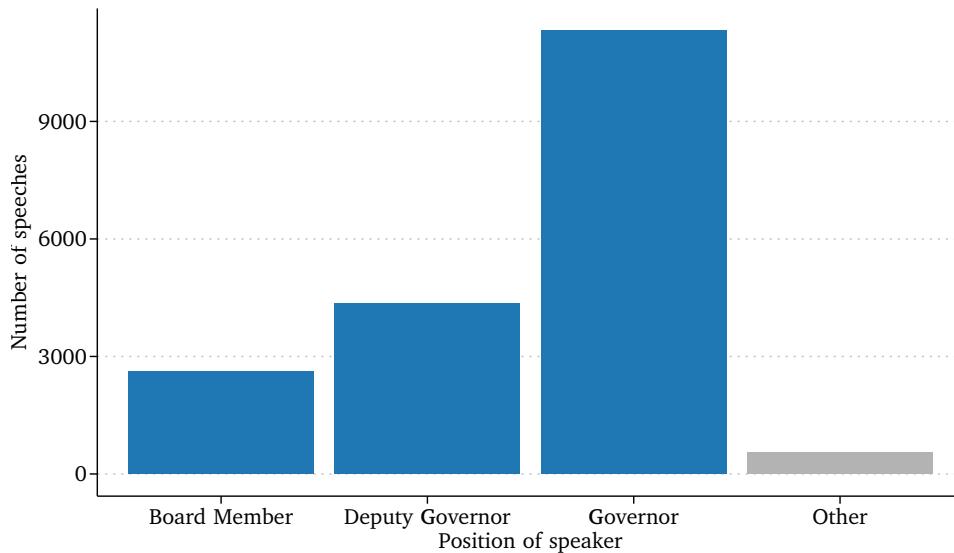


Figure 1.6: Speaker position of all speeches

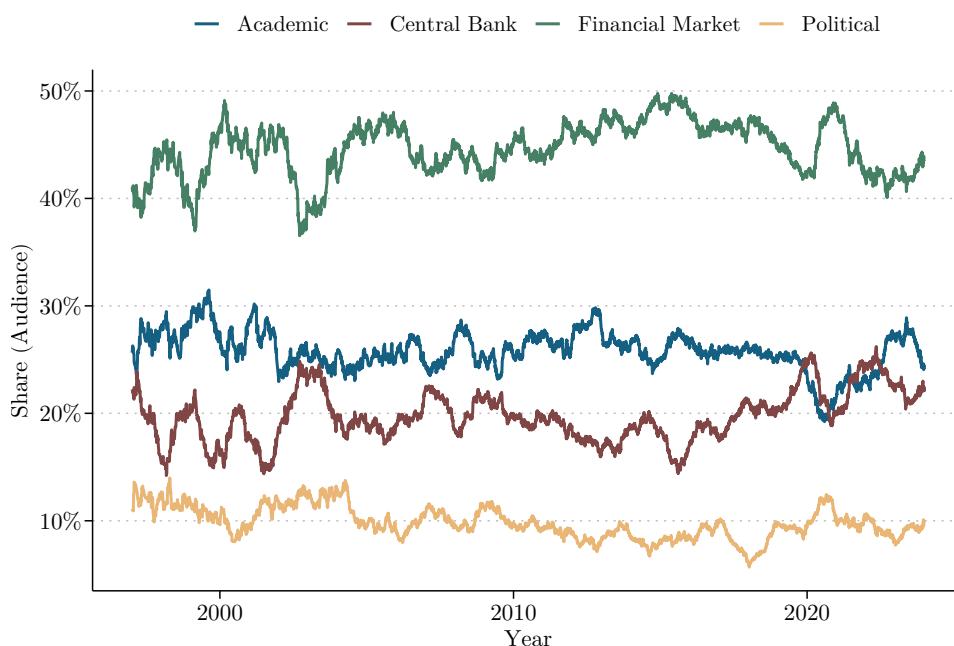


Figure 1.7: Audiences over time

1.3.2 Quantifying central bank speeches

Having collected an extensive database of central bank speeches logically leads to the next question: how best to derive insights from textual data? More fundamentally, it raises the question of how we can turn inherently “unstructured data”, i.e., a literal recoding of verbal activity structured according to the rules of language and not to serve as any form of data, into usable social science insights (Benoit, 2020). Grimmer et al. (2022) refer to text-as-data methods using a hay stack metaphor. They argue that humans are great in analysing a straw of hay, i.e., understanding the meaning of a sentence, but humans struggle with organising the haystack, i.e., describing, classifying, and scaling texts. Text-as-data tools allow for systematic latent insights into processes in central banking that have been thus far mainly analysed qualitatively, however, it does not take away work for researchers. Rather, the work comes at a later stage when turning the inherently ‘messy’ textual data into social science insights. In what follows, this should be kept in mind and will be revisited in Section 1.3.2.2.

1.3.2.1 General overview

Text-as-data methods can best be seen along the continuum of automation versus human judgment. This thesis employs and experiments with various text-as-data methods ranging from well-established methods to less established methods which only recently entered the rapidly developing field. Within this field, I use both text classification (Chapters 2 and 4) and unsupervised topic modelling (Chapters 2 and 3).

In the chapters that employ topic modelling, I use unsupervised learning due to the lack of pre-set topics, the large sample size, and the possibility to discover systemically latent trends over time. Unsupervised models are different from supervised methods since there is no separate learning step associated with inputs in the form of known classes. Specifically, there are no strong priors regarding the topics, thus, the topic modelling is

more exploratory and benefits from inductively extracting a small number of independent dimensions rather than being confirmatory which would measure pre-specified concepts. In Chapters 2 and 4, I use text classification to develop indices which identify the concepts of distinct policy linkages in monetary policy communication. Identifying these patterns from text is a complex reasoning task that requires substantial domain knowledge and is typically done through manual classification on a small scale. However, given the scale of this project – more than 18,000 speeches in which there are more than 2 million sentences which need to be coded – this is an unrealistic avenue.² Large Language Models, as I show, offer a viable solution.

I elaborate below on some challenges and pitfalls while the individual chapters provide the technical details and choices made. Figure 1.8 shows an overview of the different text-as-data methods and Large Language Models used in the main analysis of the three empirical chapters.

²To illustrate, during our manual classification of our validation set, it took us around one minute per sentence. Thus, given the sample of two million sentences, it would take around 33,000 hours or almost 4 years cumulative to classify the entire dataset manually. An automated procedure is, therefore, necessary.

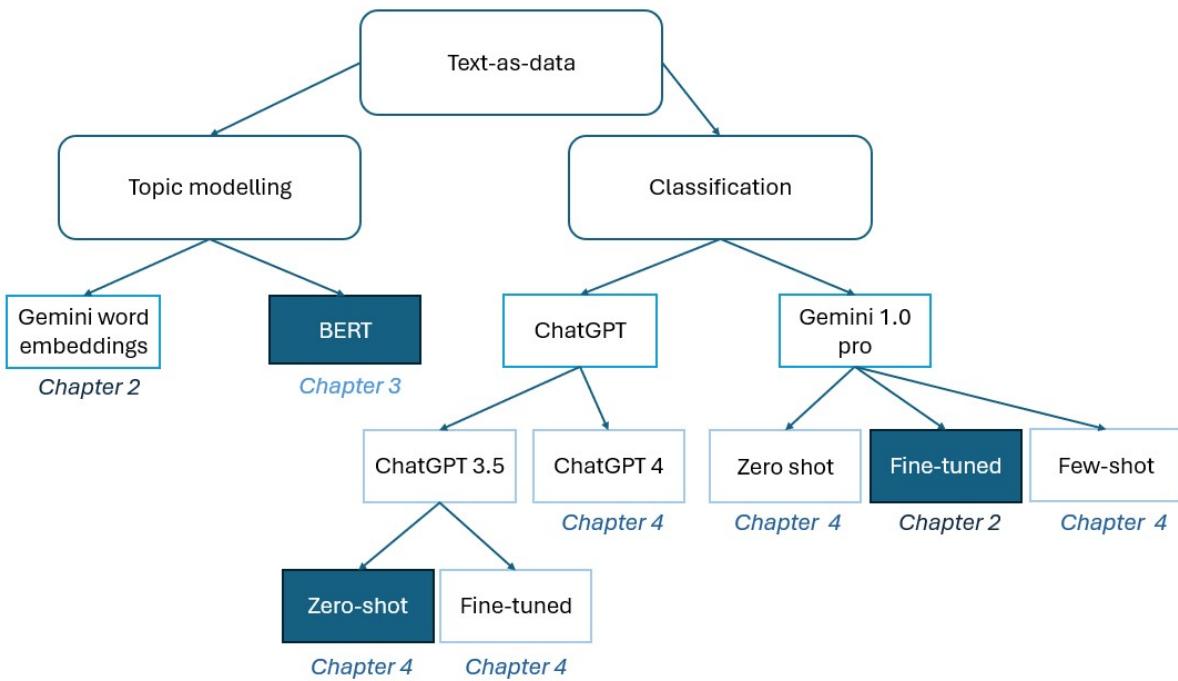


Figure 1.8: Text-as-data methods overview

Note: The dark blue boxes are the main Large Language Models used in the analysis. The other boxes with the chapter number underneath are experiments run and included in the chapters.

1.3.2.2 LLMs: challenges and pitfalls

This thesis has been marked by rapid developments in the field of text-as-data, specifically, in topic modelling and classification tasks. In all empirical chapters, due to the complexity of the tasks, I move beyond the classical “bag-of-words” based on the vector space model, which disregards grammar and word order (and was predominantly the standard when starting this thesis). (Artificial) neural-network models that use word, sentence, or document embeddings generally outperform classic bag-of-words methods on many natural language processing tasks because they capture context, meaning, and semantic relationships. Traditional bag-of-words models encode each document as a very high-dimensional sparse vector; for large corpora such as central-bank speeches, the resulting document-term matrices can contain millions of mostly-zero cells and quickly become computationally unwieldy. Embedding-based models, by contrast, map each document

into a dense vector of only a few hundred dimensions, which scales far more easily. Recent large language models (LLMs) extend this advantage: they can classify complex, abstract text at scale, allowing researchers to detect systematic trends in central-bank communication over time. The core of LLMs are transformers introduced by Vaswani et al. (2017) which allow efficient processing. Transformer models use “self-attention” to better learn the long-term dependencies in sequences of text, which then helps to structure the output in more meaningful and realistic ways.³ Recent developments made LLMs accessible for public use.⁴

The introduction of LLMs for applied social science research have made the complex classification tasks in this thesis possible. However, due to the rapid development, there is still a lack of established and best practices. As a result, the research process involved considerable time spent on experimentation, troubleshooting, and refinement. Although a brief note does not fully suffice, a couple of the challenges associated with the rapid developments in the field (both in terms of keeping up as well as there not being established practices) are discussed below.

The rapid developments have led me to change my analysis at certain points. First, while Chapter 3 of this thesis now uses a BERTopic model⁵, the first write-up included

³The original application was machine translation which uses encoder and decoder blocks. Encoders take an input and ‘encode’ it into a numerical representation, while decoding translates this input into an output. At a high level, the transformer models are a function that predicts the next token iteratively given a sequence of previous tokens. Attention weights are learned through training on very large amounts of text and with variants of gradient descent minimising a cross-entropy loss (Vaswani et al., 2017). Attention blocks allow updating each embedding as a linear combination of the embeddings in the context.

⁴For instance, GPT has strong capabilities as a “zero-shot” learner, meaning it can perform tasks without requiring additional training or fine-tuning (Laskar et al., 2023). Due to reinforcement learning from human feedback, it can produce answers based on a simple prompt and relies on the knowledge and instructions following capabilities built into the model. Thus, not requiring a substantial manually labelled dataset as a training sample.

⁵BERTopic (Grootendorst, 2022) takes advantage of the powerful pre-trained transformer-based language models and employs a clustering technique for topic extraction. BERT, in comparison to GPT and Gemini, is based on an encoder architecture: attention is backward and forward-looking whereby it weights relative to tokens before and after a given token. The BERTopic model is specifically designed for extracting topics from a corpus of text and offers more interpretable metrics. In Chapter 2, however,

Structural Topic Modelling (STM) of Roberts et al. (2014). At the time of writing this chapter, this was an improvement over the status quo in the political economy of central banking literature⁶. Despite the promised improvement of BERTopic by using contextual embeddings derived from pre-trained language models to capture richer semantic and syntactic relationships in text (bag-of-words or TF-IDF approaches are typically used in STM or LDA), it mainly led to more unstable results. After various rounds of experimenting, it became clear that the results heavily depended on the choices made throughout the pipeline. These include choices about the unit of analysis (e.g., documents vs. sentences vs. paragraphs), text pre-processing steps (e.g., tokenisation rules, handling of stopwords, or lemmatisation), the selected embedding, and the various hyperparameters required by these models (e.g., number of topics, embedding dimensionality, clustering thresholds). There is currently no universal consensus or established best practice for each of these stages, and even small changes in data cleaning or parameter selection can lead to markedly different results. Consequently, I aimed to report my methodological choices transparently and, wherever possible, undertake systematic comparisons to assess the robustness of the findings.

Second, in Chapter 4, version GPT-3.5-turbo-0301 is used⁷. At the time of writing this chapter, GPT4 was not yet available, substantially more costly and subject to restrictive rate limits. Similarly, Google’s Gemini was not yet released when the main analysis was run (it was when Chapter 2 was written). Due to rapid developments, we were only

we employ Gemini embeddings, given that we employ the architecture for the other tasks in this chapter.

⁶Most related papers used Latent Dirichlet Allocation (LDA) (Blei et al., 2003) which relies on a Dirichlet distribution and, therefore, did not allow covariates unlike STM which relies on a logistic distribution.

⁷For classification tasks, GPT and Gemini are preferable, since compared to the frequently used BERT-based text classifiers (e.g., see Bertsch et al., 2024; Bestvater and Monroe, 2023), it has two main advantages: (i) it has many more parameters (billions vs millions) resulting in a broader knowledge base, enabling better understanding of context and (ii) can directly respond to a prompt, removing the need for task-specific training data to train a classification head which is necessary to generate a classification output. However, for topic modelling BERTopic is more suitable since it is specialized in handling unsupervised topic modelling. Specifically, it has a specialised pipeline for clustering methods (e.g., using UMAP, HDBSCAN, etc.) optimised for identifying latent themes.

able to later run our validation set with GPT4 and Gemini Pro (without fine-tuning). For the main analysis of Chapter 2, I use Gemini Pro 1.0. We use Google’s Gemini in particular, because at the time of writing, it was the most capable base LLM that offered a fine-tuning interface.⁸ Again, due to the lack of established practices in fine-tuning at the time, we conducted many experiments to establish the best model settings. To find the best fine-tune based on a human coded validation sample, we construct a joint grid of hyper-parameters of the training process, different compositions of the training set and different prompt configurations. As part of the fine-tuning exercise, we also conduct light prompt engineering. However, contrary to our findings in Chapter 4, the prompt engineering was no longer crucial for the model performance. Namely, the fine-tune proved way more significant in improving the performance metrics. Further, to address the significant imbalance in our classification, we also experimented with different training set compositions (via up-sampling, synthetic sentences and randomising) which at the time did not yet improve results. Nevertheless, the main take-away is that these findings are strongly model and task dependent and thus cannot be taken as general guidelines.

There were also many failed attempts that did not make it into the thesis. For instance, I also attempted to incorporate temporal embeddings (Word2Vec, Dynamic BERT), i.e., where the model accounts for shifting word usage or context over time. However, in the preliminary experiments, these approaches yielded mixed or inconclusive results. The difficulties I encountered stemmed partly from the increased complexity of modelling temporal dynamics alongside text features, which introduced additional instability. Further, in Chapter 2, we also tried open source alternatives available at the time. Specifically, we tried to fine-tune Mistral with QLoRA⁹. Once the model was set-up,

⁸Fine-tuning involves limited additional training on top of the base LLM, which has been shown to greatly improve model classification accuracy with very little annotated data across a number of LLMs already (Alizadeh et al., 2024).

⁹QLoRA (Quantized Low-Rank Adaptation) enables resource-efficient fine-tuning of large models by combining parameter quantization (down to 4-bit precision) with low-rank adapters (Dettmers et al., 2023).

it still did not follow the output format, and the results exhibited severe class imbalance problems (e.g., the output was often ‘none’). After several fine-tuning attempts, we moved onto propensity models instead.

Another significant consideration arises from the difficulty of externally validating textual measures, particularly when applying them to central bank policies in Chapter 2 and 4. In this thesis, I mainly rely on “ground truth” derived from human-coded data to evaluate model performance, although human coders themselves often disagree – especially on complex or subjective classification tasks. In Chapter 4, we therefore extend our evaluation beyond the usual benchmarks, not only comparing model outputs to human-coding but also comparing model reliability to patterns of human disagreement. Curiously, we find that these two are related. Again, it has to be said that many questions are still open regarding the best practices, also in terms of how to deal with LLM measurement errors since LLMs generate outputs based on probability distributions over possible tokens themselves. Thus, this probabilistic nature means that different outputs across runs occur (which can be considered analogous to measurement error in experimental data). When LLM outputs are then used as inputs for further analysis (e.g., as we do for the difference-in-differences model in Chapter 2), measurement error can have downstream effects.¹⁰ Our solution included repeated sampling but other solutions could be to use ensemble methods or Bayesian methods where uncertainty is explicitly modelled.

1.3.3 General empirical strategy

Given the paper-based structure of this thesis, the empirical strategy differs for the specific chapters. The text-as-data methods mainly allow me to create various descriptive statistics.

¹⁰Battaglia et al. (2024) suggest using (i) an explicit bias correction with bias-corrected confidence intervals, and (ii) joint maximum likelihood estimation of the regression model and the variables of interest.

In Chapters 2 and 3, I combine this with additional estimation strategies. Chapter 2 leverages a natural experiment: the staggered nature of central bank independence events across the world. The difference-in-differences causal inference strategy allows us to test the causal effect of central bank independence on monetary policy communication using changes in central bank independence as ‘shocks’. Moreover, it uses the ‘exogenous’ spatial diffusion process of central bank independence as an instrumental variable to rule out endogeneity. Chapter 3 leverages the multi-level set-up for the Eurosystem which consists of both national central banks and the European Central Bank. It uses sequence analysis (first-mover graphs and Markov chains) and cross-sectional time series (TSCS) models to examine central bank issue-responsiveness, i.e., who initiates topics, how topics relate, which central banks respond to each other and under which circumstances responsiveness to outside pressures occurs. Table 1.2 provides an overview of the data, empirical methods, and text-as-data methods used in this thesis by chapter.

Chapters	Data	Empirical Method	Text-as-data Method
Chapter 2	All speeches (1997-2023)	Staggered Difference-in-differences and Instrumental Variable approach	Large Language Model (Google Gemini)
Chapter 3	NCB speeches and ECB speeches (1997-2023)	Sequence analysis (Markov chains) and cross-sectional time series	Large Language Model (BERTopic)
Chapter 4	All speeches (1997 to mid-2023)	Descriptive statistics	Large Language Model (GPT 3.5)

Table 1.2: Overview of the data, empirical methods and text-as-data methods

1.3.4 Measurement and data contributions

Besides the substantial contributions highlighted in the previous section and the chapter-specific contributions, this thesis also makes two broader measurement and data contributions.

First, this thesis uses text-as-data to examine how central banks have remained independent institutions despite various pressures. This thesis, thereby, taps into broader debates in political science, namely, how institutions interact with behavioural and policy outcomes (e.g., see Besley and Case, 2003; Cox and McCubbins, 2000; Persson and Tabellini, 2001). More traditionally, how institutions interact with behavioural or policy outcomes is studied in relation to voting behaviour or attitudes, for instance, applied to the Bank of England by Hix et al. (2007) and the Federal Reserve by Baerg and Lowe (2020). Some recent work, however, has already used communication as a proxy for central bank behaviour (e.g., Moschella and Pinto, 2019; Ferrara, 2020; Moschella and Diodati, 2020; Baerg and Lowe, 2020). Tapping into this, but going one step further, I change the unit of analysis in this thesis from individuals (e.g., in this context, governors) to central banks themselves and use sentences (and their context) to detect policy linkages and issue-responsiveness. More broadly, in this thesis, I tap into a growing application of textual methods within the field of comparative political economy and public policy research, to allow scalability and support qualitative research to systematically detect latent policy insights.

A key measurement challenge is identifying how central banks respond to external pressures using textual speech data. Commonly, text-as-data methods are used to create descriptive data without using it as a dependent variable. In Chapters 2 and 3, I offer alternatives to this. Chapter 3 employs a discrete first-order Markov chain to capture how topics and central banks transition over the sequence of speeches. A Markov chain is a stochastic process characterised by the Markov property: the probability of moving to the next state depends solely on the current state rather than the sequence of preceding states (Harary et al., 1965). Markov chains have been used sparingly in political science – thus far, mainly to model regime shifts (Klimm, 2022; Gleditsch and Ward, 2006, 1997). Nevertheless, by allocating each speech to the topic with the highest probability and using

a discrete sequence of speeches, this technique allows for systematic insights into speech topics and inter-bank relationships. Moreover, I also create quarterly and half-yearly indices using the proportions in a speech assigned to a certain topic. Using these indices in cross-sectional timeseries analysis, two and three-way interactions with the lags of NCB topic proportions on ECB topic proportions are used to examine responsiveness between topics and banks to various circumstances. Doing this also allows for enough statistical power. Similarly, Chapter 2 also develops indices for the degree of dominance in monetary policy detected. I subsequently use these indices as dependent variable in a difference-in-differences and instrumental variable set-up.

Second, all the code and data used in this thesis is made publicly available. I have modularised and heavily commented all code before making it public on GitHub. By doing this, I hope to remove barriers for researchers and practitioners alike to work with speech data and apply it to unique use cases. I have also made all data used in the thesis, including various visualisation options, accessible on the website accompanying this PhD: www.centralbanktalk.eu. Features of the website include the ability per central bank to explore key facts and topics and visualisations of the speeches frequency, the audiences and policy pressures over time. I also offer a search function for the entire speeches database, allow for exploration of our sentence and speech level LLM classifications of Chapter 2 and I offer links to the original speeches. I hope that this allows researchers and broader audiences to obtain insights into and explore the central bank speeches without having to go through the process of setting up the data themselves.

1.4 Thesis outlook

Although this thesis consists of three empirical chapters that differ in scope and specific research questions, they collectively demonstrate how central banks use communication

to address pressures on their independence - and how such communication is shaped by it. The thesis concludes with Chapter 5, which briefly synthesises the findings, discusses their limitations, and outlines both broader and more specific avenues for future research, substantively and in terms of measurement.

Chapter 2

How Central Bank Independence Shapes Monetary Policy Communication: A Large Language Model Application

Abstract

Although central bank communication is a core monetary policy and accountability tool for central banks, little is known about what shapes it. This chapter develops and tests a theory regarding a previously unconsidered variable: central bank independence (CBI). We argue that increases in CBI alter the pressures a central bank faces, compelling them to address these pressures to maintain their reputation. We fine-tune and validate a Large Language Model (Google's Gemini) to develop novel textual indices of policy pressures regarding monetary policy communication of central banks in speeches of 100 central banks from 1997 to 2023. Employing a staggered difference-in-differences and an instrumental variable approach, we find robust evidence that an increase in independence decreases the narrow focus on price stability and increases financial pressures discussed in monetary policy communication, especially in democratic and advanced economy countries. These results are not, as generally is assumed, confounded by general changes in communication over time or singular events, in particular, the Global Financial Crisis.

2.1 Introduction

Before the 1990s, central banks were known to be secretive. Conventional wisdom held that monetary policymakers should say as little as possible and, if they speak, they should do so only cryptically (Blinder et al., 2008). Over time this changed, especially with the adoption of inflation targeting (Casiraghi and Perez, 2022). It was the Reserve Bank of New Zealand in the mid-1990s which pioneered open communication and is still among the most advanced central banks worldwide in this regard (Blinder et al., 2001). Central bank transparency became common practice (Dincer and Eichengreen, 2018; Horvath and Vaško, 2013). Other central banks with various levels of independence – both in advanced and developing economies, democratic and autocratic – followed and started engaging with a broad range of topics, massively increasing the usage of communication for monetary policy over time (Woodford, 2005; Blinder et al., 2008; Hansson, 2021).

In the academic and policy-realm, this also resulted in greater attention to understanding what, to whom, and why central banks communicate regarding monetary policy (Casiraghi and Perez, 2022). Many papers thus far have focused on the consequences of monetary policy communication and specific statements made by governors (e.g., see De Haan and Sturm (2019) for a recent overview) or the effects of 'who' communicates, for instance, how certain governors or board composition impact the clarity and way of communicating (e.g., see Meade and Sheets, 2005; Baerg and Lowe, 2020; Bodea and Kerner, 2022). Recent work has argued that the Global Financial Crisis (GFC) is a major trigger of a change in communication (e.g., see Vayid, 2013), mainly due to an increase in pressing financial stability concerns for central banks (Braun, 2016). Unlike these papers, our chapter focuses on 'what' and 'when' central banks communicate in light of major institutional changes rather than singular events.

The prevalent literature has reached a consensus that monetary policy communication

is mainly used for two reasons. First, it can be used as a monetary policy tool to reach better policy outcomes since it directly impacts financial markets and can be used to control and guide inflationary expectations (e.g., Woodford, 2005; Blinder et al., 2008; Anderes et al., 2021; Haan and Sturm, 2019). This is especially the case when all other tools are exhausted, such as during the GFC. Second, increased transparency through communication is useful for both reactive democratic accountability purposes to manage their reputation (Moschella and Pinto, 2019; Haan et al., 2004; Blinder et al., 2024), as well as pro-actively to keep independence in place (Baker et al., 2024). Nevertheless, although these rationales can be expected to vary for central banks across different institutional set-ups, there is limited research, to our knowledge, examining how and through which channels central bank independence (CBI) influences monetary policy communication.¹ Does the way in which central banks communicate about monetary policy change with increases in independence? Examining the consequences of institutional independence on communication is of importance since knowing the consequences of CBI can inform policy design and reforms aimed at enhancing the effectiveness of communication in the long-term as a core tool of central banks. Especially with the recent politicisation of central banks (Tortola, 2020) and output legitimacy in terms of achieving monetary policy goals often being more difficult to achieve, communication has gained more relevance and has been targeted at a wider range of audiences (Blinder et al., 2024). Moreover, examining the systematic consequences of institutional design rather than focusing on individual cases or the consequences of singular events allows insights into the mechanisms and potentially how underlying institutions can be improved.

We argue that the degree of independence influences the degree to which central banks in their monetary policy communication talk narrowly about price stability or

¹The exception being a limited amount of papers on the relationship between transparency and CBI (Dincer and Eichengreen, 2018; Horvath and Katuscakova, 2016). Nevertheless, these papers are not primarily focused in the relationship with CBI and do not further explore the mechanisms.

address other policy pressures. Although central banks have different levels and practices of communication, with increases in independence, we can expect to see general changes in communication through two channels which enable each other. First, there are changes in inflationary pressures. Namely, the core task of central banks of bringing inflation in line with a target rate is made easier with a higher degree of independence (Alesina and Summers, 1993; Grilli et al., 1991; Garriga and Rodriguez, 2020). Instead, due to increased independence, central banks are more likely to be pressured by financial markets since financial regulation is often made more lenient and supervision is placed at more distance from central banks as a consequence of independence (Masciandaro and Romelli, 2018; Aklin and Kern, 2021). These changes in pressures are, however, not sufficient for a change in communication. The degree of independence also influences how central banks communicate. Namely, the decisions of a central bank cannot be justified by, or blamed on, the government, but central banks need to justify themselves to a range of audiences. Thus, higher independence forces central banks to justify a range of policy pressures. Overall, these two channels jointly result in a less narrow focus in monetary policy communication on price stability and an increase in communication regarding financial pressures.

To test this, we use 17,829 speeches from 1997 to 2023 of 100 central banks. We manually fine-tune² and validate a Large Language Model (LLM), namely, Google's Gemini, to identify sentences concerning monetary policy communication and to examine the portrayal of monetary policy in this communication over time. Using a LLM offers scalability to detect systematic trends and enables classification of complex and abstract concepts. These classifications are used to create novel indices of policy-pressure in monetary policy communication. Our empirical strategy to examine the consequences

²We use adapter-based tuning (PEFT) using the LoRA adapter. This fine-tune does not update all 60B base parameters of Gemini Pro 1.0. Instead PEFT adds a lightweight LoRA adapter layer (0.1% of the model size) and updates only those adapter weights on a small, labelled corpus keeping the rest frozen. In what follows, with fine-tuning we refer to this specific adapter-based tuning.

of changes in independence to these indices is twofold. First, we rely on a staggered difference-in-differences setup to leverage variations in the timing of the independence events and account for potential confounders such as common financial conditions. We also augment our event study model to examine heterogeneous effects among subgroups. Second, to address additional endogeneity concerns related to independence events, we use an instrumental variable approach. Jointly, these analyses show that, over time and pooled across all central banks, the degree of inflationary pressures addressed decreases by 25 percentage points while communication regarding financial constraints on monetary policy increases with 7 percentage points in our standardised indices ranging from 0 to 1. We find that on average this takes around five years to materialise, reflecting the time it takes for financial deregulation to result in increased pressures and credibility to be won for inflation pressures to decrease. We also examine the mechanisms, conduct a range of robustness tests, and rule out potential alternative explanations.

This chapter makes three contributions. First, this chapter develops a novel theory of how central bank independence shapes monetary policy communication across various types of central banks and over time. To do this, we combine the economics literature on the impact of communication on macroeconomic variables (e.g., Woodford, 2005) and the timing of communication (Ehrmann and Fratzscher, 2007) with the bureaucratic and political science literature on reputation and accountability of (independent) institutions (e.g., Moschella and Pinto, 2019). We focus on the core tool of central banks: monetary policy. Specifically, to operationalise this, we employ the framework of ‘monetary dominance’ and ‘financial dominance’. Although these concepts strictly refer to the degree to which there is policy subordination in central banks regarding their monetary policy conduct, i.e., ‘dominance’ (Sargent and Wallace, 1981; Davig and Leeper, 2011; Bianchi and Ilut, 2017; Brunnermeier, 2015), we broadly apply these concepts to communication. This framework reflects the key drivers and concerns of central bank’s actions regarding

monetary policy (e.g., see Schnabel, 2020). Applying these concepts to communication, we argue that when a central bank is ‘monetary dominant’ their main concern is monetary (inflationary) pressure, similarly, when a central bank is ‘financially dominated’ the main concern they reflect are pressures of financial markets. Moreover, using these established concepts allows us to better externally validate our indices with previous studies using these concepts.

Second, we undertake a novel test of the theory by creating indices based on textual data using LLMs. So far, only a few papers have used LLMs in the domain of central bank communication (e.g., see Fanta and Horvath, 2024; Hansen and Kazinnik, 2023). We, specifically, further improve classification accuracy of previous work from Leek et al. (2024) by offering an approach which allows for fine-tuning and hyper-parameter optimisation of our LLM model. Using these novel indices, we contribute to empirical work that examines the consequences of independence (e.g., see Fernández-Albertos (2015) for an overview). Thus far, the work on the consequences of independence has not considered consequences for communication while communication recently has developed into a core and essential tool for central banks (e.g., see Blinder et al., 2024).

Third, we cover a diverse set of central banks (central banks of advanced, developing, democratic and autocratic countries), allowing us to uncover heterogeneous effects of how institutions shape communication among various subgroups of central banks. Thus far, the central bank communication literature has predominantly focused on advanced economies (with the main exception being Evdokimova et al., 2023). Specifically, by extending the scope of central banks, we find that these effects mainly hold in democratic and advanced economy central banks. Moreover, we cover an extended time period which includes the time when central bank communication became established and encompasses a large number of independence events in order to detect systematic patterns over time. Given this extensive time span, we are also in the position to show that we see strong financial

considerations pre-GFC. This goes against the common consensus in the literature (e.g., see Diessner and Lisi (2020); Moschella et al. (2020) who argue that issues surrounding financial instability concerns are a post-crisis phenomenon).

2.2 A theory of central bank independence and communication

Why do central banks communicate? Around the mid-1990s central banks moved from high secrecy to transparent communication (Blinder et al., 2024; Dincer and Eichengreen, 2018). A consensus formed that effective central bank communication is a critical component of the toolkit of central banks because of two main reasons. First, communication can be used as a monetary policy tool to shape market expectations and enhance predictability of monetary policy decisions and, thereby, help achieve central banks' macroeconomic objectives (Woodford, 2005; Blinder et al., 2008; Anderes et al., 2021; Haan and Sturm, 2019). The shift towards transparency in central bank communication occurred as Casiraghi and Perez (2022) argue, especially with the adoption of inflation targeting. When an inflation target is set, central banks should clearly communicate to help the formation of expectations. This has been especially helpful during the Global Financial Crisis (GFC), when policy rates reached their zero lower bound. Communication was used to anchor and guide market expectations and thereby became one of the main tools of monetary policy (Casiraghi and Perez, 2022). This guidance took place mainly through forward guidance on the future path of interest rates (Strasser et al., 2019; Plosser, 2014). Moreover, financial stability, which has been a prime concern since the GFC, has been a target of communication to enhance predictability of central bank actions and bolster central bank credibility in macro-prudential surveillance (Vayid, 2013).

Second, communication helps build public support for potentially unpopular policies

and can be used to increase and maintain reputation and enhance democratic accountability (e.g., see Blinder et al., 2008; Moschella and Pinto, 2019; Haan et al., 2004; Baker et al., 2024) which is necessary to ensure independence of the institution. These two rationales are also related since for the effective implementation of monetary policy strong reputation is needed.

Does the independence level of a central bank influence monetary policy communication? With the independence level, we narrowly refer to CBI scores based on the codification of central banks statutes³ rather than *de facto* measures of independence or informal power relationships with governments or financial markets. Building on the above-mentioned rationales of why central banks communicate, we argue that two channels jointly influence how CBI shapes monetary policy communication. First, policy pressures both in type and strength vary - we call this the ‘policy pressure channel’. We use the term ‘policy pressure’ very broadly to reflect cases in which a central bank’s monetary policy conduct is driven by monetary, fiscal or financial considerations and pressures. Second, with increased independence, different explanations and justifications are often required regarding the central bank’s monetary policy conduct for a central bank to maintain the reputation of being an independent institution. We call this the ‘reputation channel’. The channels individually are, however, not sufficient for changes in communication. Rather, these two channels enable and reinforce each other. Increased policy pressures combined with central banks caring about their independent reputation, result in central banks having to respond to these pressures. Figure 2.1 summarises the channels on which we will elaborate below.

³The first to codify and still the foundation of most CBI indices, is Cukierman et al. (1992).

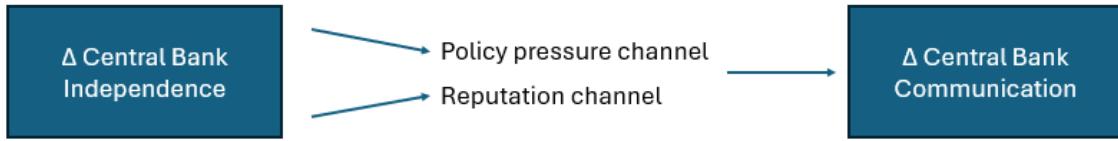


Figure 2.1: Main relationship between independence and communication

First, the policy pressure channel implies that the policy pressures that central banks have to address are different for central banks with different degrees of independence. The functional/technical aim of central bank independence is to shield monetary policymakers from political manipulations by “tying one’s hands” through delegation. This approach serves to resolve the time-inconsistency problem identified by Barro and Gordon (1983a), i.e., to facilitate governments’ credible commitment to not propagate inflation for short-term electoral gains at the expense of long-term economic gains (Giavazzi and Pagano, 1988). There is abundant evidence that the independence of central bank decisions is indeed helpful for monetary policy management (see the early work of Alesina and Summers (1993); Grilli et al. (1991) and later work of Garriga and Rodriguez, 2020). Overall, one can argue that with increased independence, the monetary policy targets are more easily reached.

Moreover, CBI theories focus on the relationship with fiscal authorities and envisioned a complete separation between governments and central banks whereby each actor performs their own task in pursuit of a strong economy. Nevertheless, formal independence from fiscal authorities says nothing directly about the degree of independence from financial markets and whether financial market pressure and financial stability are considered in monetary policy communication. This has only become clear over time. For instance, Hansen (2022) and Berger and Kißmer (2013) show that independent central banks are less responsive to banking instability due to their myopic focus on inflation resulting in higher financial instability. Moreover, Masciandaro and Romelli (2018) show

that the level of CBI is highly relevant in influencing the decision to concentrate financial sector supervision in the hands of monetary policy authorities whereby higher CBI is associated with a lower central bank involvement in supervision. Paired with this, Aklin and Kern (2021) show both qualitatively and quantitatively that governments weaken financial regulations in the aftermath of central bank independence. These weaker financial regulations allow financial institutions to increase their leverage and credit creation which is a substitute for public spending and result in privatisation of banks, entry of foreign banks and capital openness (Rajan, 2010). Thus, one could argue that an alternative time-inconsistency problem arises between short-term gains with loose financial policy when loose monetary policy is not available due to independence. This looser financial policy regulation can result in increased pressures on monetary policy which need to be accounted for by central banks to keep the financial system intact. In sum, there is strong evidence for an increase in financial pressure on monetary policy when independence increases.

We expect that the effect described above is not immediate when the independence of a central bank changes. The financial deregulation and monetary stability take time to materialise. Laeven and Valencia (2018) show this by presenting a global dataset of systemic banking crises and investigating the role of financial deregulation. They find that deregulation often leads to a delayed but significant increase in financial pressures, especially when coupled with external economic shocks. The full impact on financial pressures often unfolds over a longer period (e.g., 3-10 years), as the cumulative effect of risk-taking behaviours and structural shifts become visible. There is also a certain degree of 'strategic delay' in addressing certain pressures since as Brunnermeier (2015) argues governments and central banks will first be waiting for the other to take action to address financial stability concerns. Moreover, there is research which indicates that inflation pressures tend to ease within 3 to 7 years after CBI reforms, with faster reductions

observed in countries with strong institutional frameworks (Cukierman et al., 1992; Alesina and Summers, 1993). The timeline often depends on how quickly the central bank can gain public and market credibility through consistent and transparent policy actions.

The second identified channel is the ‘reputation channel’. This channel highlights that communication can be used as a tool for central banks to reinforce their reputation as an independent institution towards several audiences. In particular, with increased independence there is a responsibility or ‘moral corollary’ of increased accountability (Bernanke, 2010; Issing, 2014; Majone, 2001) to make up for traditional standards of input legitimacy not applying to independent agencies and being in the hands of non-elected officials (Majone, 2001; Elgie, 1998; Haan and Amtenbrink, 2000; Tucker, 2018). In line with this, Dincer and Eichengreen (2018) and Horvath and Vaško (2013) have shown that higher levels of independence are associated with more transparent communication strategies. Nevertheless, central banks still often have low levels of formal accountability and use communication to address this (especially in periods when output legitimacy is low as Shambaugh and Shen (2018) indicate was the case during the financial crisis) and thereby attempt to mitigate this democratic deficit (Puppis et al., 2014). In line with this, Gilad et al. (2015) have shown that agencies use communication to increase their popularity and protect themselves from audience-based pressures. Similarly, Baker et al. (2024) show that communication can be performatively used to reinforce independence and Hayo and Neuenkirch (2015) show that communication improves perceived central bank independence of financial market participants.

Communication about policy pressures in a wide range of issues (e.g., financial market issues) is more important for independent central banks since they often act in isolation and cannot blame or have their policies justified by governments any more. They will, therefore, have to alleviate pressures such as financial pressures that may emerge to

keep the currency afloat and maintain trust in the central bank themselves. In line with this, Moschella and Pinto (2019) find, using speeches of the Federal Reserve, that when reputational concerns of central banks are higher, they are more likely to focus on areas where its reputation is weak or not yet established, such as financial regulation (Baker et al., 2024).⁴ Construction of institutional reputation can, therefore, be achieved by proactive communication (e.g., see Baker et al., 2024; Busuioc and Rimkutė, 2020). As Braun (2016) and Baker et al. (2024) argue, central banks communicate to demonstrate awareness of the risks these issues pose to financial stability, thus demonstrating preparedness, technical capacity and knowledge. The opposite also holds: less independent central banks will focus more on justification of their narrow price stability tasks to limit the harm of damages to the government's broader reputation while still benefiting from the long-term government reputation.

2.2.1 Observable expectations

How do these altered pressures and reputation expectations translate into observable expectations in communication? Combining both channels, the policy pressures are addressed in communication by central banks to maintain their reputation as an independent institution. First, a higher degree of independence makes it easier for central banks to bring inflation in line with a target rate, thus, central banks should be less occupied with justifying monetary policy and addressing price stability pressures. Second, due to increased independence, central banks are expected to discuss financial market pressure since financial regulation is often made more lenient and supervision is placed at more distance from central banks as a consequence of independence.

⁴Contrary to this, Braun and Dusterhoff (2024) find that central banks engage in ‘strategic silence’ to defend themselves against politicisation. However, given that we limit narrowly to the main policy pressures and not the politically contentious consequences of certain policies, this is less applicable for this chapter.

As mentioned in the previous section, to operationalise these pressures, we employ the economic framework of ‘monetary dominance’ and ‘financial dominance’ referring to the degree to which there is communication about policy subordination in central banks regarding their monetary policy conduct, i.e., ‘dominance’ (Sargent and Wallace, 1981; Davig and Leeper, 2011; Bianchi and Ilut, 2017; Brunnermeier, 2015). We use a broad definition of these concepts in order to detect them in communication. The situation in which central banks are unconstrained in their policy conduct is often referred to as ‘monetary dominance’ (Sargent and Wallace, 1981). Central banks accommodating its monetary policy to financial considerations is often referred to as ‘financial dominance’ (Brunnermeier, 2015; Diessner and Lisi, 2020). Financial dominance specifically describes a situation in which there is an over-leveraged financial system and the central bank is not able or willing to tighten its policy stance as this would threaten the stability of the financial system (Brunnermeier, 2015). We choose to link our pressures to this framework since it reflects the key drivers and concerns of central bank’s actions regarding monetary policy (e.g., see Schnabel, 2020). We argue that when a central bank is ‘monetary dominant’ in communication, their main concern and topic they discuss are inflationary pressures in monetary policy. Similarly, when a central bank is ‘financially dominated’ the main concern they reflect are pressures of financial markets.

Although the original framework refers to monetary and fiscal dominance, we choose to exclude fiscal dominance from our theory since the amount of fiscal pressure expressed in communication can be expected to be low overall given that central banks do not often publicly show to be constrained by governments who have given them independence in the first instance to solve this problem. Former ECB Executive Board member, Otmar Issing, confirms this and refers to communication about fiscal policy as a “minefield for central banks” since it will be “exposed to criticism that it is exceeding its competencies” yielding to “political pressures” (Issing, 2019). Moreover, fiscal and financial dominance

often interact making it difficult to have fully exclusive categories. For example, Mabbett and Schelkle (2019) suggest that financial dominance can result from governments, either constrained by fiscal limitations or reluctant to intervene, leaving financial markets unchecked, thereby forcing central banks to accommodate them.

Combining these expected changes, one can expect to empirically observe a trade-off: financial pressures in communication regarding monetary policy replaces mainly monetary considerations in the communication. This leads to the following overarching observable hypotheses (also depicted in Figure 2.2):

Hypothesis 2.1. (*Main effect 1*): *An increase in independence of a central bank results in less focus in addressing inflationary pressures in monetary policy communication over time.*

Hypothesis 2.2. (*Main effect 2*): *An increase in independence of a central bank results in an increase in addressing financial pressures in monetary policy communication over time.*

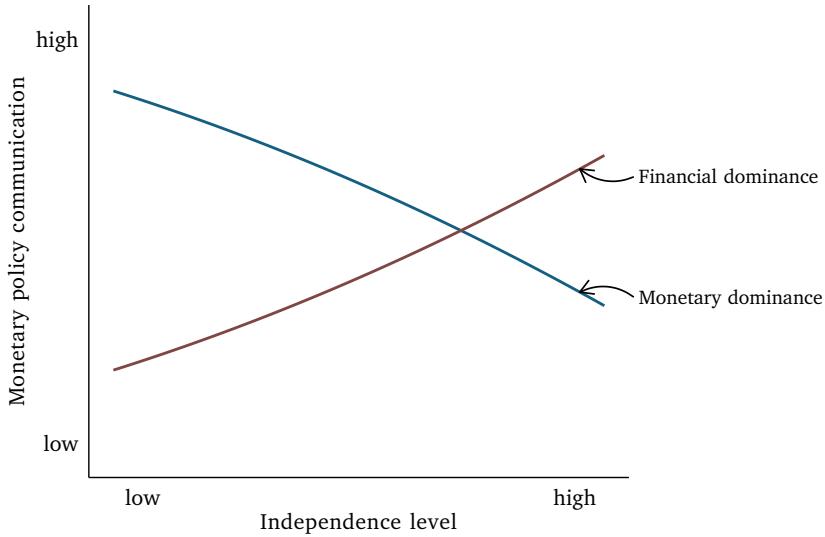


Figure 2.2: Relationship between monetary policy communication and CBI

Note: This figure summarises the anticipated monetary policy communication patterns consistent with our hypotheses. The slopes of the lines represent the direction of the changes in monetary policy communication as central bank independence increases. Our theory makes no prediction about the functional form or the intercepts of financial and monetary dominance.

This overarching hypothesis can be explained by the identified mechanism:

Hypothesis 2.3. (*Mechanism*): *A change in independence causes policy pressures to change which is addressed in central bank communication.*

We also expect certain scope conditions that ensure that the main effects are not always as strong or present depending on the country characteristics in which the central bank is embedded. For instance, as Acemoglu et al. (2008) argue, the effectiveness of economic policy reform (including CBI) are inseparable from the political economy factors responsible for distortionary policies in the first place and as Dincer and Eichengreen (2018) argue, trends in transparency are positively affected by the strength of political institutions.

First, the political regime affects the reputation channel. While communication remains

relevant for central banks in both democracies and autocracies, its role and credibility differ. As Broz (2002) argues, in autocracies where political institutions are opaque and decision-making lacks transparency, governments often adopt more transparent and externally verifiable institutional commitments - such as fixed exchange rate regimes—to anchor inflation expectations. In Broz's framework, central bank independence (CBI) is not seen as a credible solution in such settings, given the weak institutional constraints and high likelihood of political interference.

Building on this, I argue that even if CBI lacks credibility in autocratic regimes, central banks may still engage in communication, albeit under significant constraints. Specifically, their ability to speak freely on a wide range of topics is often curtailed, and dissenting from government preferences may result in political repercussions, including dismissal (e.g., Broz, 2002; Strong, 2021). Consequently, while communication may still occur, its responsiveness to changes in formal independence is expected to be more limited in autocratic settings. Therefore, the overall effect of CBI on communication is likely to be weaker in autocracies.

Second, the economic development level of a country gives rise to different types of macroeconomic pressures. As argued by Neuenkirch (2013), central bank communication is equally important for central banks of emerging economies for the effective implementation of their monetary policy. In line with this, Garriga and Rodriguez (2020) show that in developing countries CBI also reduces inflation. However, there are also different needs to address. As Evdokimova et al. (2023) point out, central banks in emerging economies communicate more often in areas in which they have more vulnerability than advanced economies. For instance, in emerging economies, there is a need for communication to support transitions to different monetary policy regimes, or to adequately explain policy decisions in the presence of multiple objectives. Emerging countries also more often experience ‘original sin’ problems, i.e., that most countries are not able to borrow abroad

in their domestic currency (e.g., see Eichengreen et al., 2023). Other examples include issues such as capital flow volatility, financial dollarisation, and more frequent economic or political shocks, in the context of overall weaker institutional capacity and policy credibility. Besides that, the financial markets are often already strongly deregulated (Pond, 2018). Therefore, financial pressures do not increase due to financial market deregulation. These consistently high pressures of financial considerations in monetary policy means that the baseline level of financial communication in emerging countries is high regardless of the level or changes in independence. In other words, the policy pressure channel is different than in advanced countries, leading again to the expectation that the main effect will not materialise.

This leads to following hypothesis regarding the scope conditions:

Hypothesis 2.4. (*Scope conditions*): *The main effect is primarily observed for central banks in democracies and advanced economies.*

2.3 Data and measurement

To construct the dataset for our analysis, we employ a seven step workflow which is summarised in Figure 3.3 below. Our aim is to create a database which includes textual indicators of pressures on the conduct of monetary policy. To this database, we also add our treatment variable, i.e., changes in CBI, and further control variables.

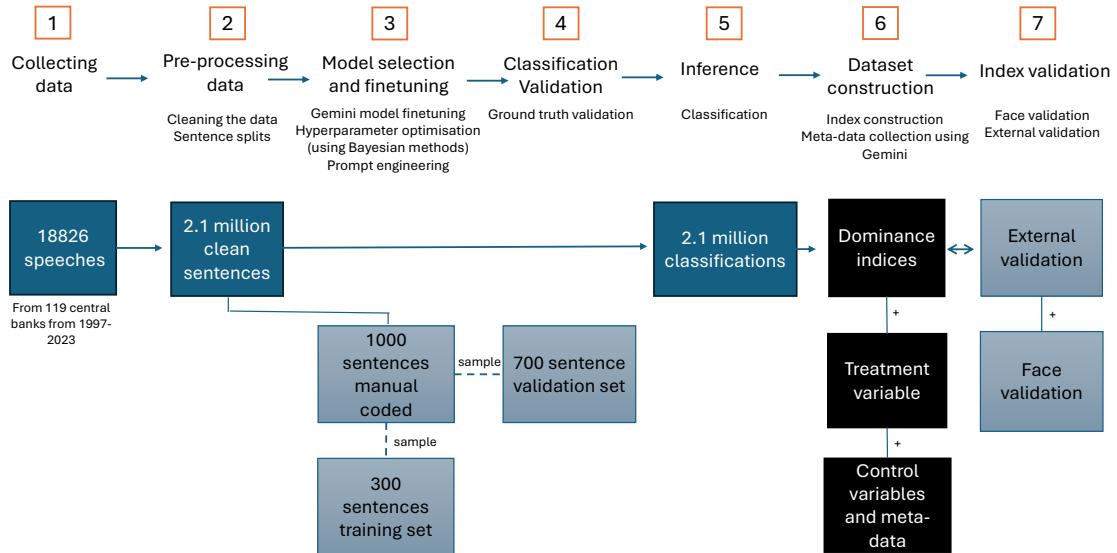


Figure 2.3: Workflow: data and measurement used to create our database

For the construction of our dependent variable, we collect speeches from the Bank of International Settlement (BIS) central bankers speeches database (step 1). This database is used by a growing number of studies (e.g., see Feldkircher et al., 2024; Armelius et al., 2020). As of March 2024, the dataset provided by the BIS contains 18,826 speeches⁵, from which we use the years 1997-2023. The speeches in the database include speeches of central bank governors and policy relevant speeches by senior central bankers. All these speeches are given on official occasions and published by the central bank itself and, therefore, can be considered as representing the stances of the bank itself. The dataset does not include all central bank speeches, however, covers diverse geographical regions including central banks in developing countries and central banks that are part of autocratic regimes (see Figure 2.4). Arguably, speeches can be seen as a most-likely case to detect the above-mentioned channels through which CBI influences central bank communication. Namely, speeches are both used as a macroeconomic policy-tool as well

⁵The entire dataset can be downloaded from the BIS website and is regularly updated: <https://www.bis.org/cbspeeches/download.htm>

as a means to reach various audiences for reputation purposes (Casiraghi and Perez, 2022). Other forms of central bank communication are more constrained in their format and often have more specific tasks (Moschella and Pinto, 2019).

After collecting the speeches, we pre-process the data (step 2). We first delete irrelevant textual artifacts (including page numbers, URLs, new page characters, footnotes, etc.) using regular expressions. We then convert the entire text corpus to natural sentences and, lastly, we apply heuristics to remove non-genuine sentences such as tables, annotations, etc. (see Appendix D2.1 for more information). After the pre-processing step, our dataset comprises of around 2.1 million clean sentences. We specifically choose to work on the sentence level, with the sentence prior and after to provide context, to achieve finer-grained indices and to facilitate more accurate validation compared to analysis conducted at the speech level.

To operationalise whether central banks are communicating about being pressured in their monetary policy conduct, we identify sentences concerning monetary policy and classify them using the concepts of dominance as described in the previous section. Specifically, we assign one of six labels to each sentence. Sentences that are unrelated to monetary policy or do not give any indication of the motives and pressures of the central bank are assigned the ‘none’ label. Sentences that are indicative of a form of dominance are classified as ‘monetary’, ‘fiscal’, or ‘financial dominance’. Following Leek et al. (2024), we also add two policy coordination categories (monetary-fiscal coordination and monetary-financial coordination) to account for instances when central banks call for coordination in a way that does not interfere with its monetary policy objectives. In this way, we ensure that instances in which central banks discuss monetary issues without responding to certain pressures are not confused with dominance. Table 2.1 contains examples from our two main classification categories of interest for our main hypotheses (see Appendix B2 for the entire codebook).

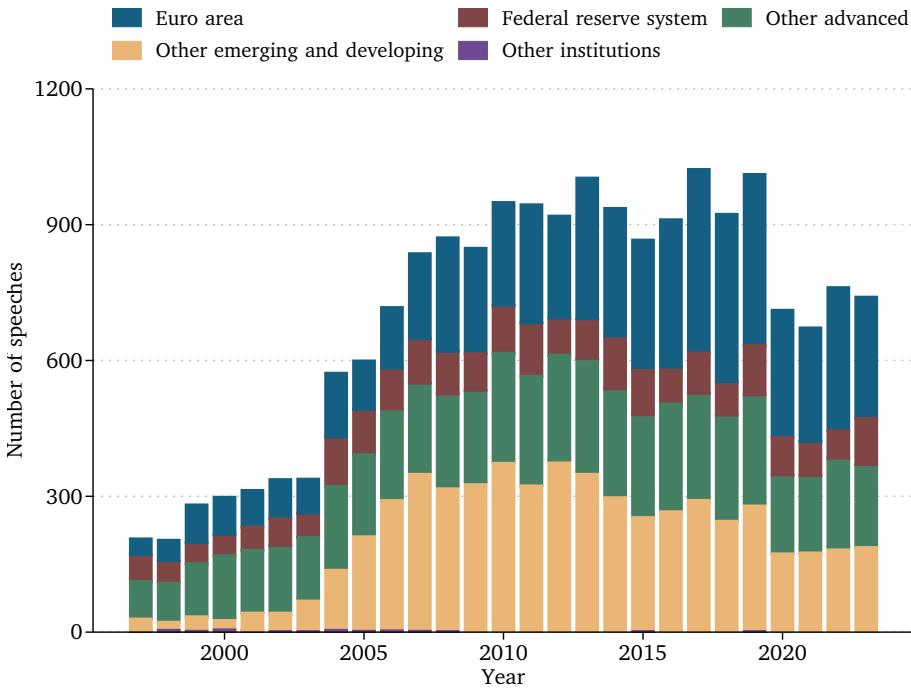


Figure 2.4: Distribution of central bank speeches according to country group

Note: The bars display the number of speeches from the BIS central bank speeches dataset, categorised by country group for the period 1997-2023. There are 11 speeches that were given before 1997 and 28 speeches from 2024, which we drop from our sample. Euro area refers to the current euro area. The classification of countries follows the IMF World Economic Outlook. The ‘Other institutions’ category comprises 35 speeches that cannot be attributed to the central bank of a country or currency area. These include speeches by regulators or intergovernmental organizations such as the BIS, IMF or the Inter American Development Bank, which we drop from our analysis.

Given the large number of sentences and the complexity of the classification task, we need a scalable method that offers a high ability to interpret text. For our measure of monetary policy communication, we use a large language model (LLM), which offers the ability to interpret text beyond other classification methods such as dictionary and word embedding models (see e.g. Widmann and Wich, 2023). We use Gemini Pro 1.0 (gemini-1.0-pro-001), which, similar to GPT, builds on the transformer decoder architecture of Vaswani et al. (2017) (Gemini Team, 2023). Compared to the frequently used BERT-based text classifiers (e.g. see Bertsch et al., 2024; Bestvater and Monroe, 2023), it has two main advantages: (i) it has many more parameters (billions vs millions)

Type	Definition	Example
Monetary dominance	The central bank prioritises to maintain price stability; its monetary policy is not subordinated to fiscal policy or to financial stability considerations.	“Furthermore, monetary policy implementation in line with the market efficiency principle would need to remain without prejudice to our primary mandate of safeguarding price stability.” (Retrieved from: The European Central Bank, 14 June 2021)
Financial dominance	The central bank accommodates its monetary policy to financial considerations; its decisions are subordinated to respond to the needs of financial markets.	“It is thus significant that our flexible and abundant provision of liquidity contained market participants’ concerns over liquidity financing.” (Retrieved from the Bank of Japan, 4 July 2002)

Table 2.1: Examples of sentence classifications

Note: The examples are taken from the validation dataset. See Appendix B2 for explanations of why examples are classified within a category. Also, see <https://centralbanktalk.eu/data-speech> to explore the classifications of the entire dataset.

resulting in a broader knowledge base, enabling better understanding of context and (ii) can directly respond to a prompt, removing the need for task-specific training data to train a classification head which is necessary to generate a classification output. We use Gemini in particular, because at the time of writing, it is the most capable base LLM that offers a fine-tuning interface. Fine-tuning conducts very limited training on top of the base LLM, which has been shown to greatly improve model classification accuracy with very little annotated data across a number of LLMs (Alizadeh et al., 2024). Also, see Hansen and Kazinnik (2023) who show this for an use-case of central bank communication.

Steps 3 and 4 cover the model selection, fine-tuning and validation for which we use the human coded validation set of Leek et al. (2024). Our LLM-based approach is built on a prompt that (after light prompt engineering) includes concise definitions of the

six categories, clear instructions on the desired output format, and the sentences to be classified with the preceding and following sentence added as additional context. From the manually coded dataset, we randomly select 300 sentences with their context for fine-tuning, leaving 700 sentences as the validation set. The fine-tuning procedure takes, as input, pairs of prompts and the desired outputs. We use the Bayesian Tree-structured Parzen Estimator Bergstra et al. (2011) implemented by the hyper-parameter optimisation framework Optuna Akiba et al. (2019) to jointly determine optimal hyper-parameters (the number of epochs, learning rate and batch size), conduct prompt engineering (sentences per prompt, temperature and format instructions) and adjust the training set composition (synthetic sentences, up-sampling factor and randomisation of sentence order across epochs).⁶ Model quality is assessed with a weighted average of the F1-macro score and accuracy⁷, whereby we assign twice the weight to the F1 score to emphasise the importance of balanced performance across all classes. For details on the exact prompt, the fine-tuning procedure and how different configurations affect model performance, see Appendix B5 and Figure B5.13. The validation in step 4 is done using the 700 held-out sentences as ground truth. These 700 sentences are a random sample of the entire corpus, ensuring that the performance metrics attained on the validation sample are an unbiased estimate of the accuracy on the full sample. We compare our model against a range of competing LLMs. Our final Gemini model outperforms all other LLMs, including GPT-3.5, fine-tuned GPT-3.5, GPT-4 and (few-shot) Gemini. In addition, the fine-tuned Gemini model comes closest to the human benchmark in that it tends to be more cautious in assigning one of the ‘dominance’ labels. (See Appendix B6.1 for more details and Table B6.9 for performance metrics).

⁶Despite strong class imbalance in our dataset, we found that up-sampling less frequent categories or adding LLM-generated sentences to the training set did not improve model performance. With appropriately set learning rates and epochs, our fine-tuned model remains sensitive to minority categories. To mitigate overfitting, we re-randomise sentences in each prompt for every training epoch.

⁷The F1-macro score assigns equal importance to all categories regardless of their frequency, and accuracy, defined as the share of correctly classified sentences.

In the fifth step, we classify the entire corpus with our final model, before we construct our indices in step 6. Figure 2.5 summarises how the classifications are distributed.

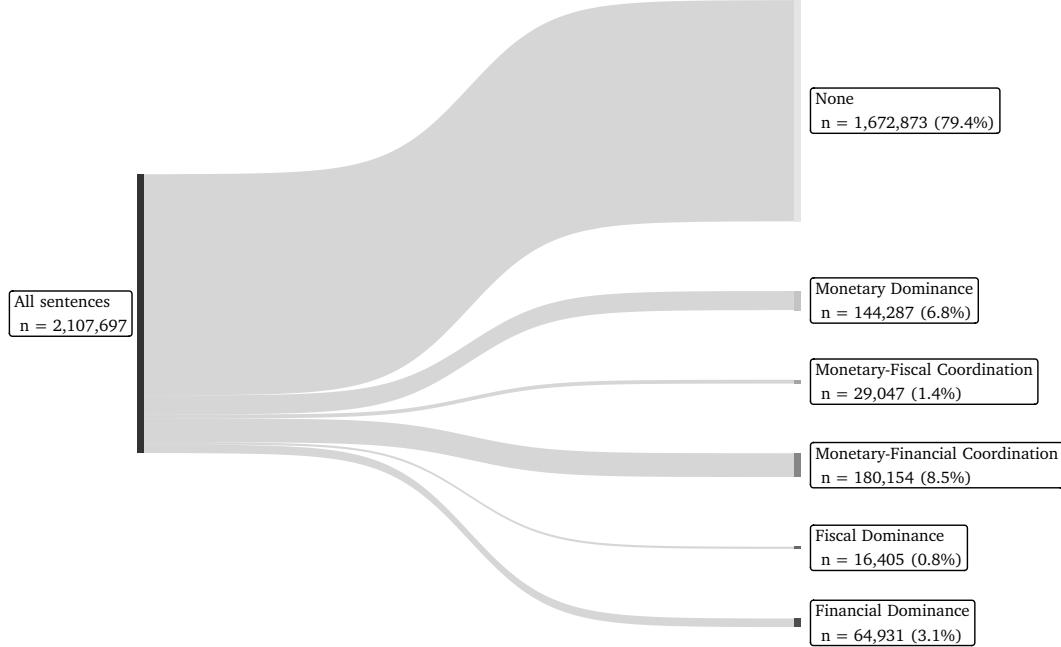


Figure 2.5: LLM classification categories breakdown

Note: The graph provides an overview of the six categories in our classification scheme and indicates how many sentences are classified under each category. The number in parenthesis represents the share of all 2,107,697 sentences.

To create our indices, we calculate a relative share ψ_i^m for each dominance and coordination category $m \in M$ for all speeches i by summing up the number of sentences belonging to each category and dividing by the number of sentences belonging to any dominance or coordination category.

$$\psi_i^m = \frac{\sum_{s \in S_i} \mathbb{1}(\text{Classification}_s = m)}{\sum_{s \in S_i} \mathbb{1}(\text{Classification}_s \in M)} \quad (2.1)$$

with $M = \{\text{Monetary dominance, Fiscal dominance, Financial dominance, Monetary-fiscal coordination, Monetary-financial coordination}\}$ and $s \in S_i$ indexing all sentences that belong to speech i . The proportions add up to 1, i.e., $\sum_{m \in M} \psi_i^m = 1 \forall i$, which creates desirable properties for cross-central bank and time comparison for two main

reasons.⁸ First, it normalises the score for the increase in central bank communication over time and accounts for strong heterogeneity between central banks in their communication frequency. Second, it creates a relative measure of the type of dominance, and since different kinds of dominances are to be interpreted as a trade-off, namely, more monetary dominance implies less fiscal and financial dominance, a relative index is appropriate.

In the final step, we assess the face and external validity of our indices. Figure 2.6 shows a 365-day moving average of our relative indicator based on all speeches included in our dataset. Communication regarding monetary dominance was the predominant form of communication until 2005 and saw a sharp decline after 2008 that corresponds with the onset of the GFC. Communication regarding financial dominance sees a first spike at the beginning of the observation period, likely explained by the Asian financial crisis in the late 1990s. Similarly, during the GFC, talk indicative of financial dominance increased quickly, followed by a slow but steady decline. All these trends show face validity, nevertheless, beyond face validity, we offer external validity of our measures through showing correspondence of our measures of monetary and fiscal dominance with a recent study of Hinterlang and Hollmayr (2022) measuring monetary and fiscal dominance in the United States using an economic modelling approach (see Appendix B6.2).

⁸In case a speech contains no sentences classified in any dominance or coordination category, i.e., all sentences are labelled as none, the indices are not defined and the speech is dropped from the estimation.

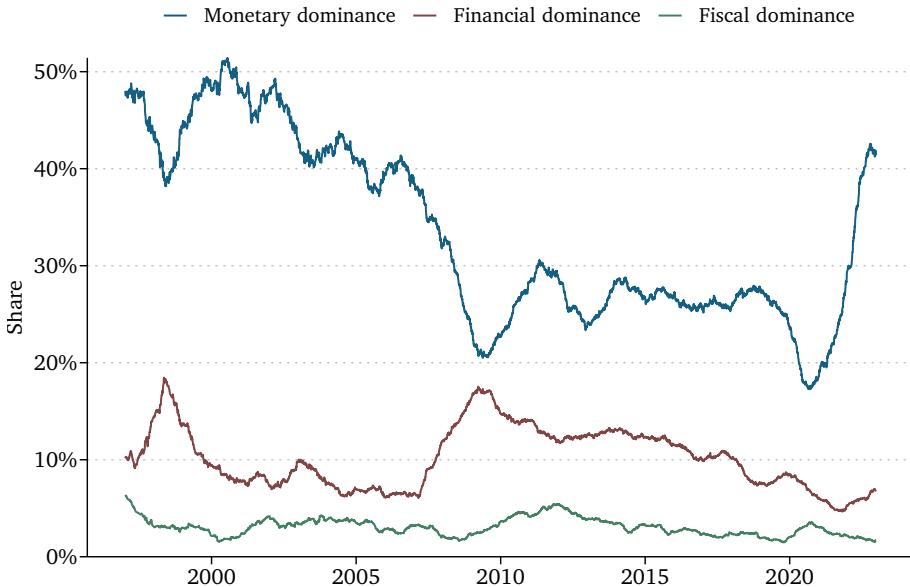


Figure 2.6: Indicators of dominance

Note: The lines represent 365-day moving averages (using a symmetric window) of the relative dominance shares in speeches, as defined by equation (2.1)

For our treatment variable of changes in CBI, we primarily rely on the CBI dataset from Romelli (2024). In Appendix B4.7, we also show results using the central bank independence dataset of Garriga (2025a) and systematically compare the measures. The dataset of Romelli (2024) provides an indicator of *de jure* central bank independence, derived from charters and legal documents of central banks. It has broad coverage, encompassing extensive time spans with data from 155 countries, typically starting from the 1980s or earlier.⁹ CBI is quantified using a continuous indicator ranging from 0 to 1. Our primary focus is on Romelli (2024)'s extended CBI Index which represents an average of six sub-indicators covering: (i) the governor and central bank board, (ii) monetary policy and conflict resolution, (iii) objectives, (iv) limitations on lending to the government, (v) financial independence and (vi) reporting and disclosure. This index also

⁹Out of the countries in our speeches database, 18 countries are missing in Romelli (2024), making up a share of 4.8% of our speeches sample. Romelli (2024) does not track the CBI indicator for the ECB, which is the central bank with the most speeches in our sample. In specifications that include the ECB, we assume that the ECB's independence remained unchanged due to the absence of simultaneous changes in CBI across euro area states.

accounts for differing mandates of central banks (which we further examine in Section 2.4.5). CBI is assessed annually and most central banks experience at least one change of independence from 1985 to 2023 (see Figure 2.7 for the coverage of the dataset and event frequency).¹⁰ After merging with our speeches dataset, the dataset consists of 17,829 complete observations with corresponding CBI data from 100 central banks.

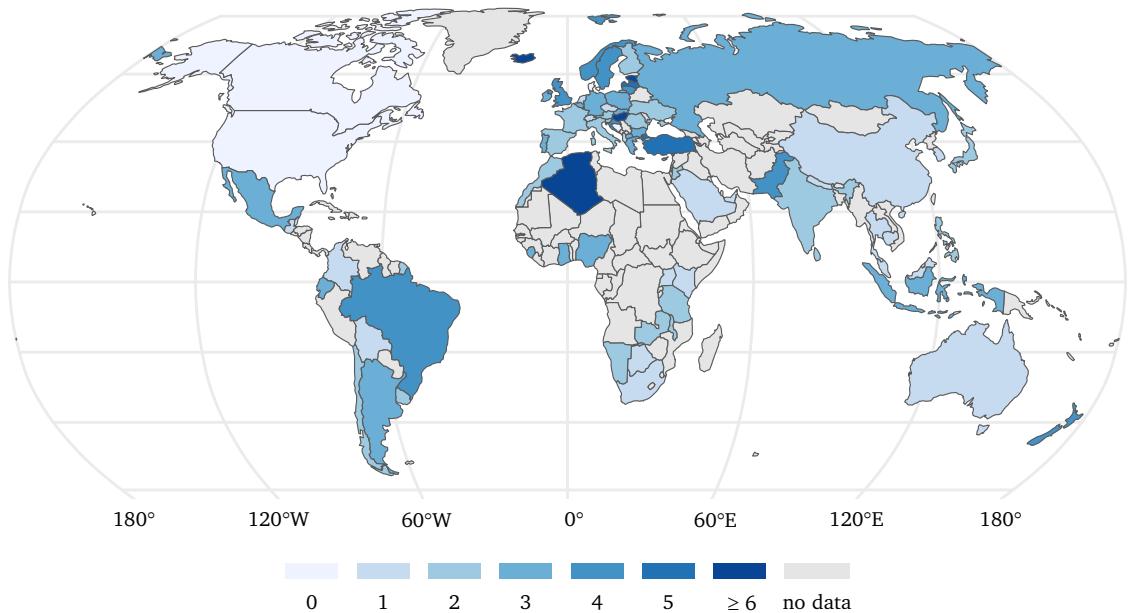


Figure 2.7: Dataset coverage and number of independence changes

Note: The colour scale indicates the number of independence changes observed in the time period 1985-2023. The countries shown are the countries which are both included in the independence dataset and in our speeches. Countries that are missing in one or both datasets are shown in grey.

The BIS speeches dataset includes short, non-standardised descriptions often containing the same speech-specific metadata. We use Gemini to process the unstructured descriptions and extract key metadata such as the central bank, location, audience, the speaker and the occasion of the speech. To conduct subgroup analysis and to control for macroeconomic and political variables, we merge additional datasets, including the

¹⁰Six central banks, most notably the Federal Reserve, exhibit a constant level of independence throughout.

April 2024 World Economic Outlook (IMF, 2024), the VDEM database (Coppedge et al., 2024), a financial stress indicator (Ahir et al., 2023), data on central banks' supervision capabilities (Masciandaro and Romelli, 2018) and currency peg arrangements (Harms and Knaze, 2021). See Appendix B1.2 for information on our meta-data extraction process and Appendix B3 for extended descriptions of the additional datasets and a table with descriptive statistics of all variables used in the paper.

2.4 Empirical analysis

2.4.1 Staggered difference-in-differences

The staggered adoption of central bank independence constitutes a natural experiment that we exploit using a difference-in-differences design. We estimate the dynamic effects of changes in CBI using a two-way fixed effects (TWFE) event study specification. To do this, we regress our constructed dominance indicators ψ_{ict}^m on leaded and lagged changes in CBI:

$$\psi_{ict}^m = \sum_{k=-5}^{k=-2} \beta_k D_{ct}^k + \sum_{k=0}^{k=12} \beta_k D_{ct}^k + \mu_c + \theta_t + \epsilon_{ict} \quad (2.2)$$

where D_{ict}^k is the event indicator, signalling if the central bank of country c associated with speech i exhibited a change in central bank independence k years from the year of observation t . μ_c and θ_t are country and year fixed effects respectively. The coefficients $[\beta_{-5}, \dots, \beta_{-2}]$ enable us to test differences in pre-treatment trends which are indicative of the plausibility of the parallel trends assumption. The coefficients $[\beta_0, \dots, \beta_{12}]$ are the dynamic treatment effects of changes in CBI. They measure the average treatment effect on the treated (ATT), i.e., the impact of changes in CBI for those countries that experienced such changes. All β_k coefficients are interpreted relative to the reference period one year before treatment whose coefficient β_{-1} is implicitly normalized to 0. The

regressions are run on the level of individual speeches, while treatment is assigned on the country level. We follow the standard practice of clustering standard errors on the level of treatment assignment (Abadie et al., 2023).

Following the recommendations of Schmidheiny and Siegloch (2023), we bin the first and last event indicator:

$$D_{ct}^k = \begin{cases} \sum_{j=-5}^{-\infty} d_{c(t-j)} & \text{if } k = 5 \\ d_{c(t-k)} & \text{if } -5 < k < 12 \\ \sum_{j=12}^{\infty} d_{c(t-j)} & \text{if } k = 12. \end{cases} \quad (2.3)$$

where $d_{c(t-k)} \neq 0$ if country c changed CBI in year $t - k$ and $d_{c(t-k)} = 0$ otherwise. The last and first of the binned event indicators D_{ct}^k , therefore, accumulate changes in independence that are known to exist outside of the event window from $k = \{-5, \dots, 12\}$. The binning of endpoints entails the assumption that the effects level off, i.e. all β_k outside of the event window would equal the coefficient at the binned endpoint. We implement suggestions by Freyaldenhoven et al. (2021) and report the p-values associated with Wald tests for levelling off of the effect $H_0 : \beta_{11} = \beta_{12}$ and for no pre-trends $H_0 : \beta_{-5} = \beta_{-4} = \beta_{-3} = \beta_{-2} = 0$.

The TWFE specification (2.2) may not yield the dynamic treatment effects of interest when treatment effects are heterogeneous over time (Sun and Abraham, 2021). Numerous alternative estimators have been developed to estimate event study models with leads and lags (e.g., see Dube et al., 2023). We report the corresponding specification using Gardner et al. (2024). This estimator uses a two-stage approach in which the fixed effects are separately estimated in a first stage using untreated observations, followed by the estimation of the dynamic treatment effects employing a dependent variable that is residualised by subtracting both fixed effects.¹¹

¹¹In the presence of staggered treatment adoption with time-varying treatment effects, part of the

In our main specification we impose the restrictions $d_{ct} \in \{0, 1\}$ and $\sum_t d_{ct} \leq 1$, implying, at most a single independence increase per country and no differentiation in intensity. We adopt this approach to maintain a clear separation of treatment and control observations, which is the standard case in the difference-in-differences literature and required for most heterogeneity-robust estimators to be applicable.¹² We aim to define events as changes in independence that are large enough to constitute a meaningful change in the institutional setup of a central bank. We, therefore, require a minimum increase in CBI of 0.05. When a country experiences multiple independence events, we only consider the largest change in CBI as an event. Figure 2.8 plots the distribution of independence changes falling into the relevant event window (1985-2023) for the estimation of equation (2.2) both in terms of magnitude (Panel A) and the number of events over time (Panel B). CBI changes excluded from the baseline due to the presence of a larger event in the same country or their magnitude falling below the defined threshold are shaded.

Limiting the analysis to a single increase of high magnitude per country introduces the potential for bias by ‘contaminating’ the control group with units that experienced CBI changes. Moreover, the occurrence of multiple treatments either before or after the selected event could influence the estimation of the dynamic treatment effects. Given the observed distribution of changes in independence, we anticipate the impact of these issues to be small. A large number of CBI changes and, in particular, decreases in CBI, are incremental adjustments to central bank regulation, with likely minimal impact. As a robustness check, to leverage the maximum variation in our dataset, we also estimate a fully flexible specification with variations in treatment intensity including reductions in

treatment effect is attributed to the fixed-effects. Removing the cleanly estimated unit and time fixed effects from the dependent variable before estimating the event-study solves the complications arising from treatment effect heterogeneity (Gardner et al., 2024).

¹²Heterogeneity-robust estimators have so far been mostly developed for the case of a single irreversible binary treatment with some very recent work generalising to more general cases (e.g., see Callaway et al., 2024; de Chaisemartin and D'Haultfœuille, 2024).

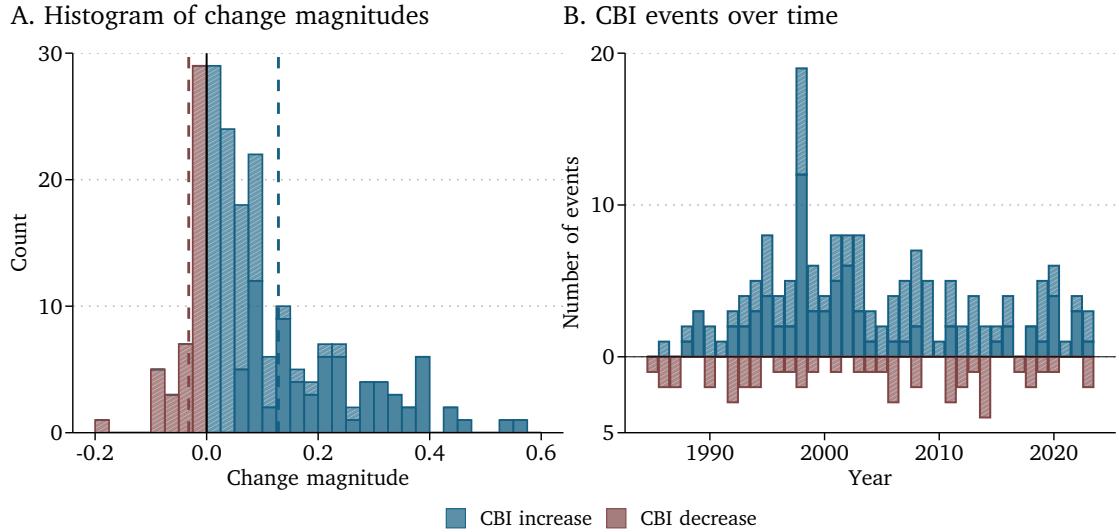


Figure 2.8: Distribution of independence changes

Note: Panel A shows a histogram of CBI changes in our merged dataset. Bars show the number of events from 1985 to 2023 in bins of 0.025. The vertical dashed lines indicate the average of increases and decreases in CBI. All events to the right of the second blue bar (change greater than 0.05) are considered independence events unless an event of stronger magnitude occurred in the same country. Panel B depicts the number of CBI events observed in each year. In both panels, the shading of the bars indicates if the CBI event is considered in our baseline specification (solid bars), or dropped (shaded bars).

CBI and multiple independence changes per country.¹³

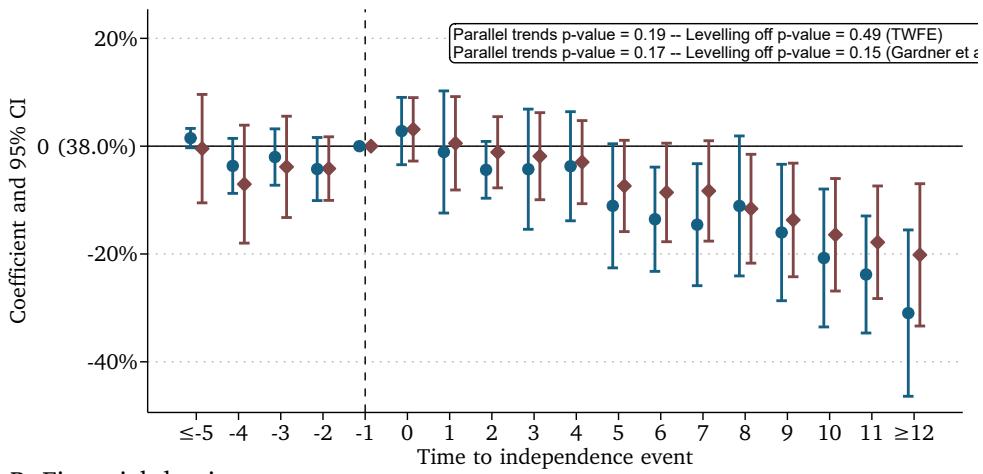
2.4.2 Results

Our baseline specification presented in Figure 2.9 shows that communication regarding financial dominance increases and monetary dominance decreases following an increase in independence. Taken together, this confirms the general direction of our main hypotheses. These results are stable across both estimators. Also in line with expectations, the plot shows that both effects only materialise after some delay, i.e., approximately after five years. The coefficient sizes are substantially relevant, i.e., the effects level off at a

¹³As shown by Schmidheiny and Siegloch (2023) the standard two-way fixed-effects specification from above with binned endpoints extends to the more general case of multiple treatments of varying intensities. In this case, the event indicators D_{ct} are no longer $\in \{0, 1\}$ but equal to the change in CBI, i.e., $D_{ct}^k = d_{c(t-k)} = \Delta CBI_{c(t-k)}$, or in case of the endpoints, the sum of changes (before) after the (last) first (lead) lag.

reduction of around 25 percentage points in the share of monetary dominance and an increase of 7 percentage points in the case of financial dominance. The specifications pass the placebo test for statistically distinguishable trends in monetary and financial dominance communication prior to the independence events, supporting the identification assumption of parallel trends that is underlying our event study approach.

A. Monetary dominance



B. Financial dominance

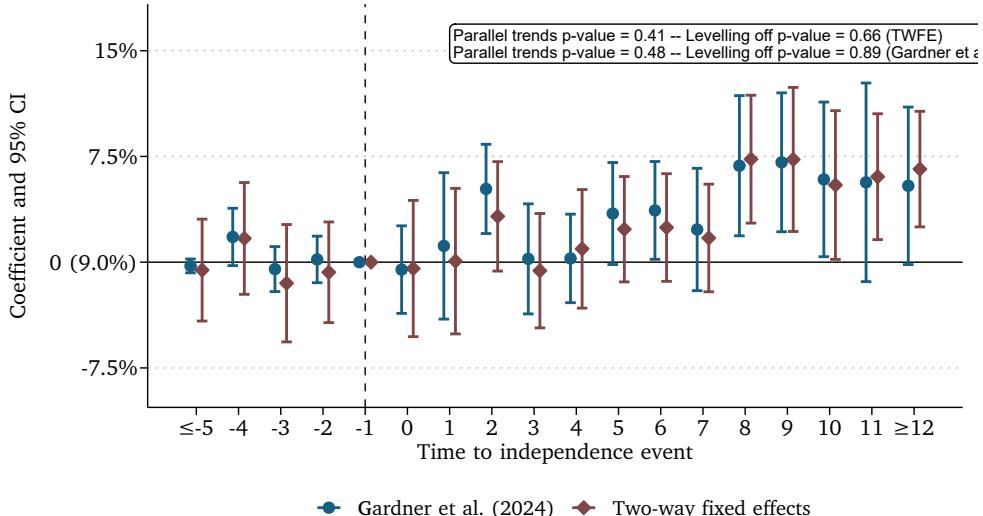


Figure 2.9: Event studies: dynamic effects on monetary and financial dominance

Note: The event-study plots show the β_k coefficients as estimated by the two-way fixed effects model (2.2) and the heterogeneity robust estimator of Gardner et al. (2024). Dynamic treatment effects are estimated relative to the year before the CBI change. The number in brackets at the zero line displays the sample average of the respective dominance measure in the reference period. The vertical bars represent the 95% confidence intervals for the estimated coefficients.

To study the differences in the effect of CBI on communication among subgroups (economic development level and political regime - in line with hypothesis 4), we augment our event study model (2.2) with additional lags that are interacted with dummies S_{ct}^j indicating membership in subgroup $j \in \{1, 2, \dots, J\}$:

$$\psi_{ict}^m = \sum_{k=-5}^{k=-2} \beta_k D_{ct}^k + \sum_{k=0}^{k=12} \beta_k D_{ct}^k + \sum_{j=2}^{j=J} \sum_{k=0}^{k=12} \delta_{jk} D_{ct}^k S_{ct}^j + \mu_c + \theta_t + \epsilon_{ict} \quad (2.4)$$

The interacted lag coefficients δ_{jk} measure differences in the dynamic treatment effect for members of subgroup $j \neq 1$ relative to the first category $j = 1$.

Table 2.2 displays the effect of CBI stratified by a number of central bank and country characteristics. For ease of interpretation and to compare effect magnitudes, we condense the effect of CBI into a single number. A natural approach would be to estimate a simple static difference-in-differences model with year and country fixed effects and one post treatment indicator.¹⁴ However, such specifications are particularly affected by the biases incurred under time-varying treatment effects.¹⁵ We, therefore, present an observations-weighted average of all post-treatment coefficients as our overall effect, which we show to be similar to alternative estimation techniques that also account for treatment effect heterogeneity over time (see Appendix B4.4 for a more detailed discussion).

In line with the results presented above, the aggregated baseline model shows an increase in financial dominance and a decrease in monetary dominance, with both changes statistically significant at the 1% level. The coefficients of subgroups by political system, i.e., autocracy and democracy, go in a similar direction as the baseline specification. However, the effect sizes are larger and more significant in central banks embedded in democracies in line with hypothesis 4. This corresponds with our expectations that central

¹⁴This would be equal to a version of regression model (2.2) with no leads and a single binned lag, i.e., $\psi_{ict}^m = \beta_1 D_{ict}^1 + \mu_c + \theta_t + \epsilon_{ict}$

¹⁵As shown by Goodman-Bacon (2021), such a model is a weighted average of all possible pairwise difference-in-differences comparisons, possibly involving negative weights and using already treated units as control observations. The event study equation (2.2) is less affected by this type of weighting issue as through its leads and lags it can already capture more heterogeneity.

banks in autocracies are more constrained in their freedom to talk about a wider range of issues regardless of the level of independence, i.e., the reputation channel. Similarly, the effect on financial dominance communication is less strong and not significant in developing and emerging countries - also in line with hypothesis 4, while only monetary dominance significantly decreases. This can, arguably, be attributed to an already high-level financial dominance being present in the communication regardless of the level of independence (e.g., see Evdokimova et al., 2023). Moreover, the argument of Masciandaro and Romelli (2018) and Aklin and Kern (2021) that financial market deregulation is an alternative for fiscal influence does not hold since autocratic governments can *de facto* still force central banks to support them with favourable monetary policy. Besides that, as mentioned, the financial markets are often already strongly deregulated (Pond, 2018). Therefore, the financial pressure portrayed in communication does not increase due to financial market deregulation.

Besides the expected scope conditions, we also test whether increases in financial dominance can be mitigated when central banks are in control of financial supervision themselves rather than the governments or another independent institution. We indeed see when central banks have higher supervision competences, the effect on financial dominance gets smaller which is likely explained by central banks being able to defend themselves against loosening financial regulation when they are themselves the financial regulator. This provides preliminary evidence that the mechanism involves increased financial pressure. We check whether the effects also hold when central banks are part of a monetary union or their currency is pegged to another currency, i.e., whether a country does not have monetary sovereignty. We observe that the financial dominance effect reduces when a central bank does not have full sovereignty over monetary policy. Arguably, when there is no full control of monetary policy by a central bank, there is a smaller change in their control of monetary policy or reputation associated when a

central bank becomes independent. Moreover, in a currency union, instead of outright complaining about pressures, coordination efforts between the various banks involved could occur (Leek et al., 2024; Diessner, 2023). Finally, we compare the effects conditional on the central bank's mandate since Baker et al. (2024), Braun and Dusterhoff (2024) and Bohl et al. (2023) all argue that the mandate influences communication patterns. We do this by distinguishing between banks whose mandates may conflict with price stability. We find no significant differences between the two groups.

2.4.3 Instrumental variable approach

Additionally, as is common practice in the literature on the effects of CBI, we use a fixed effects panel regression (Acemoglu et al., 2008; Sturm et al., 2024) paired with an instrumental variable (IV) approach to circumvent concerns of dynamic endogeneity under which the parallel trends assumption would fail. This approach circumvents biases arising from time-varying omitted variables or reverse causality, where the central bank's own communication might influence its independence (Hayo and Neuenkirch, 2015). Building on existing literature that employs instruments such as indicators of democratic governance (Crowe and Meade, 2008), lagged CBI (Jácome and Vázquez, 2008), and CBI in neighbouring countries (Garriga and Rodriguez, 2020, 2023), we use diffusion-based instruments. First, we construct an inverse distance-weighted measure of the lagged world average of CBI (Garriga, 2022), which is justified by evidence that diffusion is a key predictor of CBI (Bodea and Hicks, 2015). Second, we incorporate the lagged average level of electoral democracy in the 10 closest neighbouring countries.¹⁶ Additionally, as Acemoglu et al. (2008) argues, the strength of other institutions is related to CBI. Thus, we include the country's own judicial independence, defined as the absence of government

¹⁶We recognise that the exclusion restriction may be violated since the diffusion process can work in both directions, e.g., see Betz et al. (2018) who show that this is a general case for spatial instruments. We, therefore, choose to use this as an additional check for our difference-in-differences analysis.

	Monetary dominance	Financial dominance
Baseline		
Full sample	-0.1607*** (0.0554)	0.0548*** (0.0200)
Supervision capabilities		
Low	-0.1532** (0.0623)	0.0642*** (0.0203)
Medium	-0.1373*** (0.0457)	0.0446* (0.0228)
High	-0.2504*** (0.0724)	0.0279 (0.0189)
Political system		
Autocracy	-0.1161** (0.0459)	0.0495** (0.0195)
Democracy	-0.1747*** (0.0592)	0.0595*** (0.0187)
Monetary sovereignty		
Full monetary sovereignty	-0.1417*** (0.0532)	0.0556*** (0.0206)
Monetary union or peg	-0.2268*** (0.0732)	0.0244 (0.0229)
Economic development		
Emerging and Developing	-0.0644* (0.0354)	0.0021 (0.0289)
Advanced	-0.2229*** (0.0735)	0.0774*** (0.0170)
Mandates		
Non-conflicting with price stability	-0.1479*** (0.0506)	0.0542*** (0.0207)
Conflicting objectives	-0.2440*** (0.0700)	0.0588*** (0.0183)

Table 2.2: Effect heterogeneity by country characteristics

Note: Stars indicate significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The table presents our main event study using the two-way fixed effects specification (2.2) stratified by central bank and country characteristics. Clustered standard errors on the country level are reported in brackets. Heterogeneous effects are estimated by interacting category dummies with the lags as in regression model (2.4). The shown coefficients are aggregations of the estimated dynamic treatment effects for each subgroup. The first row ‘baseline’ reports the sample average effect. See Appendix section B3.3 for details on how the subgroups are constructed in each category.

attacks on the judiciary and arbitrary judge removals, as an instrument to capture related institutional dynamics. We estimate the following model using two-stage least squares (2SLS):

$$\psi_{ict}^m = \rho \tilde{\psi}_{ict}^m + \beta_1 \widehat{\text{CBI}}_{ct} + \beta_2 \Delta\pi_{ct} + \beta_3 \Delta u_{ct} + \theta_t + \mu_c + \epsilon_{ict}, \quad (2.5)$$

where θ_t and μ_c are country and year fixed effects. The exogenous variation $\widehat{\text{CBI}}_{ct}$ is estimated in the first stage using the three aforementioned instruments. We further control for the change in inflation $\Delta\pi_{ct}$ and unemployment Δu_{ct} as they are often used in the regional Fed governor's literature to determine central bankers' reaction functions in line with (e.g., see Chappell et al., 2008; Meade and Sheets, 2005; Malmendier et al., 2021). Further, following the empirical literature on the effects of CBI on inflation, we add a lagged dependent variable $\tilde{\psi}_{ict}^m$ to capture persistency (e.g., as in Acemoglu et al., 2008) to the regression. Because the content of each speech can differ substantially, relying only on the speech prior will not adequately capture the pressures that existed prior to the speech. To address this, we define $\tilde{\psi}_{ict}^m$ as the average measure over the central bank's last 25 speeches.¹⁷

Table 2.3 below presents the coefficient estimates from both the first-stage regression and the instrumented panel model. We find a highly relevant first stage, whereby the inverse distance-weighted world CBI is by far the strongest predictor of CBI.

¹⁷We choose 25 speeches because this is the number that maximizes the log likelihood of the estimated model. Since the number of coefficients in the model remains constant, this choice also gives the best (lowest) Bayesian (BIC) and Akaike (AIC) information criteria. Results depend little on this choice. Appendix B7.1 contains a discussion and results with alternative specifications and Appendix X provides further F statistics, exogeneity tests

	First Stage		2SLS Effect on Dominance	
	CBI		Monetary	Financial
Dependent Variables:	(1)	(2)	(3)	(4)
<i>Variables</i>				
CBI			-0.7524*	0.6921**
			(0.4365)	(0.2969)
Monetary dominance 25 prior speeches	0.0078		0.4952***	
	(0.0129)		(0.0727)	
Financial dominance 25 prior speeches		-0.0009		0.3247***
		(0.0376)		(0.0567)
Δ Inflation rate	0.0669	0.0671	0.4797***	-0.0932
	(0.0952)	(0.0955)	(0.1647)	(0.1080)
Δ Unemployment rate	0.0001	0.0001	-0.0016	0.0022
	(0.0006)	(0.0006)	(0.0039)	(0.0014)
Inverse distance weighted world CBI ₁	0.9288**	0.9182**		
	(0.3568)	(0.3653)		
Neighbour's electoral democracy index ₋₁	0.2376	0.2448		
	(0.2015)	(0.2077)		
Independence judiciary	0.0056**	0.0053*		
	(0.0027)	(0.0028)		
<i>Fixed Effects</i>				
Country	✓	✓	✓	✓
Year	✓	✓	✓	✓
<i>Fit statistics</i>				
R ²	0.97593	0.97591	0.21513	0.11198
Observations	12,205	12,205	12,205	12,205

Table 2.3: 2SLS estimation of the impact of CBI on monetary and financial dominance

Note: Stars indicate significance levels: * p \leq 0.1; ** p \leq 0.05; *** p \leq 0.01. The table presents the results of our instrumental variable specification (2.5). Clustered standard errors on the country level are reported in brackets. CBI is instrumented by the inverse distance weighted world CBI, the independence of the judiciary and the average electoral democracy index of the 10 neighbouring countries. The first and second stage control for changes in the unemployment and inflation rates as well as country and year fixed effects.

In the second stage, we obtain point estimates of -0.752 for monetary dominance and 0.692 for financial dominance. In practical terms, a 10-percentage point increase in the CBI indicator is associated with a 6.9 percentage point increase in financial dominance. For context, the median event magnitude in our event-study specification is 20 percentage points. Therefore, the results align with our event-study estimates in sign, magnitude and significance, although the effects on monetary dominance are only significant at the 10% level. We attribute this to the inherently less accurate IV estimation. Among the additional control variables, only the change in inflation is statistically significant, pointing towards a positive relationship with monetary dominance in line with our expected mechanism. We also examine heterogeneous effects by political system and economic development (see Appendix Table B7.11), where we again find that increases in financial dominance are driven by advanced economies.

2.4.4 Mechanisms

Thus far, we have found evidence for the observable outcomes of our main effect hypotheses, however, we have not examined the underlying mechanisms. To do this, we proceed in two steps. First, we establish that changes in CBI affect inflationary and financial pressures. Second, we demonstrate that these changes in pressures are, in turn, reflected in our communication-based measures. We do not test the reputation channel explicitly since we argue that the presence of this channel is a necessary condition for central banks to communicate and care about responding to these pressures in the first place.

A large body of previous research documents that increased CBI achieves lower inflation (e.g., see Alesina and Summers, 1993; Grilli et al., 1991). To test this policy pressures channel within our event study framework, we run country-level level regressions to estimate the impact of CBI on both inflation and financial stress. As shown in Figure 2.10, our event study design replicates the well-established finding that increases in CBI

lead to sustained reductions in inflation (see, e.g., Klomp and De Haan, 2010). The financial pressures estimates are not statistically significant, but the direction supports our hypothesis of increases in financial pressures following increases in CBI.

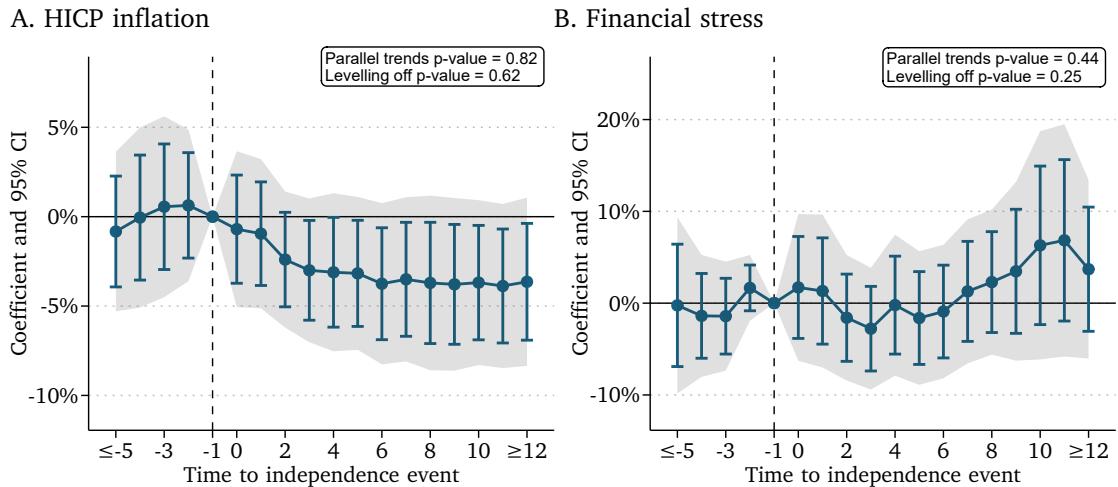


Figure 2.10: Event study estimates of the impact of CBI on policy pressures

Note: Both panels present event study estimates using our main event study specification (2.2), but with the adjustment that the regressions are performed on the country-year level instead of on individual speeches. The inflation variable is winsorized at the 99th percentile to limit the influence of extreme outliers. The grey shaded area indicates the sup-t confidence band which delimits the range of event-time paths of the effect that would still be consistent with a zero effect given the confidence level of 95%.

Second, we test whether these changes in pressures are reflected in our communication based measures. We have already seen some preliminary evidence in our instrumental variable approach that monetary dominance communication increases with higher levels of inflation. To show this more directly, in Figure 2.11 we plot our measures of monetary and financial dominance against inflation and financial pressures, respectively, to see how policy pressures are reflected in our communication-derived measures. In the region around most central banks' definition of price stability, i.e., two percent of inflation, we see a strong almost linear increase of monetary dominance communication with inflation for advanced economies (see Figure 2.11, Panel A). This suggests that lower inflationary

pressures contribute to the observed reduction in monetary dominance.¹⁸ The effect levels off with higher inflation rates. In emerging and developing economies, a similar but weaker link appears between inflation and monetary dominance. We also observe a positive relationship between financial stress as measured by Ahir et al. (2023) (who use a textual measure derived from quarterly country reports), and financial dominance in our speeches dataset with little differentiation between advanced and developing economies (see Figure 2.11, Panel B).¹⁹

2.4.5 Alternative explanations

We now address potential alternative explanations. First, various scholars argue that financial dominance is mainly driven by a global increase in financial dominance after the financial crisis (e.g., see Diessner and Lisi, 2020; Gabor, 2016; Hellwig, 2014). Following this, we test whether the effect of CBI on communication about financial dominance is explained by the financial crisis. While the year fixed effects account for these global trends, the estimation might still be skewed if the countries most affected by the crisis increased independence beforehand. To show that this is not the case, we estimate how the effect of CBI on financial dominance varies over time by running an event study with treatment cohort-specific effects, i.e., grouped by the year of independence change, using the approach of Sun and Abraham (2021). We aggregate these to yearly average treatment effects (Figure 2.12). The average effect is larger after the crisis, but remains positive and

¹⁸Examples of how high monetary pressure is reflected in communication are Bank of England governor, Andrew Bailey, who said in 2023: “Currently at 8.7% in the latest data, consumer price inflation is unacceptably high, and we must bring it down to the 2% target.” and Bank of Korea governor, Chang Yong Rhee, who said in 2022: “With regard to the future policy direction, the BOK has made it clear that it will prioritize price stability to the utmost as long as inflation remains high in the 5-6% range.”

¹⁹Examples of financial pressures being reflected are Federal Reserve vice-chair, Janet Yellen, saying in 2010: “In such situations, if emerging threats to financial stability become evident, monetary policy could be faced with difficult tradeoffs.” and Sveriges Riksbank Deputy Governor, Lars Nyberg, saying in 2008 that “The fact that monetary policy alone cannot prevent financial crises does not mean that we can only look at an inflation target and otherwise disregard what a long period of low interest rates may entail with regard to risks in the financial system.”

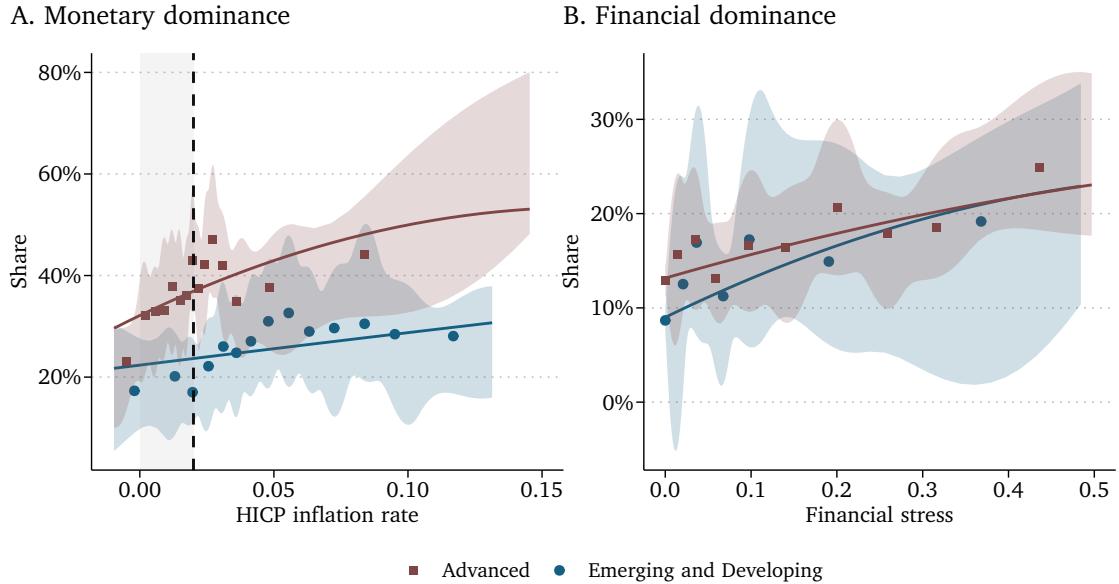


Figure 2.11: Relationship of inflation and financial pressures with dominance

Note: Each dot represents binned averages of central bank year observations. Observations with similar values on the horizontal axis are grouped together into 15 equally sized bins in Panel A, and 30 equally sized bins in Panel B. We double the number of bins in Panel B since the distribution of financial stress is heavily skewed towards zero. Without this adjustment, most bins would only contain zero values, with only a few bins containing non-zero values. The lines represent a quadratic fit. Shaded areas in red and blue represent the 95% confidence of fitted piece-wise polynomials of order two. In Panel A, the vertical line and the grey shaded area represent inflation rates below 2%, which is the common definition of stable prices.

significant before the crisis. The years with the highest global level of financial dominance (1998 and 2009-2011) do not show notably larger effect sizes.

Second, the results can be driven by the euro area countries since they constitute around 30% of our speeches sample. Specifically, we consider the adoption of the euro in 1998, which for many euro area countries coincided with their largest independence increase, and the introduction of the Single Supervisory Mechanism (SSM) in 2014. Our results remain robust when the euro area is either excluded from the estimation or the ECB speeches are added as control observation (see the first two columns of Figure 2.13). This, therefore, rules out that our results are confounded by the formation of the euro or the establishment of the SSM.

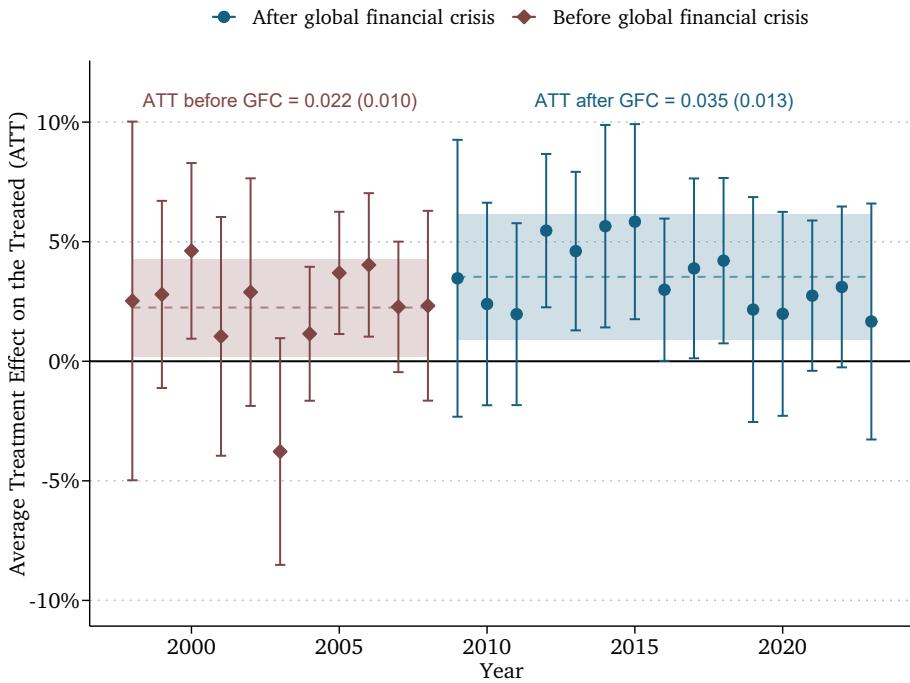


Figure 2.12: Effect of CBI on financial dominance

Note: Coefficients presented are aggregations of cohort time-specific effects, using the approach of Sun and Abraham (2021). In a given year, the average treatment effect on the treated (ATT) is calculated as the observations weighted average of all cohorts-year effects that are post treatment, i.e., among cohorts who increased independence prior to that year. The dashed horizontal lines represent the aggregated ATTs before the Global Financial Crisis (1998-2008) and after (2009-2023). The 95% confidence intervals are indicated by the shaded area above and below the line.

Third, once a central bank receives more independence for the first time, this can possibly lead to a precedent which can already have a lasting impact on the communication. For instance, once central bankers are part of an epistemic community (e.g., see King, 2005; Johnson, 2016; Adolph, 2013) and experience similar training, visit similar seminars, *inter alia*, it causes central bankers to communicate in a similar manner. Moreover, it could be that additional independence events which we do not account for in our main analysis have a significant impact on our results. To test this, instead of using the largest change in independence of a central bank as treatment, we use the first change only, dropping countries from the dataset once a second change in independence occurs. The results are similar (see the third column of Figure 2.13).

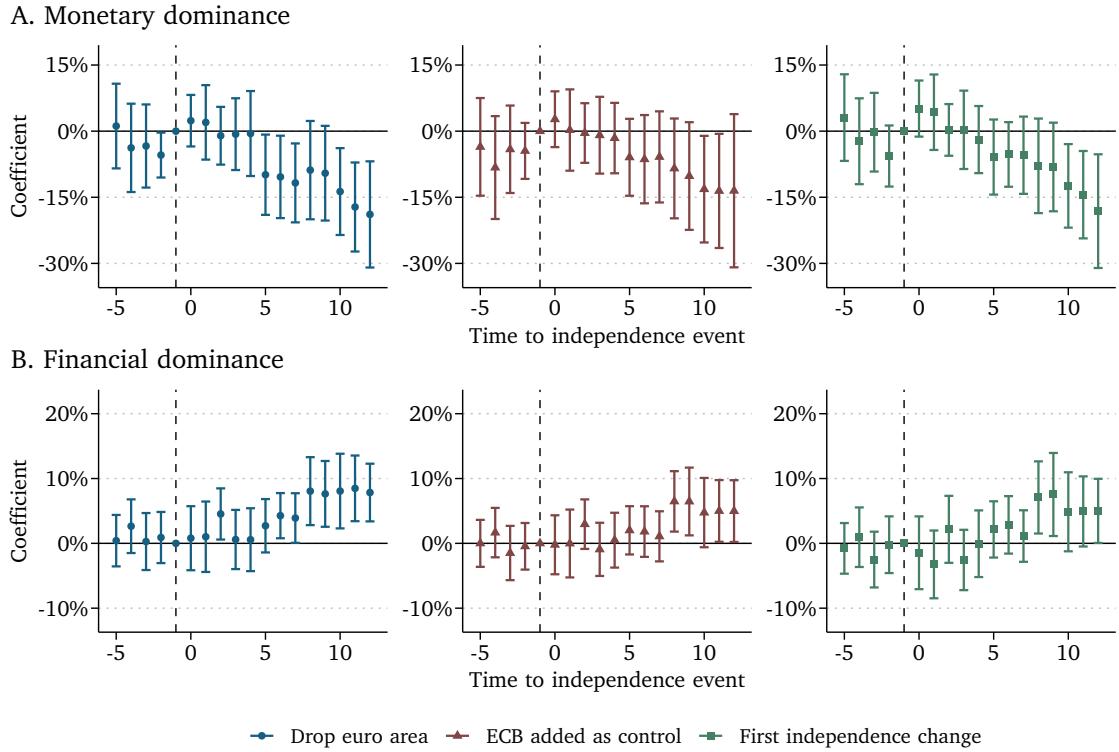


Figure 2.13: Alternative estimation samples

Note: The panels show event studies estimated using the two-way fixed effects equation (2.2). The vertical bars represent the 95% confidence intervals for the estimated coefficients. In the left column, we drop all the current euro area countries. In the middle column, we add the ECB speeches to the control group. In the right column, we modify the treatment indicator to turn on when independence changes for the first time within the event window. If two or more treatments occur, all speeches of the country are dropped, starting from the year of the second treatment.

Fourth, it has been established that independent central banks are more transparent in their communications (Dincer and Eichengreen, 2018; Horvath and Katuscakova, 2016; Liedorp et al., 2013) and have more accountability requirements to certain audiences. For instance, less independent central banks are more political accountable and more independent central banks are more technical accountable leading them to address different audiences in their communication (see Eijffinger and Hoeberichts, 2002; Elgie, 1998). We also know that central banks use communication as a tool to prepare markets for upcoming decisions (Ehrmann and Fratzscher, 2007) which is a more credible tool for

independent central banks. Thus, if increased independence resulted in communication that is more targeted toward financial markets, the observed effect may be a mere shift towards financial topics. To examine this possibility, we conduct two analyses detailed in Appendix B4.6. First, we estimate event studies with audience indicators as the dependent variable, which indicate that audiences change little following increases in CBI. Moreover, a subsample analysis confirms that our main effects also hold within each audience. Taken together, these findings rule out that our findings are driven by differences in audiences or more transparency leading to different types of communication.

Fifth, Calvo et al. (2018) and Masciandaro and Romelli (2018) have shown that financial oversight has become a greater responsibility of central banks post-GFC, namely, central banks gained more micro-prudential responsibility for banks and insurance companies, alongside new macro-prudential and resolution functions. To ensure that our findings are not merely capturing a change in the central bank's supervisory powers that coincided with increases in CBI, we conduct a placebo event study. Specifically, we re-estimate our main equation, replacing CBI changes with changes in banking supervision as the event. For this, we rely on two data sources: (i) Masciandaro and Romelli (2018), and (ii) subcomponent III of the policy dimension in the CBIE index from Romelli (2024), which indicates whether the central bank oversees the banking sector. Table 2.4 presents the aggregated treatment effect on financial dominance from this placebo analysis. We find that neither increases nor decreases in banking supervision at the central bank are significantly associated with our measure of financial dominance, reinforcing our main argument that the observed effects are due to increases in CBI.

	Two-way fixed effects	Gardner et al. (2024)
Increase		
Masciandaro & Romelli (2018)	-0.0297 (0.0188)	-0.0122 (0.0145)
CBI Policy Q3 (Romelli, 2024)	0.0093 (0.0159)	0.0017 (0.0126)
Decrease		
Masciandaro & Romelli (2018)	-0.0083 (0.0338)	0.0195 (0.0272)
CBI Policy Q3 (Romelli, 2024)	0.0026 (0.0326)	0.0323 (0.0261)

Table 2.4: Aggregated event-study estimates of supervision changes

Note: Stars indicate significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The table presents the effects of banking supervision changes using our main event study approach (2.2), estimated with both the two-way fixed effects specification and the approach of Gardner et al. (2024) and subsequently aggregated into a single treatment effect coefficient. The table contains separate estimates for increases and decreases of supervision at the central bank. Clustered standard errors at the country level are reported in brackets.

2.4.6 Robustness checks

We now assess the robustness of our results. First, the main threats to identification of our event study specification are violations of the parallel trends assumption. In Appendix B4.1 we allow for partial deviations from the parallel trends assumption. Firstly, linear country-specific trends permit varying trends between treated and untreated observations, provided they are linear and remain unchanged post-treatment (Miller, 2023). Secondly, we introduce a specification with control variables which relies on a weaker conditional parallel trends assumption. Both specifications lead to quantitatively similar results in line with our main specification (see Figure B4.1).

Second, we allow for more flexibility in the definition of our treatment indicator. Our main specification restricts independence events to increases that are uniform in their intensity and occur at most once per country. In Appendix B4.2 we present two-way

fixed effects estimates that allow for full flexibility regarding the number of treatments, the intensity and their sign. While coefficient estimates vary because of the different unit of measurement, the observed patterns remain the same (see Figure B4.2).

Third, Appendix B4.3 presents alternative event study estimators that employ different approaches to address treatment effect heterogeneity. These include the imputation based estimator of Borusyak et al. (2024), the approach of Sun and Abraham (2021) which relies on the aggregation of year and cohort specific effects, and the stacked difference-in-differences estimator used by Cengiz et al. (2019) which splits the dataset into non-staggered sub-experiments relying only on never treated observation for identification. Building on this, Appendix B4.4 compares estimates of the overall effect size including our aggregation of the event study specification (2.2), a static difference-in-differences specification and the previously mentioned heterogeneity robust estimators with the addition of Callaway and Sant'Anna (2021). Despite slight variations in effect sizes among both the event-studies and single-coefficient estimates, they confirm our main results.

Fourth, we also test the choices made in constructing our sample. In Appendix B4.5, we discuss several sample modifications, including dropping observations for which we cannot observe the full event window, pre-dating independence changes by one year to allow for anticipation and restricting the sample to central banks with independence changes. Our analysis suggests that our results do not depend on these variations (see Figure B4.5). We also report specifications that define treatments based on the sub-dimensions of the overall CBI score. Effect sizes depend little on the choice of CBI indicator as major changes to a central bank's institutional setup typically affect multiple dimensions of CBI (see Figure B4.4).

Fifth, we assess the sensitivity of our results to the construction of the CBI measure. In Appendix B4.7, we replicate our main TWFE event study specification using the

‘LVAU’ indicator from Garriga (2025a), which is the average of the four CBI dimensions originally conceptualised by Cukierman et al. (1992). Although this implies a narrower definition of CBI compared to our main dataset, we find a generally strong correlation between the CBI measures (see Figure B4.8). This also translates to largely comparable estimates with regards to pre-trends and dynamic treatment effects (see Figure B4.10). Significance tests continue to reject the null of no treatment effect for both financial and monetary dominance, underscoring that our main findings are robust to alternative constructions of the CBI measure.

Last, we introduce a falsification test to confirm that the observed effects stem from actual changes in communication, rather than being an artifact of our dataset construction or the event study setup. We implement a randomisation procedure similar to Miller et al. (2021) whereby we randomise treatment status across time and central banks, reflecting the design-based uncertainty inherent in the treatment assignment, while keeping the number of independence increases to the actual number of events in our sample. As expected, the placebo estimates of both the aggregated coefficient, as well as all of the leads and lags, are centred around zero. Our estimated overall effects lie in the extreme tails of the placebo distributions, indicating that the probability of obtaining such effects under the assumption of a zero effect is less than 1 percent (see Figures B4.11, B4.12 and Appendix B4.8 for more details).

2.5 Conclusion

This chapter has examined how the previously unconsidered variable of central bank independence shapes monetary policy communication. We have argued and shown that the effect of CBI occurs through the presence of two channels: a policy pressure and a reputation channel, i.e., central banks have to respond to changed policy pressures

to uphold their reputation of independence. This results in a trade-off when central banks become more independent: from a focus on central banks mainly addressing price stability in their monetary policy communication to increased levels of financial pressures. The findings mainly show effects for central banks in democratic and advanced economy countries. Our findings are not explained by common explanations including overall increases over time in communication, singular events such as the GFC, the adoption of the euro and establishment of the SSM, the effect being driven by the first independence event of a central bank in a country, changes in target audiences and changes in banking supervision mandates.

Our findings motivate us to reflect more broadly on the limits and constraints of delegating independent central banks the task of narrowly focusing on price stability. In this chapter, central bank communication has enabled us to examine these issues by offering the ability to systematically detect these pressures and policy-linkages since, as we show, central banks have to keep their reputation intact by communicating about these issues. First, we show that independence from fiscal authorities does not imply independence from financial markets. This supports previous work suggesting that this was the case since the financial crisis (e.g., see Diessner and Lisi, 2020; Gabor, 2016; Mabbett and Schelkle, 2019). This chapter, however, shows that this is a consistent trend over time and not per se a post-financial crisis phenomenon. Second, more independence does not imply that central banks are more focused on their narrow conduct of price stability through monetary policy. Rather, we find that central banks discuss financial concerns and pressures, implying that a narrow *de jure* mandate does not limit *de facto* what keeps central banks occupied. The underlying inter-linkages between various policy areas and monetary policy which come to the foreground with increased independence, are an important avenue for future research to ensure that central banks do not remain overburdened by being the ‘only actor in town’ (e.g., El-Erian, 2016). Research should,

again, especially focus on this being a general case, beyond only being driven by the GFC as is often argued.

This chapter also raises several other avenues for further research. While this chapter has looked at the channels which change communication when more independence is given, it has not looked at the consequences of this communication. For instance, does being transparent about policy constraints on monetary policy lead to higher trust and credibility of central banks? And does communication about constraints help in elevating policy pressures? Previous research has mixed findings about whether transparency aids trust in the institutions (e.g., see Haldane and McMahon, 2018; Horvath and Katuscakova, 2016; Hwang et al., 2023). Moreover, we examine limited sources of heterogeneous effects and scope conditions. Further research can tease out additional heterogeneity among countries (e.g., the impact of government ideology, the electoral system type, the timing of election cycles and background and training of central bankers), as previous research has shown that these aspects also matter for the monetary policy conduct (see earlier work by Beck (1987); Grier (1989) and later work by Oriola, 2023; Hix et al., 2007; Adolph, 2013; Belke and Potrafke, 2012; Ferris, 2008). Further research can also look into the extent to which the communication is driven by the agency of individual central banks in contrast to the broader trends we identify. Case studies, interviews and small-n comparisons regarding communication decision-making can aid to further examine the reasons of why these variations occur.

Chapter 3

Who Sets the Agenda? The Role of National Central Banks in the Eurosystem

Abstract

It has been well-established that central bank policy agendas are shaped by the spread of ideas, individual governors' agency and economic and political pressures. However, in the case of the European Central Bank (ECB) the influence of the multi-level set-up of the Eurosystem consisting of both the ECB and National Central Banks (NCBs) is often not considered. How does this peculiar institutional setup shape the agenda? This study argues that NCBs act as intermediaries, channeling national and public priorities to the ECB level. Using a transformer model, alongside sequence and cross-sectional time-series analyses of ECB and NCB speeches from 1997 to 2024, I find that the ECB agenda and issue-responsiveness vis-a-vis 'new' topics are driven primarily by NCBs. By revealing how NCBs shape what and when the ECB talks about certain topics, I also contribute more broadly to implications of the multi-level structure of the EU as well as how independent central banks and international organisations respond to outside pressures.

3.1 Introduction

Independent central banks were established to protect monetary policymakers from political pressures of inflationary policies driven by electoral cycles (e.g., Giavazzi and Pagano, 1988). Designed to ‘tie hands’ by delegation, independent central banks help solve the problem of time inconsistency, as noted in Barro and Gordon (1983b), by creating a credible commitment to low inflation over short-term economic gains. Historically, central banks were tasked with narrow mandates, often limited to price stability. However, in recent decades, their (informal) agendas have expanded, encompassing issues such as financial stability, climate change, digital currencies, banking supervision and fiscal policy. Even the European Central Bank (ECB), one of the most independent central banks with a limited mandate, engages in these broader policy areas. What drives this expansion of their agenda and continued engagement with a variety of topics?

Recent studies reveal that central banks respond to various interests, actor constellations, public pressures, ideas and crises in setting their agendas and addressing certain topics (Schulz, 2017; van ’t Klooster, 2022; Tucker, 2018; Braun, 2016; Moschella et al., 2020; Cross et al., 2024; Holmes, 2013). Agenda-setting in these studies involves the translation of both individual-level behaviours and macro-level changes into agenda issues of central banks. However, much of the literature overlooks how institutional aspects influence this (e.g., see Leek and Bischl, 2025).

This paper examines one such aspect, namely, the multi-level set-up of the Eurosystem. The Eurosystem forms a currency area consisting of the ECB and, since 2023, 20 National Central Banks (NCBs). The Eurosystem has been tasked to focus on supranational objectives without considering domestic interests. In comparison to other central banks, the ECB is formally the most independent central bank in the world and only votes with consensus in highly secret deliberations (ECB, 2022). Yet, the NCBs have seats and voting

rights in the ECB Governing Council together with the six-headed Executive Board (EB). NCB chairs are nationally appointed and functionally and socially embedded in the Member States. Moreover, recent research has shown that national perspectives are raised in the Council (Hayo and Méon, 2013). Following from this, this paper examines how the Eurosystem's institutional structure shapes the ECB agenda, exploring (i) how key topics are introduced onto the ECB's agenda, and (ii) how the ECB's issue-responsiveness in topics operates within this multi-level setup once it has reached the policy-agenda.

I focus in particular on the role of NCBs as intermediaries that transmit public and national priorities to the ECB. I argue that the ECB draws on the informational advantages of NCBs and uses their agenda and issue-responsiveness in speeches as signals in response to domestic pressures. To test this, I use central bank communication which is not just 'cheap talk' but is argued to set political directions and lead to actual responses (Cross and Greene, 2020; Blinder et al., 2008; Hayo and Méon, 2013; Moschella and Pinto, 2019). This paper defines ECB agenda-setting as the introduction of key topics into the ECB's public communications for the first time, typically through a salient speech or high-level statement. This first articulation functions as an act of prioritisation that shapes the institutional agenda. Issue-responsiveness, by contrast, refers to the subsequent sequencing and recurrence of those topics, capturing how the ECB responds to evolving conditions and pressures once the topic has entered the agenda. Using a transformer model (BERTopic of Grootendorst, 2022), I operationalise agenda-setting and responsiveness by determining the topics of the speeches of the EB and the 'big 5' (Germany, France, Italy, Spain and the Netherlands) NCB speeches from 1997 to 2024 ($n=4659$). I employ sequence analysis (first-mover graphs and Markov chains) to examine how issues are brought onto the central bank agenda and subsequently examine topic specific short-term transitions between central banks. I find that NCBs are crucial in setting 'new' issues on the agenda. Next, I examine responsiveness to societal and

economic pressures using cross-sectional time-series analysis. I find mixed evidence for the ECB responding directly to these pressures. To further explore this, I examine scope conditions and find that responsiveness is higher in salient periods and when coalitions of NCBs are formed. However, responsiveness is lower after large scale events such as Sintra. I rule out the alternative explanation of responsiveness being driven by issue-linkages but do find that contextual factors influence within topic responsiveness.

I make three contributions. First, I contribute to the broader research studying agenda-setting and responsiveness in technocratic institutions and international organisations. This paper is one of the first to offer empirical evidence of institutional impacts on agenda-setting and issue-responsiveness for the ECB, one of the most independent central banks in the world. The ECB, as a supranational institution, is more isolated from domestic politics, domestic media and voter attention than other (national) central banks and thus offers the least-likely case to see these influences. I draw on debates which argue that publics have become a target audience of central banks and that salience has increased for central banks (e.g., see Blinder et al., 2022). Unlike standard theories of delegation, I expect central banks to care about the legitimacy of their institution. I argue that being responsive to these new issues is not considered as going against their independence but, rather, as a functional form of responsiveness needed to show awareness to broader societal changes increasing the central bank's legitimacy. Specifically, due to a lack of other signaling devices, speeches can be used to signal their opposition to publics when voting cannot be used by national central bank governors. With this, I also test whether insights from the legislative politics literature hold for technocratic institutions (e.g., Herzog and Benoit, 2015).

Second, I show that the role of NCBs as intermediaries between the public and the ECB within the Eurosystem is key in understanding the ECB's dynamic independence. I thereby contribute to the literature that the ECB adjusts the topics it discusses under public pressure

(Moschella et al., 2020) and that economic and domestic political considerations matter for communication disagreements among the Governing Council members (Moschella and Diodati, 2020). More broadly, these findings also display the complex multi-level set-up common to other EU institutions which have been studied more frequently in the context of responsiveness (e.g., Meijers et al., 2019). Specifically, this study shows that the Eurosystem, seen as the most technocratic of these EU institutions and separated the most from the member states, even shows limited issue-responsiveness.

Third, methodologically, I introduce an innovative application of Markov Chains applied to topic modeling to uncover responsiveness on certain topics between central banks. Markov chains can be useful for determining responsiveness but are, thus far, very sparsely used in political science research and when used it is mainly in the context of identifying political regime changes (Klimm, 2022; Gleditsch and Ward, 2006). I apply this approach specifically to demonstrate that certain central banks are more likely to follow one another, i.e., are directly responsive.

3.2 Central bank agenda-setting and issue-responsiveness

Central banks have traditionally been viewed as stable non-majoritarian entities designed to shield monetary policy from political pressures. This perspective aligns with early political economy theories that promoted central bank independence (CBI) as a safeguard against the time-inconsistency problem – where governments might pursue short-term, inflationary policies to achieve electoral gains at the expense of long-term economic stability (e.g., Barro and Gordon, 1983b; Giavazzi and Pagano, 1988). This technocratic view of central banks, which assumes limited responsiveness to political or public interests, frames central banks as primarily focused on their narrow mandates with minimal

interaction with broader societal or political dynamics (e.g., Cukierman et al., 1992).

However, scholars have increasingly argued that central banks are not purely technocratic actors. Examples include Adolph (2013)'s work on the 'shadow principals' which shows that career theories of central banker behaviour explains differences in various economic decisions to outside pressures. Another example is the work of Kirshner (2018) on the 'inescapable' politics of money, which argues that monetary phenomena are always and everywhere political. Throughout his book, he and others argue that, rather than an economic logic, political factors are essential to explain fundamental choices about money. In line with this, Moschella (2024) argues that the political nature of central banks lies at the heart of central bank transformations due to central banks having to manage their reputation among their audiences.

Taking as a starting point that central banks are political in their nature, recent studies have expanded the scope of central bank research into more traditional political science research to account for agenda-setting and issue-responsiveness. Policy agendas in this context can be defined as certain issues gaining the attention of an institution over time (Baumgartner and Jones, 1993). The concept of responsiveness is most commonly used in political science to study the relationship between majoritarian institutions and those they represent (Eulau and Karps, 1977) or between non-majoritarian institutions and political or public pressures (e.g., Blauberger et al., 2018). Responsiveness often refers to policy congruence of aggregated policy output and aggregated public opinion (e.g., Soroka and Wlezien, 2010). However, the focus in this paper is on issue-responsiveness in particular which regularly refers to whether politicians respond to the issue priorities of the public (Jones and Baumgartner, 2004; Neundorf and Adams, 2018). And if there is responsiveness, whether or not they reflect the policy preferences on these issues (Stimson, 1991; Soroka and Wlezien, 2010). Unlike policy responsiveness, which entails concrete policy adjustments, issue-responsiveness is primarily about whether an institution signals

awareness of an issue and incorporates it into its communications (Baumgartner and Jones, 2009; Green-Pedersen and Walgrave, 2014). This research adjusts these broad concepts to central bank issue-responsiveness. I define central bank issue-responsiveness very broadly as the process by which central banks choose which topics to prioritise in their communication based on external pressures. Central banks being issue-responsive is not considered as going against their independence or being pressured into these issues by NCBs, but, rather, as a functional form of responsiveness needed to show awareness to broader societal changes increasing the central bank's legitimacy and, therefore, reducing potential challenges to its autonomy (e.g., also see Baker et al., 2024). In this sense, a certain degree of responsiveness to 'new' issues will not 'punish' the ECB (e.g., in terms of reputation, see Moschella and Pinto, 2019; Moschella, 2024).

It has already been found that independent central banks have recently begun incorporating issues such as climate change, financial stability and digital currency innovation onto their agenda, reflecting an expansion beyond their traditional narrow mandates. For instance, the post-crisis consensus in the politics of central banking literature has been that monetary policy is also driven by financial stability concerns (Braun, 2016; Mabbett and Schelkle, 2019; Diessner and Lisi, 2020). Another topic that recently reached the agenda (not without a battle) is climate (Siderius, 2023; Deyris, 2023). Recent studies have also revealed that central banks respond to various interests, actor constellations, public pressures, ideas and crises in setting their policy agendas. Examples include Schulz (2017) who shows that the ECB is driven in its agenda by a battle of ideas rather than conflicts of interest between various member states. A similar argument is made for the case of macro-prudential policy by Thiemann (2019). van 't Klooster (2022) argues that ideational changes accelerated by recent crises have led to a paradigm change within the agenda of central banks. Research by Cross et al. (2024) has shown that within the ECB, individual behaviour shapes institutional level outcomes

such as the policy-agenda. Jointly, this body of work reflects a nuanced understanding of central banks as institutions deeply embedded within the socio-political context, responsive to both internal dynamics and external pressures.

3.3 The Eurosystem set-up

The case of the ECB offers a distinctive application of agenda-setting and responsiveness of central banks due to its unique multi-level set-up. Unlike national central banks, which operate under single sovereign authorities, the Eurosystem includes both the supranational ECB and 20 NCBs of the Eurozone member states. Decisions are made by the ECB's Governing Council, comprising the ECB's Executive Board and the NCB governors, which plays a pivotal role in shaping the ECB's policy agenda. Although the governors wear a 'supranational hat' and have to act in the eurozone-wide interest, the presence of NCB governors in the Governing Council introduces channels through which national concerns can influence ECB discussions, making the Eurosystem more complex than other central banks in terms of responsiveness. Scholars such as Hayo and Méon (2013) and Meade and Sheets (2005) have shown that NCB governors often bring national economic data and specific domestic concerns into ECB deliberations. National governors, appointed by domestic political authorities, are often embedded within their respective national policy contexts, thereby acting as liaisons between national interests and the ECB's broader mandate.

Besides voting in the Governing Council, the NCBs and their governors have various other roles in the Eurosystem. For instance, the implementation of monetary policy is based on the principle of subsidiarity, under which the operations of the Eurosystem are normally carried out by the NCBs. Other tasks of the NCBs are to ensure the monetary and financial stability of the system through operations with credit institutions, open

market operations, standing facilities and management of required reserves. Some NCBs also act as a supervisory authority under the Single Supervisory Mechanism (SSM). The NCBs have relatively large research departments and extensive operational experience in financial and banking markets, publish statistics, conduct research and often provide advice to governmental organisations. The number of staff in large NCBs are also often independently larger than the corresponding departments at the ECB¹. The economic analyses carried out by the NCB staff are thus an important source of information to the Eurosystem's technical committees and working groups in which the NCBs and ECB come together. Moreover, the competence to address country specific political economy circumstances needed for effective policy will partly depend on the ability of the ECB to draw on the staff at the NCBs.

Besides a pure informational and technical need, responsiveness to certain 'new' issues raised by NCBs can be used as a channel to defend the ECB's autonomy and legitimacy by showing social awareness. Certain longer-term issues which are salient in the member state publics can be brought up by NCBs, such as climate, inequality or geopolitics, while they are not part of the original 'script' of the ECB. NCB governors are in the position to do so since the member states publics and governors often appear in national media.

In sum, NCBs provide an important informational and public response function. Embedded in their national contexts, they are thus able to channel domestic concerns and priorities upward to the ECB, introducing and discussing topics that resonate with national interests. Since NCBs are more closely connected to national priorities and directly affected by local economic conditions, they are well-positioned to highlight issues reflecting these concerns. As first movers, NCBs can bring attention to matters that may

¹Estimates are that the ECB employs 4,500 employees of which about half work on SSM related tasks while the larger NCBs have 8,000-10,000 employees doing the same work. For instance, the Bundesbank in 2022 had 10,294 core staff full-time equivalents according to their own statistics and the Banque de France had 12,125 in 2015 according to Statista.

later be adopted by the ECB, especially when these topics have high domestic relevance. Therefore, I expect that:

Hypothesis 3.1. *NCBs are more likely than the ECB to be first movers in setting the agenda on ‘new’ topics.*

The ECB pays attention to the public since they are dependent on public support to operate and survive over time (Tucker, 2018). High public trust helps anchor inflation expectations (Christelis et al., 2020) and generates acceptance of central bank independence among the public at large (van der Cruijsen and Samarina, 2023). NCBs can be essential for creating this trust since they are closer to their national publics. As Delors famously said in 1992: ‘Not all Germans believe in God, but they all believe in the Bundesbank’. Public pressure has also increased over time. Similar to the EU in general, the ECB has also been increasingly salient in national politics, arguably ending the ‘permissive consensus’. The ECB has recently been increasingly involved in the political realm (e.g., through the Troika and directly influencing national policies by sending letters, by undertaking quasi-fiscal policy through their unconventional monetary policies, introducing the Outright Monetary Transactions program and *de facto* acting as lender of last resort (de Grauwe, 2020)) which, in turn, resulted in increased politicisation. Schneider (2018) argues, regarding the EU Council, that governments undertake ‘electioneering’ in the EU, which implies that EU representatives demonstrate responsiveness. She argues that this can also be applied to EU institutions more generally if they are being politicised. In the case of politicisation, the ECB will feel pressured to publicly respond to these pressures, especially asymmetrically to the countries with more pressing publics and concerns. This is in line with the general finding that issue-responsiveness of governments or presidents is stronger the more salient an issue becomes (Soroka and Wlezien, 2009; Canes-Wrone and Shotts, 2004).

Issues with high public salience create greater visibility and demand attention, particularly when they align with the ECB's objectives. I, therefore, expect that the ECB adjusts its agenda in response to heightened public salience, which is consistent with previous findings that central banks use communication as a tool to signal attentiveness to societal concerns (e.g., Alexandrova et al., 2020). Leading overall to the following hypothesis:

Hypothesis 3.2. *The ECB's responsiveness to NCBs increases under conditions of high national public salience.*

The ECB can also be influenced by economic pressures. The eurozone consists of 20 diverse countries with diverse political economies and diverse needs. This is a well-known issue in the economics and EU political economy literature. For instance, discussion among the supporters and critics of the Optimal Currency Area (OCA) have shown that the eurozone falls short in many aspects of what an optimal currency area constitutes (de Grauwe, 2020). Especially since the eurozone has a wide diversity of institutional variation in terms of political economies, which also has implications for the currency area. For instance, Coordinated Market Economies (CMEs) and Liberal Market Economies (LMEs) have different ways of adjusting to demand shocks. LMEs can use discretionary monetary and fiscal policies while CMEs have more difficulties (Soskice, 2007). The Eurozone countries also have different growth models, labour markets, pension systems and mortgage structures which cause different reactions at different speeds to a one-size-fits-all monetary policy (de Grauwe, 2020). Certain information provided and topics raised by NCBs may portray some 'tactical' projections which the ECB does not always have the capacity to check. Given the diverse political economies the ECB covers, each with their own needs, the NCBs can try to asymmetrically influence and pressure the ECB to publicly announce and commit to the optimal policies for them instead of eurozone

averages.

Thus, when NCBs face significant economic pressures domestically, they are more likely to push these concerns onto the ECB agenda. Therefore, I expect that the ECB will be responsive to these conditions, especially given that domestic economic stability across the Eurozone aligns with the ECB's broader mandate. Thus, I hypothesise:

Hypothesis 3.3. *The ECB's responsiveness to a NCB increases when economic pressures are higher within individual member states.*

3.4 Data and methodology

In this section I describe the data, measurement strategy and variable operationalisation to test the three hypotheses.

3.4.1 Speeches-as-data and topic modeling

For the dependent variable, I create a measure of responsiveness and agenda-setting using the topics discussed in governor speeches. The EB and NCB governors often write their speeches in liaison with a communications team and, therefore, represent the institutional topics and narratives they wish to raise. I choose speeches and not Q&A sessions or announcements since governors are freer to choose the topic, format and audience of their speeches, which reveals important information. Using speeches-as-data is common in the central banking literature to study disagreement and preferences of governors under high secrecy and limited public information (e.g., Baerg and Lowe, 2020; Bennani and Neuenkirch, 2017; Moschella and Diodati, 2020). Moreover, as is found in the legislative politics literature, speeches can be used to signal to local constituents when voting behaviour cannot be observed or diverge (Herzog and Benoit, 2015). One possible downside is that depending on the audience, the speeches can be intended to

merely convey information (e.g., lectures given by governors at universities). Although those speeches seem to merely convey information, the preferences in terms of topic choice can still be observed, e.g., some governors choose to talk about structural reforms instead of monetary policy issues.

The speeches dataset covers the establishment of the ECB until March 2024 and are all translated by the Bank of International Settlements into English. The corpus consists of 4,654 speeches of which 2,167 are NCB (Germany, Italy, France, Spain and the Netherlands) speeches and 2487 EB speeches (see Appendix C1 for more information). I include the largest five NCBs in the eurozone and all founding euro members and, therefore, the most important in terms of responsiveness and sufficient speeches to meaningfully detect responsiveness. Moreover, all five NCBs have banking supervision tasks as part of the SSM, ruling out that the communication about certain topics is driven by a difference in NCB mandates.

Figure 3.1 presents the text-as-data workflow from raw speeches to the final dependent variable which I discuss below.

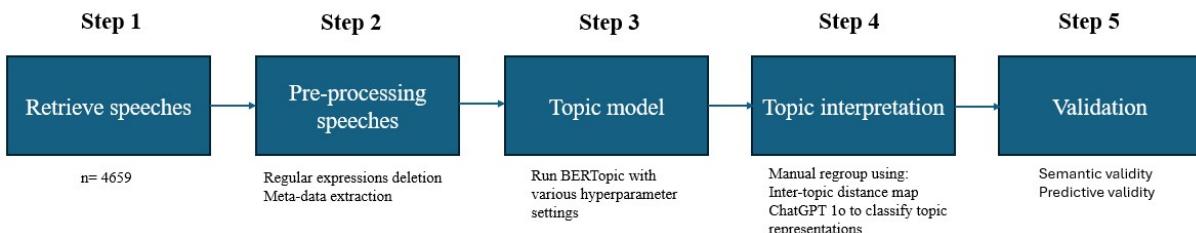


Figure 3.1: Text-as-data workflow steps

Determining the topics of speeches can be done in various ways ranging from hand-coding to unsupervised machine learning. Due to the lack of pre-set topics, the large sample size and the possibility to discover systemically latent trends over time, quantitative text analysis is most appropriate (Grimmer and Stewart, 2013). Since there are no strong priors regarding the topics, the topic modeling is more exploratory and will benefit from

inductively extracting a small number of independent dimensions rather than being confirmatory, which would measure pre-specified concepts. Thus, a fully unsupervised method is most appropriate (Grimmer and Stewart, 2013). Rather than using topic modeling algorithms based on a vector space model and the classical “bag-of-words” assumption which disregards grammar and word order, I use embeddings. Embeddings generally lead to much better results since they capture the context, meaning and semantic relationships between words and text documents.

Very recently, developments made Large Language Models (LLMs) accessible to use by the public. One such LLM is BERTopic (Grootendorst, 2022) which takes advantage of the powerful pre-trained transformer-based language models and employs a clustering technique for topic extraction. The BERTopic model is thus specifically designed for extracting topics from a corpus of text and offers interpretable metrics and, therefore, most preferable for topic modeling tasks as these. The BERTopic framework is completely modular and integrates transformer-based embeddings (I use the “all-MiniLM-L6-v2” embeddings, dimensionality reduction via UMAP and cluster using a c-TF-IDF approach). For further information about the parameters used and the reasons behind choosing certain modules, see Appendix C2.

Before running a joint topic model for all central banks on the speech level, I conducted text pre-processing steps following standard methodology (see Gentzkow et al., 2019) (see Appendix C2). Running a joint topic model at the speech level allows easy interpretation of the same low-dimensional space and increases the number of unique n-grams which helps reduce noisy topics. Following the output of BERTopic, I manually clustered the topics into broader coherent and meaningful topic categories. Although there is the risk of reducing the number of topics too much, I also needed enough observations per topic to ensure statistical power in the cross-sectional time-series interaction terms, so I reduced it to 10 clear categories. To aid manual interpretation, I used two methods: (i) visual tools

(e.g., static and interactive intertopic distance maps, representative word clouds and c-TF_IDF scores) and (ii) I prompt ChatGPT 1o to classify the topic representations within the 10 manually determined topics (see Appendix C2 for more information). ChatGPT was prompted to assign BERTopic-generated topics to these higher-level categories based on their representative keywords and descriptive summaries. This step provided an additional layer of validation and is recommended by, for instance, Grootendorst (2022).

Descriptive figures can used as a means of validation (following Quinn et al., 2010). The top scoring words for each topic demonstrate that the topics have a coherent meaning (semantic validity) (see Appendix C2.3 for topic specific wordclouds and a representative regrouping map). Topic usage also corresponds to external events (predictive validity) showing that spikes in their probability distribution can be matched to relevant political events. Figure 3.2 provides a face validity check using a smoothed local function to show that the topic proportions correspond to expected trends over time. For example, the speeches concerning crisis management and climate have increased over time, in line with expectations. Further, there is heterogeneity in topics discussed across central banks. For instance, France talks more about crisis management and the ECB talks proportionally more about monetary policy and economic indicators. The topic of national economy is an issue mainly ‘owned’ by Spain and Italy which is also unsurprisingly following the Euro crisis dynamics (e.g., the letter of Trichet sent to Italy regarding drastic rebalancing of the economy and Spain).

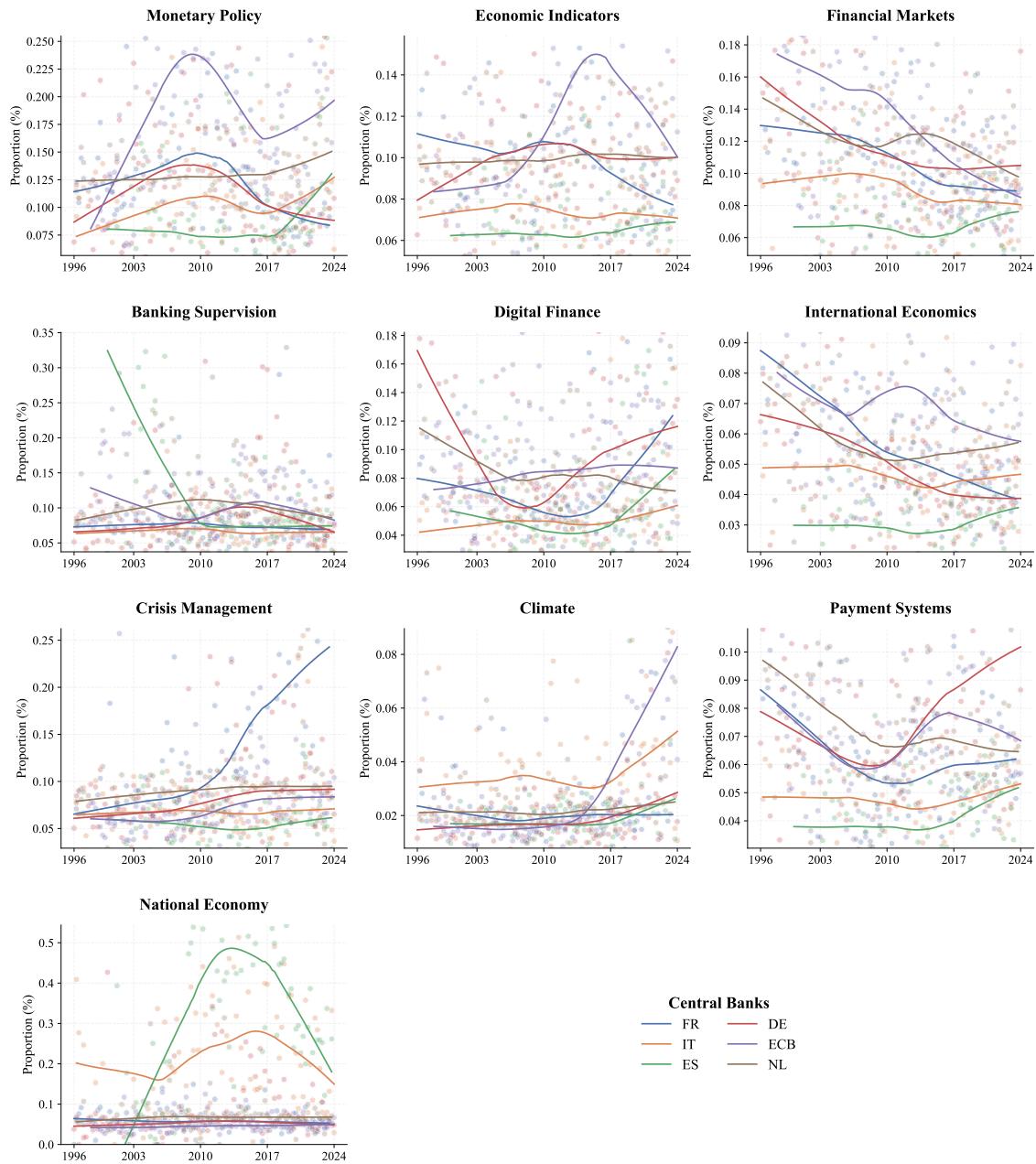


Figure 3.2: Smoothed topic proportions over time

Note: To smooth the topic proportions I use LOWESS.

3.4.2 Empirical strategy

The empirical strategy is broken down into two parts: (i) a sequence analysis part and (ii) a cross-sectional timeseries part. First, I use first-movers graphs to examine how a topic

reaches the agenda. Next, I use a Markov chain approach with rolling 30-day windows to track the probability of central bank responsiveness. This sequence analysis thus allows me to observe direct reactions to discussing certain topics and direct responses between central banks in terms of speech sequence. Second, I use a cross-sectional time series model (with three month and half-year lags) to test whether these short-term sequences accumulate into systematic issue-responsiveness trends. This two-fold strategy allows me to distinguish short-lived reactions from more structural issue-responsiveness, thereby providing a comprehensive picture of how and when governors talk about certain topics. Figure 3.3 provides an overview of the empirical strategy on which I will elaborate below.

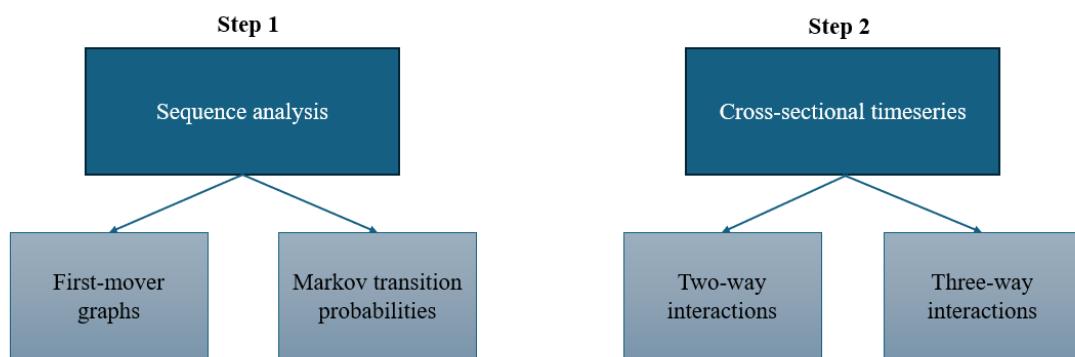


Figure 3.3: Empirical strategy overview

The first-movers are identified as follows. To pass the threshold to be considered a first-mover, a bank needs to talk about the topic in least 25% of the speech and at least three other central banks need to discuss the topic in the subsequent year. These conditions allow me to identify that a topic is one of the main topics in the speech and remains on the agenda and not to detect noise or just mentions without follow-up. To determine the transitions between topics and central banks, I use a discrete first-order Markov transition process. A Markov chain is a type of stochastic process that fulfills the Markov property, i.e., the probability of the next state depends only on the present state. Owing to this property, Markov chains can be useful for determining responsiveness. I

apply this approach specifically to demonstrate that certain central banks are more likely to follow one another, i.e., are directly responsive. For simplicity, I limit myself to the highest topic probability identified by BERTopic in the speech and use the sequence, rather than a continuous time parameter; i.e., I use a discrete time Markov chain.

The transitions between central banks are conditional on the topic and, therefore, defined as follows: let $\mathcal{B} = \{1, 2, \dots, M\}$ denote the set of central banks. At each time t , the central bank issuing a communication is $CB_t \in \mathcal{B}$. For each topic $k \in \mathcal{T}$, I define

$$m_{ab|k} = \text{count}\{t : CB_t = a, CB_{t+1} = b, T_t = k\},$$

where $a, b \in \mathcal{B}$. This measures how often central bank a is immediately followed by central bank b given that the current topic is k . I define the conditional transition probability as:

$$P_{ab|k} = \frac{m_{ab|k}}{\sum_{c \in \mathcal{B}} m_{ac|k}},$$

so that for each fixed k , the matrix $[P_{ab|k}]_{a,b \in \mathcal{B}}$ is row-stochastic, i.e., sums up to 1. This formulation captures the likelihood that, under a specific topic, one central bank's communication will be followed by that of another. To allow for the possibility of central bank transitions change over time, I use time-indexed transition probabilities, which also excludes potential non-stationary effects in the data. Specifically, in the main analysis, I use a 30-day moving average to smooth fluctuations, capture short-term dynamics and ensure that there are sufficient observations for inference. Moreover, as argued by Blinder et al. (2008), taking a 30-day moving average also mimics the monetary policy communication cycles. Appendix C3.2 also examines alternative windows of 7, 14, 60 and 90 days.

The sliding window of length W is implemented as follows: Indexing each window by w , I define

$$m_{ab|k}^{(w)} = \text{count}\{t \in \text{window } w : CB_t = a, CB_{t+1} = b, T_t = k\} \quad \text{and} \quad P_{ab|k}^{(w)} = \frac{m_{ab|k}^{(w)}}{\sum_{c \in \mathcal{B}} m_{ac|k}^{(w)}}.$$

For the second part of the empirical analysis, I employ cross-sectional time series in line with the literature on issue-responsiveness (e.g., Hobolt and Klemmensen, 2008).² In these models, the topic proportions of the Executive Board devoted to the 10 topics serve as the dependent variable; i.e., I do not take the highest probability topic as I do in the previous step. Given the frequency with which central bankers meet, a lag length of three months should be sufficient to capture broader responsiveness to external circumstances. This choice is in line with framing cycle theory and previous research on ECB communication (Müller and Braun, 2021). Robustness tests are also run with half-year lags.

The potential issues associated with the frequency distribution being right-skewed and non-stationarity (e.g., see Appendix C1) are resolved by using a proportional topic index, which I construct as follows: let $I_{b,q,t}$ be the index of topic proportion for: central bank $b \in \{1, 2, \dots, 6\}$, quarter $q \in \{1997_1, 1997_2, \dots, 2024_3\}$ and topic $t \in \{1, 2, \dots, T\}$, where T is the total number of topics.

The index can then be defined as:

$$I_{b,q,t} = \frac{1}{|S_{b,q}|} \sum_{s \in S_{b,q}} p_{s,t}, \quad (\text{Index})$$

where: $S_{b,q}$ is the set of all speeches given by central bank b in quarter q , $|S_{b,q}|$ is the number of speeches in $S_{b,q}$, $p_{s,t}$ is the proportion of topic t in speech s .

²An alternative is to use Vector Autoregression (VAR) models, with the primary objective of identifying long-run relationships by calculating maximum eigenvalue and trace statistics. However, for simplicity of interpretation and to accommodate various interaction terms, I employ cross-sectional time series analysis.

To account for topic-specific effects, I introduce the following dummy variable:

$$D_{p,j} = \begin{cases} 1, & \text{if the observation corresponds to topic } j, \\ 0, & \text{otherwise} \end{cases}$$

for $j = 2, \dots, T$. Using these dummies, I estimate the following models:

Model A is the baseline model to test general issue-responsiveness of the ECB to NCBs (Hypothesis 3.1) which regresses the topic proportion for the Executive Board for topic p at time t on the lagged measure from national central banks for the same topic:

$$EB_{t,p} = \alpha_p + EB_{t-1,p} + \beta_1 NCB_{t-1,p} + \epsilon_{t,p}, \quad (\text{Model A})$$

where $p = 1, \dots, T$ indexes topics, t indexes time, α_p is the topic-specific intercept, β_1 is the common slope for the lagged NCB measure, and $\epsilon_{t,p}$ is the error term.

Model B adds topic-specific slopes for NCB_{t-1} to the baseline specification allowing to test topic heterogeneity in responsiveness between ‘new’ and ‘old’ topics (Hypothesis 1). To allow the effect of NCB_{t-1} to vary across topics, I include interaction terms:

$$EB_{t,p} = \alpha_p + EB_{t-1,p} + \beta_1 NCB_{t-1,p} + \sum_{j=2}^T \beta_{2,j} (D_{p,j} \times NCB_{t-1,p}) + \epsilon_{t,p}. \quad (\text{Model B})$$

Here, for each topic $j \geq 2$, the coefficient $\beta_{2,j}$ captures the differential effect of NCB_{t-1} compared to the baseline.

Model C includes three-way interactions to examine the circumstances (i.e., pressures) under which the ECB is more responsive to NCBs (to test Hypotheses 2 and 3). Let C_t denote the pressure at time t . To allow for both a direct effect of C_t and its moderation

of the effect of NCB_{t-1} in a topic-specific manner, I specify:

$$\begin{aligned}
EB_{t,p} = & \alpha_p + EB_{t-1,p} + \beta_1 NCB_{t-1,p} + \gamma C_t \\
& + \sum_{j=2}^T \beta_{2,j} (D_{p,j} \times NCB_{t-1,p}) \\
& + \sum_{j=2}^T \beta_{3,j} (C_t \times NCB_{t-1,p} \times D_{p,j}) \\
& + \sum_{j=2}^T \beta_{4,j} (C_t \times D_{p,j}) + \epsilon_{t,p}.
\end{aligned} \tag{Model C}$$

In this specification: γ represents the direct effect of the pressure variable C_t , $\beta_{3,j}$ captures the topic-specific moderation of the effect of NCB_{t-1} by C_t , and $\beta_{4,j}$ reflects the topic-specific direct effect of C_t .

In these models, $EB_{t-1,p}$, i.e., the lagged topic proportion of the Executive Board is included to control for the topic persistence which is detected in the Markov chains. $NCB_{t-1,p}$ represents the corresponding lagged topic proportion for the national central banks. The coefficient β_1 captures the degree of responsiveness using the lag from the previous period. As mentioned, to test the rationales underlying responsiveness, interaction terms are included: T denotes topic-specific dummies, and C_t represents proxies for the pressure rationale.

The proxies for pressures are the following (Appendix C1.2 gives further information, lists all variables used, their data sources and summary statistics). Real GDP growth is used as a proxy for economic pressure, because sluggish or negative growth typically signals heightened economic strain, whereas robust growth indicates a more favourable environment. Further, inflation is used to represent inflationary and deflationary pressures. To proxy public pressure, I use two measures. First, I use a question about trust in the ECB from the Eurobarometer. The answer options include: “tend to trust”, “tend to not trust” and “I don’t know”. I subtract “tend to not trust” from “tend to trust” and omit

“I don’t know” (since it does not fluctuate a lot over time). See appendix C1.2.2 for more information. Second, Google trend searches of the term “ECB” are used to assess the salience of the ECB in the various countries. Google trends are often used as proxy for salience in research (Hölzl et al., 2025). The values represent search interest relative to the highest point on the chart for the given region and time and ranges from 0 to 100. See Appendix C1.2.1 for more information on internal and external validity.

Additionally, in all models, I include a range of control variables at $t = 0$ and their interactions with the lags of NCB topic proportions. Specifically, inflation and real GDP growth are included as control variables following the regional Federal Reserve governor literature which uses them to determine central bankers’ reaction functions (e.g., Chappell et al., 2008; Meade and Sheets, 2005; Moschella and Diodati, 2020). Panel-corrected standard errors are employed given the small number of topics and the large number of time points (Beck and Katz, 2006).

The counterfactual to NCB responsiveness can take three forms. First, it can be that NCBs and EB talk about the same topics in the same period without sequentially responding to each other (i.e., only contemporaneous effects). Second, it can be that there is no correlation at all between the topics discussed or, third, it can be that the NCBs are responsive to the ECB, i.e., the reverse relationship. To be clear, the presence of contemporaneous effects and reverse responsiveness do not rule out responsiveness of the ECB to NCBs. Nevertheless, close to contemporaneous effects can be examined in the Markov chain robustness tests with shorter time-windows in Appendix C3. Reverse (temporal) causality is tested by replacing the lags by leads, i.e., whether NCBs respond to the ECB, in Appendix C4.24.

3.5 Results

3.5.1 Sequence analysis

Figure 3.4 displays which central bank first introduced a given topic between 1998 and 2024. In line with the first hypothesis, NCBs appear as first-movers more frequently than the ECB in these emerging topics. The introduction of particular topics by certain NCBs is also in line with expectations from previous research, e.g., the Netherlands leading the discussion on climate. Following an informational logic, one can expect the Dutch central bank to drive the agenda since they have developed and employed sophisticated analytical tools and frameworks for assessing climate-related financial risks, positioning itself as a leading technical authority (Boneva et al., 2022; Siderius, 2023; Deyris, 2023). France leading the discussion on crisis management (e.g., in line with Brunnermeier et al., 2016's discussion of the battle of the euro) also aligns with expectations. See Appendix C3.1 for more details on the development of the topic of crisis management and climate over time. The ECB does introduce two more technical (not *per se* new) topics: economic indicators and payment systems. Curiously, with the exception of international economics, the German Bundesbank, seems to be rather in the background in terms of agenda-setting despite being prominently present in the establishment of the ECB (Kaltenthaler, 2005).

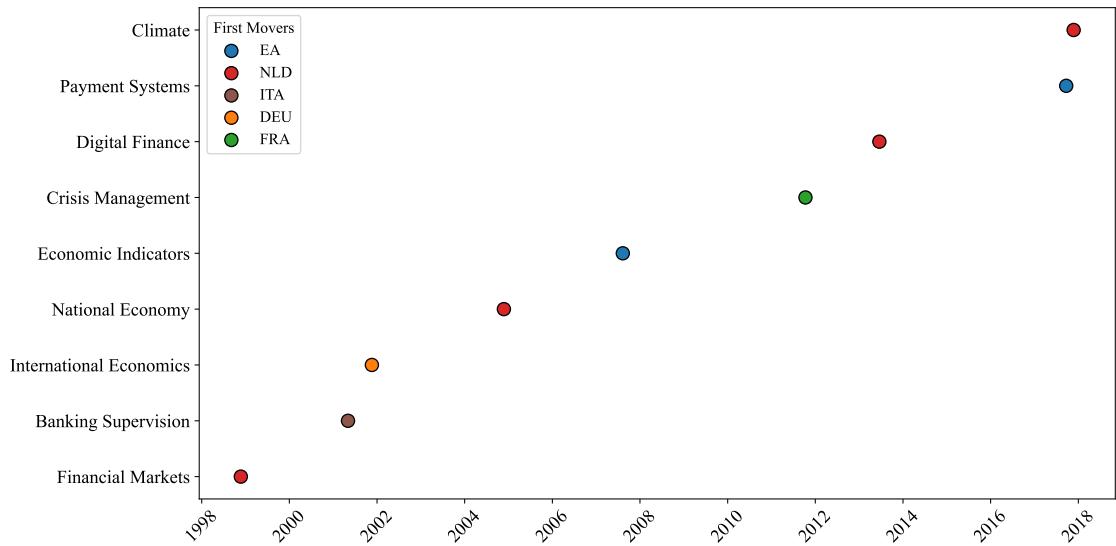


Figure 3.4: First movers on emerging topics

Next, I examine direct responsiveness by examining transitions between central banks conditional on the topic. Figure 3.5 displays how a speech of a given central bank (rows) is followed by another central bank (columns) for each of the topics, within a 30-day window. Overall, one can clearly see that NCBs are first movers within this short-time period. Namely, the ECB (EA) in most topics is responding to the various NCBs (all topics except national economy), again supporting the first hypothesis and responsiveness of the ECB to NCBs. The ECB not responding to national economy discussions shows that the ECB remains impartial regarding certain NCBs and their particular issues. Across the different panels, nevertheless, also noticeable variation emerges per topic. For instance, the ECB is very responsive in terms of economic indicators while the opposite was the case in the 30-day Markov chain results. Thus, rather than just responding directly, the ECB follows-up only after a time lag. Another curious observation is that the ECB is often responding to Germany and the Netherlands in terms of monetary policy. This can indicate that the ECB is either responding to criticisms or that the ECB giving in to certain pressures. This will be further explored in Section 3.5.2. Appendix C3.2

shows similar plots for 7, 14, 60 and 90 day windows, showing that Markov transition results are robust to different time-windows.

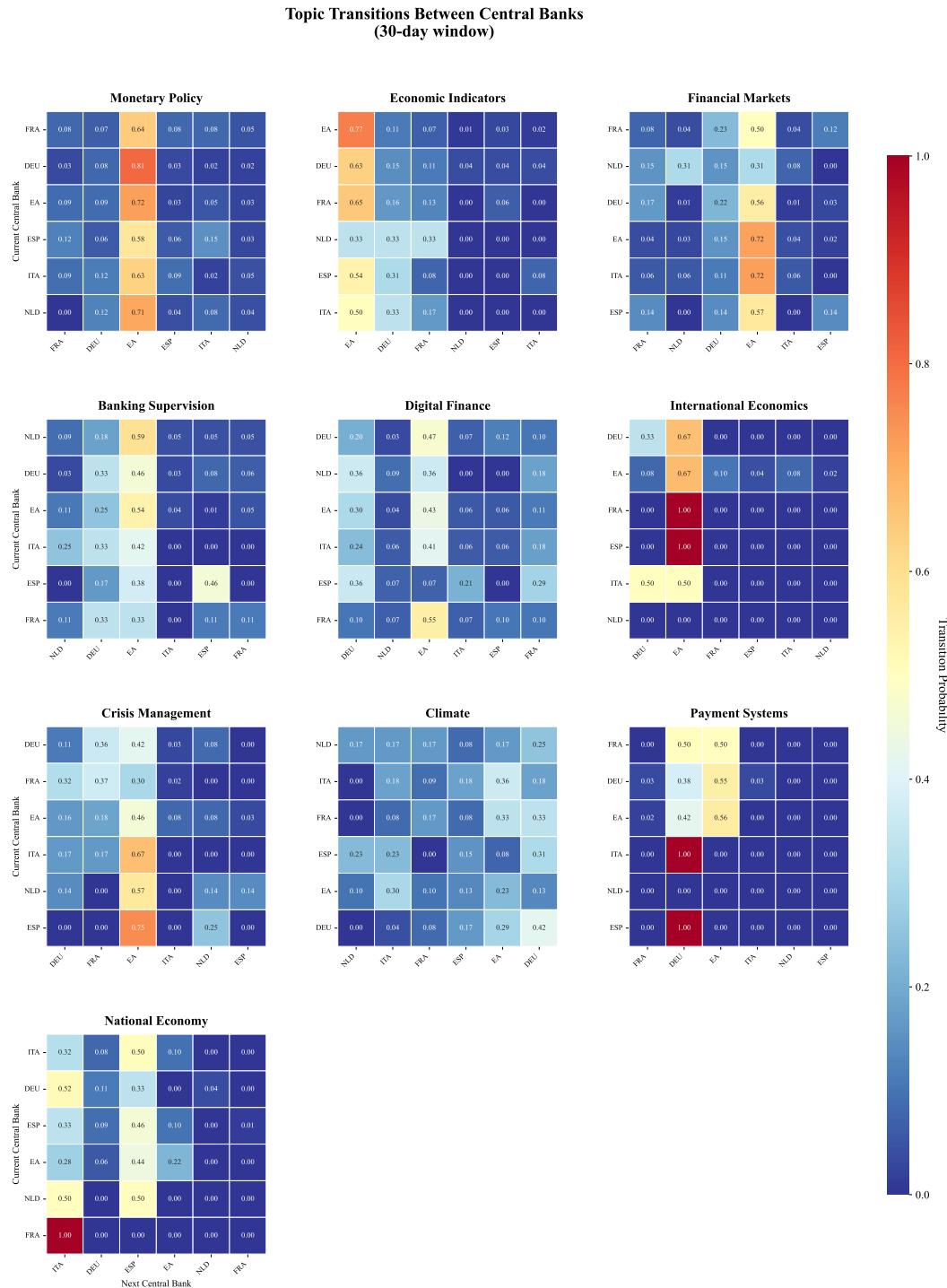


Figure 3.5: Markov Transition Matrix of central banks conditional on the topic

One possible critique could be that these patterns are just artefacts of random sequencing. To rule this out, the robustness of all transition claims was assessed with a speech-level permutation test that randomises topic labels while preserving each bank's speech calendar. The procedure, detailed in Appendix C6.3, shows that the observed matrices deviate from the empirical null at $p < 0.001$, confirming that the patterns are not artefacts of random sequencing.

3.5.2 Cross-sectional timeseries

To further examine trends in responsiveness, Figure 3.6 (based on Model B) depicts the two-way interaction coefficients of the various topics interacted with the NCB lag coefficient (three months). The results again show strong topic heterogeneity. Several striking examples stand out. First, regarding climate, a few NCBs, particularly France and Germany, display coefficients that are both statistically significant and substantial in magnitude. Thus, while the Netherlands was effective in bringing it onto the agenda, the further discourse was led by other countries which shows that introducing topics onto the agenda is not the same as continued agenda-setting. Banking supervision is heavily driven by Italy which is expected given Italy's active role in the Single Supervisory Mechanism. Crisis management is also heavily driven by certain NCBs including the Netherlands, Spain and Italy with several opposing views on how to deal with the crisis management. This latter finding also aligns with the findings of the short-term responsiveness in the previous section since on crisis management there also was responsiveness to the different sides of the debate by in crisis-times being responsive to both Germany and Italy within a 30-day time frame. Certain lags are also negative pointing towards if a NCB discusses a topic, the ECB talks less about it. For instance, in banking supervision, if the Netherlands and France have discussed it already, the ECB refrains from discussing it subsequently. Appendix C4.1 shows the results pre and post crisis, with leads instead of lags and with

half-year instead of quarterly lags. The pre and post crisis results slightly differ but not in a specific direction and the change of the lag structure deems robust.

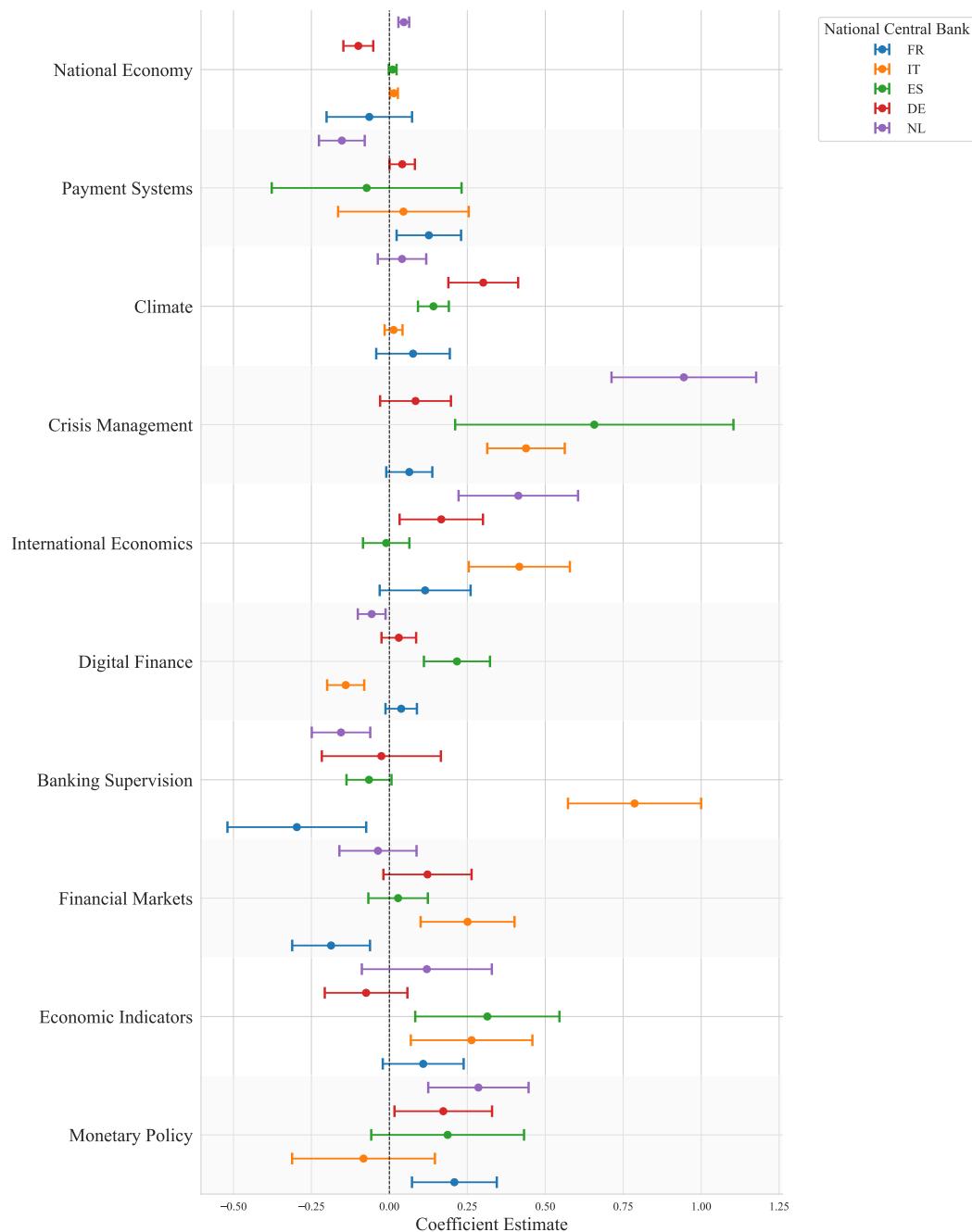


Figure 3.6: Coefficient plot using three-month lags

Note: Panel clustered standard errors are employed. The coefficients include 95% confidence intervals.

Having shown a degree of issue-responsiveness and agenda-setting of the NCBs, I explore whether there is more responsiveness to certain NCBs under certain conditions, i.e., whether the ECB responds to public and economic pressures. Figure 3.7 shows an interaction grid of all the topic, bank and pressure combinations to show responsiveness to public and economic pressures following Model C. The figure shows both the sign and whether the three-way interaction terms are significant using binary dummies for higher or lower than average within country pressures. If public pressure is a driving factor behind responsiveness (Hypothesis 2), there should be significant positive responsiveness observed when the ECB is highly salient and trust is low in the ECB. If economic pressure is a driving factor behind responsiveness (Hypothesis 3), then there should be significant positive responsiveness observed with low growth and high inflation. The interaction grid does not show systematic evidence of either of these being the case. What can be observed is that in most countries (except France), there is positive significant responsiveness to crisis management under economic pressures, offering some preliminary evidence for economic pressure responsiveness (Hypothesis 3). In the same topic, however, one can observe negative responsiveness to public salience (and significant in Italy and the Netherlands), i.e., the ECB responds less to public pressure in this case.

Other topics also give mixed results pointing towards country-specific dynamics being at play. Overall, these country-specific findings can be seen as a sign that the multi-level Eurosystem does selectively respond to certain countries, however, not systematically. Appendix C4.3 shows that the results are robust to measuring pressures as continuous instead of binary and using half-yearly lags instead of quarterly lags.

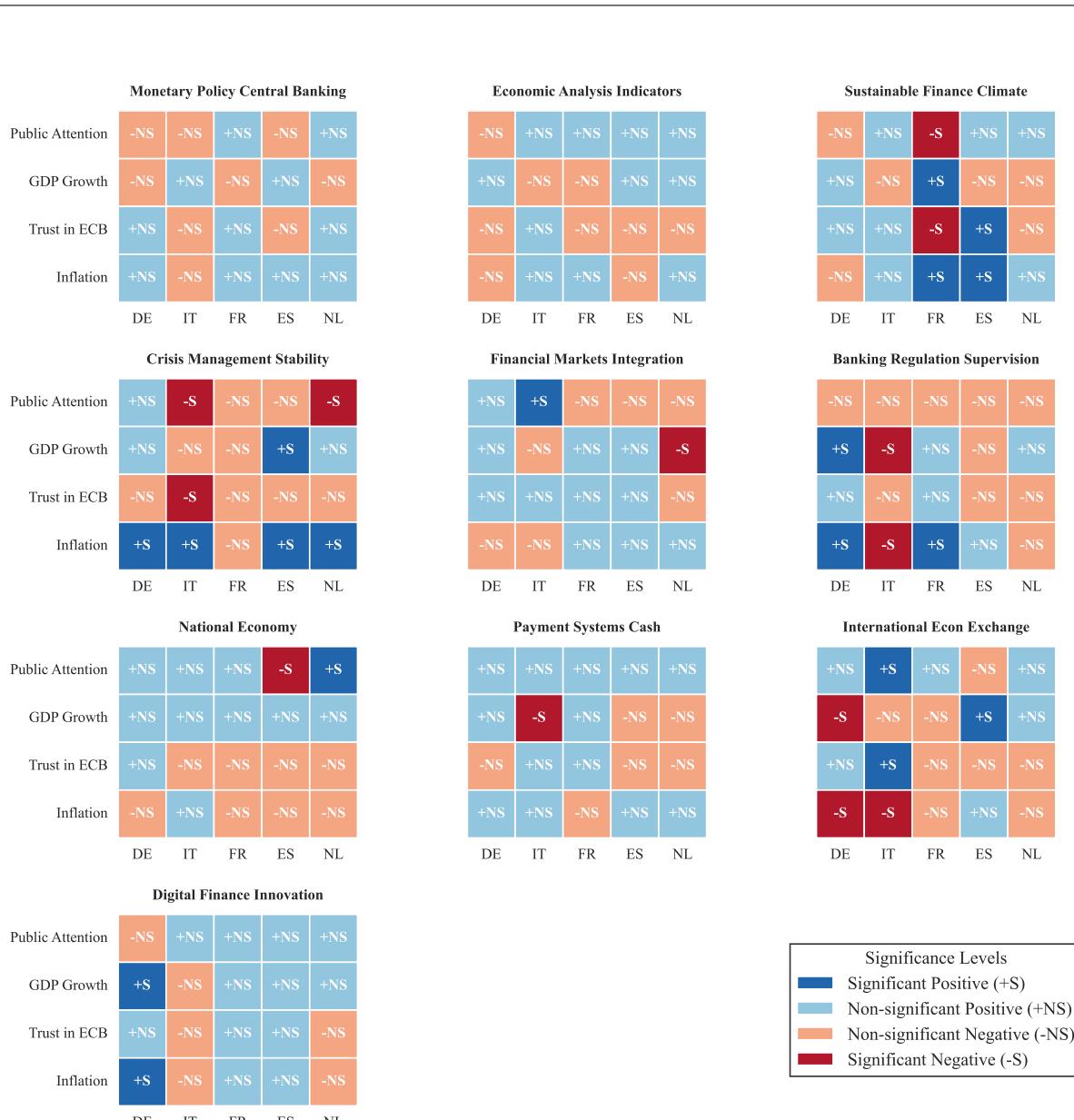


Figure 3.7: Interaction grid of three-way interactions

Note: A p-value of 0.05 is employed for significance.

3.5.3 Heterogeneous effects

Having established mixed support for Hypothesis 2 and 3, I now examine three possible scope conditions. First, I examine whether, under very salient conditions, the responsiveness dynamics change. I take the case of the euro crisis, given that it was the most salient period of the eurozone thus far (Appendix C1 Figure C1.5 depicts a

salience based measure based on FT articles). I subset the dataset for only the euro crisis years (2010-2015) and re-estimate the transitions between central banks conditional on the topic. Figure C5.32 displays that the responsiveness of the ECB to NCBs is even stronger on monetary policy, financial markets, international economics and national economy than in non-crisis times. In terms of crisis management, the ECB is in particular responsive to Germany and Italy. These two countries held opposing points of view in the euro crisis: namely, Germany tended to endorse a strict adherence to fiscal discipline, lower inflation and austerity-oriented crisis management; on the other side of the debate were countries more severely affected by the sovereign debt crisis (Spain and Italy), which advocated for flexibility in fiscal policies, additional financial support and mechanisms of intra-European solidarity (e.g., Eurobonds, rescue funds, or even more accommodative monetary policy). Thus, given that the ECB responds to both and does not choose a side, it confirms both a high degree of responsiveness to public pressure and a certain degree of collegiality and responsiveness of the ECB in salient times. Appendix C5.1 shows that there is also a slight increase in 3 month lag responsiveness over time (post-crisis) in most topics.

Second, besides the ECB being responsive to individual NCBs, the ECB can more strongly respond to coalitions of countries with similar interests. The more strong the block of countries, the more difficult it is for the ECB to ignore. The most common distinction of central bank interests are whether a bank has “hawkish” orientations, emphasising strict inflation control or “dovish” orientations, prioritising growth and employment alongside price stability. Within the Eurosystem, historically hawkish countries (e.g., Germany, the Netherlands) often champion low-inflation policies, while dovish countries (e.g., France, Italy, Spain) may exhibit greater tolerance for inflation or stronger concerns about economic growth. These NCBs, therefore, often share a similar stances and may coordinate, thus exerting collective influence on the ECB’s communication agenda. In

Appendix C5.2 the interaction grid results are shown when the individual NCBs are replaced with coalition dummies. I find that the results go into a similar direction and that ECB responsiveness is at some points even stronger towards the coalitions.

Third, major events or high-profile gatherings that bring together ECB and NCB governors can also impact responsiveness. One such event is Sintra, which is widely regarded as the ECB's equivalent of the Federal Reserve's Jackson Hole symposium, where central bankers, economists and market participants gather annually to exchange views and debate policy. This conference thus provides a focal point for agenda-setting and potential consensus-building among Eurozone policymakers and I, therefore, expect responsiveness (the NCB lag) interacted Sintra dummy to be negative. Appendix C5.3 shows that the effects of Sintra are significant and mainly negative, pointing towards responsiveness being lower post-Sintra conference. Specifically, there is less responsiveness to Italy and Germany on financial markets topics, banking supervision to Italy, climate to Spain and national economy to France and the Netherlands. This implies that coordination at the conference can possibly replace responsiveness through speeches or that there has been coordination that NCBs will be giving speeches on their own domestic topics without the ECB having to respond. One can, therefore, see a similar strategic use of communication as in the case of Herzog and Benoit (2015) in Irish legislative austerity debates, i.e., speeches are used to show dissent when backbenchers cannot show dissent in their voting behaviour.

3.6 Alternative explanations

Possible alternative explanations of the found results are now explored. First, central banks can respond to certain topics by linking and talking about a closely related topic, e.g., the ECB could respond to financial stability concerns of NCBs with highlighting

the importance of price stability. If this occurs systematically, it could undermine the responsiveness findings. To examine direct responsiveness (within a 30-day window) in discussions between various topics across all central banks, I look at the topic sequences presented in the Markov transition matrix in Figure C6.34 in Appendix C6.1. Overall, most transitions between topics generally occur with low probability, pointing towards the volatile nature of speech topics and no systematic patterns of certain topics responding to other topics. It can also be seen that the longer the window length, the more monetary policy responds to other topics. This indicates that monetary policy is still a consistently core topic that is discussed in response to various other topics.

Second, although issue-responsiveness operationalised in the way mentioned above allows to explore whether similar topics are discussed across central banks, i.e., between central bank responsiveness, it does not capture how specific aspects and stances taken in that issue evolve in the speeches, i.e., within topic responsiveness. For instance, the NCBs and the ECB might focus on monetary policy but with different emphases – one stressing inflation targeting, the other detailing unconventional instruments. Using text similarity measures, figures C6.36 and C6.37 show the level of agreement on monetary policy and climate (see Appendix C6.2 for more information). The figures show different patterns, i.e., the disagreement after being initially stable, decreases over time for monetary policy while remains consistent for climate. A decrease in disagreement in monetary policy can indicate that the NCBs and the ECB after a long-period of crisis disagreement, are converging in their views and approaches possibly due to increased coordination, shared economic data or a mutual recognition of similar economic challenges. This can potentially reflect the contextual situation, i.e., in low inflationary times there were disagreements without many further consequences for markets, while in higher inflationary times the markets required one united voice to stabilise. Conversely, disagreement remains stable in the case of climate which may imply that the discussions are more contentious or less

well-defined reflecting the different stances of publics in the member states. Given that the climate speeches are mainly targeted to the public and not the financial markets, one united voice is not per se necessary. This further can point towards the ECB in ‘new’ topics, rather than following certain NCBs and their pressures, instead responding to a general epistemological consensus on discussing certain policy issues using their communication strategic.

3.7 Conclusion

This paper examined agenda-setting and issue-responsiveness in the case of the Eurosystem. Unlike recent studies that focus on either individual-level behaviours or structural macro-changes, I focus on the institutional set-up of the Eurosystem. I argue and show using speeches from 1997 to 2024, that NCBs play an important role in initiating certain issues and, thereby, act as a transmission mechanism between national publics and the supranational ECB. Specifically, ‘new’ and salient topics are brought onto the ECB agenda by NCBs, showing informational rationales and functional social awareness of the ECB.

I find mixed evidence on whether the ECB is more responsive when NCBs experience economic and public domestic pressures. Specifically, the evidence shows that responsiveness patterns are not systematic and, if present, often respond to both sides of the debate. This can imply that the ECB rather than responding to national pressures, instead embraces emerging epistemological consensus or policy issues. Exploring scope conditions has shown that responsiveness is higher in salient periods and when coalitions between countries are formed, pointing towards public pressures mattering. In contrast, Sintra reduces responsiveness pointing towards coordination occurring and responsiveness being strategic. I rule out that responsiveness is driven by issue-linkages, but I do find

that contextual factors influence within topic responsiveness in certain topics pointing towards that communication is used strategically by central banks.

This research has limitations which raise avenues for further research. First, by mainly focusing on ‘between’ topic responsiveness, I do not in detail examine ‘within’ topic changes (other than the similarity scores) which can be crucial to understand the full extend of NCB influence. Second, I do not directly test the extent to which NCBs are interlocutors between the ECB and Member States. Do NCBs represent the national interest directly in their communication; are NCBs self-interested or do they follow the political positions of the Member States? I have argued that there are good reasons to assume the former, but this can still be systematically tested over time. Third, the Markov chains and cross-sectional timeseries employed in this paper rely on many assumptions (e.g., the Markov chain is only conditional on the previous speeches and the timeseries does not imply that countries are responding to each other *per se*). Nevertheless, this paper has still offered the first systematic empirical evidence specifically for the role of NCBs and offers avenues to follow-up with case studies.

Chapter 4

Introducing Textual Measures of Central Bank policy linkages Using GPT-3.5

Abstract

While institutionally independent, monetary policy-makers do not operate in a vacuum. The policy choices of a central bank are intricately linked to government policies and financial markets. We present novel indices of monetary, fiscal and financial policy linkages based on central bank communication, namely, speeches by 118 central banks worldwide from 1997 to mid-2023. Our indices measure not only instances of monetary, fiscal or financial dominance but, importantly, also identify communication that aims to coordinate monetary policy with the government and financial markets. To create our indices, we use a Large Language Model (GPT 3.5-0301) and provide transparent prompt-engineering steps, considering both accuracy on the basis of a manually coded dataset as well as efficiency regarding token usage. We also test several model improvements and provide descriptive statistics of the trends of the indices over time and across central banks including correlations with political-economic variables.

4.1 Introduction

Monetary policy is intricately connected to fiscal and other government policies, financial market developments and regulation. Since the 1990s, a broad consensus prescribes monetary policy to be delegated to independent central banks to ensure that monetary policy is able to keep inflation under control (Rogoff, 1985). Yet, monetary policy does not operate in a vacuum. First, monetary and fiscal policy are closely linked, given their joint objective of macroeconomic stabilisation (cf. the macroeconomic policy-mix literature, e.g., see Davig and Leeper, 2011). Second, monetary policy is not detached from financial stability considerations. On the one hand, monetary policy is one of the key determinants of financial cycles (Miranda-Agrippino and Rey, 2020). On the other hand, regulation and macro-prudential policy are key determinants for stable financial markets, which are, in turn, a prerequisite for the conduct of a price stability-oriented monetary policy (BIS, 2016).

This chapter presents novel indices that identify monetary, fiscal and financial linkages in central bank communication. These policy linkages can conceptually be differentiated into (i) monetary, fiscal and financial dominance and (ii) monetary-fiscal and monetary-financial coordination. First, monetary dominance is defined as a regime by which government policies accommodate the monetary policy objective of maintaining price stability. Fiscal dominance, however, implies that monetary policy accommodates governments (see Sargent and Wallace, 1981). Accordingly, financial dominance implies that the monetary policy authority subjugates its inflation targets to financial market pressure (see Brunnermeier, 2015; Schelkle, 2023). Second, we introduce an additional regime, namely, policy coordination. In this regime, central banks aim to coordinate their monetary policy with the government or financial markets to improve the macro-financial policy mix.

We use a Large Language Model (LLM), specifically, a Generative Pre-Trained Transformer (GPT 3.5-0301), to identify these linkages in speeches of officials from 118 central banks from 1997 to mid-2023. Our textual analysis consists of three independent coding stages. We optimise this multi-stage classification task using prompt engineering for both accuracy on the basis of a manually coded dataset and efficiency regarding token usage. We provide transparency regarding all the decisions and experiments we conduct for the benefit of future use cases.

Our measure of central bank policy communication brings several improvements vis-a-vis previous studies. First, focusing on speeches helps to overcome a key difficulty in detecting the nature of policy linkages ('dominance' and 'coordination'). More specifically, using text rather than more macroeconomic data combined with economic modelling (e.g., Favero and Monacelli, 2005; Afonso and Toffano, 2013) or machine learning methods (Hinterlang and Hollmayr, 2022) allows us to go beyond binary assessment of monetary and fiscal dominance and provides more qualitative information about the interactions. For example, our measure, unlike traditional measures, also captures unconventional monetary policy instruments and their interaction with fiscal or financial policies. Second, our multi-level manually validated and prompt engineered GPT-3.5 measurement goes beyond word-frequency counts and topic modelling approaches. In contrast to these measures, GPT-3.5 has a natural understanding of text and can interpret words in their context. This allows us to measure more abstract concepts and relationships previously difficult to detect. Third, our Large Language Model scales beyond what is feasible with manual human coding in terms of sample size. This is necessary to create indices across time and central banks, especially given the increase of central bank communication in the last two decades (Blinder et al., 2022). This also allows further applications for policymakers to use an early warning gauge of when there are pressures on monetary policy.

Nevertheless, recent advances in the application of LLMs in political science and economics have rapidly transformed the field of text-as-data. Scholars increasingly rely on LLMs such as GPT, LLaMA, and fine-tuned domain-specific variants to classify political texts, policy documents, and institutional communication. Methodological contributions have clarified best practices for using LLMs in classification tasks, including prompt engineering, few-shot learning, and chain-of-thought prompting (e.g., see Törnberg, 2024; Gilardi et al., 2023). This paper contributes to these rapid developments underscoring the importance of transparent workflows and multi-stage validation strategies to ensure reproducibility.

This chapter is structured as follows. Section 4.2 discusses the different forms of interaction between monetary and other government and financial policies – policy coordination and policy dominance – and how these relate to central bank communication. Section 4.3 discusses the construction of our indices, its validation and prompt engineering. Section 4.4 presents the results of the database describing the trends in the index over time and across central banks. Section 4.5 concludes.

4.2 Monetary, fiscal and financial policy linkages in central bank communication

The central bank independence paradigm posits a strict separation of monetary policy being delegated to a politically independent authority while the government remains in charge of fiscal and other policies (e.g., see Cukierman et al., 1992). Independence is meant to ensure monetary dominance, forcing fiscal or financial authorities to subordinate their policies to the central bank's price stability objective (Sargent and Wallace, 1981). To be more specific, the central bank is expected to control its target variable – typically the short- or medium term HICP – independently, with governments adjusting their policies

to the monetary policy objective. Monetary dominance thus implies that monetary policy in principle should not be available for any other policy purpose but to steer the rate of inflation.

Central bank independence is meant to protect the central bank from fiscal dominance. Fiscal dominance describes the situation in which the central bank subordinates its price stability objective to fiscal policy objectives, i.e., the situation in which monetary policy is primarily driven by fiscal considerations rather than maintaining price stability (Sargent and Wallace, 1981). Fiscal dominance may imply outright monetary financing of government debt and intervening directly in the sovereign bond market to stabilise government debt by the central bank. Another example of fiscal dominance is the case that the central bank accommodates the fiscal authority budget deficits. In normal circumstances, the central bank would be required to raise its interest rates to combat these inflationary pressures. However, under a fiscal dominance regime, the central bank does not raise rates in response to these inflation pressures. Low nominal interest rates help the government to sustain its deficit while high inflation erodes the real value of the debt stock (Leeper, 1991).

More recently, scholars have also pointed out that central banks face risks of ‘financial dominance’ (e.g., Fraga et al., 2003; Brunnermeier, 2015; Diessner and Lisi, 2020). Following the logic of fiscal dominance, financial dominance describes a situation in which the central bank is not able or willing to tighten its policy stance as this would threaten the stability of the financial system given the over-leveraged financial system. Brunnermeier (2015) argues that central banks see themselves forced to come to the rescue of financial institutions when facing contagion across banks and a more systemic crisis. This can be driven by concerns about negative feedback effects from the financial markets (such as contagion and doom loops) and central banks may be forced to bail-out or recapitalise banks. In other words, systemic financial risks are pressuring monetary

(and fiscal) authorities to ensure the survival of the financial system, subordinating the inflation target to the financial stability objective.

In their seminal piece: "Some Unpleasant Monetarist Arithmetic", Sargent and Wallace (1981) describe monetary and fiscal policy interaction as a competitive game: "Which authority moves first, the monetary authority or the fiscal authority? In other words, who imposes discipline on whom?" (p.7). However, the interaction between the monetary authority and the government does not necessarily resemble Sargent and Wallace's competitive game. In practice, governments do not continuously and as a matter of principle aim at undermining monetary dominance and exerting fiscal dominance. This allows room for coordination between central banks, governments and financial markets. In this setup, agents benefit from coordinating their activities by making decisions that complement each other and thus maximise the payoffs for all.

We use the term 'monetary-fiscal coordination' as referring to the governments and central banks coordinating their policies to offer an optimal or stabilising policy mix and, therefore, accommodate each other in a non-hierarchical manner. For example, in the case of macroeconomic shocks, it may be desirable to coordinate monetary and fiscal policy to ensure optimal output smoothing while keeping inflation at target. This may be especially prevalent when monetary policy approaches the zero lower bound (ZLB) since fiscal policy in that case has to take a larger role in macroeconomic stabilisation. Specifically, to the extent the ZLB reduces the effectiveness of monetary policy relative to fiscal policy in stimulating demand, this may justify governments to raise demand when negative shocks to economic activity occur (Haldane, 2020).

In addition to monetary-fiscal coordination, we introduce the term, 'monetary-financial coordination' to refer to the situation in which there is coordination of monetary and financial policy and regulation. Central banks are well placed to internalise these complex

interactions between monetary and macro-prudential policies since monetary policy affects credit growth which has implications for the health of the financial system (Maddaloni et al., 2022). A pure monetary dominance regime would require the central bank to obstinately pursue its price stability objective with no regard for financial stability considerations and, in case the financial market bubbles burst, ‘mop up’ afterwards (ECB, 2020). Alternatively, it can coordinate its policies with financial regulators by pushing for adequate capital and liquidity buffers. For example, it can push for stability-oriented financial market regulation, which, in turn, may help prevent financial exuberance and require less forceful and economically potentially less damaging monetary policy interventions.

We suggest to measure these policy linkages using central bank communication. There has recently been a rise in the use of central bank communications in research, in particular speeches, to determine stances of central banks under high secrecy and limited public information (e.g., Baerg and Lowe, 2020; Schonhardt-Bailey, 2013; Bennani and Neuenkirch, 2017; Moschella and Diodati, 2020; Ferrara, 2020). While the literature on monetary policy communication and its predictive power is abundant together with the literature on the (macro)economic effects of central bank communication (e.g., Hansen and McMahon, 2016; Swanson, 2021), the literature on central bank communication on fiscal and financial policy is more scarce. The focus in this literature is primarily on measuring the intensity of central banks’ fiscal communication, that is, how much central banks talk about fiscal policy and to quantify the direction (e.g. “hawkish” vs “dovish”) of that communication (Marozzi, 2021). Heinemann and Kemper (2021), for instance, looks (manually) at whether governors and board members take hawkish, neutral or dovish positions on fiscal policy. Moreover, Aruoba and Drechsel (2022), building on foundational work of Romer and Romer (2004) show the usefulness of natural language processing by using the documents of the Fed’s Federal Open Market Committee meetings

to predict monetary policy shocks. These studies conclude central bank ‘talk’ is not necessarily cheap but sets policy directions and can lead to actual responses.

The existing scholarship on monetary policy-interactions in the economic literature mainly focuses on monetary versus fiscal dominance. Most of these studies rely on economic models with New-Keynesian elements, most often Dynamic Stochastic General Equilibrium (DSGE) models (e.g., Davig and Leeper, 2011; Bianchi and Ilut, 2017). These models capture the market interactions of households, firms, the government and the central bank and can be used to conduct counterfactual policy experiments to determine the effect of policy rules on outcomes like inflation. This literature often uses the terminology of Leeper (1991) describing historical periods with either “active” or “passive” fiscal and monetary policy. An active monetary policy regime is as a scenario where the central bank prioritizes the control of inflation through adjustments in nominal interest rates without accommodating fiscal deficits, thus signalling monetary dominance, whereas active fiscal policy is indicative of fiscal dominance. Markov-switching regressions (Favero and Monacelli, 2005) are commonly used to estimate regime changes endogenously. These models introduce the ability to capture structural changes in the behavior of economic agents to represent periods of fiscal and monetary dominance by, for example, altering the central bank’s reaction to a deviation from its inflation target. Closely related to this are a number of papers that model dominance by integrating Markov-switching processes with DSGE models (e.g., see Bianchi and Ilut, 2017; Davig and Leeper, 2011; Chen et al., 2022). Other approaches involve testing for Granger causality between fiscal and monetary variables in a vector autoregression setup (VAR) (Sabaté et al., 2015), as well as testing for the significance of fiscal variables directly inside policy rules Ahmed et al. (2021), such as the Taylor rule. Even more recent approaches use machine learning techniques to classify an unobserved economic state using simulated DSGE data to detect periods of fiscal and monetary dominance (Hinterlang and Hollmayr, 2022).

We argue that using a central bank communication approach can complement and even improve on the existing measurement approaches. First, existing approaches typically focus on government deficits and the policy rate of the central bank to identify monetary or fiscal policy regimes. However, these indicators can be influenced by many other variables and do not take account of the rapid expansion of non-standard monetary policy tools such as quantitative easing and monetary policy transmission safeguards. In this context, central banks also resort to communication as a policy-making tool in itself, namely, through making statements about the likely future path of interest rates (forward guidance).¹ Second, these models rely heavily on simplifying assumptions about the economy, such as a limited number of distinct actors and perfect foresight (e.g., see Stiglitz (2017) and see Lindé (2018); Christiano et al. (2018), for different perspectives on DSGE modelling). Traditional DSGE models often do not consider the financial sector at all or have a simplified representation of the financial sector which does not capture the impact of the pressure from financial markets. Third and most importantly, in the existing literature on monetary policy interactions, coordination as a form of policy linkage is overlooked. Coordination cannot be adequately represented in standard DSGE setups or related models. However, policy linkages between independent central banks and other policy agents are not only described by monetary, fiscal or financial dominance, but also by coordination between these three policies.

4.3 Constructing and validating our indices

In this section, we will first describe our rationale for our method of constructing our indices using a Large Language Model (LLM), in our case GPT 3.5. Second, we will

¹Forward guidance is in particular used if the central bank can no longer cut policy rates because they already have reached their zero lower bound. Via forward guidance, central banks provide additional information regarding their likely response to economic developments, which can anchor expectations about future policy rates and reduce uncertainty (Strasser et al., 2019).

describe our textual corpus and pre-processing steps. Third, we will describe our multi-level coding scheme and provide examples. Fourth, we present the validation of our indices including our usage of prompt engineering and decisions made regarding token efficiency. We also test a range of other large language models. Fifth and last, we will discuss how we construct our policy linkage indices from our textual corpus.

4.3.1 Choice of GPT-3.5

The purpose of this chapter is to develop indices which identify policy linkages in central bank communication. These indices not only identify policy linkages but also determine their nature (policy coordination or dominance) over time and across central banks. Identifying these patterns from text is a complex reasoning task that requires substantial domain knowledge. It is typically done through manual coding. However, given the scale of this project – more than 18000 speeches in which there are more than 2 million sentences which need to be coded – this is an unrealistic avenue.² An automated procedure is therefore necessary.

Traditional text-as-data methods – for example, counting the occurrence of certain words or determining the topic of a speech using topic modelling (e.g., using bag of word or word embedding methods) are not suited to determine the often subtle portrayals of policy dominance and coordination. This can be illustrated by the following excerpt: "Accordingly, and as recently confirmed by the ECOFIN Council and the European Council, the fiscal policy provisions of the Maastricht Treaty and the Stability and Growth Pact should continue to be applied fully. The fiscal rules are one of the indispensable pillars of EMU and the single currency, which must remain firmly in place so as not to undermine the confidence in fiscal sustainability. Finally, the current situation calls

²During our manual classification of our validation set, it took us around one minute per sentence. Thus, given the sample of two million sentences, it would take around 33,000 hours or almost 4 years cumulative to classify the entire dataset manually.

for ensuring the high quality and timeliness of statistical information on government interventions to ensure the transparent and accountable use of public funds.” from an ECB Press Conference by Jean-Claude Trichet on November 6 2008, which discusses fiscal rules that protect the central bank from fiscal dominance and thus implies monetary dominance. GPT classifies this excerpt correctly, while an algorithm without contextual understanding would not be able to identify the implication for monetary policy contained in this statement. We overcome these two empirical challenges using a (close) to state of the art large language model, namely, GPT-3.5.³

GPT functions as a “zero-shot” learner. This means it can perform classification tasks without requiring additional training or few shot learning. It produces its answers based on the instructions and input text contained in the prompt and solely relies on the knowledge and instructions following capabilities built into the model. The crucial advantage of this is that we can conduct our analysis without requiring a substantial manually labelled dataset as training sample. Moreover, GPT has demonstrated strong performance across a number of natural language processing tasks in zero-shot learning setups (Laskar et al., 2023). GPT’s capabilities are based on a Generative Pre-trained Transformer (GPT) model.⁴ GPT is fine-tuned to closely follow instructions using a process called reinforcement learning from human feedback (RLHF). This method enhances the GPT model by having humans review its responses. Based on their evaluations, the model’s parameters are adjusted to generate more desirable responses (Ouyang et al., 2022). The prompt-following ability makes GPT easier to use than next word prediction models and readily applicable for classification tasks.

³We use version gpt-3.5-turbo-0301. At the time of writing this chapter GPT-4 was not yet available to us, substantially more costly and subject to restrictive rate limits. Similarly, Google’s Gemini was not yet released when the main analysis was run. We do run our validation set with GPT-4 and Gemini Pro.

⁴GPT belongs to a family of transformer models which use the self-attention mechanism developed by Vaswani et al. (2017). The self-attention mechanism is well parallelisable in training and is therefore more scalable than the previously used recurrent neural networks, allowing for larger neural networks with more parameters (Wolf et al., 2020).

In recent months, LLMs have seen widespread uptake in economics and political science, both as general-purpose text classifiers and as domain-specific tools. Several methodological reviews and benchmarking studies have highlighted the reliability and limits of zero-shot classification using GPT, especially compared to traditional supervised machine learning and fine-tuned transformer models (Gilardi et al., 2023; ?). While concerns about reproducibility and model updates persist - particularly with closed-source models - prompt-based zero-shot classification remains a powerful tool when paired with transparent workflows and domain-informed prompt engineering.

In the context of central bank communication, applications of LLMs remain limited but are beginning to emerge. Recent (working) papers include Veyrune (2025), Bertsch et al. (2024), Hansen and Kazinnik (2023) and Gambacorta et al. (2024). These papers have tested and used LLMs for forecasting monetary policy decisions from text, interpreting ECB press conferences, and comparing GPT-based topic inference to human-coded classifications. Our approach builds on and expands these early applications by providing a systematic, multi-stage classification framework for our task at hand (classifying dominance and coordination) applied across 118 central banks and over two decades of speeches.

4.3.2 Corpus and pre-processing

Our corpus of central bank communication is scraped from the Bank of International Settlement website. It covers mostly speeches, but also press conferences and interviews of central bank officials from 118 central banks and monetary institutions over the period from January 1997 to July 2023. This dataset totals around 18,000 documents (see Figure 4.1 for the geographic distribution of the frequency of speeches and see the Appendix Table D2.1 for more details on the speeches). The speeches are downloaded in PDF format from the BIS website, converted to text and cleaned using various pre-

processing steps following standard methodology (e.g., removing page numbers, footnotes headers, chart titles, new page characters, URLs, headers and line breaks). Subsequently, we break up the speeches into single sentences.⁵ We also remove anything less than 6 tokens, anything more than 200 tokens, less than 20 characters, sequences of whitespace characters and all sentences consisting of less than two thirds ASCII characters. These are conservative heuristics to remove text that is very unlikely to contain relevant information. We retain information on the ordering which will be used in further steps.

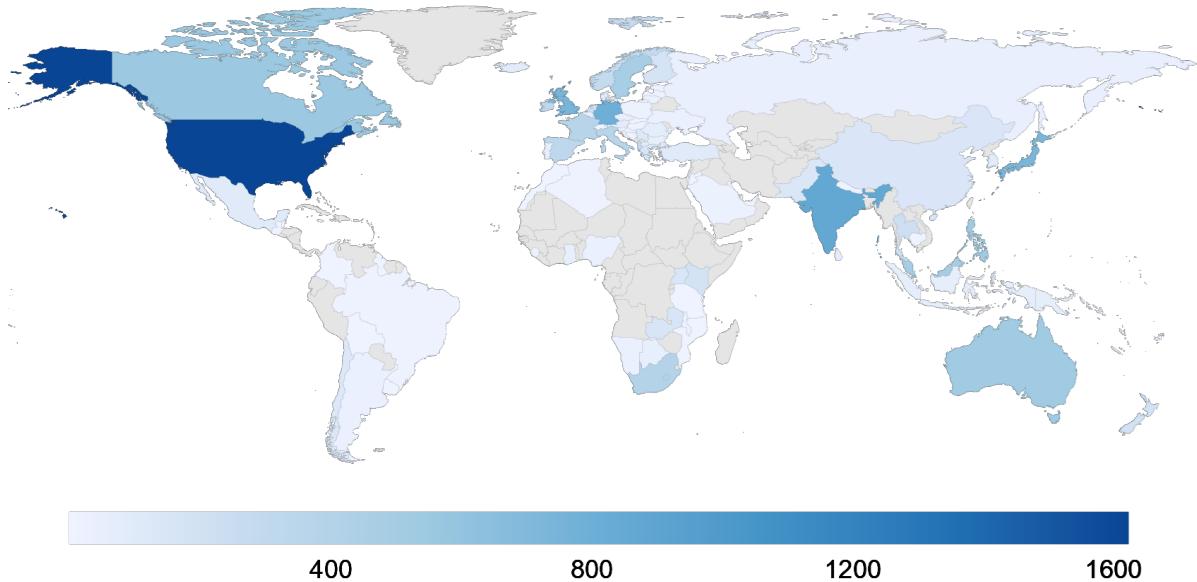


Figure 4.1: Geographical spread of the number of speeches available per central bank.

Note: The colour scale indicates the number of speeches per country contained in our dataset. See our Appendix Table D2.1 for the full list of central banks and the number of speeches per central bank. The central bank with the most speeches is the ECB with 2377 speeches (not shown on the map in order to be able to see which National Central Banks within the Eurosystem are included).

⁵Since the source PDFs do not follow an entirely standardized layout, information about paragraphs cannot be retained. Also owing to the PDF format, we cannot fully clean the content of the speeches; thus, infrequent footnotes, chart annotations and citations remain. We rely on GPT to classify these as “other” and “none”, which are not considered for our final indicators.

4.3.3 Identifying policy linkages

Our classification scheme consists of three steps. In the first stage (level 1), we determine the topic of the single sentence: “monetary”, “fiscal”, “financial”, “climate”, “macro”, “international” or “other”. We determined these topics by inductively examining 100 randomly drawn sentences from the entire corpus. Although there are many more topics (e.g., see Hansson, 2021), the first level mainly aims at examining the frequency of topics addressed by central banks. Including also other policy topics creates a database for further analyses (e.g., to examine whether mentioning certain topics is highly correlated with dominance and whether certain topics more often result in dominance or coordination). In the second step (level 2), we determine whether the sentence is descriptive or normative. That is, we prompt GPT-3.5 using the API to determine whether the sentence simply describes monetary, fiscal or financial developments or policies or offers a value judgement. Very often central bankers use descriptive sentences to portray statistical information such as inflation or growth rates. Alternatively, the sentence can be classified as normative if it prescribes a certain policy action, be it to the central bank itself or to other institutions. Again, the data from this stage can be used for further analysis, namely, are certain topics more normative or are there certain central banks that talk more normative while other central banks talk more factual? In the third step (level 3), we classify the nature of the policy linkages. To operationalise the definitions of monetary, financial and fiscal dominance and coordination, we rely on an actor-centred approach. Dominance places one actor hierarchically above another actor (e.g., in the case of fiscal dominance, the monetary authorities accommodate the fiscal authorities). In the case of coordination the actors are placed on an equal footing.

In contrast to levels 1 and 2, level 3 categories are less clear cut, generally require more complex reasoning steps and often depend on contextual information. For instance, accurately interpreting the intent behind a speech often requires considering both a

factual statement and a subsequent value judgement together, as only their combination provides a clear understanding of the speakers intentions. Therefore, we add the sentence before and after to classify 3-sentence excerpts in level 3 only. We do this for level 3 only since it is computationally intensive as it effectively triples the text that needs to be classified. Figure 4.2 below summarises the three level coding scheme.

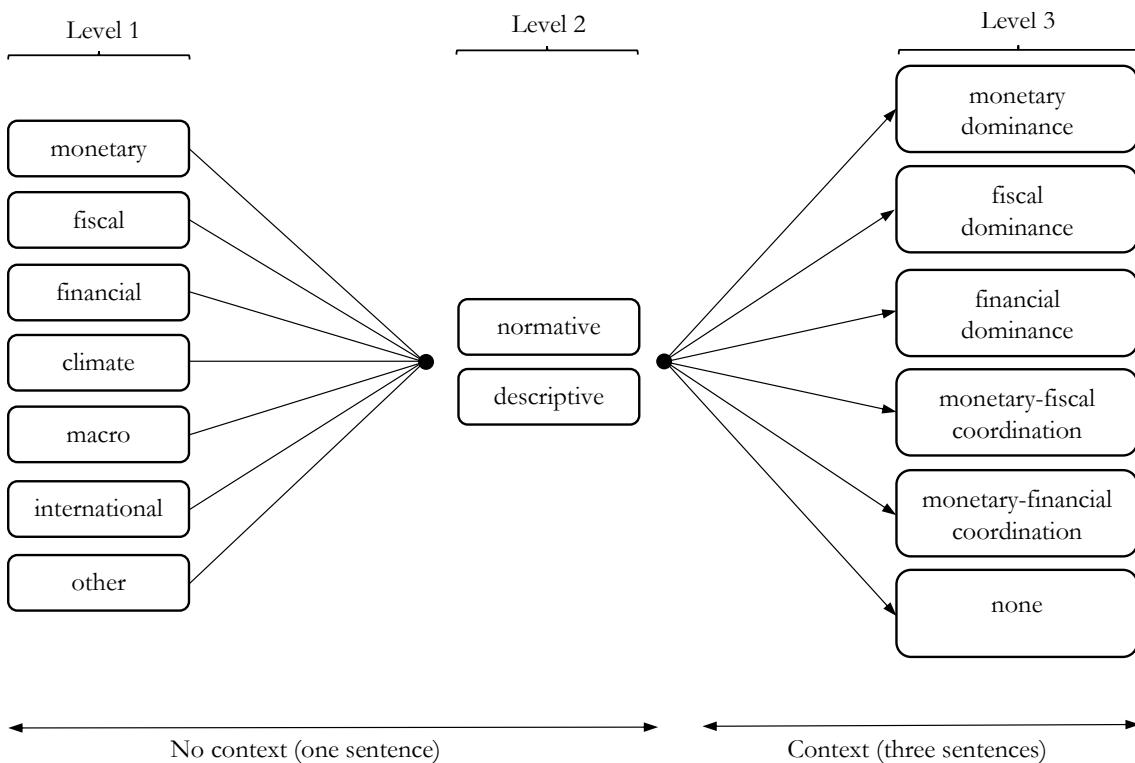


Figure 4.2: Three stage coding scheme to identify dominance and coordination.

To illustrate the various forms of policy linkages we obtain with the level 3 classification, Table 4.1 provides a number of examples and brief explanations. These sentences are taken from a 1000 sentence random sample created to validate our GPT results. See the Appendix sections D3.2 and D3.3 for guidelines to classify ambiguous sentences and examples of all categories of level 1 and 2.

Classification	Example
Monetary Dominance	<p>"Furthermore, monetary policy implementation in line with the market efficiency principle would need to remain without prejudice to our primary mandate of safeguarding price stability." (Retrieved from the European Central Bank, 14 June 2021)</p> <p><i>Explanation:</i> The topic concerns a monetary topic and they emphasize their primary mandate of price stability being above other priorities. Therefore, this sentence can be classified as monetary dominance.</p>
Fiscal dominance	<p>"Moreover, although most of the resources administered by the BIS are invested in financial assets of top quality at international level and their exposure to the various risks are managed conservatively, a greater portion of such funds could be spent toward the direct purchase of debt denominated in local currencies of emerging countries or to the use of them as collateral of certain bond issuance of countries with limited depth of their financing markets in local currency."</p> <p>(Retrieved from the Central Bank of Argentina, 9 July 2008)</p> <p><i>Explanation:</i> This sentence refers to funds being spent towards the direct purchase of debt (=monetary financing) instead of considering pure price stability considerations, thus we consider this sentence to be fiscal dominance.</p>
Financial dominance	<p>"It is thus significant that our flexible and abundant provision of liquidity contained market participants' concerns over liquidity financing." (Retrieved from the Bank of Japan, 4 July 2002)</p> <p><i>Explanation:</i> This sentence states that monetary policy is accommodating financial markets by providing liquidity, thus showing that financial markets are a consideration for the bank in conducting their monetary policy.</p>
Monetary-fiscal coordination	<p>"Since restarting our strategy review, we have introduced a new work stream on monetary-fiscal interactions precisely to address such questions." (Retrieved from the European Central Bank, 30 September 2020)</p> <p><i>Explanation:</i> This sentence refers to the monetary-fiscal interactions which is a key policy in the monetary-fiscal coordination.</p>

Classification examples (*continued*)

Classification	Example
Monetary-financial coordination	"If market participants are willing to continue to work together, then we can safely achieve the transitions needed to create a better and more robust system that will help to ensure our ongoing financial stability." (Retrieved from the Board of Governors of the Federal Reserve System, 7 November 2017) <i>Explanation:</i> this sentence shows that the bank wants coordinate with market participants to ensure financial stability.

Table 4.1: Classification examples.

4.3.4 Validation and prompt engineering

Employing GPT-3.5 via the API as our classifier requires some efforts in prompt engineering. In contrast to supervised machine learning algorithms, there is no training dataset and hardly any tuneable parameters to optimise. The behaviour of the model and thereby also the accuracy with which it can classify sentences is entirely determined by how it is instructed, i.e. the prompt that is given to the model. At the time of writing, there are no established best practices yet on how to write an optimal prompt. Academic contributions that analyse the trade-offs in designing prompts do not exist to the best of our knowledge yet.

The aim of this section is to describe and offer guidelines on the steps and experiments in determining the optimal prompt for our use case. This may also serve as a reference for future research. Since there are an almost infinite number of prompt variations, it is not feasible to test every possible modification to the prompt. Systematically testing prompts is further complicated by the fact that GPT-3.5 is not entirely deterministic. GPT-3.5 occasionally hallucinates output categories or does not follow the prescribed output format, which requires manual intervention to calculate accuracy metrics. Finding

our final prompt was thus an iterative process of making incremental changes to the prompt until further modification no longer resulted in substantial improvements in terms of output stability; that is the models adherence to instructions, and accuracy vis-a-vis a manually validated dataset which we describe below.

4.3.4.1 Manual validation

Our first step to evaluate different prompts is to manually classify a validation sample, that we treat as ground truth and compare GPT-3.5’s results against. We determined the quality of a prompt based on a number of validation metrics. Our manual validation sample consists of 1000 sentences that were randomly drawn from the entire sample of two million sentences, making the validation sample representative of the whole dataset with regard to the frequency of each category. All three levels were coded by the three authors independently, using the definitions provided in the prompts and the ambiguous sentence guidelines presented in the Appendix. All three coders coded the same first 400 sentences for two reasons. First, in this way we can calculate coder reliability scores to determine how consistent the assigned labels are across coders. This is especially important given the high level of judgement and abstraction needed to classify some of the sentences. Second, we can use this information to examine how the accuracy and uncertainty in GPT-3.5’s classification correlates with agreement among human coders. All three coders subsequently coded 200 extra sentences each, expanding our validation set to 1000 sentences. We validate our model using both the full sample and a agreed sample, where all human coders made the same assessment.

Agreement among the independent coders is high. The inter-coder reliability score measures the consistency of coding results (see O’Connor and Joffe (2020) for more information). The average correspondence in coding for level 1 is 84%, 81% for level 2 and 76% for level 3. Appendix section D3.4 includes tables that show the pairwise

correspondence of the coders, as well as standard intercoder reliability scores (Krippendorff's alpha and Cohen's kappa). The slightly lower (but still considerably high) score for level 3 indicates that the classification task is difficult even for human coders trained as (political) economists. Most disagreement in the manual classification comes from the distinction between the coordination and dominance categories, i.e., whether the relationship between the actors is hierarchical or not. In case of disagreement between coders, a majority rule is used to determine the “correct” label.⁶

We used five standard validation metrics (Accuracy, F1 macro, F1 weighted, Precision and Recall) to determine the best prompts to run on the entire corpus.⁷ When determining the best prompt, we pay particular attention to the macro average F1 score. The macro average F1 score is computed by first calculating precision and recall for each class, deriving the F1 score for each class as the harmonic mean of precision and recall, and then taking the unweighted average of these per-class F1 scores.⁸ This score is sensitive to changes in prediction accuracy in the less frequent categories. This is important given that we are especially interested in being able to identify specific kinds of dominance, particularly fiscal and financial dominance which are relatively infrequent.

4.3.4.2 Prompt engineering steps

In a second step, we designed and optimised the prompt, that is, the instructions given to the GPT-3.5 API. This prompt engineering exercise had three dimensions, namely (i) which and how much information the prompt should optimally contain, (ii) the

⁶In case there is no majority, i.e., all three coders classified the sentence differently, the assigned label is randomly chosen. This affects 5 sentences on level 1 and 7 sentences on level 3. The binary classification of level 2 ensures that there is always a majority label.

⁷Scores were calculated using the scikit-learn package in Python (see Pedregosa et al. (2011) and the Appendix section D3.5 for a description of these measures

⁸Precision and Recall are defined as Precision = $\frac{TP}{TP + FP}$, , and Recall = $\frac{TP}{TP + FN}$, where TP, FP and FN stand for the number of observations that are true positive, false positive and false negative. The F1 score itself is $F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$

optimal temperature settings and (iii) the optimal sentence count within a prompt (also considering efficiency). Subsequently, we tried several of the techniques outlined in OpenAI’s prompt engineering guidelines.⁹

The first and seemingly most critical choice is what information to provide GPT-3.5 regarding the classification task it should conduct. Do we need to give exact definitions and examples of, for instance, fiscal dominance? Or can we rely on GPT’s understanding of what fiscal dominance is, encoded by its billions of parameters?¹⁰ However, correctly defining concepts does not imply that GPT can operationalise these to classify sentences. Our experiments suggest a middle ground – between providing no extra information and very detailed instructions and examples – works best for our classification of level 3. Both a prompt with minimal instructions as well as a more elaborate prompt perform worse than our optimal prompt (see Table D4.9 and section D6 in the Appendix for validation metrics and the prompt variations). This indicates that initially providing more information can aid the model, but an excess of details might hinder its ability to differentiate between the main instructions and additional information. Performance does not increase through including overly specific instructions to correct infrequent mistakes. Providing the model with example sentences does improve performance, but as evidenced by the lower accuracy of our more elaborate prompt, the information that is presented to the model should be chosen carefully. Overall, prompts should be concise, but elaborate enough to provide essential instructions.

Second, temperature settings need to be chosen for the prompt. The responses from

⁹See <https://platform.openai.com/docs/guides/prompt-engineering> for the guidelines.

¹⁰When asked for a definition of “Fiscal Dominance” GPT-3.5 replies accurately. More specifically, GPT-4 replied to “Can you define the concept of fiscal dominance in three sentences?” with “Fiscal dominance occurs when a government’s fiscal policy—particularly its borrowing and spending habits—dictates monetary policy decisions. In such a scenario, the central bank may be pressured to finance government deficits, often by purchasing government securities or through other means that can lead to inflation. This situation typically arises when a government’s debt levels are so high that normal market mechanisms for debt financing are no longer effective, forcing the central bank to step in.”

GPT-3.5 are never fully deterministic. The degree of variation in the responses can, to a certain extent, be controlled with the temperature settings. A higher temperature leads to a less stable, but more diverse and creative responses from GPT-3.5. To validate the temperature setting, we ran temperature levels between 0 and 1 with 0.25 increments in the temperature five times on the validation set. As expected, the accuracy, measured as overlap between the human and GPT-3.5 classification, varies more across runs as temperature increases (see Figure 4.3). Average accuracy is very similar for temperature settings smaller or equal to 0.5. Since choosing the lowest temperature settings does not seem to incur a cost in terms of average accuracy, we select a temperature of zero. Higher stability of results across runs facilitates the prompt engineering process as less runs are required to determine the effect of changes to the prompt. Also, lower temperature is helpful for replicability, although perfect replicability is currently not possible with GPT-3.5. Typically, more than 97-99% of sentences are classified identically between two runs with temperature 0 directly after each other.¹¹. In addition, server-side changes to the model configuration are outside of our control and can potentially affect the classification.¹²

¹¹OpenAI introduced a option to set a seed as a beta feature in November 2023. However, passing the seed to the API does not seem to have an effect with our version of the model.

¹²Since November 2023, the API returns a system fingerprint that can be used to trace when changes occur. However, it is still unclear what was changed and whether this affects model behavior. Complete replication of results is only possible when there is unrestricted access to the model, similar to what is available with certain open-source large language models.

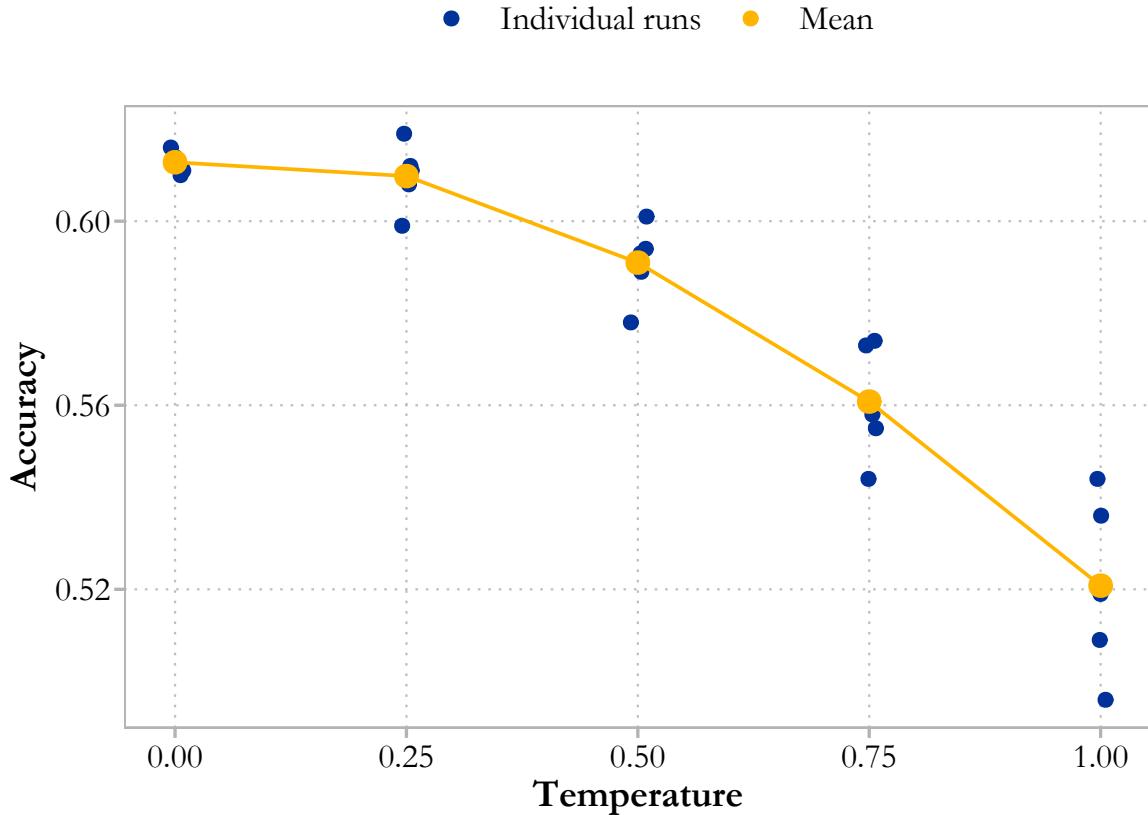


Figure 4.3: Variation in accuracy with different temperature settings on level 3 classification.

Note: Every temperature setting was run five times. Blue dots show the accuracy of each run. Yellow dots indicate the average of each temperature setting.

A third consideration is how many sentences to add within a single prompt. We let this choice depend on two criteria: accuracy and efficiency. The effective use of tokens is of importance since the utilization of commercially operated LLMs is billed by token usage and can quickly become cost-intensive, especially given the size of our dataset.¹³ LLMs like GPT-3.5 operate on tokens instead of words whereby on average a word consists of around 1.3-1.5 tokens. Increasing the number of sentences per prompt reduces the number of tokens that are required to classify the entire sample as the instructions

¹³Our final level 3 classification used 2,791,384,568 tokens. Increasing this number by even small factor would lead to a substantial rise in expenses.

need to be restated fewer times. The theoretical maximum is given by the maximum context length, that is the number of tokens a LLM can process in a single prompt. In the case of GPT 3.5-0301, the context length is 4,096 tokens, which is equivalent to roughly 100 sentences plus instructions. However, as the sentence count increases, token savings decline quickly since the instructions only make up a small share of the tokens compared to the sentences to classify. Accuracy is highest when classifying 3 sentences inside a single prompt (see Figure 4.4 for level 3). Including 5 or 10 sentences reduces accuracy slightly, while significantly lowering the token count. Including 10 or 25 sentences inside a prompt yields more accurate results than 5 sentences. A possible explanation is that showing multiple sentences at the same time, helps GPT-3.5 to distinguish the categories. On the other hand, if the number of sentences gets too large, its ability to correctly classify sentences decreases. We find a similar bell-shaped relation for level 1 and level 2 (see Appendix section D4). Taken these points into consideration, we opt for 10 sentences per prompt on all three levels.

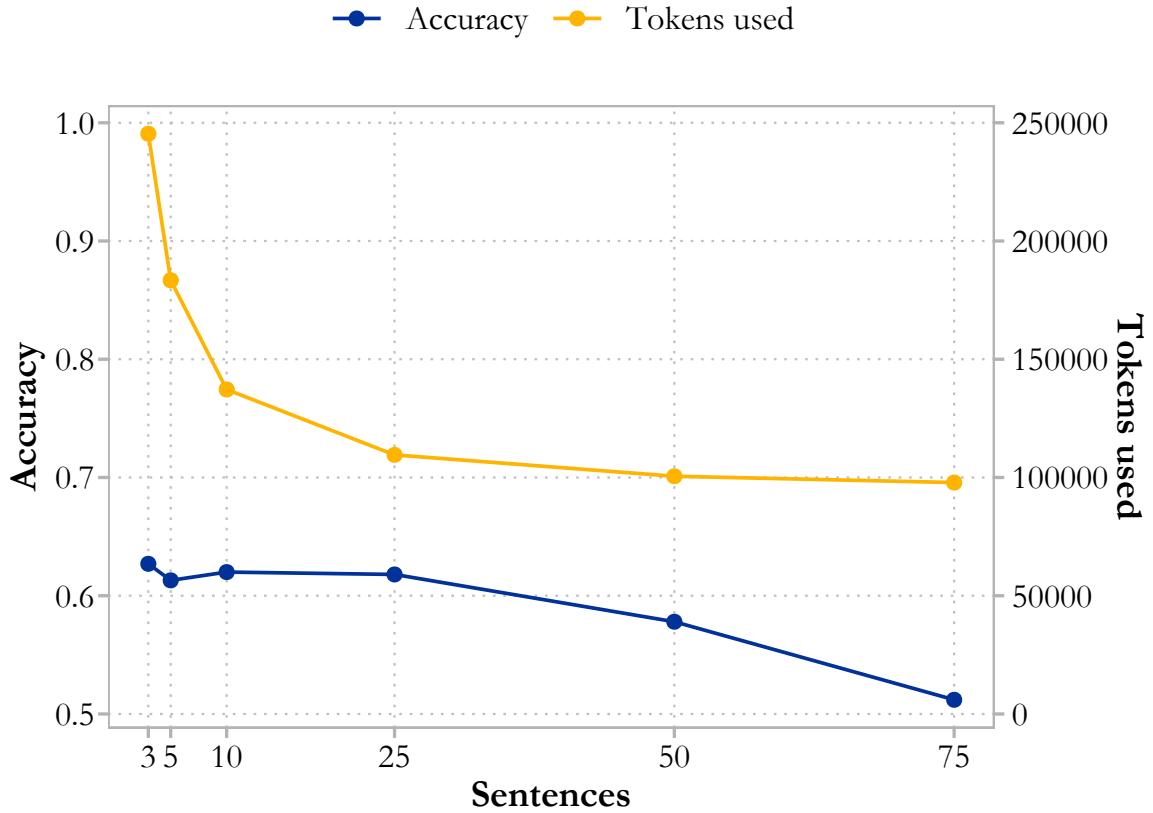


Figure 4.4: Relationship between sentence count, accuracy and token usage.

Note: The yellow (blue) line illustrates how token usage (accuracy) varies with the number of sentences that are included in a single prompt.

Lastly, we test the OpenAI prompt engineering guide. The main takeaway from OpenAI's prompt engineering guide is to conduct systematic testing to the degree the unstable output of GPTs allows for it. Our final prompt went through many iterations testing numerous of the suggestions from OpenAI's prompt engineering guide. We find that changing the system message – which should govern the overall behaviour of the model – has little impact (see the Appendix Table D4.8 for the results of runs with different system prompts). Moreover, clearly structuring the prompt with delimiters helps to stabilise the model output and instructions regarding the output format are best

placed at the end of the prompt.¹⁴ The final prompts for levels 1 to 3 are given in the below.

Level 1

Classify sentences from a central bank speech into the following categories:

Monetary if: • The sentence addresses monetary issues (e.g., inflation, price stability, interest rates).

Fiscal if: • The sentence addresses fiscal issues (e.g., sovereign debt, budget balance, fiscal governance, taxes, pensions).

Financial if: • The sentence addresses financial issues (e.g., banking supervision, financial instability, credit risks).

Climate if: • The sentence addresses climate issues (e.g., environmental issues, CO₂, climate change, sustainable development goals).

Macro if: • The sentence addresses macroeconomic issues (e.g., GDP, economic growth, unemployment, productivity, economic outlook).

International if: • The sentence addresses international economic issues (e.g., trade, exchange rates, capital mobility, tariffs).

Other if: • The sentence does not relate to the topics above. • This category should be used as the default category.

Reply only with the sentence number and the assigned label.

Level 2

Classify sentences from a central bank speech as either normative (value judgment) or descriptive.

These are the sentences:

1. <Sentence 1>

2. <Sentence 2>

...

¹⁴We did not experiment with techniques that elicit a reasoning process in the model output like chain-of-thought prompting (see, Wei et al., 2023) as this would increase the output tokens by at least 50 times compared to directly prompting for the label.

Level 3

Classify excerpts from central banker speeches into one of the following categories:

None if: • The excerpt contains no reference to monetary, financial, or fiscal developments. • The excerpt describes these developments without a normative statement (e.g., "the deficit is expected to reach 2.5% of GDP in 2020").

Monetary Dominance if: • The excerpt states that the central bank subordinates fiscal or financial policies to price stability (e.g., "the role of the central bank is not to ensure financial stability or fiscal sustainability but to maintain price stability").

Monetary-Financial Coordination if: • The excerpt suggests cooperation between the central bank and financial regulators (e.g., "higher capital requirements will increase the resilience of the banking system and support the transmission of monetary policy").

Monetary-Fiscal Coordination if: • The excerpt suggests cooperation between fiscal authorities and the central bank (e.g., "the deficit should remain below 3% of GDP not to further increase inflationary pressures").

Financial Dominance if: • The excerpt states that the central bank subordinates itself to financial stability considerations (e.g., "lower interest rates will ensure the stability of the banking system").

Fiscal Dominance if: • The excerpt states that the central bank subordinates itself to fiscal authorities (e.g., "lower interest rates will ensure that public finances remain sustainable").

Classify each excerpt individually. Reply only with the excerpt number and the assigned label.

These are the excerpts: ...

4.3.4.3 Validation scores

Table 4.2 displays the classification shares in the validation sample for the humans, which are treated as the ground truth, and GPT-3.5. For reference, the last column contains the shares in the whole dataset consisting of 2,034,313 sentences. Compared to the human classification, GPT-3.5 is too much of a “yea-sayer” that assigns dominance and coordination labels to sentences that are in fact none.

Further exploring this, Figure 4.5 below shows the confusion matrices, i.e., a matrix visualisation of the distribution of predicted labels per category. The confusion matrices can be used to determine which classes are most frequently confused with each other, which classes the model predicts most accurately, and where the model's weaknesses lie in distinguishing between classes. Taking the human-coded classification as the ground truth, GPT3.5 assigns the correct label with the highest probability across all three levels with the exception of financial and fiscal dominance on level 3. In our level 3 classification, the the most frequent mistake is always the related dominance/coordination category except for financial dominance which is most often confused with "none". (e.g., a sentence that was classified as monetary-fiscal coordination in the validation set has a 52% probability to be correctly classified and a 33% probability to be classified as fiscal dominance). Looking at the excerpts that should not be categorized under dominance or coordination, we find that our GPT classifier has a tendency to over-represent financial dominance and coordination. This could be explained by the fact that financial dominance and coordination are not as thoroughly discussed in existing literature compared to the monetary and fiscal counterparts, which likely means that these concepts are less represented in GPT's training data, reducing its ability to accurately identify dominance and coordination with regard to financial markets. With regards to our five validation metrics, level 1 and 2 perform well in all metrics with at least 83% of sentences matching the human classification, and F1 scores of close to or above 80% (see Table 4.3) in the full sample. GPT-3.5's ability to mimic the human coding is lower for level 3 but can still be considered relatively high given the complexity of the task, the room for interpretation, and the degree of disagreement also present among human coders. The accuracy in classification at Level 3 is greater in the sample where human coders agree compared to the entirety of the dataset. For Levels 1 and 2, however, the discrepancies in classification between the agreed sample and the full dataset are small.

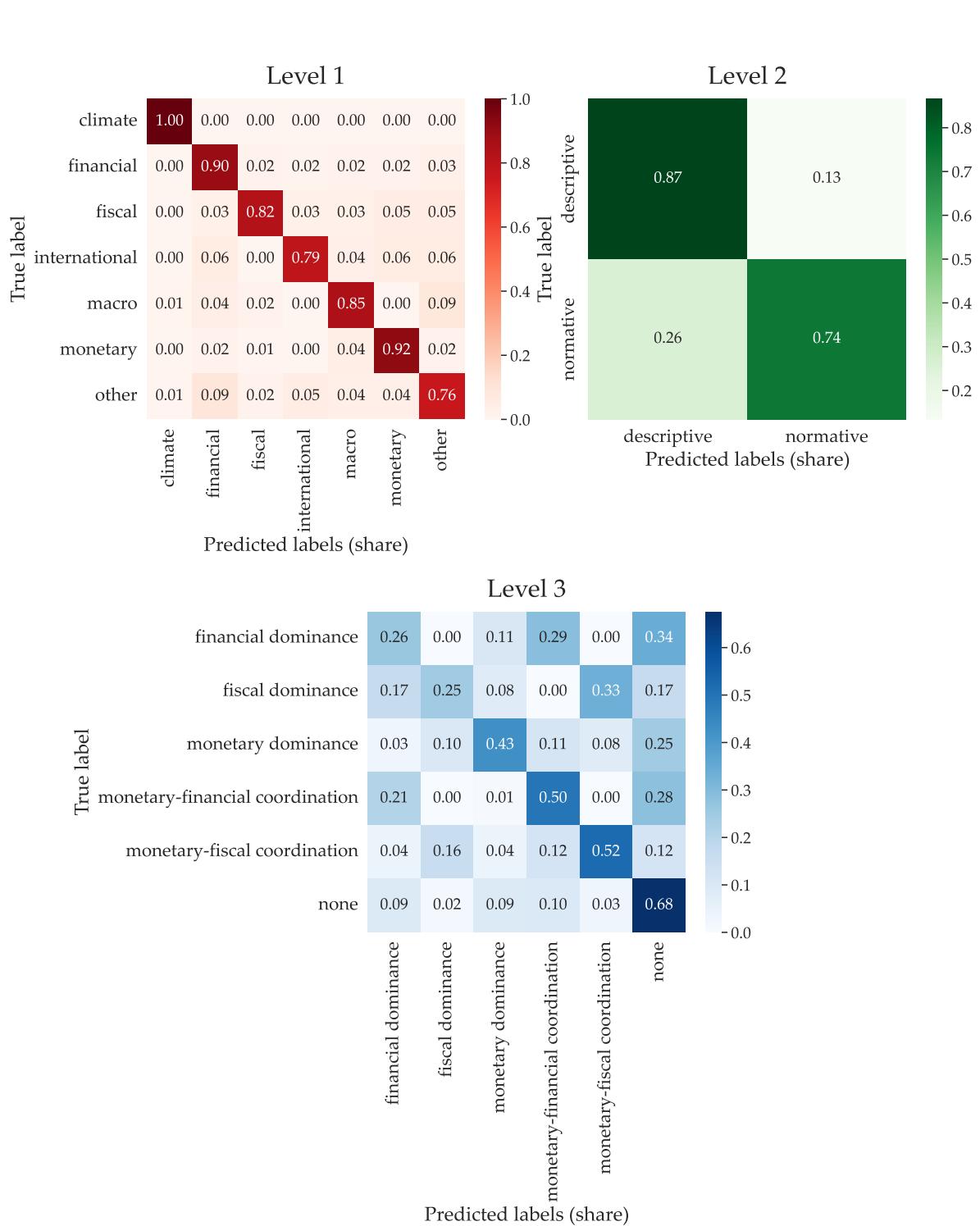


Figure 4.5: Confusion matrix

Note: The confusion matrices display the correspondence between actual categories and predicted labels. For instance, among the sentences categorized as 'financial' by the human coders in the validation set, 90% were accurately labeled by GPT-3.5. The remaining 10% were misclassified, with their distribution being relatively uniform across other categories, with the exception of the climate category. The confusion matrix is based on the full validation sample.

Table 4.2: Classification Share (in Percent)

Classification	Validation (Humans)	Validation (GPT-3.5)	Full dataset
NA	0.0%	0.0%	0.1%
financial dominance	3.8%	10.7%	9.7%
fiscal dominance	1.2%	2.6%	2.9%
monetary dominance	6.1%	10.1%	10.1%
monetary-financial coordination	10.9%	14.7%	14.8%
monetary-fiscal coordination	2.5%	4.5%	3.6%
none	75.5%	57.4%	58.9%

Note: This table reports the distribution of classification labels assigned to 1,000 randomly sampled validation sentences by human coders (treated as ground truth) and by GPT-3.5 via API. The final column shows the label distribution in the full dataset of 2,034,313 sentences. The comparison highlights GPT’s tendency to over-assign dominance and coordination labels relative to the human-coded ground truth. Label proportions in the validation and full samples are roughly similar due to random sampling.

	Accuracy	F1 (weighted)	F1 (macro)	Precision (macro)	Recall (macro)
B. Agreement sample					
Level 1	0.83	0.83	0.78	0.75	0.84
Level 2	0.83	0.84	0.75	0.74	0.77
Level 3	0.68	0.74	0.35	0.32	0.46
B. Full sample					
Level 1	0.85	0.85	0.83	0.80	0.86
Level 2	0.83	0.84	0.79	0.78	0.80
Level 3	0.62	0.66	0.36	0.34	0.44

Table 4.3: Validation metrics of final model.

Note: The agreed sample includes 316 sentences for level 1, 295 sentences for level 2, and 307 sentences for level 3. For inclusion in the agreement sample, all coders must have coded identically on the respective level. Precision and recall are macro averages taking the unweighted average of the precision and recall score of all categories.

4.3.4.4 Improvements and uncertainty measures

In the following, we examine a range of other LLMs, which include more recent versions of GPT-3.5, a GPT variant with few shot learning, and Google’s Gemini Pro. We begin

by comparing our main GPT model (gpt-3.5-turbo-0301) to the most recent iterations of GPT-3.5 (gpt-3.5-turbo-0125) and GPT-4 (gpt-4-turbo-0125). We use the same validation set and the prompt we engineered for our main model. Overall, we find considerable improvements, especially when using GPT-4 for level 3. Note that the validation metrics in Table 3 do not always move in the same direction. When given the same prompt, GPT-4 is much less likely to classify in any of the dominance categories, aligning more closely with human classification. As a result, we find much higher precision and a somewhat lower recall for GPT-4. We also noticed that GPT-4 always replied with one of our predefined labels, which makes evaluating prompts much easier. This illustrates that the choice of model plays a more significant role in influencing validation metrics compared to incremental refinements of the prompt. Due to token limits on the runs in the Open-AI platform and due to GPT-4 being at least 10 times the cost per token, we still opt for the older GPT3-based version.

Metric	Level 1			Level 2			Level 3		
	3.5-0301	3.5-0125	4-0125	3.5-0301	3.5-0125	4-0125	3.5-0301	3.5-0125	4-0125
A. Agreement Sample									
Accuracy	0.79	0.83	0.79	0.85	0.83	0.84	0.75	0.68	0.86
F1 (weighted)	0.79	0.83	0.79	0.85	0.84	0.85	0.79	0.74	0.85
F1 (macro)	0.77	0.78	0.75	0.76	0.75	0.78	0.40	0.35	0.39
Precision	0.77	0.75	0.73	0.77	0.74	0.76	0.37	0.32	0.40
Recall	0.78	0.84	0.81	0.76	0.77	0.80	0.58	0.46	0.41
B. Full Sample									
Accuracy	0.79	0.85	0.80	0.82	0.84	0.82	0.65	0.62	0.78
F1 (weighted)	0.79	0.85	0.81	0.82	0.84	0.83	0.68	0.66	0.76
F1 (macro)	0.77	0.83	0.77	0.75	0.79	0.78	0.38	0.36	0.43
Precision	0.78	0.80	0.78	0.77	0.78	0.76	0.36	0.34	0.59
Recall	0.76	0.86	0.78	0.74	0.80	0.80	0.46	0.44	0.41

Table 4.4: Validation metrics for GPT 3.5-0301, 3.5-0125, 4-0125.

Note: The agreed sample includes 316 sentences for level 1, 295 sentences for level 2, and 307 sentences for level 3. For inclusion in the agreement sample, all coders must have coded identically on the respective level. Precision and recall are macro averages taking the unweighted average of the precision and recall score of all categories.

Next, we explore few shot learning and replacing GPT with a competing model (Gemini developed by Google). For the exercise of few shot learning, we split our sample of 1000 manually annotated sentences into a training sample of 300 sentences and 700 sentences for evaluation. Few shot learning involves further prompting of the model using input prompts and the expected outputs. Utilising the OpenAI interface with default settings, using gpt-3.5-turbo-1106 we input 60 samples presenting 5 sentences in each sample with the same prompt as before. The resulting model improves significantly upon the vanilla (i.e., zeroshot) gpt-3.5-turbo model and exceeds GPT-4 on the macro F1 score. Including the few shot learning, gpt-3.5-turbo becomes much less likely to classify the infrequent dominance categories, while retaining relatively high recall on dominance

and coordination, making it the overall best performing model. While few shot learning yields significant enhancements, we did not opt for it since it also markedly increases the cost of running inference. Running prompts using few shot learning with gpt-3.5-turbo incurs 6 times the cost per input token compared to the standard gpt-3.5-turbo as of February 2024.

	GPT-3.5			Gemini	
	gpt-3.5-0301	gpt-3.5 few shot	gpt-4-1106	Pro 1.0	Pro 1.0 Few Shot
Accuracy	0.64	0.77	0.79	0.78	0.79
F1 (weighted)	0.69	0.78	0.78	0.73	0.75
F1 (macro)	0.35	0.43	0.40	0.36	0.40
Precision (macro)	0.33	0.40	0.48	0.44	0.50
Recall (macro)	0.43	0.49	0.40	0.34	0.36

Table 4.5: Few shot learning and gemini

Note: Both GPT-3.5 (gpt-3.5-turbo-1106) and Gemini were evaluated using the same validation set of 700 sentences. The few-shot prompt for each model was based on the same 300 labelled examples, which were included in the conversation history prior to classifying new excerpts.

Architecturally similar to GPT, Gemini are a family of decoder-only transformer models that are developed by Google. The mid-tier version called Gemini Pro is found to perform in between GPT-3.5 and GPT-4 on a number of LLM evaluation benchmarks (Gemini Team, 2023).¹⁵ Our results in Table 4.5 confirm this finding showing high performance of Gemini Pro 1.0 closer to GPT-4 than GPT-3.5. The large context window of Gemini Pro, which spans 32,768 tokens, significantly more than the 4,096 token limit of GPT-3.5-0301, enables the inclusion of the 300 training sentences directly within the prompt history, acting as a “few shot” learner. This approach further boosts performance, making Gemini almost perform on GPT-4 levels. The fact that fine-tuned GPT-3.5,

¹⁵As of February 2024 only Gemini Pro 1.0 is available via the API. There are also Gemini Ultra, which is the most capable Gemini model, and Gemini Nano that is optimized for memory usage and computation speed.

GPT-4 and Gemini can outperform 3.5 on a prompt that was engineered using GPT 3.5 is further evidence that model characteristics play a larger role than prompt engineering can.

Finally, we can show that human and machine uncertainty about sentence classification correlate. For this, we test whether disagreement among human coders is related to GPT-3.5’s variation across runs. We document that GPT-3.5’s classification is more uncertain, meaning it varies more between different attempts across runs, for sentences that lack consensus among humans. Figure 4.6 plots the stability of GPT-3.5’s classification, measured as the share of sentences which GPT classifies in its most frequent category across 25 runs, against the number of human coders who agree. Higher agreement among coders is associated with a more stable GPT classification.¹⁶ This variation could be used to construct an uncertainty measure. However, due to the size of our dataset, running the entire dataset multiple times is not practical due to cost and token limit constraints.

4.3.5 From raw data to constructing indices

To construct our full dataset, we run the final prompts on our pre-processed dataset consisting of 2,034,313 sentences. We randomly reshuffle the dataset to ensure that the sentences presented to GPT-3.5 inside the prompt are not subsequent sentences inside the speeches. We adopted this approach for three reasons (i) to mitigate the risk of GPT-3.5 categorizing sentences in relation to one another within a single speech, (ii) to ensure that accuracy observed in our random validation set is indicative of the whole dataset, and (iii) to decorrelate the classification of a speech with eventual fluctuations in model behavior over time. GPT-3.5 sometimes diverges from the specified output categories or formats in our prompts. We identified the most common deviations and categorized

¹⁶Unlike our main analysis, this test was run with temperature set to 0.25, which is still far below the default value of 1, as otherwise the variation is very small.

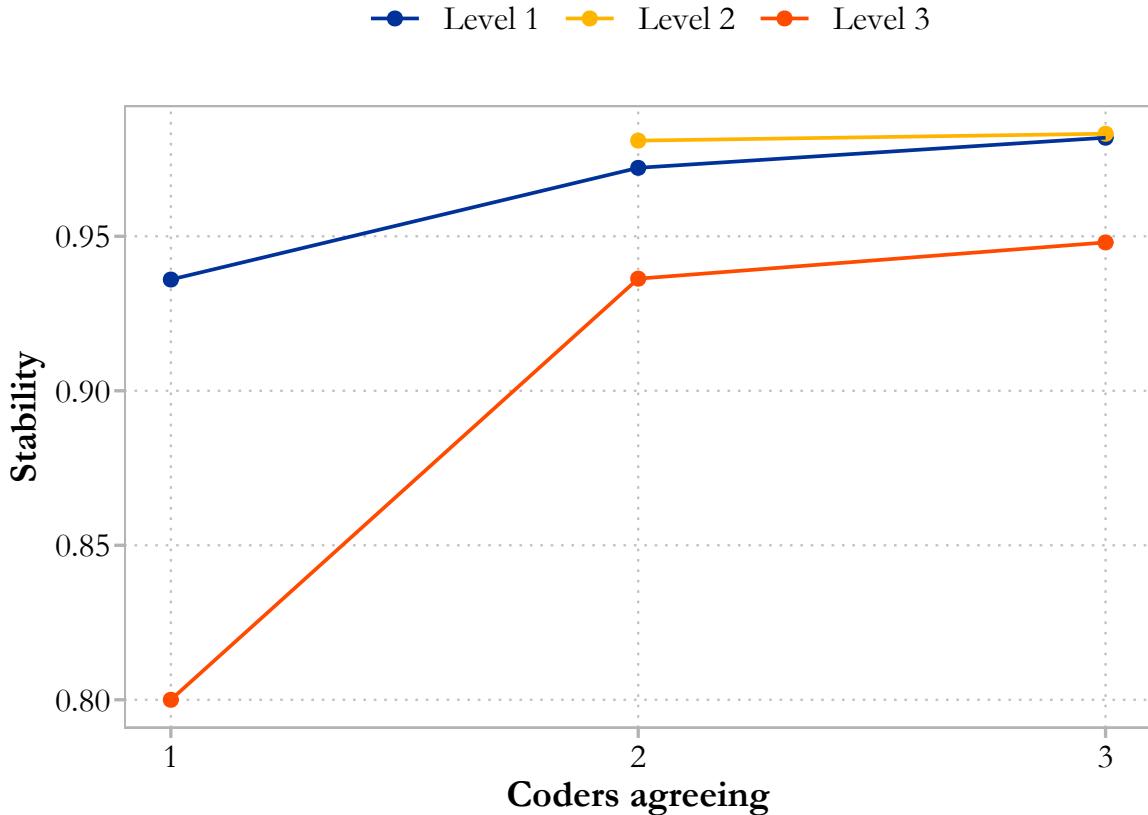


Figure 4.6: Share of sentences classified with most frequent label by agreement of human coders.

Note: Stability is measured as share of sentences that are classified identically when running gpt-3.5-turbo-0301 25 times with temperature 0.25 and our final prompts

them accordingly when the classification was unambiguous. Sentences that were not successfully classified are tried a second time. After the two attempts, the number of sentences that remain unclassified is negligible across all three levels, with level 3 having the highest count of unclassified sentences at 1,433 instances, which represents only 0.07% of the dataset.

Our classified sample is tagged with the date of the speech and the central bank. We aggregate on the central bank \times year level. Our indices of fiscal, monetary and financial dominance and coordination are turned into a proportional index ψ . We calculate the relative proportion of each of the categories $d \in D$ for all central bank $c \in C$ and year

$t \in T$ combinations:

$$\psi_{c,t,d} = \frac{\sum_{j \in J} 1(\text{Classification}_j = d)}{\sum_{j \in J} 1(\text{Classification}_j \in D)}$$

with

$$D = \left\{ \text{Monetary Dominance, Fiscal Dominance, Financial Dominance, Monetary-fiscal coordination, Monetary-financial coordination} \right\}$$

$$C = \left\{ \text{European Central Bank, Bank of England, ...} \right\}$$

$$T = \left\{ 1997, \dots, 2023 \right\}$$

where J is the set of all sentences belonging to the central bank \times year combination (c, t) . The proportions together add up to 1 which creates desirable properties for further analysis and cross-central bank and time comparison for three reasons. First, it normalises the score for the increase in central bank communication over time. Second, it normalises the score taking into account the strong heterogeneity between central banks in their frequencies of communication. Third, it creates a relative measures of policy linkages, disregarding non-relevant communication. Different kind of policy linkages – dominance and coordination – are to be interpreted as a trade-off, namely more monetary dominance implies less fiscal and financial dominance and coordination. A relative index is thus an intuitive way to describe the linkage of monetary and other policies.

4.4 Descriptive variation of the indices

In the following, we first present the descriptive results of all three coding levels, including level 3 which provides us with indices of central bank policy linkages across time and

central banks. Second, we offer some first analytical insights from simple correlations of our indices with various political-economic variables, including (i) levels of inflation, (ii) economic development levels and political regimes as well as (iii) fiscal crises.

Figure 4.7 below shows the raw frequency of the number of sentences of our level 1 topic classification over time. Plotting and examining this frequency data allows for face validity of our produced classification. For example, given the role of central banks as counterparts of the financial sector - providers of liquidity, lenders of last resort and – in many – cases banking supervisors, central bankers refer to financial developments more often than any other policy area. Notably, the frequency of financial references picked up considerably after the Global Financial Crisis (GFC) and has only recently receded. Similarly, references to climate change have recently gained traction in central bank communication from 2016 onwards, in line with known trends.

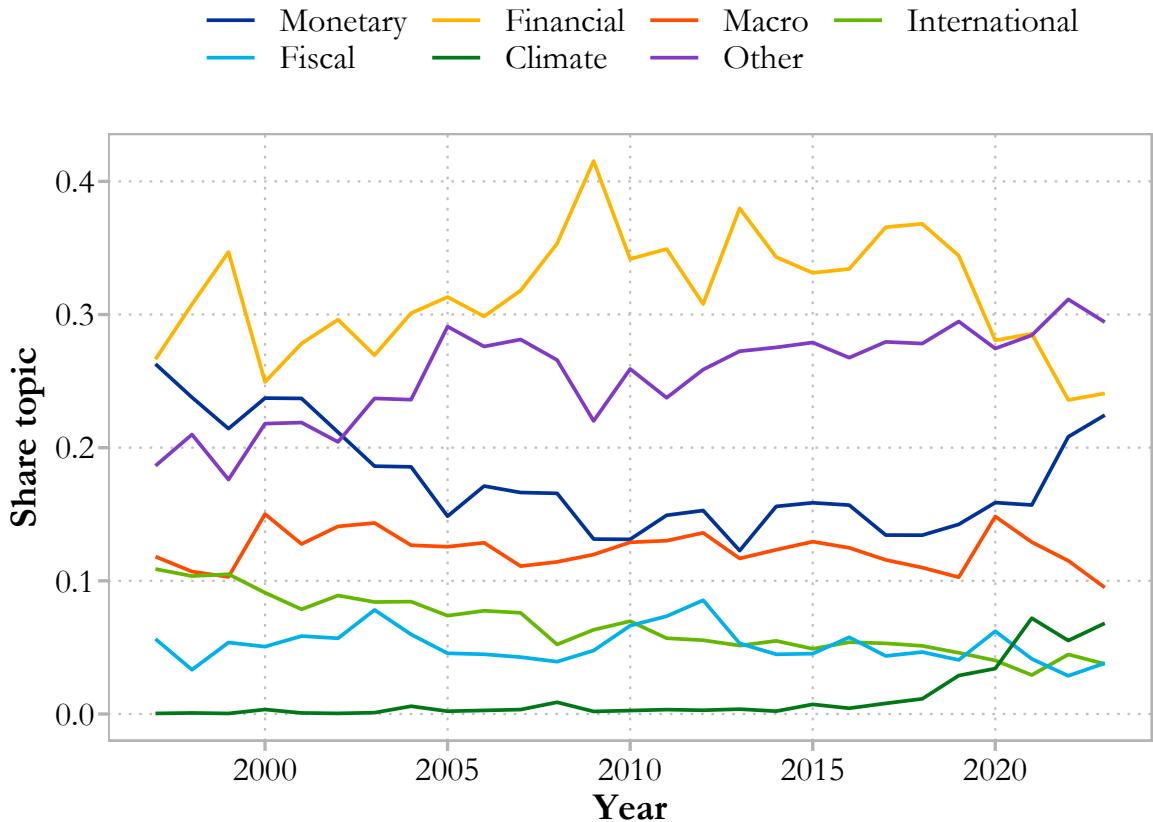


Figure 4.7: Level 1 classification results of the topic model.

These topic frequencies differ strongly among sub-groups of countries (see Table 4.6). First, central bank officials of developing and emerging economies – following the classification of IMF (2023) – talk less about monetary topics, on average, and more about financial topics. This may partly be driven by earlier financial crises in developing countries, such as the 1997 Asian financial crisis and higher financial pressure in general. In contrast, central bank officials of advanced economies communicate more about monetary policy. Nevertheless, advanced economy central banks also make frequent reference to financial topics albeit less than developing and emerging countries. Second, similarly, we find that central banks in democratic countries talk considerably more about monetary topics and less about financial topics, compared to autocratic countries.¹⁷

¹⁷We follow ? whether a country is a democracy (“liberal and electoral”) or autocracy (“closed and

Third, one can examine the relationship with macroeconomic variables such as inflation (we discuss additional macroeconomic variables in the Appendix). The frequency with which central bankers from countries with high levels of inflation refer to monetary policy or macroeconomics is somewhat lower than in central banks in low inflation countries. This can be partly due to central banks of high inflation countries having a range of other issues putting pressure on them being able to communicate about their core tasks.

Country group	Monetary	Financial	Macro	International	Fiscal	Climate	Other
Advanced	20.1%	32.0%	13.5%	5.5%	5.1%	1.3%	22.3%
Emerging and Developing	12.5%	37.4%	10.7%	5.9%	4.0%	1.2%	28.2%
Democracy	19.4%	32.2%	13.4%	5.6%	5.1%	1.0%	23.1%
Autocracy	8.8%	42.2%	9.1%	5.8%	3.2%	2.0%	28.9%
High inflation	14.0%	35.4%	10.8%	5.7%	4.3%	1.0%	28.6%
Low inflation	17.5%	34.2%	13.4%	5.7%	4.6%	1.5%	23.1%

Table 4.6: Distribution of topics

Note: Classification of advanced economies according to IMF (2023). Democracy indicator from the VDEM Dataset(Coppedge et al., 2024). High inflation countries are defined as countries with median inflation higher than the median inflation of the entire dataset from 1997-2023 (3.1%).

Figure 4.8 shows the level 2 classifications portrayed as the share of sentences classified as normative by topic. Central bank communication is most normative regarding fiscal policy. This may be due to the nature of the fiscal communication which will be uncovered in level 3. Communication on domestic and international macroeconomic developments is least normative which may reflect that when central bankers talk about macroeconomics they are often merely describing economic conditions.

electoral").

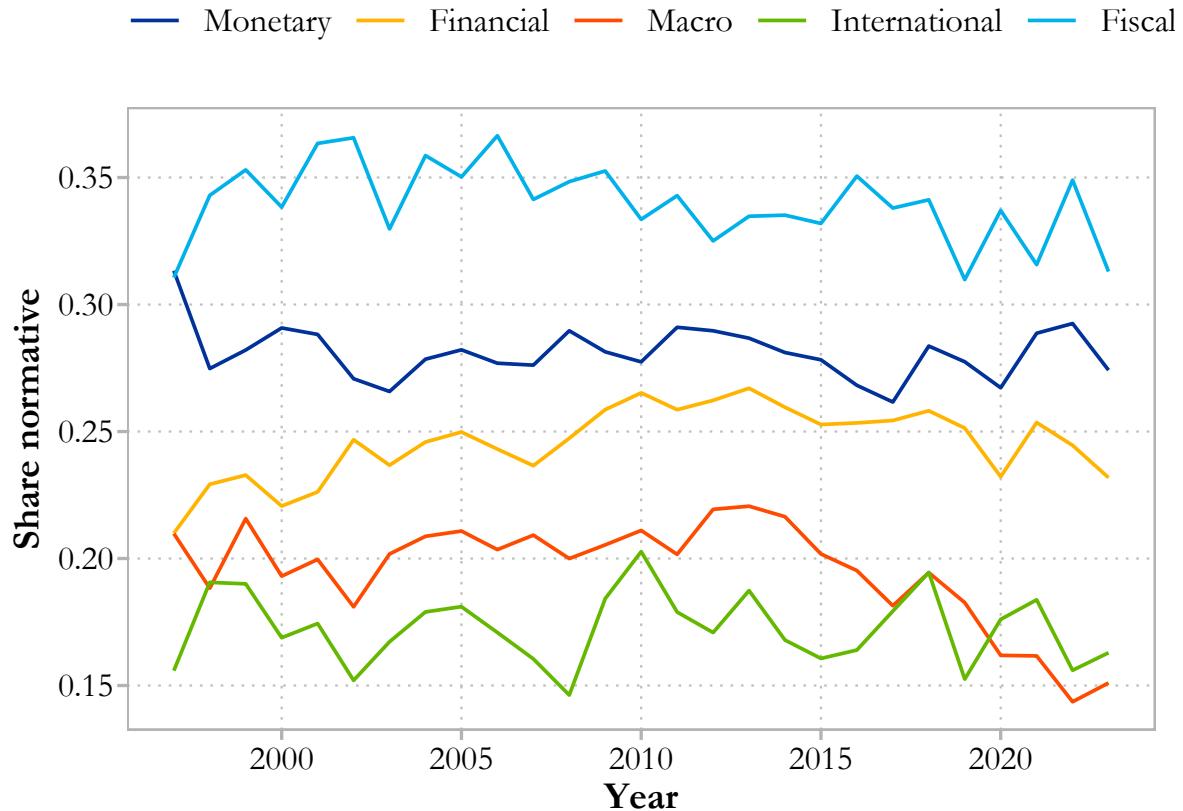


Figure 4.8: Level 2 classification results of the descriptive/normative classification.

Note: The Figure shows the share of normative sentences by topic over time. The topics 'other' and 'climate' are excluded. For 'climate' the share of normative sentences cannot be meaningfully calculated as communication on climate before 2016 was virtually zero.

Figure 4.9 below shows the classifications of level 3, which provides us with indices of linkages between monetary, fiscal and financial policy and their nature. Pooled across all central banks, our indices cover roughly three periods: (i) the pre-GFC period, (ii) the global financial crisis period and its aftermath, and (iii) the period from 2020 onward including the pandemic and the energy crisis. During these periods, one can observe three trends. First, prevalence of monetary dominance communication changed significantly over time. Namely, the pre-crisis period is characterised by a high degree of monetary dominance while following the GFC in 2007 there is marked drop in such rhetoric, which again strongly rebounds from 2020 onwards. Second, the chart also shows the indices for

monetary-financial linkages. The post-GFC period shows a pickup of rhetoric that point both to monetary-financial coordination as well as financial dominance communication. Both receded, however, during the pandemic and energy crisis. Third, the chart shows communication on monetary-fiscal linkages which remains for both fiscal coordination as well as monetary-fiscal coordination relatively muted and stable over time.

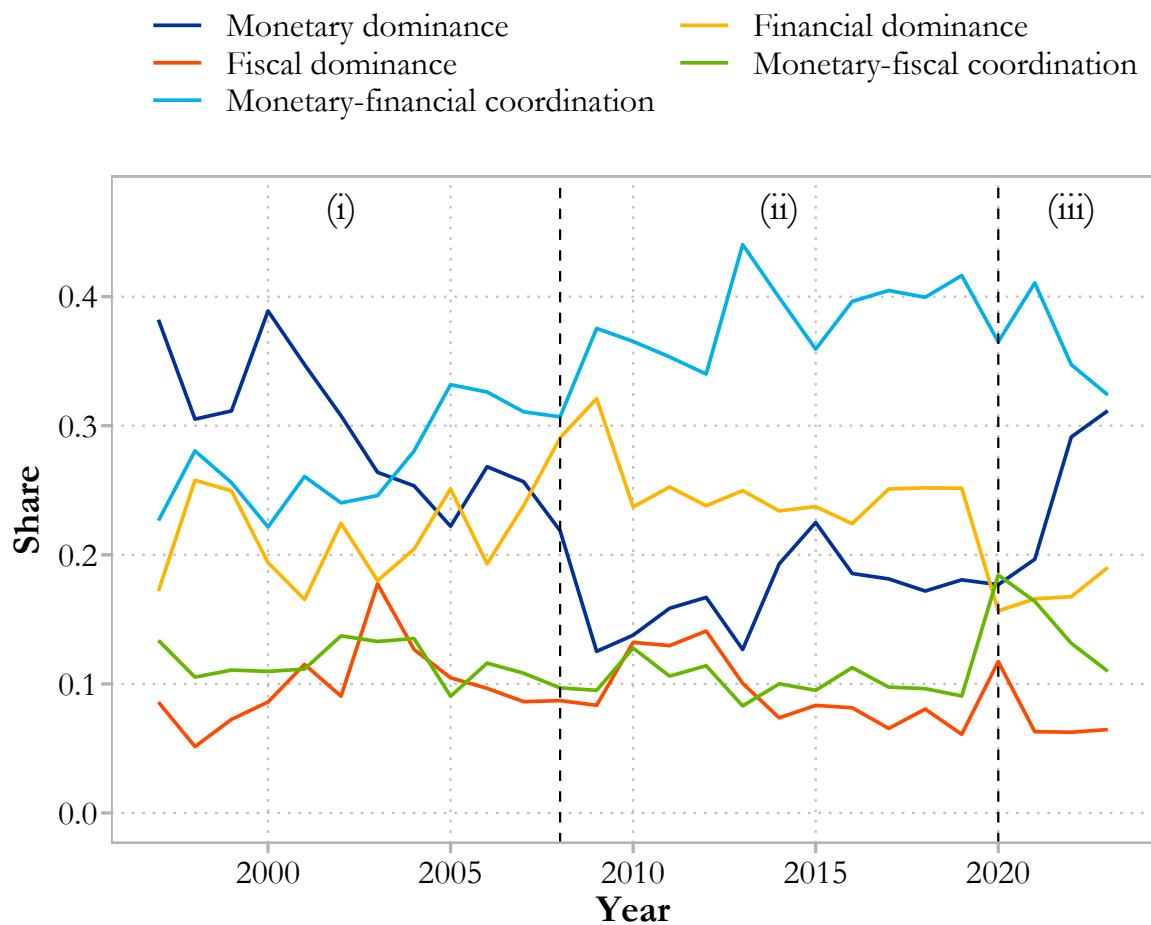


Figure 4.9: Level 3 classification results of policy coordination and dominance.

Note: The share shown is the relative index of dominance and cooperation as defined in section 4.3.5. The figure is divided in 3 time periods: (i) the pre-GFC period, (ii) the global financial crisis period and its aftermath, and (iii) the period from 2020 onward including the pandemic and the energy crisis

To highlight possible applications of the indices in economic or political economy

research we examine some basic correlations with (i) levels of inflation, (ii) economic development levels and political regimes and (iii) fiscal crises. Some first indicative results emerge that may motivate further research, three of which we present briefly in the following.

First, there is a positive relationship between the headline inflation rate and monetary dominance communication, particularly in advanced economies. Figure 4.10 shows this in the form of a scatterplot of HICP inflation and central bank-year observations of our monetary dominance indicator. This correlation may explain the observed variation of monetary dominance rhetoric over time. Namely, the post-GFC period was marked by low inflation and even deflationary pressures in many countries. Strong rhetoric on monetary dominance may have seemed out of place at the time or even counter productive where central bank experienced an undershooting of their inflation target. Supporting this, there has been a marked pickup in inflation since the pandemic and energy crisis (the third period in our data), which has gone hand-in-hand with a significant revival of monetary dominance rhetoric. Thus, in times of high inflation, central bankers may resort to strong anti-inflationary rhetoric to bolster their credibility. The heterogeneity between central banks of advanced and developing economies may be explained by Figure 4.11 which shows that monetary dominance communication is more prevalent in democratic than in autocratic countries which is highly correlated with advanced and non-advanced economies, respectively. In this context, one could argue that given the more limited *de facto* independence of central banks in autocratic and less developed countries – there is less leeway for central bankers to use strong, anti-inflationary rhetoric in these countries. Additionally, Figure 4.10 also shows that higher inflation slightly correlates with more monetary-fiscal coordination, both in advanced and developing economies. Thus, one could argue that in addition or instead of monetary dominance, central banks may also attempt to coordinate monetary with fiscal policy when inflationary pressures are higher.

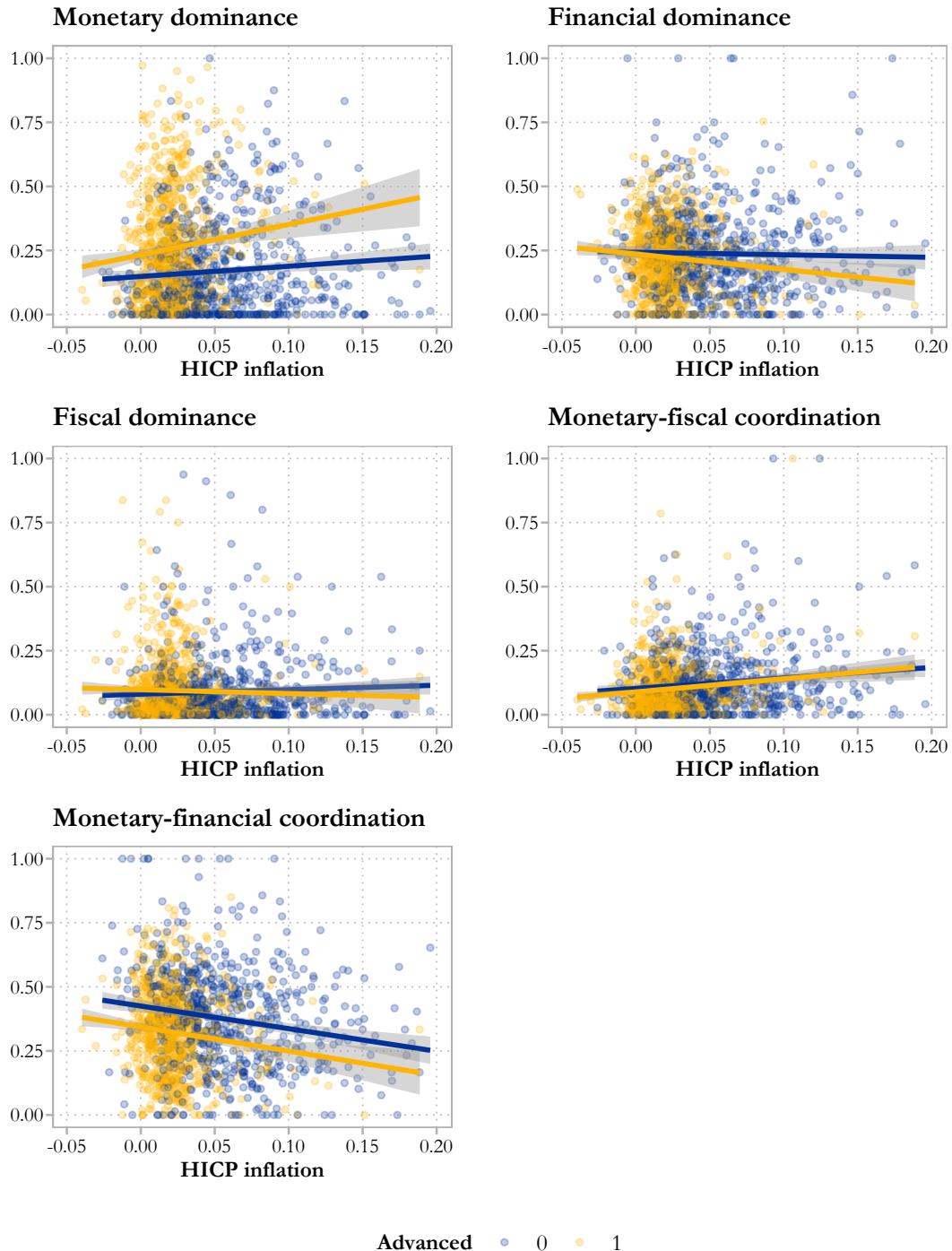


Figure 4.10: Inflation and policy linkage indicators

Note: Scatterplots of the levels of inflation and their relationship with our policy linkage indicators. Dots show central bank-year observations. The solid line indicates indicates a linear fit by advanced/emerging and developing economies. The shaded region around the regression line is the 95% confidence band.

Second, monetary-financial linkages play a larger role in autocratic and developing countries compared to in advanced and democratic countries. The scatterplots in Figure 4.11 for our various policy linkage indicators clearly show this. This should not come as a surprise since central banks of autocratic, developing and emerging economies are often highly vulnerable to external financial crises, ranging from systemic banking, currency and sovereign debt crises which originate from the financial markets (see Valencia, 2018). Moreover, developing countries with limited borrowing capabilities and small government sectors are constrained in their ability to react to financial stress. Namely, they often cannot make credible guarantees on bank liabilities or bail out defaulting financial institutions. In the literature, financial dominance has mainly been modelled formally (e.g., see Farhi and Tirole, 2012) following the idea that private borrowers and banks increase leverage because they anticipate that in a crisis the central bank will rescue them. In line with this, Gros and Shamsfakhr (2021) put forward multiple indicators to recognise financial dominance including indicators of excessive leverage or credit. Arguably, most of these indicators are more prevalent in autocratic and developing and emerging economies. This can also explain why the term "financial dominance" was first used by Fraga et al. (2003) in the context of emerging market economies with inflation-targeting regimes. Nevertheless, financial dominance is also present, albeit to a lesser degree, in developed and democratic countries, especially after the GFC. For instance, Diessner and Lisi (2020) shows this for the case of the European Central Bank. Similar trends hold for monetary-financial coordination. Notable is also that these differences mainly are present before the financial crisis while there seems to be a slight convergence since the GFC. This can be explained by the fact that following the GFC, advanced economies saw themselves confronted with similar problems developing and autocratic countries had already experienced. One can see that this convergence is less strong for monetary-financial coordination, namely there is a continuously higher need to

coordinate for non-democracies. This latter finding could also indicate that autocracies have to do more efforts in general to contain market pressures.

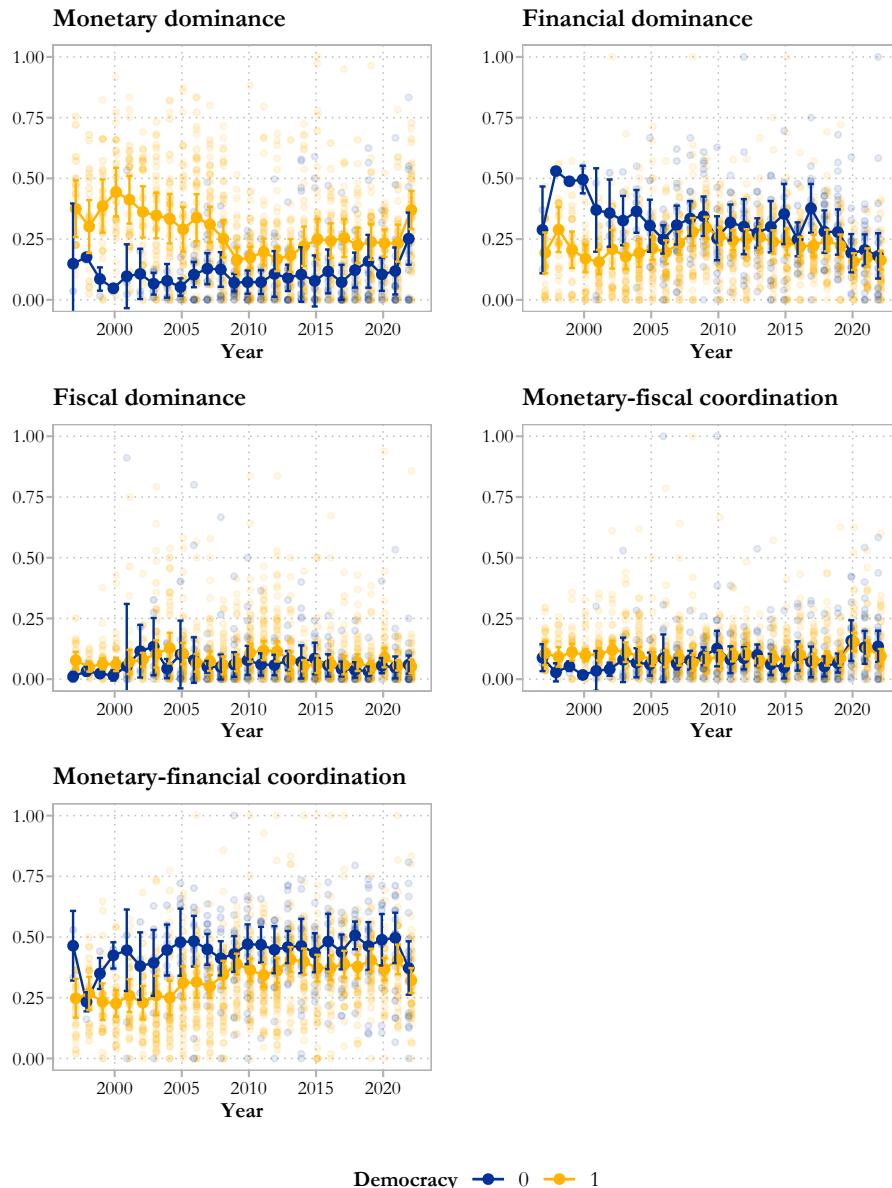


Figure 4.11: Evolution of policy linkage indicators in political regimes over time

Note: Scattered line plot of the evolution our policy linkage indicators in democracies and autocracies over time. Small dots show central bank-year observations. Observations with more speeches are more opaque. The solid line indicates the average of the category-average weighted by the number of speeches. Error bars indicate the 95% confidence interval of the weighted mean.

Third, fiscal crises have led to an increase in fiscal dominance rhetoric, especially in central banks of advanced economies. This is shown in Figure 4.12 for which we have merged our dataset with the binary fiscal crisis indicator from Xu (2017). These trends are again unsurprising when one examines the literature. In the literature, there have been multiple conditions identified under which fiscal dominance is more likely to occur including, for instance, high net public liabilities, whether a fiscal authority has exhausted its fiscal capacity already and deliberate fiscal action of flooding the bond market to force central banks to undertake action (Mengus et al., 2021). Therefore, communication regarding fiscal dominance can be expected to be higher under these circumstances. Although fiscal dominance has been identified as a phenomenon in both advanced and non-advanced countries (e.g., Sabaté et al. (2006) for an advanced country and see Makochekanwa (2008); Ersel and Özatay (2008) for two examples of non-advanced countries), in advanced countries fiscal crisis events are less frequent and, therefore, the reaction in communication can be expected to be stronger if it does occur. For instance, in the aftermath of the GFC, European countries were faced with severe tensions in the sovereign debt markets and instances of sovereign debt crisis which resulted in strong increases in fiscal communication. In contrast, Figure 4.12 shows that monetary-fiscal coordination does not show strong differences between crisis and non-crisis times and advanced and non-advanced (developing and emerging) countries. This can point towards monetary-fiscal interactions taking place both in normal and in crisis times.

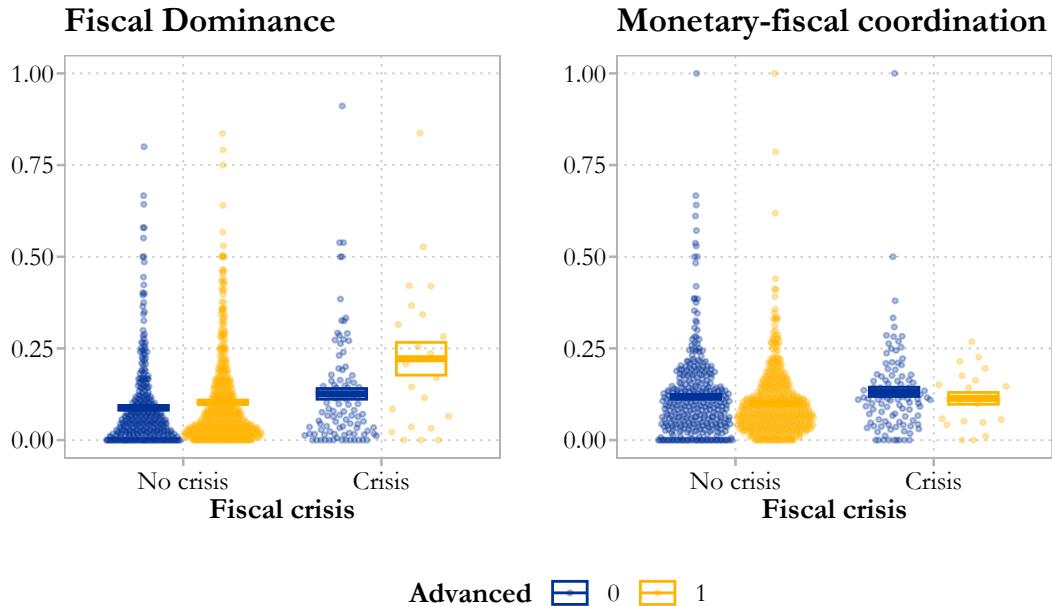


Figure 4.12: Dominance and monetary-fiscal coordination during crisis and non-crisis times

Note: Violin plot of dominance and monetary-fiscal coordination during crisis and non-crisis times. Fiscal crisis indicator from the IMF's fiscal crisis dataset (Xu, 2017). Each dot represents a central bank-year observation. Thick horizontal bars are estimates of the mean in each group. The box around the bars indicate the 95% confidence interval.

4.5 Conclusion

This chapter has developed and (manually) validated novel textual indices using GPT-3.5 to examine central bank policy linkages with governments and financial markets expressed through monetary, financial and fiscal dominance and policy coordination from 1997 to mid-2023 for 118 central banks. Using these indicators, we then provided some descriptive statistics and correlations. Overall, this chapter shows that Large Language Models, in particular GPT-3.5, are useful for classification tasks. In contrast to existing tools (such as word counts), these tools allow to pursue complex classification tasks that are challenging even for human coders. We provide various results from our prompt engineering and reflect on efficiency choices which can be used as a stepping stone for similar classification

tasks given the scarce number of use cases of Large Language Models in political economy and economics thus far. Beyond this, we also offer evidence that newer models can offer potential improvement in future use cases.

Given that our indicators are based on central bank communication, further research can examine the relationship with actual central bank policies or other model approaches (e.g., DSGE models) which traditionally have been used to detect monetary and fiscal dominance regimes. Moreover, although this chapter is mainly about measurement, we lay the groundwork for exploring relationships of the presented indices with various other variables. For instance, further research can examine how political variables specific to democratic and autocratic regimes, such as the impact of ideologically extreme governments, weak coalitions and military coups influence these indices. One can also examine the relationships with institutional variables, for instance, future research can tease out the relationships between changes in central bank independence and the indicators of dominance and coordination.

Chapter 5

Conclusion: More Than Just Talk?

5.1 Summary

This thesis has shown how central banks reconcile their independent status with extensive public communication. Central bank communication has not merely been used as a tool for information transmission but is shaped by and employed to address pressures on their independence. In three empirical chapters, I have shown how central banks across the world have used communication as a dynamic tool to address societal (to limited extent) and financial pressures and show issue-responsiveness for issues not foreshadowed in the original formulation of central bank independence. Specifically, Chapter 2 has shown how changes in formal independence alter the way central banks talk about monetary policy, demonstrating that communication is deployed to address emerging macro-economic pressures and uphold autonomy. Chapter 3 extends this analysis to the Eurosystem context, where national central banks are strategically used by the ECB to be (under certain conditions) responsive to member states - a process that likewise reinforces institutional standing. Finally, Chapter 4 provides a systematic method, via large language models, to detect how policy linkages (monetary, fiscal, and financial) are woven into speeches, exposing the deliberate calibration of communication.

I have also shown that the way in which central bank communication is used strate-

gically, heavily depends on the institutional and contextual set-up of the central bank (e.g., the level of independence, the political regime of the country and being part of a currency union). Using a comprehensive speech dataset with novel meta-data and recent developments in the applied use of Large Language Models, I have shown that these trends in central bank communication are not limited to a certain time, certain policies or certain central banks.

As is the case for any thesis, this thesis has made many assumptions due to time and space constraints. In this concluding chapter, I recognise several limitations of this thesis and, on the basis of this, also propose future research. Subsequently, I place my findings and already highlighted contributions in a broader perspective. Overall, this chapter, therefore, reflects on whether the findings in this thesis extend beyond just ‘talk’.

5.2 Limitations and further research

Below, I discuss five limitations covering substantive issues, measurement and data. I also offer suggestions on how these limitations can potentially be addressed.

First, I do not make any normative statements in this thesis regarding what the optimal level of responsiveness and communication in response to external pressures is. I only claim that a balancing act between responsiveness and independence is needed. There is already some evidence that more communication and transparency is not always better. Lohmann (2003) argues that well-designed institutions generate commitment that is both credible and flexible. Core therein is a task division between the various audiences in line with ‘audience cost’ theory in which it is posited that complex institutions, such as central banks, speak to multiple audiences in an informationally segmented way. Each audience monitors the performance of the institution on some dimensions and ignores other dimensions. The audience costs associated by these audiences shapes

the functioning and performance of the institution, thus resulting in central banks to strategically target their communication, rather than just increasing communication. Further research can further examine empirically how these central banks target various audiences in a segmented way.

Further research can also tease out whether too much or too little communication in general has negative consequences for fending off pressures and building reputation. For instance, democratic theorists can approach questions of responsiveness and communication by examining issues such as how institutions should balance expertise with public input and how transparency bolsters legitimacy. Democratic theorists can strive to clarify why a particular level of responsiveness matters for upholding fundamental values such as representation, accountability and social justice. Formal modellists can complement this by applying game-theoretic frameworks to examine the strategic dynamics of communication and reputation-building. They can establish how actors respond to pressures under varying conditions of uncertainty, resource constraints, or reputational risks. Their models can pinpoint tipping points where too much or too little communication may become suboptimal.

Second, this thesis has used central bank communication to track the transformation of central banks. Although various literature has argued that central bank communication is not ‘cheap’ talk but has actual policy implications (e.g., exerting tangible effects on market perceptions, policy predictability and economic outcomes), this can still be more thoroughly examined. Throughout the empirical chapters, I undertake several internal validity checks and externally validate the communication measures I employ with the results of previous studies. For instance, in Chapter 2 we compare the results of our policy linkage indicators with previous studies which measured the same concepts with non-textual data and methods with which we find high correspondence. By testing the mechanisms, the main measures are also further validated. In Chapter 3, various

quantitative metrics are used to validate the topics, together with manual checks of the topics.

Notwithstanding these validation efforts, future research could delve even deeper into the relationship between communication and concrete policies. For example, when central banks discuss environmental or financial stability concerns, is this discourse purely anticipatory or reflective of policy shifts already in progress? Systematic longitudinal analyses could examine whether new communications precede, coincide with, or follow tangible policy actions. This also sheds light on whether public statements on topics, such as climate change, serve primarily as a reporting mechanism (reactive) or as forward-looking signals (pro-active). Moreover, further research can also examine whether there are differences in how policy relates to communication between developing and developed economies and autocratic and democratic countries. As I already showed in Chapter 2, the type and effectiveness of communication also heavily depends on the political economic situation and institutions in place in the country.

Third, further examination can be conducted on the multi-level set-up of the Eurosystem. For instance, in Chapter 3, I examined ‘between’ topic responsiveness, i.e., whether the topic itself are copied by other central banks. I thereby overlook ‘within’ topic changes, i.e., whether the stance on the topic discussed is also copied. For instance, certain central banks can bring topics on the agenda, but the stances of other influential central banks can be followed within the debate. One could also examine whether this has changed over time with the Euro maturing as a currency. Further, in Chapter 3, I also do not directly test the full extent to which NCBs are interlocutors between the ECB and Member States. Do NCBs represent the national interests, are NCBs self-interested or do they follow the political positions of the Member States? I have argued that there are good reasons to assume the former, but this can still be systematically tested over time. Besides that, more comparative research can also be conducted in how the multi-level

set-up of central banks influences actions and views of them as (independent) institutions. For instance, it can be argued that the Eurosystem findings of the multi-level set-up are similar to the experience of other decentralised central banks, such as the US Federal Reserve consisting of regional Federal Reserve Banks which provide localised economic insights, possibly influencing the national monetary policy (e.g., see Meade and Sheets, 2005; Conti-Brown, 2017; Kocherlakota, 2017; Ainsley, 2021).

Fourth, this research has primarily relied on LLMs for textual analysis. As noted above, this is a promising avenue that opens up a range of methods to answer many questions, both qualitatively and quantitatively. However, for research purposes, I identify three issues. Firstly, most of the LLMs used in this thesis (Gemini and GPT) have the downside that they are proprietary and black boxes with APIs run by commercial third-parties. Commercial APIs are easy to use for (early-career) researchers since they have low startup costs whereby users only pay for the tokens they use. These models are trained with proprietary data and computational resources far beyond what researchers would have at their disposal. Reproducibility is close to impossible with these models. In my own experiments, I did not only notice random differences between classification runs in the results but the models were also updated without formal announcements offering different results over time. Secondly, various biases still persist in these models (e.g., see Acerbi and Stubbersfield, 2023) which has led various researchers (e.g., Boelaert et al. (2024) to conclude that Generative LLMs cannot replace humans yet(!)). These biases are partly due to the nature of its training data, namely, LLMs are trained on vast datasets containing texts (and, increasingly, images and videos) from various sources. These datasets often reflect and amplify societal biases, stereotypes and misinformation present in their original content. Thirdly, privacy is another concern with commercial APIs, since querying them typically requires sending data to the service itself. As a result, when using highly sensitive or personally identifiable data, using commercial APIs for

LLMs has risks.¹

Open source/open-weight models may offer solutions to all three problems since the data, code and hardware used to train the models tend to be publicly available.² Since the models are run on private machines and not given to a third party, data privacy concerns can be minimised. Another advantage of using open-source models is the possibility to access the raw probabilities output by the model, allowing both interpretability and for quantifying the uncertainty inherent to generation by LLM. However, having access to the raw probabilities does not per se allow for reproducibility of the results either. Namely, even if the log probabilities are accessible, the results will only be stochastically reproducible since there are hidden features that even with the most deterministic settings (e.g., setting the top-p, temperature, etc.) may still vary. Moreover, running open source models, locally or on a remote server can present several difficulties for researchers. Namely, LLMs still require computing power for inference speeds to be fast enough to be practical for applied research. These problems are exacerbated as model sizes continue to increase. Moreover, in terms of environmental costs, open source models also remain very costly, which has caught the attention of policymakers to the usage of LLMs in social sciences and beyond (e.g., see Faiz et al. (2024)). There is ongoing research to improve both limitations, namely, combined efforts by the open-source community have been able to reduce the hardware requirements to run LLMs locally through reducing the size of model weights via quantisation and offloading LLM layers to CPUs. Additionally, recent developments show that techniques such as knowledge distillation can further enhance the accessibility and efficiency of open-source LLMs.³ In terms of the environmental costs,

¹I circumvented this during my time at the European Central Bank by running the models in a protected Microsoft Azure environment. Other alternatives researchers can consider is using Amazon Web Services, but it must also be noted that both still require major set-up costs.

²For instance, Ollama is an open-source tool that runs LLMs directly on a local machine which circumvents problems with data control and privacy since full data ownership is maintained and the potential security risks associated with cloud storage is avoided.

³For example, the Chinese AI start-up DeepSeek has attracted attention by employing distillation which is a process by which a large, resource-intensive “teacher” model transfers its learned knowledge

new initiatives such as the environmental/carbon model cards may offer a starting point in creating awareness provided by Huggingface. Debates have also just started to kick off in the academic community to make about model cards (e.g., when submitting to journals).⁴

Fifth, this thesis relies on the speeches provided by the Bank of International Settlements. Although this database is very comprehensive, some central banks are better covered than others over time. For instance, the Bank of Bangladesh is completely missing from the database. I did undertake various tests to examine whether there is systematically data missing by manually checking websites of central banks and manually translating a random selection of the speeches. Although the random checks showed that some speeches were missing, there were no indications that this offered systematic problems or resulted in any biases. Nevertheless, further research can include direct contact with various central banks to create a more complete speeches database. Moreover, further checks and manual searches can be done on the websites and archives of central banks themselves. For instance, methods such as the Way Back Machine can be used to see whether central banks have excluded speeches and further archival research can improve the coverage of speeches (especially in emerging and developing countries) in

to a smaller, more computationally efficient “student” model to develop competitive models at a fraction of the traditional cost. However, DeepSeek’s use of distillation has also sparked controversy, with allegations of unauthorised replication of proprietary model outputs raising ethical and intellectual property concerns.

⁴The idea of a “model card” was first (formally) introduced by Mitchell et al. (2019) as a lightweight, standardised disclosure covering a model’s intended use, data provenance, performance and ethical considerations. Since then the discussion has broadened in three directions. (i) Community uptake: HuggingFace now ships a model card template (and a web-based writing wizard) and encourages every public model to include it; the template also supports an optional “CO2 section” that reports training and inference emissions. (ii) Environmental extensions: HuggingFace and others have begun attaching automatic carbon-tracking metadata - sometimes called environmental model cards - to improve transparency around energy use. Tools such as CarbonTracker, CodeCarbon and GreenAlgorithms offer the services to calculate the carbon usage. (iii) Towards mandatory disclosure: the ACL Responsible NLP Checklist explicitly links to model card guidance, and several venues (e.g., NeurIPS submission checklist) now ask authors to include cards or equivalent documentation. Outside computer-science venues, there seems to thus far only be conference talk to move from “nice-to-have” transparency artefacts to an emerging norm (and possible requirement) in peer-reviewed research.

the early years.

5.3 Broader implications

Building on the contributions provided in Chapter 1, Section 1.3.4, I now speculate on four broader implications following from the insights of this thesis. Specifically, the reality of seeing CBI as fundamentally dynamic, the changing importance of financial considerations, the implications of scope expansion and the need for rethinking relationships with other authorities.

First, I have argued that CBI should be seen as something requiring an ongoing effort rather than a static principal-agent relationship. While canonical theories often depict CBI as a fixed legal arrangement designed to insulate monetary authorities from political interference, empirical evidence from varied contexts reveals that independence is constantly renegotiated. This dynamic perspective implies that even when statutory independence appears robust, it can be eroded by changes such as reduced political will, economic crises or changing public expectations. For instance, in many contexts, such as Turkey, Argentina and Lebanon, formal independence has recently been undermined by political pressures, volatile markets, and acute macroeconomic vulnerabilities. Turkey's central bank under president Erdogan, has since 2016 experienced multiple dismissals of central bank governors who resisted political pressure to lower interest rates (Jazeera, 2021). Similarly, Argentina's experience with recurrent inflationary episodes shows how government demand for expansionary measures can conflict with central bank efforts to stabilise prices. Argentina's president Milei even campaigned to scrap the central bank, declaring that the local currency "is not worth crap" and describes the central bank of Argentina as "one of the greatest thieves in the history of mankind." While the dissolution of the central bank has not occurred, Milei's administration has implemented

measures to encourage the use of the U.S. dollar alongside the peso, effectively promoting a spontaneous dollarisation of the economy (Bloomberg, 2024). Another pertinent case is the Banque du Liban, Lebanon's central bank, which has been subject to factions exerting influence over financial institutions, leading to short-term political gain decisions rather than long-term stability. This has gone paired with various allegations of corruption against Riad Salameh, who served as central bank governor for close to three decades. The major financial crisis amidst political turmoil has eroded public trust in CBI, leading to the Lebanese pound losing over 98% of its value (Malsin, 2024; Times, 2024).

Second, financial pressures, such as sudden capital outflows, market turbulence or systemic banking crises, often fall outside the traditional principal-agent script that grants autonomy to central banks for the narrow task of price stability. The canonical delegation model tends to focus on mitigating political interference in monetary policy, typically by insulating monetary authorities from day-to-day electoral or legislative imperatives. However, when financial vulnerabilities escalate, e.g., through bank failures, currency volatility or liquidity shortages, the central bank often expands well beyond the original scope of inflation control. As we have seen play out, this expansion can include emergency lending, unconventional monetary policy interventions or regulatory actions that place the central bank at the heart of broader financial stability efforts. We have shown in Chapter 4 that new forms of dependence arise compelling cooperation between monetary authorities, treasuries and (international) financial institutions. As also discussed in Chapter 2, traditional central bank mandates started to include financial considerations, i.e., following the GFC, many central banks have taken on an explicit financial stability role in addition to their traditional inflation-targeting mandate. For example, after the GFC, the ECB expanded its role to supervise banks (through its integration of macro-prudential functions) and many US agencies through frameworks such as the Financial Stability Oversight Council established under Dodd-Frank have

similarly embraced a dual focus on price and financial stability (Crisanto et al., 2018). Further, global efforts such as the Basel III framework and its subsequent “Endgame” reforms, aim to harmonise capital and liquidity requirements across jurisdictions (Board, 2024).

Third, the expansion of the geographical scope of central bank communication research to encompass central banks beyond high-income democracies marks only a start. The results presented in this thesis show that different processes are at play. As Acemoglu et al. (2008) has extensively studied, the effectiveness of CBI reforms more broadly depends on the broader institutional frameworks. Thus, it is necessary to acknowledge the institutional and contextual heterogeneity that exists across diverse countries, thereby challenging the assumption that insights regarding central bank communication from advanced democracies readily apply elsewhere. A very recent example of this can be found in Nigeria, where Nigeria’s Central Bank governor, Olayemi Cardoso, announced to issue a diaspora bond in the US in 2025 to boost remittance inflows. This proactive communication to attract foreign investment would never have occurred in an advanced economy. Despite that, the effectiveness of such communication is influenced by Nigeria’s unique conditions, including reliance on remittances to restore investor confidence (Strohecker and George, 2024). Furthermore, external constraints such as volatile capital flows and exchange-rate pressures add additional considerations and complexity not encountered in more stable economic systems. Incorporating a wider range of economic situations and models also raises new methodological and theoretical challenges: from the limited availability of high-quality data in these countries to adopting context-sensitive comparative frameworks that capture the nuances of socioeconomic and political environments.

Fourth, the need for enhanced coordination with governments arises from the recognition that monetary policy alone cannot adequately address the full spectrum of macroe-

economic and financial challenges, particularly when confronted with issues such as public debt overhangs, structural unemployment or growth stagnation. In these scenarios, relying solely on interest rate adjustments or liquidity provision often proves insufficient, prompting central banks to work closely with fiscal authorities to implement complementary measures. For instance, the Reserve Bank of India under its new president Sanjay Malhotra has aligned closely with Prime Minister Modi's administration to enhance economic growth (e.g., by cutting interest rates and delaying certain banking regulations as Bhat and Dugal (2025) indicate). However, the extension of central bank functions into domains traditionally associated with government policy, such as inequality, can expose monetary authorities to heightened political pressures. This politicisation may not only erode their independence, but also complicate accountability structures, as lines of responsibility become blurred. Hence, while the current levels of coordination can enhance policy effectiveness and provide a more holistic response to economic problems, it also underscores the balance between necessary cooperation and preserving central bank autonomy. Future research can examine how formalised frameworks, legal safeguards and transparent communication channels could sustain the benefits of policy synergy while simultaneously addressing the risks associated with politicised decision-making.

5.4 Final remarks

Overall, this thesis has shown how the rise of 'central bank talk' across the world has gone paired with and reinforced central bank independence despite various pressures on central banks. Central banks have used communication to address financial and societal pressures not foreseen by the original delegation rationale, by going beyond using central bank communication as an informational tool and instead using it to address various pressures they face. Rather than acting in complete isolation, the institutional set-up of central banks has allowed for coordination with other policies and issue-responsiveness to

the publics they serve. Thus, allowing for a dynamic form and understanding of central bank independence may explain that central bank independence as an institutional set-up is lasting longer than scholars and pundits alike might have expected.

Appendices

Appendix Introduction

A1 Extra information

This appendix offers extra information regarding Figure 1.3, Table 1.1 and Figure 1.4.

A1.1 Figure 1.3

Figure 1.3 panel uses Gemini for topic modelling in two steps. First, a simple prompt is run to extract the top 10 most prevalent topics from each speech:

1 Prompt:

I'll provide you with the text of a central bank speech. Your task is to extract the 10 most prevalent topics discussed within the text. Topics should be specific but rarely more than 3 words.

Reply with a list in the form: ["topic 1", "topic 2", "topic 3", ..., "topic 10"]

Order the topics by their importance within the speech, i.e., the larger the share of the speech is devoted to a topic and the more emphasis is put on a topic, the earlier it should appear in the list.

This is the speech:

<Text of the speech>

In total 31,524 unique topics are extracted this way. The most frequent topic is, unsurprisingly, ‘monetary policy’ with 8,372 occurrences followed by ‘financial stability’ and ‘inflation’. The top panel refers to mainly topics part of the primary mandate while

the bottom panel depicts topics that were often not meant to be part of the original delegation script.

A1.2 Table 1.1

Table 1.1 uses an similar topic modelling approach as Figure 1.3. Topics are ranked based on the ratio of their relative frequencies in the speeches. For example, the topic ‘structural reforms’ appears in 14.2% of speeches from central banks with high independence and only 2.3% of speeches when CBI is low, resulting in a ratio of 6.3. The table displays the top and bottom 10 topics using this frequency ratio, considering only topics that appear more than 100 times in speeches from both groups. The relevant cut-off to be considered high CBI is 0.8 based on the data of Romelli (2024).

A1.3 Figure 1.4

Figure 1.4 uses the same speeches database with the same underlying topics as Figure 1.3. To study variation in the topics over time 256 dimensional vector representations of the topics, so called embeddings, are retrieved using the Gemini API using the “text-embedding-004” model with task-type “similarity”.

Two approaches are taken to describe diversity in the topics addressed by central banks. First, pair-wise cosine similarities ($1 - \text{cosine distance}$) of all the topics (10 per speech) within the relevant time frame are calculated. If a topic appears multiple times, it is included multiple times in the similarity matrix. Each topic’s similarity with itself is always 1. This setup ensures that frequent discussion of the same topic, for example, repeated references to monetary policy, is accurately reflected in the similarity scores.

Second, as an alternative measure of topic diversity, clusters of topics are used. This procedure first reduces the dimensionality of the embeddings to five using the UMAP

algorithm McInnes et al. (2020) and then uses the HDBSCAN Campello et al. (2013) to form clusters. Both dimensionality reduction and clustering use the same settings that are also employed by the topic modelling used in BERTopic of Grootendorst (2022).

To generate the time series displayed in the graph, the average cosine distance and average number of topic clusters are first calculated for each central bank, incorporating all speeches within a symmetric one year window. Subsequently, averages are taken for high and low independence central banks using a cut-off value of 0.8 from the dataset of Romelli (2024). The blue and red lines refer to the averages of high and low CBI banks, respectively. To limit the influence of outliers with few speeches, central banks are weighted by their number of speeches.

Appendix Chapter 2

B1 BIS dataset processing

B1.1 Speeches corpus pre-processing

All the text pre-processing steps are undertaken in python and available in the replication files (see `speech_preprocessing.py` in the `codes/l1m` folder of the replication codes).

The pre-processing follows the following steps:

1. Regular expressions

We examined the raw speeches text for recurring patterns which we remove using appropriate regular expressions. These are:

- Page numbers
- Page headers
- New page characters
- Footnotes
- URLs
- Subsequent whitespace characters

2. Conversion to sentence level

We convert the entire corpus to the sentence level using the *Punkt* sentence tokenizer from the Natural Language Toolkit (NLTK) python package Bird et al. (2009). We also tried the sentence extraction from spacy's *en_core_web_lg* model, which we found to produce similar results while being much slower.

3. Sentence level heuristics

After segmenting the corpus into individual sentences, the dataset still contains

entries that do not constitute genuine sentences of the primary text. Instead, these entries include tables, annotations or what is likely binary data belonging to e.g., a graph that is erroneously encoded in the text. To address this issue, we implement conservative rules aimed at filtering out clearly irrelevant entries:

- Remove sentences with less than 2/3 ASCII characters
- Remove sentences that consists of less than 6 tokens or more than 200 tokens
- Remove sentences with less than 20 characters

After pre-processing, we obtain a dataset that consists of 2,107,697 sentences. A small share of sentences do not contain relevant text, i.e., some chart annotations, references etc. remain. We are reluctant to more aggressively delete sentences as our LLM approach will effectively discard irrelevant text anyways by classifying them into the ‘none’ category.

B1.2 Metadata extraction

The BIS dataset as downloaded from the BIS website⁵ contains the text of the speech, the date on which it was given, the author, and a non-standardised description string which contains metadata on the speech such as the location, the speaker and the occasion of the speech. In few cases the description or the author names are clearly invalid or missing. In this case, we manually correct the description by looking up the PDF file of the speech on the BIS website and add the missing information.⁶

The metadata contained inside the description does not follow a standardised format and varies from speech to speech. Crucially for our analysis, we link each speech to the central bank the speaker is affiliated with. This is complicated by the fact that the

⁵<https://www.bis.org/cbspeeches/index.htm>

⁶There are 23 speeches with a (partially) missing description. In the case of 30 speeches there was a missing or clearly incorrect author.

description may contain multiple central bank’s names, e.g. when a member of the ECB Executive Board delivers a speech at a conference organised by Banque de France. We rely on a large language model with a few-shot prompting approach, providing appropriate instructions and examples to correctly interpret the description. Specifically, we use the Gemini 1.0 Pro to process the descriptions and extract the following metadata that is commonly found inside the descriptions:

1. Speech Identifier (e.g. b050203)
2. Type of Text (e.g. Speech, Introductory statement or Introductory remarks or Interview)
3. Name of the speaker (e.g. Jean-Claude Trichet)
4. Central bank of the speaker (e.g. Bank of England)
5. Position of the Speaker (e.g. President of the Federal Reserve Bank of Kansas City)
6. Occasion (e.g. 30th Economics Conference ‘Competition of Regions and Integration in EMU’)
7. Venue (e.g. London School of Economics and Political Science)
8. Location (e.g Frankfurt or Vienna)

In the few-shot prompt we include 6 examples to help the LLM better understand the task and the output format. See Appendix B8.2 for the exact prompt and the examples we included.

The now structured metadata still contains variations in the exact spellings of central banks, names, locations etc. We take the central bank name extracted by the Gemini and match it against a dictionary of central bank names, which we extend until all

extracted central banks match. In 27 cases, we manually assign speeches for which the description does not contain a central bank or the extracted central bank does not match the dictionary of central bank names. We clean the location and speaker columns by trying to detect non-unique spelling (e.g. Frankfurt, Frankfurt (Main), Frankfurt am Main) of locations and speakers with fuzzy string matching and LLMs to assign a unique name for each entity.⁷ The description, and in particular the venues at which speeches take place, provide information about the audience in front of which central bankers speak. We run another prompt to classify the audience of a speech as one of the following:

1. ‘Academic’, if the audience is likely to be academics
2. ‘Political’, if the audience is likely to be politicians, government officials or elected representatives
3. ‘Financial market’, if the audience is likely to be financial market actors or representatives,
4. ‘General central bank’, if the audience are central bankers or a not further specified general audience.

See Appendix B8.3 for the exact prompt and included few-shot examples. Finally, we augment the location data by assigning geographic coordinates to the extracted locations also using the Gemini LLM. The prompt and examples can be found in Appendix B8.4. Summary statistics of the selected metadata are presented in Table B1.1. 1.2

⁷Detecting which names refer to the same entity is difficult to automate. We used string similarity measures, as well as Gemini Pro 1.5 and to GPT-4 to look for potential duplicates. Both LLMs and fuzzy string matching produce false matches. We, therefore, manually built a dictionary of names based on the suggestions of the two approaches. Likely a small number of duplicates remain in the dataset.

Variable	N	Missing (%)	Unique values	Mode	Mode freq.
Institution	18787	0	124	European Central Bank	2458
Date	18787	0	6369	2014-11-17	17
Country	18734	0.28	108	Euro area	2458
Speaker	18787	0	888	Jean-Claude Trichet	478
Audience	18787	0	4	Financial market	8469
Location	17348	7.66	1385	Frankfurt	1110
Longitude	17317	7.82	2665	8.6821	1108
Latitude	17317	7.82	2270	50.1109	1106

Table B1.1: Speech metadata

Note: Each speech is associated with a central bank, date, and speaker. A small number of speeches (0.28%) is given by regulators or intergovernmental organizations like the BIS, IMF or the Inter American Development Bank. Also each of the regional Feds is considered a separate central bank with country USA. Data on geographic location, i.e. longitude and latitude are generated from the location information using Gemini.

B2 Codebook

We categorise the sentences as a form of dominance, coordination or none, in line with Leek et al. (2024). We use our categories very broadly to reflect either constraints on the conduct of monetary policy (dominance) whereby there is a form of hierarchy between monetary and fiscal institutions and financial markets while when the actors are on an equal footing we use coordination categories. See Appendix B8.1 for our exact prompt. The sentence before and after each sentence is added as context. To illustrate, Table D3.5 provides examples of sentences from our validation set and their classification including an explanation.

Classification	Example
Monetary dominance	<p>“Furthermore, monetary policy implementation in line with the market efficiency principle would need to remain without prejudice to our primary mandate of safeguarding price stability.” (Retrieved from: The European Central Bank, 14-06-2021).</p> <p><i>Explanation:</i> The topic concerns a monetary topic and they emphasize their primary mandate of price stability being above other priorities. Therefore, this sentence can be classified as monetary dominance.</p>
Fiscal dominance	<p>“Moreover, although most of the resources administered by the BIS are invested in financial assets of top quality at international level and their exposure to the various risks are managed conservatively, a greater portion of such funds could be spent toward the direct purchase of debt denominated in local currencies of emerging countries or to the use of them as collateral of certain bond issuance of countries with limited depth of their financing markets in local currency.” (Retrieved from the Central Bank of Argentina, 09-07-2008.)</p> <p><i>Explanation:</i> This sentence refers to funds being spent towards the direct purchase of debt (=monetary financing) instead of considering pure price stability considerations, thus we consider this sentence to be fiscal dominance.</p>
Financial dominance	<p>“It is thus significant that our flexible and abundant provision of liquidity contained market participants’ concerns over liquidity financing.” (Retrieved from the Bank of Japan, 04-07-2002)</p> <p><i>Explanation:</i> This sentence states that monetary policy is accommodating financial markets by providing liquidity, thus showing that financial markets are a consideration for the bank in conducting their monetary policy.</p>
Monetary-fiscal coordination	<p>“Since restarting our strategy review, we have introduced a new work stream on monetary-fiscal interactions precisely to address such questions.” (Retrieved from the European Central Bank, 30-09-2020).</p> <p><i>Explanation:</i> This sentence refers to the monetary-fiscal interactions which is a key policy in the monetary-fiscal coordination.</p>

Classification examples (*continued*)

Classification	Example
Monetary-financial coordination	"If market participants are willing to continue to work together, then we can safely achieve the transitions needed to create a better and more robust system that will help to ensure our ongoing financial stability." (Retrieved from the Board of Governors of the Federal Reserve System, 07-11-2017). <i>Explanation:</i> This sentence shows that the bank wants coordinate with market participants to ensure financial stability.

Table B2.2: Classification examples

B3 Extended dataset description

B3.1 BIS Speeches database

Our initial speeches dataset that we classify using the LLM approach contains 18,826 speeches of which 18,787 fall in the time period 1997-2023. In this time period, at least 200 speeches are included each year. 2,511 speeches cannot be uniquely assigned to a country's central bank. The vast majority of these speeches belong to the European Central Bank (2,458), which is also the central bank with the most speeches in our sample. The remaining 53 unassigned speeches belong to the Eastern Caribbean Central Bank (18 speeches), a number of intergovernmental institutions (including the Bank of International Settlements, the International Monetary fund and the Inter American Development Bank) or national regulators.

B3.2 CBI data

For our difference-in-differences analysis of the impact of institutional changes, we merge our labelled speeches dataset with Romelli (2024)'s dataset on the *de jure* independence of central banks. The extended Central Bank Independence score (CBIE) of Romelli

(2024), which we refer to as CBI in the main text and is underlying our main analysis, is measured as a continuous indicator ranging from 0 to 1. The two datasets have a large overlap. Out of the 107 countries contained in our speeches sample, 18 are not included in the CBI dataset.⁸ These countries account for 4.8% of the speeches in our sample. After dropping countries for which the CBI indicator is not observed, 17,829 speeches remain for the difference-in-differences analysis.

Our event study specification includes 12 lagged event indicators which requires CBI to be observed from 1985 (12 years before the first speeches) to use all speeches for the estimation. While in most cases the CBI observation window starts before 1985, for 24 countries, the CBI data begins only after 1985. These countries typically have comparatively few speeches. The 16 countries with the most speeches are completely observed. When the event window is not fully observed, we assume no change in independence prior to the inclusion in the dataset. Similarly, the lead event indicators require the treatment variable to be observed after the speech is given. Since our specification includes 5 leads, speeches given after 2018 would require independence changes occurring after 2023 to be observable, which lie in the future and are thus of course outside the dataset's coverage. In the baseline we forward fill the treatment indicator with zeros, assuming no independence changes in the future. In Appendix B4.5 we show that results are robust to dropping observations with incompletely observed CBI indicators.

Table B3.3 presents a overview of the countries included, its number of speeches and the coverage of Romelli (2024)'s CBI dataset. Further, the table also reports the number of CBI events as well as the relevant CBI increases for our main specification, i.e., a country's largest CBI increase with a change in the CBI score of at least 0.05.

⁸The missing countries ordered by their number of speeches are Hong Kong, Fiji Islands, Serbia, Israel, Barbados, Papua New Guinea, Kosovo, Serbia, Cayman Islands, Mozambique, Vanuatu, Armenia, Solomon Islands, Netherlands Antilles, Curaçao, Solomon Islands, Samoa, Mozambique, Cayman Islands, Guyana, Vanuatu, Aruba, Armenia, Belize.

Countries with no such event are marked with an asterisk. In more than three-fourths of the countries, the largest CBI event is either the only CBI increase or its magnitude is more than twice that of the second largest event.

1.2

Country	Speeches	CBI coverage	Number of events since 1985		Largest event since 1985	
			Increases	Decreases	Magnitude	Year
USA*	2223	1951	0	0	-	-
IND	905	1934	2	0	0.08	2016
DEU	836	1957	3	0	0.14	1994
GBR	781	1946	2	2	0.21	1998
JPN*	751	1957	1	1	0.05	1997
CAN*	565	1954	0	0	-	-
PHL*	546	1948	1	1	0.02	1993
AUS*	544	1959	1	0	0.03	1998
MYS*	525	1982	1	0	0.02	2009
SWE	494	1966	4	0	0.21	1999
CHE	405	1953	1	0	0.38	2003
ZAF	402	1956	1	0	0.09	1989
FRA	399	1936	2	0	0.39	1993
ITA	397	1948	2	0	0.57	1998
ESP	355	1962	2	0	0.44	1994
IRL	340	1942	2	1	0.37	1998
SGP	309	1991	1	0	0.10	2007
ALB	299	1992	1	0	0.21	1997
NOR	287	1966	3	1	0.25	2003

Table B3.3: Dataset coverage (*continued*)

Country	Speeches	CBI coverage	Number of events since 1985		Largest event since 1985	
			Increases	Decreases	Magnitude	Year
THA	226	1942	1	0	0.23	2008
NLD	207	1948	1	0	0.47	1998
NZL	200	1933	2	2	0.09	1989
KEN	182	1984	1	0	0.07	1995
FIN	179	1966	1	1	0.44	1998
MUS	162	1966	1	1	0.15	2008
GRC	161	1959	3	0	0.14	1994
ZMB	157	1971	2	0	0.08	2022
UGA*	154	1966	0	1	-	-
CHN*	137	1995	1	0	0.03	2003
PAK	135	1972	3	1	0.34	2022
CHL	128	1953	2	0	0.14	1989
DNK*	109	1942	0	0	-	-
TTO*	107	1964	0	1	-	-
TUR	98	1970	3	2	0.29	2001
MEX	96	1960	2	1	0.15	1993
KOR	94	1950	1	0	0.15	1998
ISL	90	1966	5	2	0.23	2001
MKD	90	1992	1	2	0.28	2003
PRT	85	1962	3	0	0.21	1998
AUT	80	1955	1	1	0.24	1998
ROU	69	1991	2	0	0.15	2004

Table B3.3: Dataset coverage (*continued*)

Country	Speeches	CBI coverage	Number of events since 1985		Largest event since 1985	
			Increases	Decreases	Magnitude	Year
LKA	68	1953	2	0	0.25	2023
IDN	65	1953	2	1	0.53	1999
MLT	59	1994	2	1	0.38	2002
CZE	58	1991	1	0	0.18	2000
GHA	57	1975	2	1	0.20	2002
BHR*	51	1973	0	1	-	-
BGR	46	1991	2	1	0.19	1997
BWA	45	1975	1	0	0.06	1996
RUS*	39	1992	3	0	0.02	1995
BEL	38	1948	2	1	0.30	1998
LUX	38	1983	2	1	0.31	1998
NAM	35	1990	1	1	0.07	2020
ARG	34	1935	1	2	0.31	1992
NGA*	33	1969	2	1	0.05	1991
LTU	30	1994	2	2	0.05	2015
MAC*	29	2000	0	0	-	-
UKR	29	1991	2	0	0.39	1999
SAU	28	1957	1	0	0.09	2020
MWI	25	1989	2	0	0.16	2019
EST	21	1993	5	1	0.10	2011
JAM	20	1992	1	0	0.20	2020
POL*	20	1997	3	0	0.03	2008

Table B3.3: Dataset coverage (*continued*)

Country	Speeches	CBI coverage	Number of events since 1985		Largest event since 1985	
			Increases	Decreases	Magnitude	Year
SYC	17	1986	4	0	0.08	2011
BHS	17	1974	2	0	0.11	2000
BIH*	15	1997	0	0	-	-
NPL	15	1955	1	0	0.28	2002
HRV	12	1991	3	1	0.40	2001
SLE	12	1963	2	1	0.09	2000
BRA	11	1964	3	1	0.25	1988
HUN	11	1991	5	1	0.24	2001
LVA	11	1992	5	0	0.15	1998
SVN	11	1991	2	0	0.34	2002
ARE	10	1980	1	0	0.07	2018
CYP	9	1963	2	0	0.31	2002
DZA	7	1962	4	2	0.08	2022
COL	6	1923	1	0	0.35	1992
KWT*	6	1968	0	0	-	-
MAR	6	1959	2	0	0.32	2006
SVK	5	1992	2	1	0.20	2001
MDV	4	1982	3	0	0.10	2020
KHM	2	1954	1	0	0.16	1996
BOL	1	1945	1	0	0.36	1995
ECU	1	1957	1	2	0.09	1998
GMB	1	1971	2	0	0.14	2018

Table B3.3: Dataset coverage (*continued*)

Country	Speeches	CBI coverage	Number of events since 1985		Largest event since 1985	
			Increases	Decreases	Magnitude	Year
GTM	1	1959	1	0	0.14	2002
JOR	1	1971	1	1	0.14	2016
TZA	1	1966	2	0	0.08	1995
URY	1	1938	2	0	0.40	1995

Table B3.3: Dataset coverage

Note: Countries marked with an ‘*’ are part of the never treated group. These countries either had no independence increase in the relevant time period (1985-2023) or their largest independence increase was below the threshold of 0.05.

B3.3 Supplementary datasets for subgroup & mechanism analyses

To conduct subgroup analysis, control for macroeconomic and political variables and study the mechanisms underlying our main analysis, we rely on additional datasets. First, we take yearly observed macroeconomic indicators and the classification of economic development into ‘advanced’ and ‘emerging and developing’ economies from the April 2024 IMF (2024) World Economic Outlook (WEO). The IMF WEO covers the full time period of our speeches sample and all 89 countries for which we have both speeches and independence indicators.

Second, for democracy indicators and the independence of the judiciary that we use in our instrumental variable approach, we rely on the VDEM database Coppedge et al.

(2024). We classify a country as a democracy if the ‘v2x_regime’ variable in the VDEM dataset is coded as either 3 or 4, and as an autocracy otherwise. This dataset also covers the full time horizon of our speeches and all countries except for Macao and the Bahamas (44 speeches).

Third, to study financial stress, we rely on the indicator of Ahir et al. (2023) which has an overlap of 64 countries with our main dataset and covers the years until 2018. Fourth, to divide our dataset into countries with free floating currencies, pegged currencies and members of a currency union, we use data from Harms and Knaze (2021). We extended the original dataset’s coverage from 2000-2021 to 1997-2023 to align with our speeches dataset. This extension involved manually updating currency areas and forward- and backward-filling for currency pegs. Currency pegs are typically long-term arrangements, so filling with the closest observed instance should lead to little error. We also verified several countries and did not find any of which its currency peg status differed from the closest observed data point in Harms and Knaze (2021).

Fifth, to analyse the role of supervision capabilities, we merge the dataset of Masciandaro and Romelli (2018) from which we obtain an ordinal supervision capabilities indicator for 71 countries that is observed until 2013. We categorize supervision capability into three levels based on the index of Masciandaro and Romelli (2018): low (1-2), medium (3), and high (greater than 3) and carry forward the last observation (from 2013) to future years. This rests on the assumption that there were no major changes to supervision capabilities after 2013. We make this assumption because most reforms took place as a reaction to the global financial crisis and have most likely been implemented by 2013. This is also backed up by looking at subcomponent III of the CBI policy dimension from Romelli (2024) which, although less granular, categorizes central bank oversight of the banking sector as either not involved, shared responsibility, or solely responsible. Out of the countries included in the sample of Masciandaro and Romelli (2018), four (Chile,

Estonia, Finland and Malta) have a change to this indicator after 2013. These countries in total contribute 462 (3.4% of sample) speeches of which 30 (0.2% of sample) were after 2013. We therefore expect the error from this assumption to be minimal.

Last, we partition central banks into two groups based on their mandate. We differentiate between mandates that have a price-stability mandate (possibly alongside other non-conflicting objectives) and mandates that could conflict with stable prices. We assign central banks to the ‘conflicting mandate’ group, if its value on objectives dimension of Romelli (2024) has a value of less than 0.5, and ‘non-conflicting’ otherwise. Our dataset comprises 9,380 speeches from non-conflicting central banks and 5,991 from conflicting ones.

To summarise the data we use throughout the study, Table B3.4 provides the descriptive statistics of all variables used in this paper.

Table B3.4: Descriptive Statistics of Main Variables

Variable	Mean	Min	Max	Std. Dev.	Observations
Textual Pressure Indicators					
Monetary Dominance	0.31	0.00	1.00	0.33	17122
Financial Dominance	0.10	0.00	1.00	0.15	17122
Fiscal Dominance	0.03	0.00	1.00	0.09	17122
Financial and Price Stability Indicators					
Inflation	0.03	-0.05	1.33	0.04	16199
Financial Stress	0.06	0.00	1.11	0.18	11727
Control Variables					
Unemployment Rate	7.17	0.70	36.02	5.08	14266
Δ Inflation Rate	0.00	-0.38	0.61	0.02	16199
Δ Unemployment Rate	-0.16	-8.57	16.12	1.17	14250
Central Bank Independence and Treatment					
Central Bank Independence Index	0.62	0.14	0.93	0.20	15371
Treated	0.52	0.00	1.00	0.50	15371
Year of Treatment	2000.74	1988.00	2023.00	8.57	9517
Ever Treated	0.62	0.00	1.00	0.49	15371
Instruments					
Inverse Distance Weighted World CBI	0.69	0.50	0.82	0.07	15371
Neighbour's Electoral Democracy Index	0.66	0.25	0.90	0.18	15371
Independence Judiciary	1.40	-3.65	2.79	0.87	16145

Note: Monetary, financial, and fiscal dominance are our textual indicators of central bank pressures. Inflation is the year-on-year change in the consumer price index from the IMF WEO. Financial stress is taken from ?. Unemployment Rate represents the share of the labor force unemployed; Δ Inflation Rate and Δ Unemployment Rate are first differences.

The Central Bank Independence Index is the extended *de jure* CBIE index from Romelli (2022). *Treated* is a binary indicator for whether the central bank associated with the speech experienced an increase in independence; *Year of Treatment* marks when this occurred; and *Ever Treated* indicates whether such an event ever took place.

Instruments include the Inverse Distance Weighted World CBI (weighted by distance from the central bank headquarters), Neighbour's Electoral Democracy Index (average of 10 closest neighbors), and Judicial Independence (V-Dem). Observations refer to the number of speeches for which each variable is observed. All statistics are computed at the speech level.

B4 Event study robustness checks

B4.1 Relaxing the parallel trends assumption

For our event study specification, to yield valid estimates of the causal effect of CBI on our communication measures, the treated and untreated units must follow parallel trends

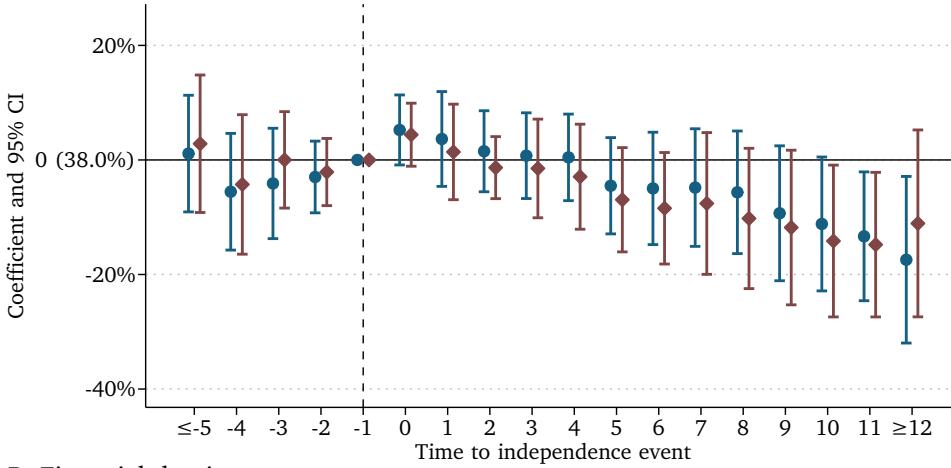
in the absence of treatment. While the generally small and statistically insignificant coefficient estimates on the pre-treatment trends lend credibility to this assumption, the parallel trends assumption remains untestable as only the treated state is observed. We can, however, relax the parallel trends assumption to some degree. For this we add a vector of time-varying control variables $\mathbf{x}'_{ict}\gamma$ to our initial event study specification:

$$\psi_{ict}^m = \sum_{k=-5}^{k=-2} \beta_k D_{ct}^k + \sum_{k=0}^{k=12} \beta_k D_{ct}^k + \mu_c + \theta_t + \mathbf{x}'_{ict}\gamma + \epsilon_{ict} \quad (\text{B4.1})$$

First, we allow for country-specific linear trends. This is achieved by adding an interaction of the country dummies that estimate the unit fixed-effects with a time-to-treat variable, i.e. $\mu_c \times (t - z_c)$, where z_c is the year in which the country changed independence.

Second, we add a set of macroeconomic time-varying control variables that could potentially confound our estimation if CBI changes are correlated with macroeconomic circumstances which are themselves not the outcome of the changes in CBI. We control for real GDP growth, HICP inflation, and the structural balance as percentage of GDP. In this case, it is enough if parallel trends hold conditional on these variables. Results of these specifications are shown in Figure B4.1.

A. Monetary dominance



B. Financial dominance

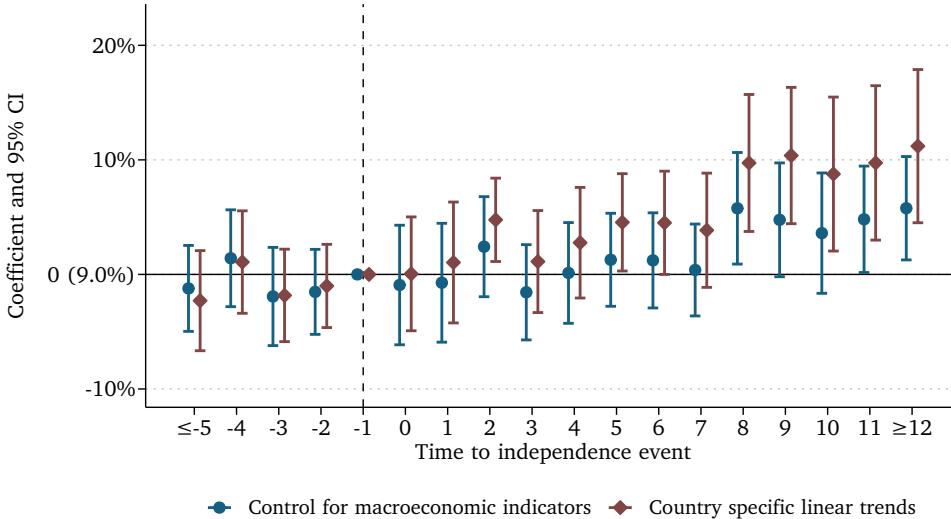


Figure B4.1: Controlling for linear trends and macroeconomic controls

Note: The event-study plots show the beta coefficients as estimated by the two-way fixed effects model (2.2) with either macroeconomic controls or country specific linear trends added. Dynamic treatment effects are estimated relative to the year before the CBI change. The number in brackets at the zero line displays the sample average of the respective dominance measure in the reference period.

After accounting for linear country-specific trends and controlling for macroeconomic conditions, the results largely align with our baseline specification. The inclusion of linear trends enhances the statistical significance of financial dominance. Consistent with the mechanisms discussed in section 2.4.4 and the coefficient of changes in inflation on

monetary dominance in our IV specification (equation 2.5), controlling for inflation leads to a slightly smaller decrease in monetary dominance. This is likely due to part of the effect of CBI operating through lower inflation.

B4.2 Accounting for treatment intensity and multiple treatments

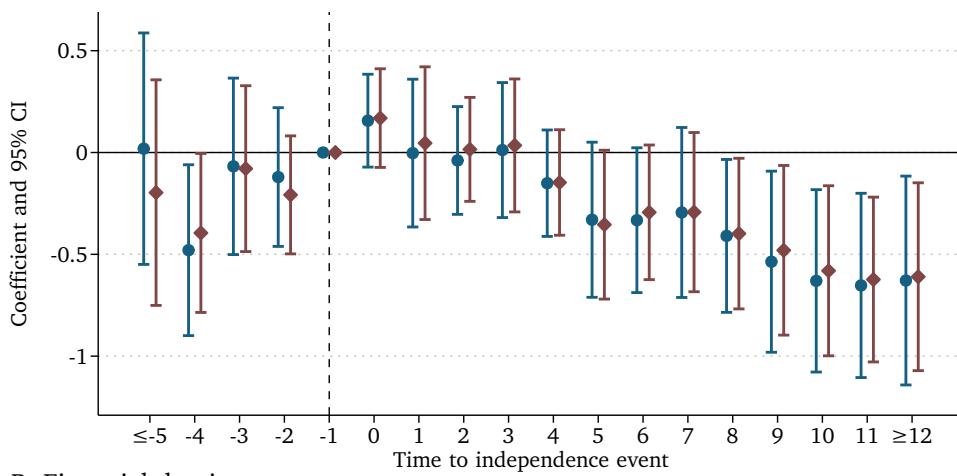
Our main specification limits independence changes to events of equal intensity and at most one per country. As briefly mentioned in section 2.4.1 and discussed in Schmidheiny and Siegloch (2023), our main event study model with binned endpoints can handle multiple treatments of varying intensities. Figure B4.2 presents results from two specifications that account for the magnitude of the independence change: (i) using the same events as in our main specification, taking into account their intensity, and (ii) taking into account all independence changes, allowing for multiple per country and decreases in CBI.

Due to the different scale on which the independence changes are measured (changes in CBI vs. 0/1 event dummies), the coefficients are of different size, but the key result from our main specification remains unchanged. Increases in CBI are associated with a decrease in monetary dominance which is substituted by increasing financial dominance.

B4.3 Heterogeneity robust estimators

In the main text, we show estimates of our study specification using the standard two-way fixed effects (TWFE) approach and the two-stage approach of Gardner et al. (2024). We also estimate comparable event study specifications using the approaches of Sun and Abraham (2021), Callaway and Sant'Anna (2021), Borusyak et al. (2024), and the stacked difference-in-differences approach popularized by Cengiz et al. (2019). The estimator of

A. Monetary dominance



B. Financial dominance

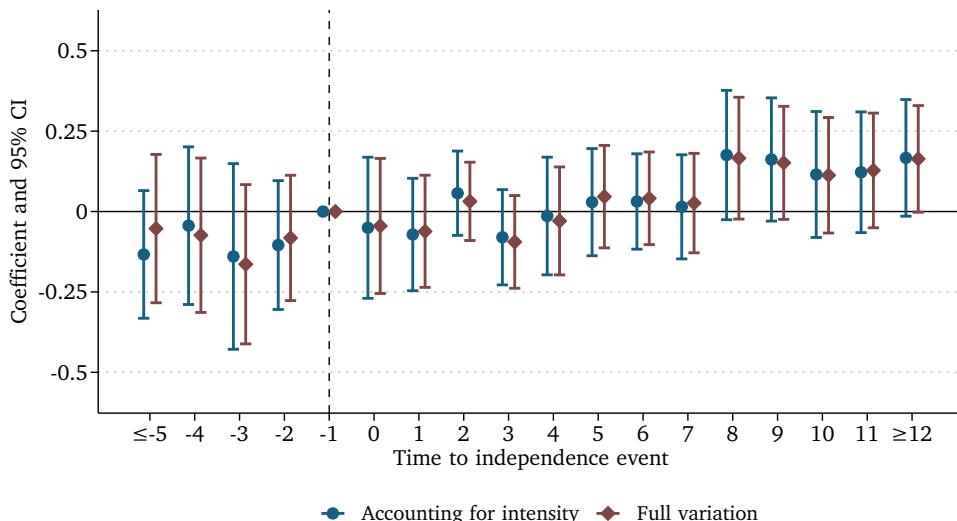


Figure B4.2: Event study with continuous treatments

Note: The event-study plots show the beta coefficients as estimated by the two-way fixed effects model (2.2), allowing for multiple independence changes and variations in intensity.

Borusyak et al. (2024) follows a similar approach to Gardner et al. (2024) and produces identical point estimates to Gardner et al. (2024) in post-treatment periods. Callaway and Sant'Anna (2021) and Sun and Abraham (2021) rely on estimating group- and time-specific treatment effects which can be aggregated and normalised to resemble our main specification. The stacked difference-in-differences (stacked DiD) approach is a variation of the standard two-way fixed effects event study estimator. It constructs sub-samples for each treatment year cohort and then stacks the individual datasets to jointly estimate dynamic treatment effects using cohort-specific two-way fixed effects. Each of the sub-samples contains only the observations treated in a given year and the never-treated observations. Therefore, identification of the effects is solely based on the comparison of treated units against never-treated units. This avoids the problematic comparisons of units that received treatment at different times that are the root cause of the bias of two-way fixed effects under heterogeneous treatment effects Borusyak et al. (2024); Goodman-Bacon (2021).

In Figure B4.3 we report our main event study design estimated using Borusyak et al. (2024), Sun and Abraham (2021) and the stacked difference-in-differences estimator. In the case of Sun and Abraham (2021), we bin coefficients outside of the event window to maximise comparability with the other estimates. The results obtained with these alternative estimators closely resemble the TWFE and Gardner et al. (2024) specifications reported in the main text. Table B4.5 reports aggregated estimates. The table also includes Callaway and Sant'Anna (2021)'s estimate.⁹

⁹The full event study design can also be estimated using the Callaway and Sant'Anna (2021) approach. However, the bootstrapped standard error estimation does not work well with our relatively spaced out treatments and the option to set a ‘universal’ base period, which is necessary to make the estimated coefficients comparable to the other estimators and is currently not functioning correctly when the panel is unbalanced. We, therefore, only report the aggregated coefficient.

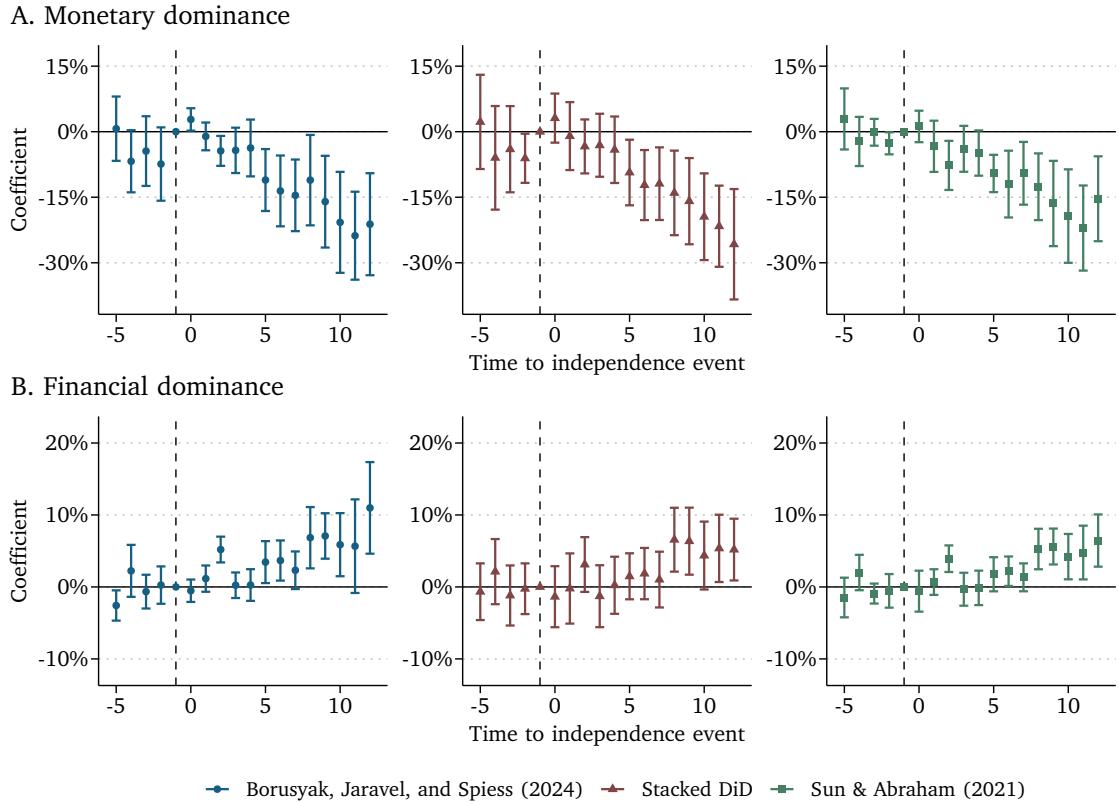


Figure B4.3: Alternative event study estimators

Note: The columns show heterogeneity robust event study estimators which correspond to same setup as in the TWFE equation (2.2). The vertical bars represent the 95% confidence intervals for the estimated coefficients.

B4.4 Dynamic treatment effect aggregation

To illustrate differences in effect magnitude between sub-groups it is often desirable to condense the effects of changes in central bank independence into a single number. Since the effect builds up over time, it is not obvious which coefficient(s) should be used to infer the ‘overall’ effect. A straight forward option is to run a simple static difference-in-differences (DiD) estimator with a single treated indicator which turns on after the independence change. However, as shown by Goodman-Bacon (2021) such a specification can be problematic when treatment adoption is staggered as it may not be an intuitive weighted combination of all possible two-group/two-period difference-in-differences estimators. To avoid the weighting problems specific to the this approach, we

aggregate the estimates from our two-way fixed effects event study specification by taking means of all post-treatment coefficients. We aggregate coefficients using two weighting schemes: (i) with equal weights of all post-treatment coefficients and (ii) weighted by the number of observations that are in our dataset at each lag. Standard errors are averaged accordingly, taking into account the covariance structure of the estimated coefficients. Differences stem mainly from the much higher weight that is assigned to the last (binned) lag when weighting by the number of observations. We prefer this approach as it produce estimates that are more similar to the fully heterogeneity robust estimators. Table B4.5 compares the two-way fixed effects aggregation against the static DiD estimator, as well as a set of estimators that were developed to address the weighting problems occurring under treatment effect heterogeneity. We use the approaches of Borusyak et al. (2024) and Gardner et al. (2024) to estimate the static difference-in-differences specification with a corrected treatment indicator. In addition, we report single coefficient estimates using the estimators from Callaway and Sant'Anna (2021) and Sun and Abraham (2021) which themselves rely on aggregations of group and time specific estimates. The table also reports an aggregation of the ‘stacked’ two-way fixed effects estimator which divides the sample into sub-experiments by year of treatment (see section B4.3 for more details).

We find overall similar effect sizes among the heterogeneity-robust estimators. The coefficient estimates from our two-way fixed effects average are somewhat smaller, especially for monetary dominance. This is expected as the average assigns relatively little weight to the long-run effect, i.e the periods after the event window, because the long run effect is captured by the last coefficient which receives no larger weight than the coefficients that only capture a single period. The static estimate is much smaller than the other estimates. Given the substantial number of independence events in the earlier years of our sample, a likely explanation is that high weight is placed on early versus later treated comparisons, which are the source of the bias of the static estimator under

	Effect on dominances	
	Monetary	Financial
Two-way fixed effects		
Static difference-in-differences	-0.0165 (0.0379)	0.0192* (0.0112)
TWFE aggregation (equal weights)	-0.0818** (0.0380)	0.0327* (0.0195)
TWFE aggregation (observation weighted)	-0.1607*** (0.0554)	0.0548*** (0.0200)
Stacked DiD (equal weights)	-0.1067*** (0.0333)	0.0251 (0.0183)
Stacked DiD (observation weighted)	-0.1963*** (0.0496)	0.0411** (0.0199)
Aggregation based		
Sun & Abraham (2021)	-0.1195*** (0.0354)	0.0329*** (0.0112)
Callaway & Sant'Anna (2021)	-0.1720*** (0.0572)	0.0504** (0.0251)
Imputation based		
Borusyak, Jaravel, and Spiess (2024)	-0.2019*** (0.0514)	0.0436** (0.0188)
Gardener (2022)	-0.2019*** (0.0593)	0.0436** (0.0206)

Table B4.5: Single coefficient estimates

Note: Stars indicate significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. This table compares estimators which aggregate the post-treatment effect into a single number. The two-way fixed effects (TWFE) aggregation is used to aggregate the specification underlying the subgroup specific treatment effects reported in Table (2.2)

heterogeneous treatment effects Goodman-Bacon (2021).

B4.5 Sample and treatment indicator variations

In this section, we perform robustness checks on our baseline sample, assessing the impact of various decisions made during its construction. We look at the effects of omitting observations with non-observable leads, test for anticipation effects, exclude never-treated observations and define independence events based off sub-dimensions of CBI.

First, we address the treatment status of observations with incompletely observed CBI indicators. Given that our speeches cover the years from 1997 to 2023 and that there are 5 leads and 12 lags included in our event study specification, the full event window for which treatment indicators are required extends from 1985 to 2028. Incomplete observations of treatment status occurs due to two factors: (i) Romelli (2024)'s CBI coverage sometimes begins after 1985, though this primarily affects countries with fewer speeches (see Appendix B3) and (ii) speeches recorded from 2018 onwards have lead indicators extending beyond the independence dataset's 2023 limit. Our baseline specification utilises all available speech data, filling unobservable past and future CBI change indicators with zeros. This entails the assumption that independence increases occur past 2023, which, while speculative, is justified given that most central banks currently exhibit high independence levels, making substantial increases unlikely. Nevertheless, we provide an alternative specification here that excludes observations with incompletely observed event windows.

Second, we account for anticipation of the independence changes and alternatively estimate a specification where all independence changes are shifted to occur one year before the independence change is recorded in Romelli (2024). In the event study plots, effects are, therefore, reported relative to the period two years before the actual independence change. Given that our effects only materialize with a delay in the main specification, we expect the effect of accounting for anticipation to be small.

Third, we consider the comparison groups for the event study estimation. About one-third of the speeches are from central banks that have never undergone an independence change. This most notably includes the central banks of the Federal Reserve System, which contribute over 2,000 speeches. One might argue that these never treated central banks are fundamentally different from those experiencing independence changes, suggesting that the latter may serve as a better comparison group. To test this, we estimate a

specification that drops all never treated observations.

Figure B4.5 displays all three sample variations estimated using the two-stage approach of Gardner et al. (2024). Dropping not fully observed observations and pre-dating independence changes result in quantitatively similar estimates to the main specification (see Figure 2.9). Relying only on never treated units confirms the findings from the other specifications regarding monetary and financial dominance. In addition, this specification is able to identify a pronounced increase in fiscal dominance following central bank independence. Estimated pre-trends are somewhat more stable around zero in this specification as well, hinting at the fact that observations that are eventually treated may serve as better control observations. Due to the large loss in sample size, we keep the never treated observations in our baseline specification.

Last, we consider changes in sub-components of the CBI indicator as the relevant treatment variable. We impose the same restrictions as in the baseline specification, namely that only increases of at least 0.05 are considered. Should there be more than one such event for a country in the time period 1985-2023, only the largest increase is considered. Figure B4.4 plots the overall effects as estimated with Gardner et al. (2024) and our observations weighted TWFE aggregation for each CBI sub-indicator contained in Romelli (2024)

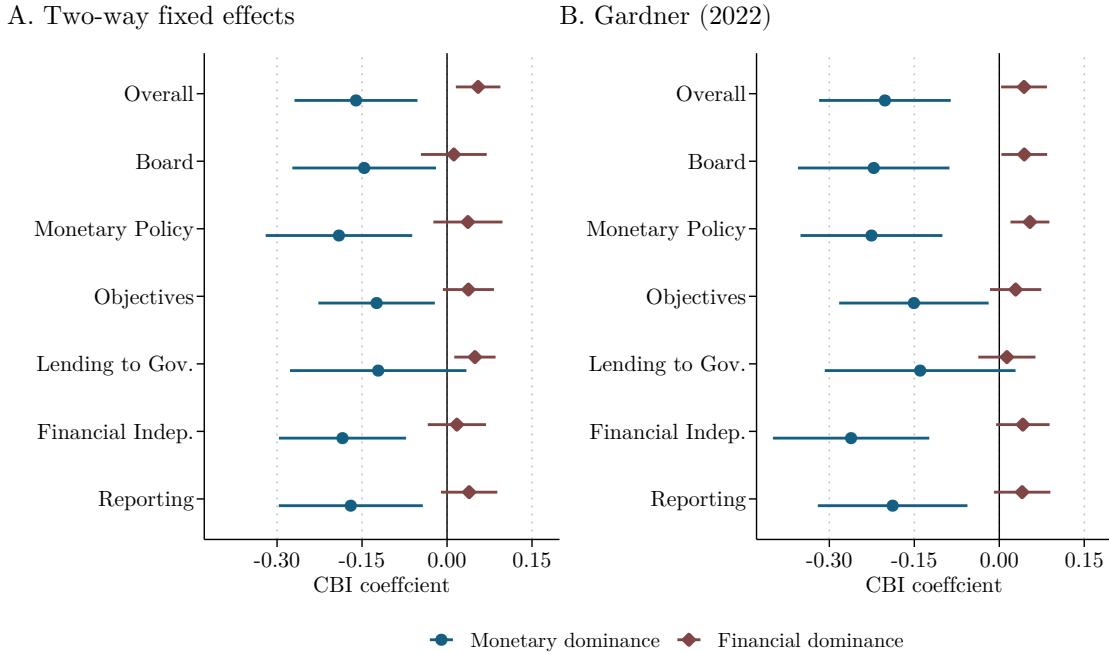


Figure B4.4: Effect of CBI sub-indicator changes on dominances

Note: Panel A plots the overall effect as estimated with the TWFE event study specification and aggregated into a single coefficient by weighting the dynamic treatment effects by the number of speeches observed at each treatment time. Panel B estimates a static difference-in-differences model with a single post-treatment coefficient using the two-stage procedure of Gardner et al. (2024). Error bars indicate the 95% confidence interval.

We find similar coefficient estimates across most dimensions of CBI, explained by a relatively high correlation among the CBI sub-indicators. Based on our definition of an independence event using the overall CBI indicator, we observe that typically (in the median case) an independence event entails changes in three out of the six CBI sub-indicators.

B4.6 Effect of CBI on audiences

In the main text, we provide a theory and empirical evidence that CBI alters the policy pressures central banks face, which then also result in these pressures being reflected in communication. In particular, we provide evidence that CBI leads to increased financial

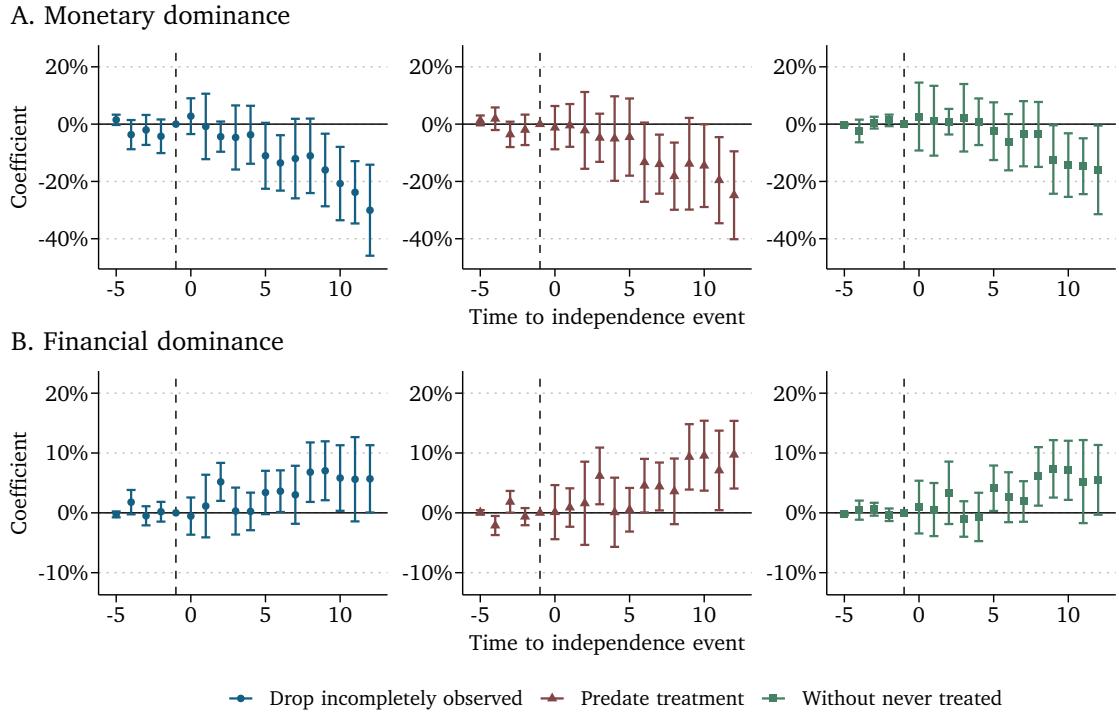


Figure B4.5: Robustness sample construction

Note: The panels show the three variations in the estimation sample discussed in Appendix B4.5. The event studies are estimated using the two-stage procedure of Gardner et al. (2024). The vertical bars represent the 95% confidence intervals for the estimated coefficients.

pressures. In section 2.4.5, we rule out several alternative explanations such as the effect being driven by the financial crisis or supervisory changes. Another confounding channel could be that more independent central banks enjoy more freedom in which audiences they choose to address. To rule out the possibility that the observed effect is merely driven by who central banks speak to, we use our event study design to determine the effect of CBI on the audience of the speech. We classify the audience of the speech based on the description in the BIS dataset as one of ‘Academic’, ‘General Central Banking’, ‘Financial Market’ and ‘Political’ (See section B1.2). Below, in Figure B4.6, we report four event studies where the dependent variables are binary indicators corresponding to each audience.

The graphs provide evidence that central banks do not target their speeches towards

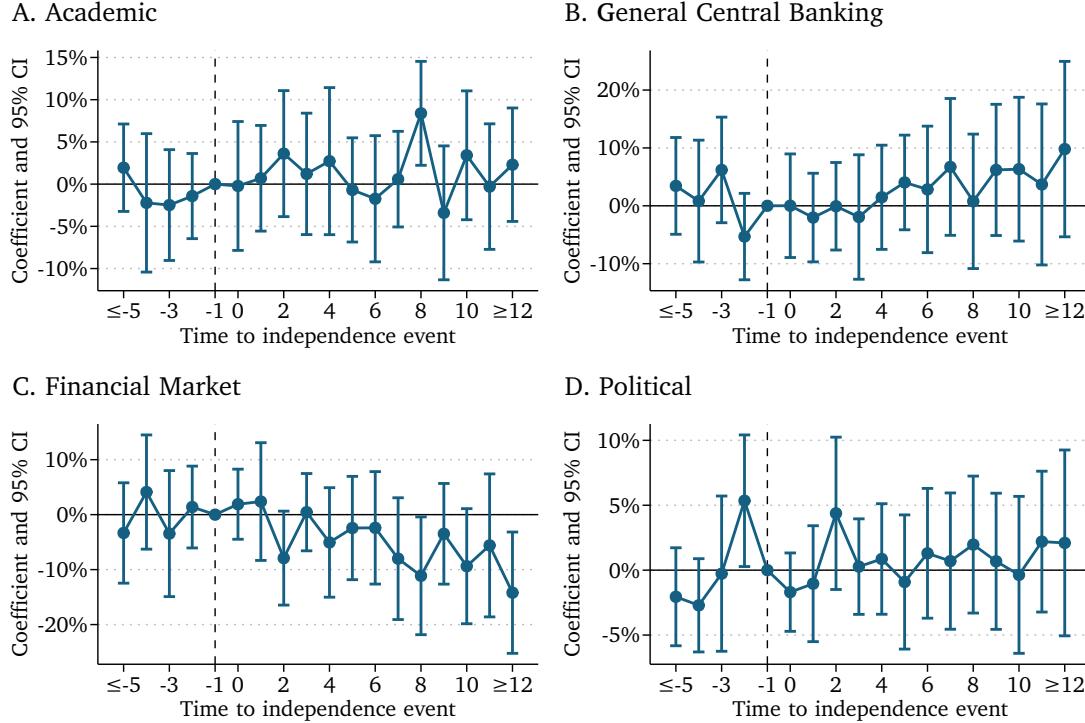


Figure B4.6: The effect of CBI on audiences

Note: The panels show our event study design (2.2) estimated using TWFE with a binary indicator indicating the audience of the speech as dependent variable. The vertical bars represent the 95% confidence intervals for the estimated coefficients.

different audiences after obtaining independence. If there is any effect, point estimates point towards a reduction in communication directed to financial markets. It is therefore not plausible that our observed effects are explained by different audiences, and in particular not by an increase in speeches directed at financial markets.

In addition, we estimate heterogenous effects by audience (as we do in the main text with equation (2.4)). Results are reported in Table B4.6. Point estimates have the same signs and are of a similar magnitude throughout. With the exception of political audiences, we also find significant effects in all subsamples. However, with only 1775 speeches this is the least addressed audience and is likely suffering from low power to detect an effect. Thus, we conclude that audiences do not seem to play a large role in

	Monetary dominance	Financial dominance
Academic	-0.1426** (0.0559)	0.0783*** (0.0197)
General Central Banking	-0.1419** (0.0579)	0.0514** (0.0204)
Financial Market	-0.1917*** (0.0553)	0.0461** (0.0204)
Political	-0.1047* (0.0586)	0.0218 (0.0208)

Table B4.6: Effect of CBI by audience

Note: Stars indicate significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The table contains estimates of the effect of CBI on monetary and financial dominance by audience of the speech estimated using equation (2.4)

our analysis.

B4.7 Alternative CBI dataset

In this section, we repeat our estimation using the CBI dataset of Garriga (2025a). The author provides two CBI measures, “LVAW” and “LVAU”, which are weighted and unweighted averages of the traditional four dimensions of CBI identified by Cukierman et al. (1992), namely (i) personnel independence, the (ii) objectives, (iii) independence in policy formulation and (iv) limitations on lending to the government. These four dimensions are common to our main dataset Romelli (2024), which in addition also tracks reporting and transparency (v) as well as financial independence (vi). To facilitate comparisons between the datasets, we choose “LVAU”, the unweighted measure from Garriga (2025a), as Romelli (2024)’s extended CBI index “CBIE” is also a unweighted average over all its dimensions.

There are slight differences in the coverage of countries, with Garriga (2025a) covering 14 more countries from our speeches dataset that together account for 557 speeches (see note below Figure B4.7). Only one central bank, Macao, is tracked by Romelli (2022)

but missing in Garriga (2025a), with a total of 29 speeches.

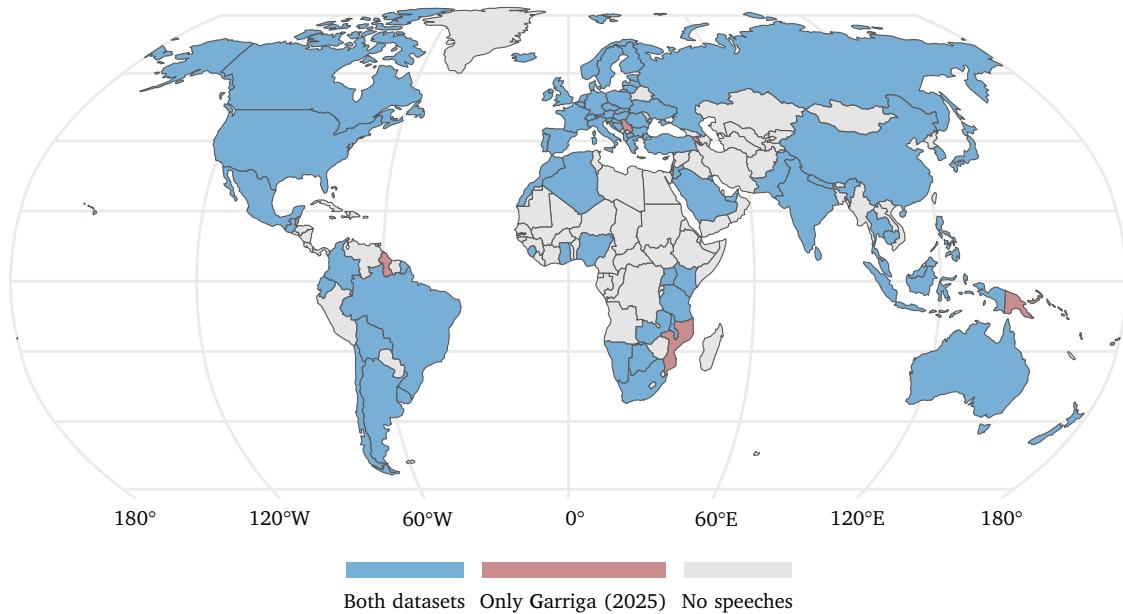


Figure B4.7: Dataset coverage and number of independence changes

Note: This map compares the coverage of the dataset used for our empirical analysis Romelli (2024) and Garriga (2025a). Countries shown in red are only available in Garriga (2025a). These are: Aruba, Armenia, Belize, Barbados, Cayman Islands, Fiji, Guyana, Israel, Mozambique, Papua New Guinea, Solomon Islands, Serbia, Vanuatu, Samoa. Together these countries account for 557 speeches with Fiji, Israel and Serbia each contributing more than 100 speeches. Countries for which we have no speeches are shown in grey, no matter their inclusion in any of the two datasets.

To focus the comparison of the differences in the construction of the CBI score, the estimates below use the intersection of the two samples. However, note that including the additional speeches leaves the estimates practically unchanged. We first descriptively look at how the CBI scores correlate in Figure B4.8.

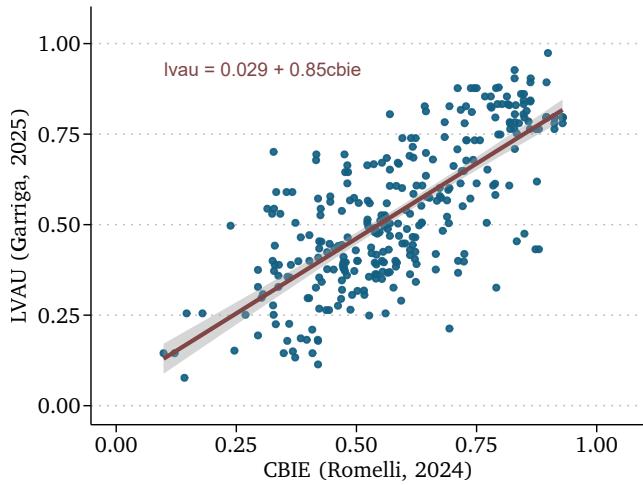


Figure B4.8: Correlation of CBI scores

Note: The plot shows all unique CBI \times country \times year combinations with observations in both datasets. The red line represents a linear fit of a OLS regression of LVAU on CBIE.

As expected, there is generally a strong correlation between the two measures. Since they are both normalised to a 0 to 1 scale, we apply the same approach described in section 2.4.1 of the main text to construct independence events from Garriga (2025a) dataset. That is, we take the largest increase in CBI from the time period 1985–2023 that has a magnitude of at least 5 percentage points. Figure B4.9 displays the number of CBI changes per year in both datasets. Events that are not considered for the event study are shaded. This includes events with a magnitude below 0.05 and events for which a larger independence increase is observed within in the same central bank during the period 1985–2023.

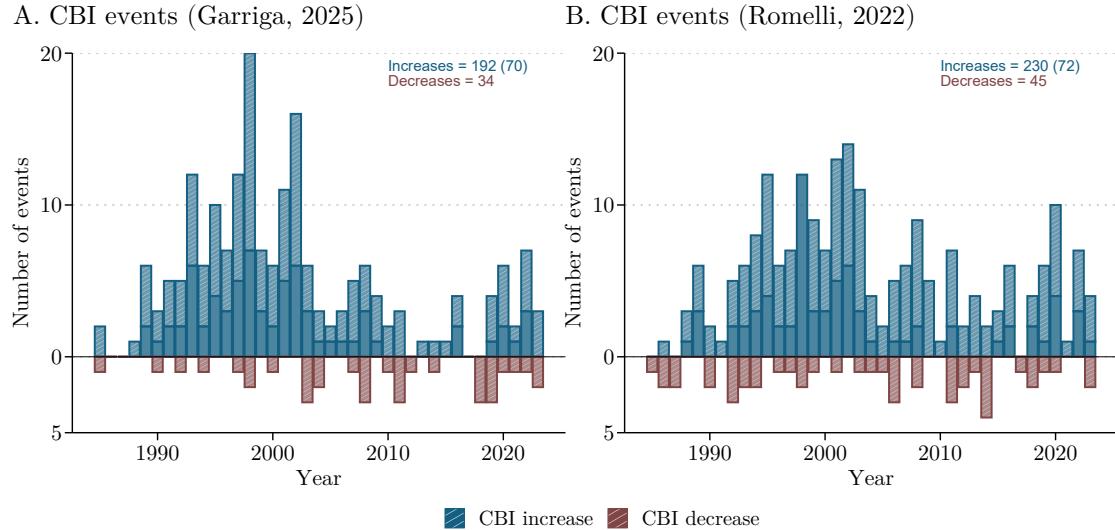


Figure B4.9: Comparison of CBI events

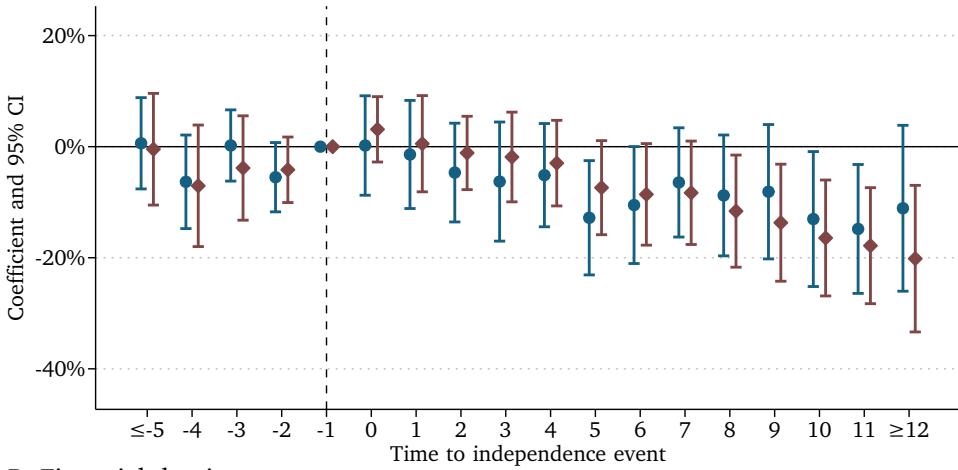
Note: Both panels display the number of CBI events over time, categorised into increases (blue bars) and decreases (red bars). Events excluded from the event study design are shaded. In the top right, the total number of events is shown with the number of relevant increases for the event study in brackets.

Again, both datasets are largely similar. Romelli (2024) observes slightly more CBI events in both directions. CBI increases are most frequent around the year 2000 with the year 1998. The formation of the ECB sticks out in both datasets as the year with most high magnitude events. In Romelli (2024), the year of the establishment of the ECB, is considered the largest event for all countries that experienced an event in that year. However, in Garriga (2025a), a large proportion of the independence events in the same year are not considered in the event study either because the independence increase was too small or another event of larger magnitude was observed. This suggests that, relative to other instances of independence increases, the formation of the ECB is perceived as a comparatively smaller step towards independence.

Next, in order to assess whether our main results are robust to the alternative CBI indicator, we compare estimates of our main TWFE event study specification (2.2) using both datasets. The results are shown in Figure B4.10. We observe largely similar estimates

both in terms of pre-trends and the effects of independence. Using Garriga (2025a)'s dataset, we find slightly smaller dynamic treatment effects. In particular, the levelling of coefficients β_{12} are (barely) not significant on the significance level $\alpha = 5\%$. However, we can still reject the joint null of no treatment effect, i.e. $H_0 : \beta_0 = \beta_1 = \dots = \beta_{12} = 0$, for both financial ($p = 0.0325$) and monetary dominance ($p = 0.0151$). Our main results are therefore robust to the alternative CBI indicator.

A. Monetary dominance



B. Financial dominance

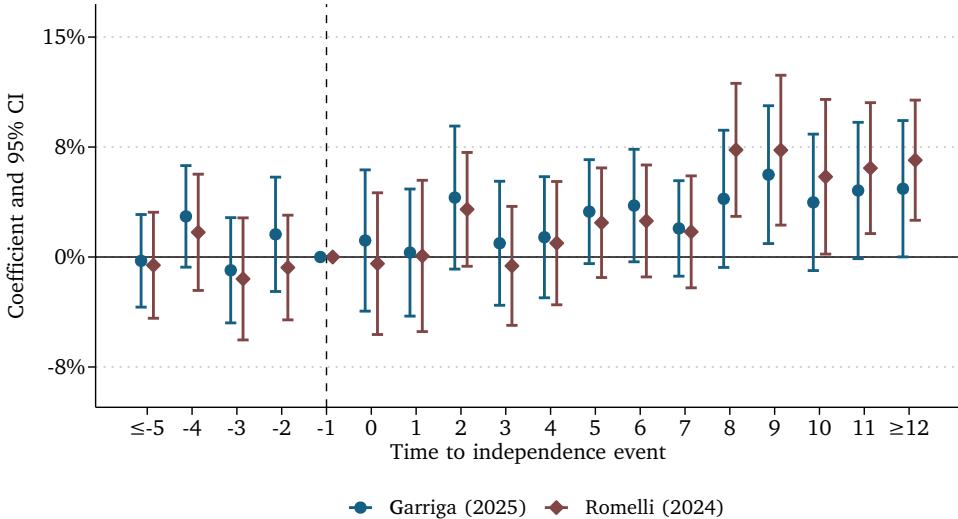


Figure B4.10: Event study estimates with different CBI datasets

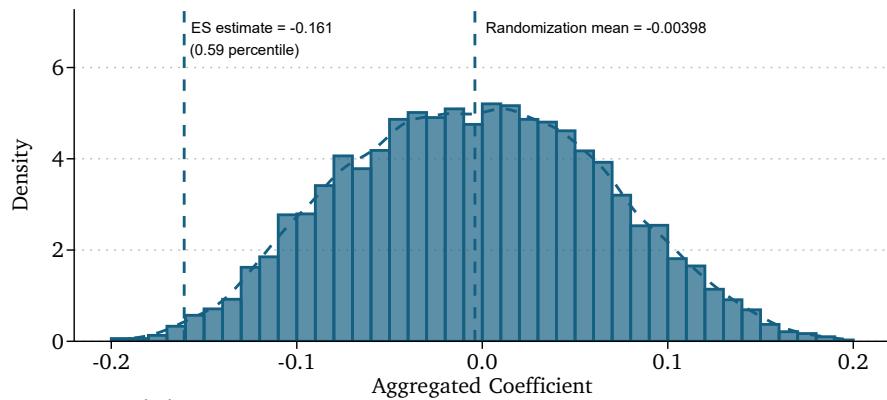
Note: The event-study plots show the beta coefficients as estimated by the two-way fixed effects model (2.2) for the datasets of Garriga (2025a) and Romelli (2024). Dynamic treatment effects are estimated relative to the year before the CBI change. The vertical bars represent the 95% confidence intervals for the estimated coefficients.

B4.8 Randomisation tests

To ensure that our observed treatment effects are indeed the result of differing patterns in central bank communication after independence events and are not driven by the definition of our treatment indicators, dataset construction or the event study setup,

we conduct a placebo randomisation test where both the treated countries and their respective years of treatment are randomised. The overall number of events are the same as in our main specification. Similarly, we impose the restriction of at most one treatment per country. Since the set of countries remains the same, and we are randomising the treatment assignment, the distribution of estimates will reflect the design based uncertainty in the estimates Abadie et al. (2020). Chart B4.11 shows the distribution of the aggregated event-study coefficient with 10,000 replications.

A. Monetary dominance



B. Financial dominance

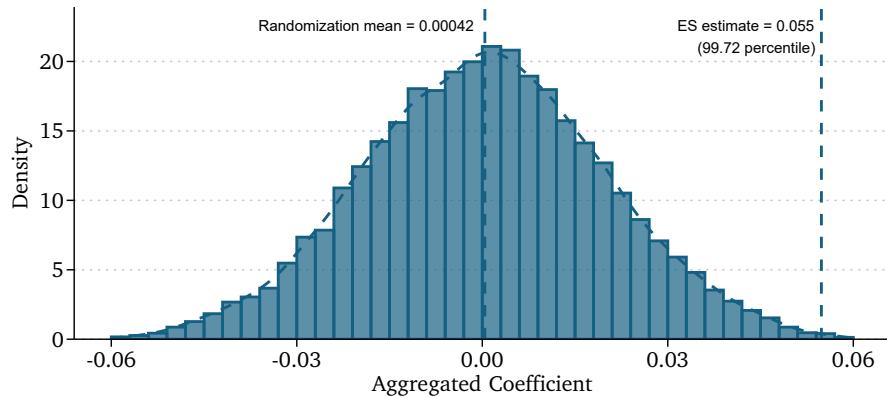


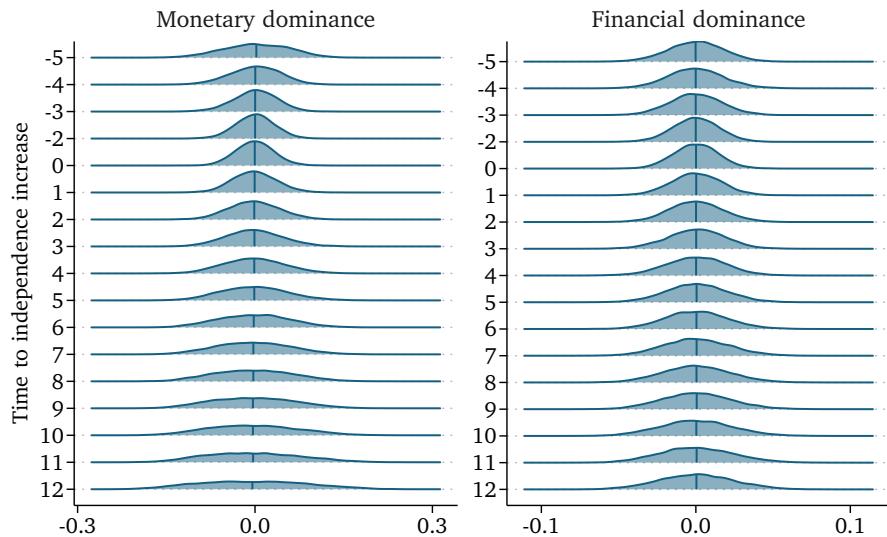
Figure B4.11: Treatment randomisation placebo (aggregated)

Note: The three panels illustrate the aggregated coefficients from our event study specification, based on 10,000 randomisations of the treatment countries and years. Vertical lines mark the means of the randomisation exercise and our estimated coefficients from the baseline specification

The distribution's mean is practically zero in line with the zero effect expectation of

the placebo test. Our estimated coefficient fall into the the 0.59% and 99.72 % percentiles of the placebo distributions of monetary, financial and fiscal dominance. Further, Chart B4.12 plots the distributions of the coefficient estimate and the t-statistic for the dynamic effects estimated with our main two-way fixed effects model (2.2). Again, we find that randomising treatments across countries and time, leads to an average coefficient estimate of virtually zero for all leads and lags and t-statistics also centred around zero.

A. Event study coefficient



B. t-statistic

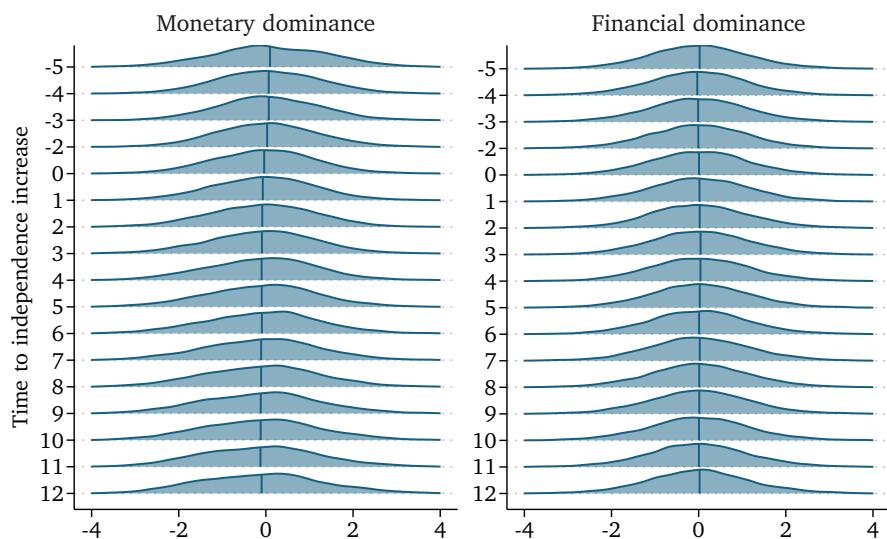


Figure B4.12: Treatment randomisation placebo (dynamic effects)

Note: Panel A shows the distribution of the event study coefficients. Panel B shows the associated t statistics. Both Panels are the result of a placebo exercise whereby the treatment is randomised across central banks and time. The vertical line indicates the distribution's average.

Coefficients towards the edges of the event study window display a higher variance, yet the means are narrowly spread around zero.

B5 LLM finetuning

To classify sentences into dominance and coordination categories we fine-tune a Gemini Pro 1.0 model. In contrast to zero-shot classification, where the model’s behaviour can typically only be influenced by measures of variation in the output (such as temperature) and the prompt itself, the model training offers several adjustable parameters. Moreover, the training data selected for fine-tuning ultimately determines the characteristics of the final model. The basis for our fine-tuned model is Google’s Gemini 1.0 Pro which we instruct to classify sentences into one of these six categories:

1. Monetary dominance
2. Financial dominance
3. Fiscal dominance
4. Monetary-fiscal coordination
5. Monetary-financial coordination
6. None

The full prompt is in Appendix B8.1. These categories correspond to the human-annotated sample of Leek et al. (2024), which consists of 1,000 sentences (see also Appendix B2 for further explanations of the classifications). We use a subset of 300 randomly selected sentences for model fine-tuning, while the remaining 700 sentences are kept for model evaluation. To find the best fine-tune, we construct a joint grid of hyper-parameters of the training process, different compositions of the training set and different prompt configurations. See Table B5.7 for the options we consider.

Parameter	Description	Possible values
Optimisation settings		
Epochs	Number of training cycles.	[1, 10]
Learning rate	Size of the steps taken in the model parameter space during optimisation.	[0.0001, 0.01]
Batch size	Number of training samples utilised in one iteration of model updating.	{2, 4, 6, ..., 16}
Prompt engineering		
Sentences per prompt	The number of sentences to be included in one prompt	{5, 10, 25}
Temperature	Parameter controlling the variation in generation output.	[0, 0.9]
Format instructions	Include instructions on output format.	{yes, no}
Dataset composition		
Synthetic sentences	Add AI generated sentences to training sample.	{yes, no}
Up-sample factor	A factor governing the degree of up-sampling, where a factor of 1 would result in a fully balanced training set.	[0, 1]
Randomise epochs	Re-randomise sentences included in prompts for each epoch.	{yes, no}

Note: All parameters were sampled using the Bayesian optimization techniques built into Akiba et al. (2019) using a uniform prior, with the exception of the number of epochs where we set a log-uniform prior to reduce training time.

Table B5.7: Hyper-parameter tuning options

The hyper-parameters are standard configuration options typically available when training machine learning models. As part of the fine-tuning exercise, we are also conduct light prompt engineering. The primary parameter is the number of sentences included in each prompt. Additionally, we may incorporate optional instructions regarding the desired output format and optimally select the model’s temperature setting.

To address the significant imbalance in our classification, we experimented with different training set compositions. We incorporated an up-sampling factor, which, when set to 1, oversamples the minority categories such that perfect balance is achieved. To introduce more variety especially in the less frequent categories, we experiment with

adding 60 LLM generated examples (10 per category) to the training set. Additionally, to ensure the model does not learn patterns based on sentence order inside the prompt (recall that we classify multiple sentences per prompt to save tokens), we implemented an option that re-randomises the sentence order for each training epoch.

Since fine-tuning is computationally expensive, we can only evaluate a very limited number of configurations. We use the Bayesian Tree-structured Parzen Estimator (TPE) Bergstra et al. (2011) implemented by the hyper-parameter optimisation framework Optuna Akiba et al. (2019) to narrow down the search space on parameter settings that are likely to deliver a good model and quickly rule out sub-optimal parameter settings. Model quality is assessed with a weighted average of the F1-macro score and accuracy, whereby we assign twice the weight to the F1 score to emphasise the importance of balanced performance across all classes. We concluded our fine-tuning process after completing the training of 108 models. Correlations of the parameter settings of the trained models and their F1 macro score evaluated on the evaluation set are shown in Figure B5.13. It is important to keep in mind that the model parameters are not sampled randomly but by the Bayesian algorithm. If a specific parameter or dataset configuration is sampled more frequently, it suggests that the algorithm considers that setting to be likely to produce better models with a higher evaluation metric on average.

The fine-tuning exercise did not yield a single best model but resulted in several models with comparable performance. Based on the parameters that produced the highest evaluation metrics paired with some researcher’s judgement, we selected the parameters presented in Table B5.8 for our final model.

Regarding the batch size, the learning rate, and the up-sampling factor, we observed a broad range of parameters producing similar evaluation metrics. For our final model, we opted for conservative values, choosing low learning rates and batch sizes, paired

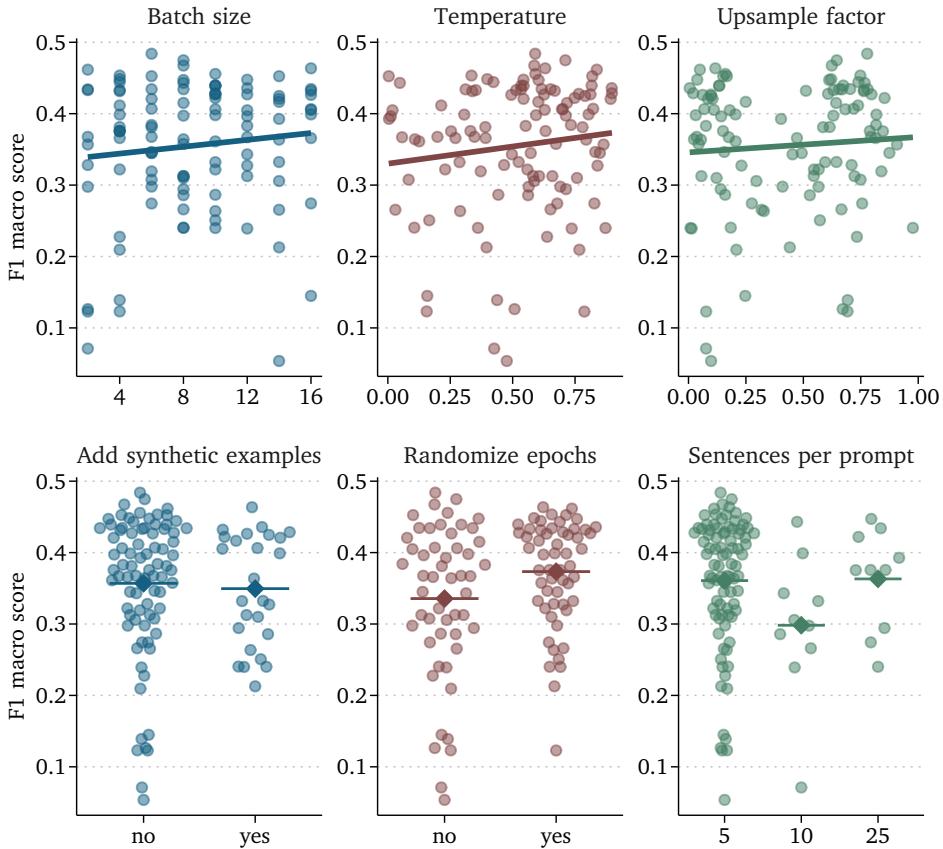


Figure B5.13: Hyper-parameter influence on validation metrics

Note: The six panels plot hyper-parameter settings against our main validation metric, the F1 macro score calculated out of sample. In total, 108 successful fine-tunes were trained. Each dot corresponds to a fine tune. The scatter plots in the first row are meant to illustrate the correlation between the hyper-parameter and the observed F1 score. The second row shows the distribution of F1 scores for categorical parameters. The horizontal lines indicate the average F1 score for each setting.

with a higher number of epochs. We decided against up-sampling categories as the evidence for its performance improvement was inconclusive, and the model already showed sufficient sensitivity to minority categories without up-sampling. Adding synthetic examples appeared to degrade classification quality. Randomising sentences in each epoch slightly increased accuracy and led to more stable model performance. For the number of sentences per prompt, we chose five sentences, as this configuration clearly provided the best performance, even though using more sentences could have significantly reduced

Parameter	Final value
Optimization settings	
Epochs	7
Learning rate	0.0005
Batch size	2
Prompt engineering	
Sentences per prompt	5
Temperature	0
Format instructions	yes
Dataset composition	
Synthetic sentences	no
Upsample factor	0
Randomise epochs	yes

Table B5.8: Final model parameters

Note: The table contains the parameters used to fine-tune our final Gemini 1.0 Pro model that we use to classify our speeches corpus.

computational costs by substantially lowering the number of processed tokens. In Figure B5.13, we observe a slight positive correlation between temperature and F1 scores. However, in individual testing we found that higher temperature settings only marginally improved model performance, if at all. We, therefore, selected the minimum temperature of 0 for less variation in classifications.

B6 LLM validation

B6.1 Policy pressures

To validate our main Gemini classification model, we compare classifications against the ‘ground truth’ established by the validation set that was independently coded by 3 human coders in Leek et al. (2024). Since we follow the same classification scheme, we can benchmark our fine-tuned Gemini model to the Large Language Models tested by Leek et al. (2024). Table B6.9 reports standard evaluation metrics for the Gemini and GPT based LLMs evaluated in Leek et al. (2024) with our novel Gemini fine-tune

added as additional column in bold. With the exception of base gpt-3.5, models are generally highly accurate with close to or above 80% of the sentences correctly classified. Our fine-tune attains the highest accuracy and F1-score (both macro and observations-weighted averages) of all tested models. Our fine-tuned model surpasses the other models in particular on the F1 (macro) which places high emphasis on the relatively infrequent categories.

	GPT			Gemini Pro 1.0		
	gpt-3.5	gpt-3.5-fine-tune	gpt-4	Base	Few Shot	Fine-tune
Accuracy	0.64	0.77	0.79	0.78	0.79	0.81
F1 (weighted)	0.69	0.78	0.78	0.73	0.75	0.79
F1 (macro)	0.35	0.43	0.40	0.36	0.40	0.47
Precision (macro)	0.33	0.40	0.48	0.44	0.50	0.49
Recall (macro)	0.43	0.49	0.40	0.34	0.36	0.45

Table B6.9: Validation metrics

Note: All columns are taken from Leek et al. (2024) except the bold fine-tune column, which shows the validation metrics of our Gemini 1.0 Pro fine-tune based on 300 sentences. The validation scores are calculated on the holdout sample of 700 sentences.

To further understand model performance, we report the confusion matrices of our Gemini fine-tune and the zero-shot GPT-3.5 used by Leek et al. (2024) (see Figure B6.14).

The matrices highlight similarities and differences between the two models. Both models are least accurate on the financial and fiscal dominance category. This results from the related coordination categories, which can be challenging to differentiate, also for human coders. Our fine-tuned Gemini model is much more reluctant to assign dominance and coordination categories. It assigns considerably more sentences to the ‘none’ category (see last columns), which is consistent with the much higher precision metric in Table B6.9. Conversely, Gemini tends to be less sensitive in the categories of dominance, aligning more closely with human coders. Overall, Gemini provides better classifications with

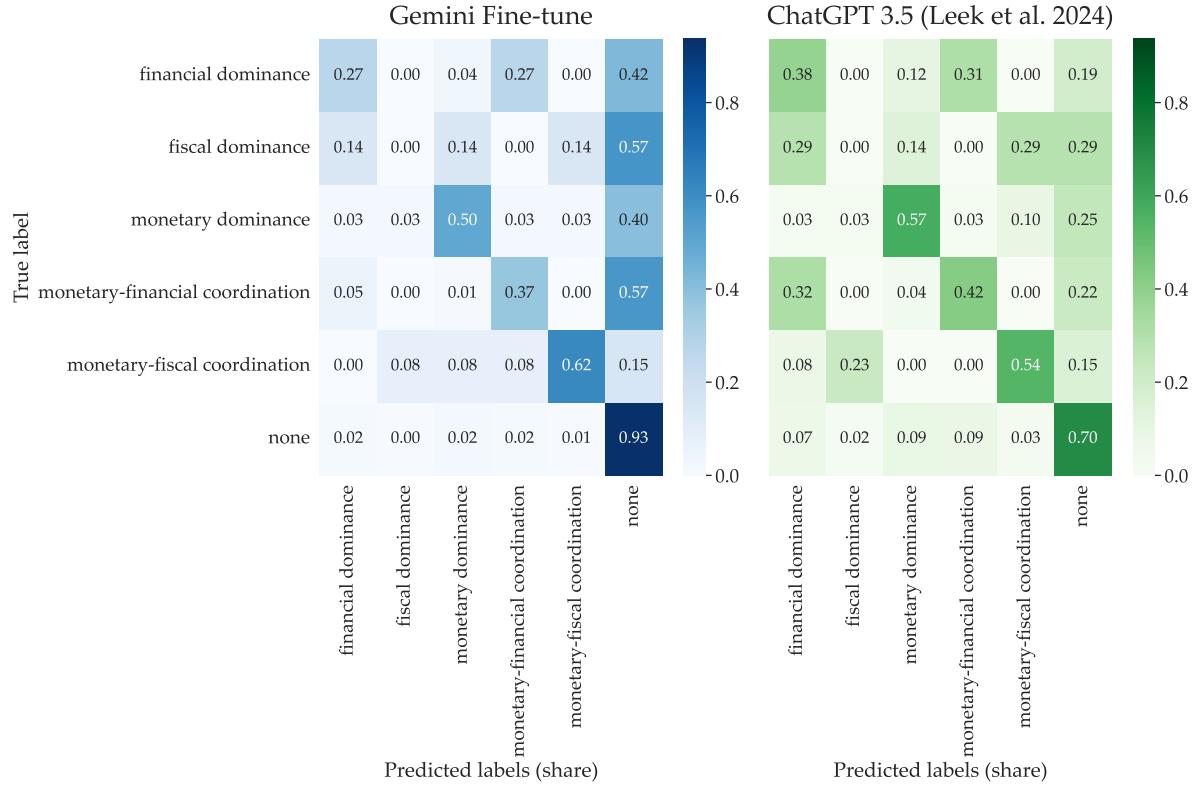


Figure B6.14: Confusion Matrices of LLM classifiers

Note: The confusion matrices plot the distribution of predicted labels by the ‘true’ label from the validation sample which consists of 700 sentences. It does not include the 300 sentences that were used for training the Gemini classifier. The left hand confusion matrix displays the Gemini model used in this chapter to classify our sample. On the right, the zero shot GPT 3.5 model used by Leek et al. (2024) is shown.

comparable ability to identify dominance and cooperation, but much fewer ‘bad’ mistakes, where a dominance or coordination category is erroneously assigned.

B6.2 External validation

Beyond face validity and comparisons against a manually coded validation sample, we offer external validity for our indices. We compare our measures of monetary and fiscal dominance against a recent model based measure of monetary and fiscal dominance from Hinterlang and Hollmayr (2022). The authors first simulate a Markov Switching DSGE model with a monetary dominance and a fiscal dominance regime, characterized through

different parameterisations of the government spending and Taylor rules inside a DSGE model. In the second step, a tree-boosting machine learning classifier is trained to predict the regime based on the simulated macroeconomic data. Finally, the classifier is applied to predict the policy regime using actual macroeconomic data of the US from 1968 to 2017. Figure B6.15 plots a dichotomised version of our relative dominance indicator which only considers fiscal and monetary dominance for the US. We find that our text based indicator correlates with the model based classification. Specifically, the two largest spikes of fiscal dominance in our indicator fall in times classified as fiscal dominance (shaded areas in Figure B6.15) and sentences are substantially more likely to be classified as fiscal dominance during the fiscal dominance episodes identified by Hinterlang and Hollmayr (2022).

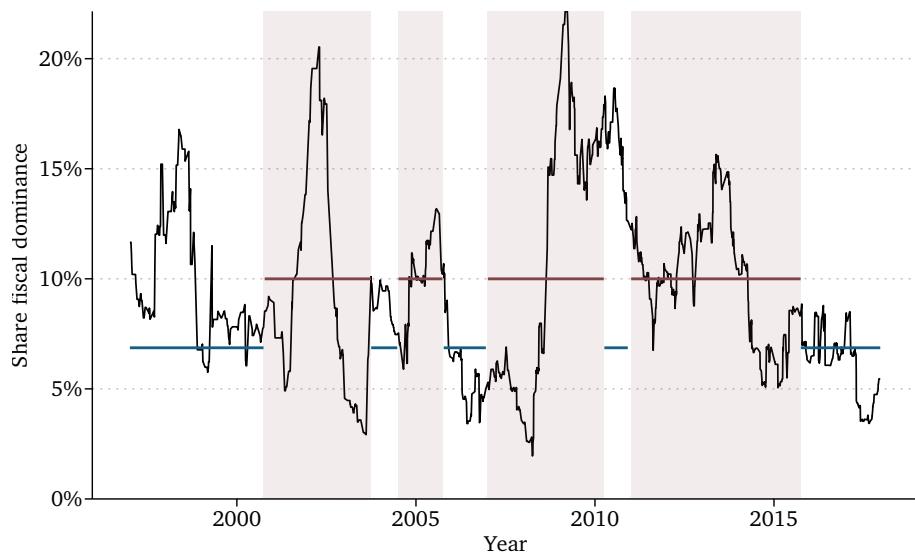


Figure B6.15: Monetary and fiscal dominance against a model based measure

Note: Lines indicate a 365 day moving average (symmetric window) of the relative shares of fiscal dominance vis-a-vis monetary dominance in the US. The red line plots the average of the fiscal dominance share in the time periods identified as the US economy operating under fiscal dominance (red shaded) by Hinterlang and Hollmayr (2022). The blue line is the sample average outside of the fiscal dominance episodes.

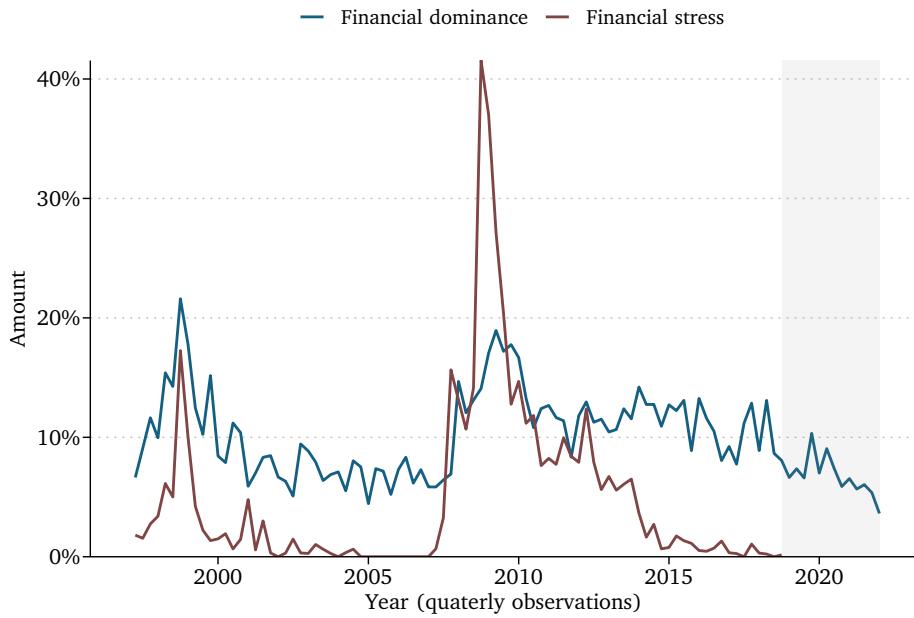


Figure B6.16: Financial pressure and financial dominance

Note: Lines plot financial stress indicator of Ahir et al. (2023) and our textual measure of financial dominance at quarterly frequency.

Similarly, in Figure B6.16 we document that our measure of financial dominance correlates with a financial stress indicator constructed by Ahir et al. (2023). We observe similarly timed peaks in both indicators around the Global Financial Crisis and before 2000. Taken together, these comparisons demonstrate that our textual measures align with real-world pressures identified in other research, lending external validity to our approach.

To further illustrate the measures, we can manually examine an example of a central bank that underwent a significant independence change (an increase of 25 percentage points) during our period of study (2003), namely, the central bank of Norway. On June 7, 2001, Central bank of Norway governor Svein Gjedrem said: “The long-term objective of monetary policy is to contribute to low and stable inflation. Price stability is the best contribution monetary policy can make to economic growth and prosperity. A nominal anchor is also a necessary precondition for stable financial markets and property markets.

We cannot achieve higher employment in the long run by accepting higher inflation.” In total, this speech contained 84.8% monetary dominance and 6.3% financial dominance. Whereas on 27 April, 2015, long after the independence event, Øystein Olsen in his capacity of central bank governor of the Central bank of Norway said: ”On the other hand, we do know that interest rates affect house prices and debt. This suggests that monetary policy should take into account the risk of financial imbalances” This speech overall contained 16.7% monetary dominance and 38.9% financial dominance. Comparing these two speeches with stark differences in monetary and financial dominance illustrates the shift in considerations after an independence event.

B7 Instrumental variable design

B7.1 Choice of lagged dependent variable

As briefly discussed in the main text, it is common practice to include a lagged dependent variable (DV) in panel regression models when the dependent variable is likely to exhibit persistence. For macroeconomic indicators such as inflation, this procedure is straightforward; one can simply use the previous observation. However, in our speeches dataset the situation is less clear-cut. Speeches are often topic-specific, and our measures of dominance fluctuate considerably between speeches, yet we still expect persistence in the sense that current dominance patterns are correlated with these in prior speeches. To account for this dynamic behaviour, we incorporate an average $\tilde{\psi}_{ict}^m$ of past speeches as the lagged dependent variable in the estimation equation (Equation 2.5 in the main text):

$$\psi_{ict}^m = \rho \tilde{\psi}_{ict}^m + \beta_1 \text{CBI}_{ct} + \beta_2 \Delta\pi_{ct} + \beta_3 \Delta u_{ct} + \theta_t + \mu_c + \epsilon_{ict}, \quad (\text{B7.2})$$

with $\Delta\pi_{ct}$ and Δu_{ct} denoting the change in inflation and unemployment. We experiment with different definitions of the lagged dependent variable. In particular, we consider two

approaches: (i) using calendar-based windows, in which we average each central bank's speeches delivered within a specified timeframe before the speech, and (ii) averaging over a fixed number of the central bank's prior speeches.

Table B7.10 reports the coefficient estimates for the effect of central bank independence on monetary and financial dominance, along with goodness-of-fit measures. The results indicate that including the lagged dependent variable substantially improves model fit, as evidenced by higher log-likelihood values and adjusted R^2 statistics across specifications (compared to the benchmark without lagged dependent variable). Importantly, the precise definition of the lagged DV has little impact on the magnitude of the observed CBI effects. For our baseline instrumental variable specification in the main text, we adopt the definition that best fits the data, that is including the lagged DV and averaging over the central bank's last 25 speeches.

DV definition	Goodness of fit		Effect on dominances	
	Loglik	adj. R^2	Monetary	Financial
No lagged dependant variable	1593	0.121	-1.5637* (0.7926)	0.9370** (0.3829)
Past speeches				
5 speeches	1727	0.141	-1.2556* (0.6830)	0.8298** (0.3684)
10 speeches	1791.8	0.150	-1.0767* (0.5817)	0.7717** (0.3452)
25 speeches	1859.1	0.159	-0.8682* (0.4508)	0.6872** (0.3182)
50 speeches	1844.6	0.157	-0.8505* (0.4527)	0.7084** (0.3283)
100 speeches	1792.5	0.150	-0.9022* (0.5036)	0.7932** (0.3548)
Calendar days				
50 days	1382.7	0.131	-1.6125* (0.8842)	1.1009** (0.4713)
150 days	1656	0.143	-1.1882* (0.6848)	0.9226** (0.3916)
300 days	1769.9	0.151	-1.0031* (0.5607)	0.7896** (0.3384)
500 days	1815	0.155	-0.9382** (0.4635)	0.7395** (0.3228)

Table B7.10: Lagged dependant variable definition

Note: The table shows the coefficient of $\widehat{\text{CBI}}_{ct}$ in the second stage of the IV panel model (B7.2), Stars indicate significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The log-likelihood and adjusted R^2 columns report the average of the respective measure over the monetary and financial dominance estimates. Standard errors are clustered on the country level.

B7.2 Heterogeneity estimates using the IV approach

To further support our findings regarding country differences, we run the IV approach presented in the main text with (instrumented) interactions of CBI and dummies for advanced economies and democracies. The second stage for the advanced economies interaction is given by:

$$\psi_{ict}^m = \rho\tilde{\psi}_{ict}^m + \beta_1\widehat{\text{CBI}}_{ct} + \beta_2(\widehat{\text{CBI}}_{ct} \times \text{Advanced}_c) + \beta_3\Delta\pi_{ct} + \beta_4\Delta u_{ct} + \theta_t + \mu_c + \epsilon_{ict} \quad (\text{B7.3})$$

with $\Delta\pi_{ct}$ and Δu_{ct} denoting the change in inflation and unemployment. The first stage contains interaction terms for each of the instruments:

$$\begin{aligned} \text{CBI}_{ct} = & \gamma_1 \text{Inverse distance weighted world CBI}_{c(t-1)} + \\ & \gamma_2(\text{Inverse distance weighted world CBI}_{c(t-1)} \times \text{Advanced}_c) + \\ & \gamma_3 \text{Neighbour's electoral democracy index}_{c(t-1)} + \\ & \gamma_4(\text{Neighbour's electoral democracy index}_{c(t-1)} \times \text{Advanced}_c) + \\ & \gamma_5 \text{Independence judiciary}_{ct} + \\ & \gamma_6(\text{Independence judiciary}_{ct} \times \text{Advanced}_c) + \\ & \gamma_7\Delta\pi_{ct} + \gamma_8\Delta u_{ct} + \delta\tilde{\psi}_{ict}^m + \kappa_t + \lambda_c + \nu_{ct} \end{aligned} \quad (\text{B7.4})$$

κ_t and λ_c denote year and country fixed effects. $\tilde{\psi}_{ict}^m$ is the lagged dependant variable. ν_{ct} is the residual of the first stage. A separate first stage equation for the interaction term ($\text{CBI}_{ct} \times \text{Advanced}_c$) with identical independent variables is estimated. The estimation equations for democracies vis-a-vis autocracies are analogously defined using a binary democracy variable.

Similar to our event study estimates, the results show that advanced countries drive the effect for monetary and financial dominance.

Interaction	Advanced		Democracy	
Dependent Variables: Model:	Monetary (1)	Financial (2)	Monetary (3)	Financial (4)
<i>Variables</i>				
CBI	0.1663 (0.3776)	-0.2488 (0.2838)	-0.5233 (0.4166)	0.7176** (0.2779)
$\widehat{\text{CBI}} \times \widehat{\text{Advanced}}$	-1.272** (0.5316)	1.142** (0.5392)		
$\widehat{\text{CBI}} \times \widehat{\text{Democracy}}$			-0.0646 (0.0444)	0.0292 (0.0355)
Lagged DV	0.4854*** (0.0760)	0.3047*** (0.0616)	0.4952*** (0.0733)	0.3254*** (0.0573)
Δ Inflation rate	0.4553** (0.2049)	-0.0639 (0.0688)	0.4636*** (0.1694)	-0.0922 (0.1095)
Δ Unemployment rate	-0.0012 (0.0041)	0.0020 (0.0013)	-0.0020 (0.0039)	0.0022 (0.0015)
<i>Fixed Effects</i>				
Country	✓	✓	✓	✓
Year	✓	✓	✓	✓
<i>Fit statistics</i>				
R ²	0.21344	0.11058	0.21732	0.10880
Observations	12,205	12,205	12,205	12,205

Table B7.11: Instrumental variable regressions with interaction term

Note: Stars indicate significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The tables show the second stage estimates of the 2SLS equations (B7.3) and (B7.4). Standard errors are clustered on the country level.

B8 Prompts

This Appendix contains the prompts of all LLM tasks that were run using Gemini for this chapter.

B8.1 Policy pressure channel

We used the following instructions for both the fine-tuning and the final model run on approximately 2.1 million sentences. The prompt is written in terms of ‘excerpts’ since to classify each sentence, we incorporate the sentence before and after the sentence we intend to classify as additional context.

1 Prompt:

You will be provided with excerpts from a central bank speech. If the information contained in the excerpt allows for it, label these excerpts according to the policy approach regarding monetary policy. Choose from one of the following policy approaches:

\monetary dominance", i.e., the central bank prioritizes to maintain price stability, and its monetary policy is not subordinated to fiscal policy or to financial stability considerations.

\fiscal dominance", i.e., the central bank accommodates its monetary policy to fiscal considerations, and its decisions are subordinated to meet the demands fiscal policy.

\financial dominance", i.e., the central bank accommodates its monetary policy to financial considerations, and its decisions are subordinated to respond to the needs of financial markets.

\monetary-fiscal coordination", i.e., the central bank suggests to cooperate with the governments to better mutually align policies.

\monetary-financial coordination", i.e., the central bank suggests to cooperate with financial market participants to better mutually align policies.

If the excerpt discusses topics unrelated to the interaction of monetary policy with fiscal policy and financial markets or is purely descriptive without implying a policy approach label as: \none", i.e., the excerpt does not contain information regarding the policy approach taken by the bank.

Assign one of the three dominance categories if a hierarchy of actors is apparent in the excerpt, i.e., monetary dominance if the central bank prescribes actions to others or acts irrespective of the needs of other actors. Financial or fiscal dominance if the central bank is subordinating its monetary policy. Assign one of the two coordination categories if the central bank suggests to coordinate monetary policy with fiscal or financial policy to achieve a better policy mix.

<optional instructions on output format>

These are the excerpts:

1. <First excerpt>
2. <Second excerpt>
- ...
3. <Fifth excerpt>

The prompt contains <optional instructions on output format>, which are the following:

1 Reply only with the number and the label of each excerpt. Use the following format: [1] label of first excerpt
[2] label of second excerpt

...

and included in our final prompt.

B8.2 Metadata extraction

To process the metadata, we passed 6 speech description in a single prompt. We run a few shot prompt with the following 6 examples that are real examples from the BIS speeches dataset:

1. b170831: Welcome address by Dr Ernest Addison, Governor of the Bank of Ghana, at the Bank of Ghana's 60th Anniversary Lectures and Exhibition, Accra, 18 August 2017.
2. a000303: Introductory statements given by Mr Willem F Duisenberg, President of the European Central Bank, and Mr Christian Noyer, Vice-President of the European Central Bank, held in Frankfurt on 2 March 2000.
3. b050203: English translation of an interview of Mr Jean-Claude Trichet, President of the European Central Bank, conducted by Ms Françoise Crouïgneau and Mr Thibaut Madelin (Les Echos), 20 January 2005.
4. h970403: Text of the London School of Economic and Political Science Annual Lecture on Central Banking delivered by the Deputy Governor of the Bank of England, Mr. Howard Davies, on 4/3/97.
5. e140331: Speech by Dr Andreas Dombret, Member of the Executive Board of the Deutsche Bundesbank, at the Harvard Law School Symposium on "Building the Financial System of the 21st Century - An Agenda for Europe and the United States", Armonk, New York, 28 March 2014.

-
6. e220708: Speech by Ms Lael Brainard, Member of the Board of Governors of the Federal Reserve System, at the Bank of England Conference, London, 8 July 2022.

The expected outputs were passed in the same JSON format that Gemini was asked to use:

1.

```
{  
    "speech_identifier" : "b170831",  
    "type_of_text" : "Welcome address",  
    "speaker" : "Ernest Addison",  
    "central_bank" : "Bank of Ghana",  
    "position" : "Governor of the Bank of Ghana",  
    "occasion" : "Bank of Ghana's 60th Anniversary Lectures and  
Exhibition",  
    "venue" : "Bank of Ghana",  
    "location" : "Accra"  
}
```

- 2.

```
{  
    "speech_identifier" : "a000303",  
    "type_of_text" : "Introductory statement",  
    "speaker" : "Willem F Duisenberg;Christian Noyer",  
    "central_bank" : "European Central Bank;European Central Bank",  
    "position" : "President of the European Central  
Bank;Vice-President of the European Central Bank",
```

```
        "occasion" : null,
        "venue" : null,
        "location" : "Frankfurt"
    }

3.

{
    "speech_identifier" : "b050203",
    "type_of_text" : "Interview",
    "speaker" : "Jean-Claude Trichet",
    "central_bank" : "European Central Bank",
    "position" : "President of the European Central Bank",
    "occasion" : "Interview conducted by Ms Françoise Crouïgneau  
and Mr Thibaut Madelin (Les Echos)",
    "venue" : "Les Echos",
    "location" : null
},

4.

{
    "speech_identifier" : "h970403",
    "type_of_text" : "Lecture",
    "speaker" : "Howard Davies",
    "central_bank" : "Bank of England",
    "position" : "Deputy Governor of the Bank of England",
    "occasion" : "London School of Economic and Political Science  
Annual Lecture on Central Banking",

```

```
        "venue" : "London School of Economics and Political Science",
        "location" : "London"
    }
```

5.

```
{  
    "speech_identifier" : "e140331",
    "type_of_text" : "Speech",
    "speaker" : "Andreas Dombret",
    "central_bank" : "Deutsche Bundesbank",
    "position" : "Member of the Executive Board of the Deutsche
Bundesbank",
    "occasion" : "Harvard Law School Symposium on "Building the
Financial System of the 21st Century - An Agenda for Europe and
the United States"",
    "venue" : "Harvard Law School",
    "location" : "New York"
}
```

6.

```
{  
    "speech_identifier" : "e220708",
    "type_of_text" : "Speech",
    "speaker" : "Lael Brainard",
    "central_bank" : "Board of Governors of the Federal Reserve",
    "position" : "Member of the Board of Governors of the Federal
Reserve System",
}
```

```
        "occasion" : "Bank of England Conference",
        "venue" : "Bank of England",
        "location" : "London"
    }
```

The prompt contained the following instructions: 1 **Prompt:** I'll provide you with metadata on central bank speeches from which you should extract the following information:

1. Speech Identifier (e.g. b050203)
2. Type of Text (e.g. Speech, Introductory statement or Introductory remarks or Interview)
3. Name of the speaker (e.g. Jean-Claude Trichet)
4. Central bank of the speaker (e.g. Bank of England)
5. Position of the Speaker (e.g. President of the Federal Reserve Bank of Kansas City)
6. Occasion (e.g. 30th Economics Conference 'Competition of Regions and Integration in EMU')
7. Venue (e.g. London School of Economics and Political Science)
8. Location (e.g Frankfurt or Vienna)

Reply with a JSON list with one object for each speech. Please use the following keys for the metadata:

1. "speech_identifier"
2. "type_of_text"
3. "speaker"
4. "central_bank"

5. "position"

6. "occasion"

7. "venue"

8. "location"

Additional instructions:

- * The central bank should always be the central bank with which the speaker is affiliated. E.g. if the the governor of the Bank of England delivers a speech at an event organized by the ECB you should assign the Bank of England as the central bank.
- * If the central bank is part of the US federal reserve system be specific whether it is the board of governors of the Fed or a regional Fed like the Federal Reserve Bank of Chicago.
- * If a speech contains the names of two speakers separate the names with a semicolon. Do the same for their positions and central banks. See the second example given below.
- * Not all fields are always contained in the data. In this cases you can leave missing values. See the second example which does not contain a occasion and venue. And the third example which does not contain a location.
- * Occasion should contain the occasion on which the speech was given. This should contain most of the information that is given in the metadata. E.g. if it was part of a session of a particular conference on a topic it should include all that
- * Venue is more narrow and should be a place like an institute, university, ministry or central bank or a particular congress/forum. Don't put a city as venue. The city should go to location.

-
- * Location should always be a city. If the city is not directly given you can infer the city from e.g. the university.
 - * Sometimes abbreviations are used such as Ass. for association, Conf. for conference or Econ. for economics. Please spell out these abbreviations.

Extract the metadata from the following lines. Each line starts with the speech identifier followed by a string that contains the metadata that should be extracted.

<First description>

<Second description>

...

<Sixth description>

B8.3 Audience classification

To generate the audience for each speech we used the following prompt

1 Prompt:

I'll provide you with metadata from central bank speeches from which you should infer the audience of the speech. Please assign one of the following labels for each speech:

"academic", if the audience is likely to be academic, e.g. a speech at a university or a conference at a research institute "financial_market", if the audience is likely to be financial market actors or representatives, e.g. a speech at a financial markets association

"political", if the audience is likely to be politicians, government officials

or elected representatives, e.g. an address in front of parliament or a ministry

"central_bank", if the audience are central bankers or a general central bank audience, e.g. interviews with newspapers or central bank press conferences. Also assign this category if none of the other categories fit.

Reply with a JSON list with one object for each speech. Each object should contain two entries:

1. "identifier"
2. "audience"

Extract the audience from the following lines. Each line starts with a speech identifier, followed by a colon, followed by the metadata on the speech.

<First line>

<Second line>

...

<Fifth line>

The prompt was run as a few-shot prompt with the following example descriptions from the BIS speeches dataset:

1. b060714: Speech by Mr Jean-Claude Trichet, President of the European Central Bank, at the 57. Jahresversammlung des Ifo Instituts für Wirtschaftsforschung an der Universität München, Munich, 29 June 2006.
2. a151202: Speech by Mr Amando M Tetangco, Jr, Governor of Bangko Sentral ng Pilipinas (BSP, the central bank of the Philippines), at the Launching of the

Paranaque City Credit Surety Fund, Manila, 3 November 2015.

3. e180108: Address by Mr Rameswurlall Basant Roi, Governor of the Bank of Mauritius, at the annual dinner for major economic stakeholders, Flic-en-Flac, 17 November 2017.
4. a170831: Opening statement by Dr Andreas Dombret, Member of the Executive Board of the Deutsche Bundesbank, at the press conference presenting the results of the low-interest-rate survey conducted by the Bundesbank and BaFin, Frankfurt am Main, 30 August 2017.
5. a170418: Introductory statement by Mr Ignazio Visco, Governor of the Bank of Italy, at an "Open coordinators meeting" of the ECON Committee (European Parliament) for an exchange of views on the economic and financial situation of Italy and prospects for economic governance in the European Union, Brussels, 11 April 2017.

For the few shot prompt example, we included the following JSON array as the desired output:

```
1
[
{
  "identifier": "b060714",
  "audience": "academic"
},
{
  "identifier": "a151202",
  "audience": "financial_market"
```

```
  },
  {
    "identifier": "a170831",
    "audience": "central_bank"
  },
  {
    "identifier": "e180108",
    "audience": "financial_market"
  },
  {
    "identifier": "a170418",
    "audience": "political"
  }
]
```

B8.4 Geographic location

To retrieve the geographic coordinates for the locations we retrieved from the speech descriptions we used the following prompt. We queried 6 locations per prompt.

1 Prompt:

I'll provide you with identifiers and associated locations. I want you to find the geographic coordinates associated with each location. Reply with a JSON list with one object for each identifier. Each object should contain the following keys:

1. identifier
2. latitude
3. longitude

Extract the metadata from the following lines. Each line starts with the identifier, followed by a semicolon, followed by the location for which you should find the coordinates.

```
<First identifier>;<First location>
<Second identifier>;<Second location>
...
<Sixth identifier>;<Sixth location>
```

The prompt was run as a few-shot prompt with the following example locations:

1. a981012;Brussels
2. c090907;Buenos Aires
3. c050609;Frankfurt

And the expected output:

```
1
[
{
  "identifier": "a981012",
  "latitude": 50.8465,
  "longitude": 4.3517
},
{
  "identifier": "c090907",
```

```
"latitude": -34.6037,  
"longitude": -58.3692  
},  
{  
    "identifier": "c050609",  
    "latitude": 50.1109,  
    "longitude": 8.6821  
}  
]
```

Appendix Chapter 3

C1 Data

Below, first the speeches dataset is described, followed by the additional data and the data sources used in the paper.

C1.1 Speeches-as-data

The speeches dataset as downloaded from the BIS website¹⁰ contains the text of the speech, the date on which it was given, the author, and a non-standardised description string which contains metadata on the speech such as the location, the speaker and the occasion of the speech. Below, Table C1.1 shows the number of speeches over time, Table C1.2 shows the number of speeches per national central bank and Figure C1.1 visualises the frequency of speeches per central bank over time.

Years	NCB presidents/chairs	Executive Board
≤ 2008	434	865
≤ 2010	749	1126
≤ 2012	903	1387
≤ 2014	1045	1624
≤ 2016	1272	1822
≤ 2018	1528	2071
≤ 2020	1938	2287
≤ current	2167	2487

Table C1.1: Overview of the number of speeches cumulative over time

¹⁰<https://www.bis.org/cbspeeches/index.htm>

National Central Bank	Number of (scraped) speeches
Deutsche Bundesbank	838
Bank of Spain	355
Bank of France	400
Bank of Italy	400
Netherlands Bank	207
European Central Bank	2487

Table C1.2: Overview of the number of speeches per national central bank

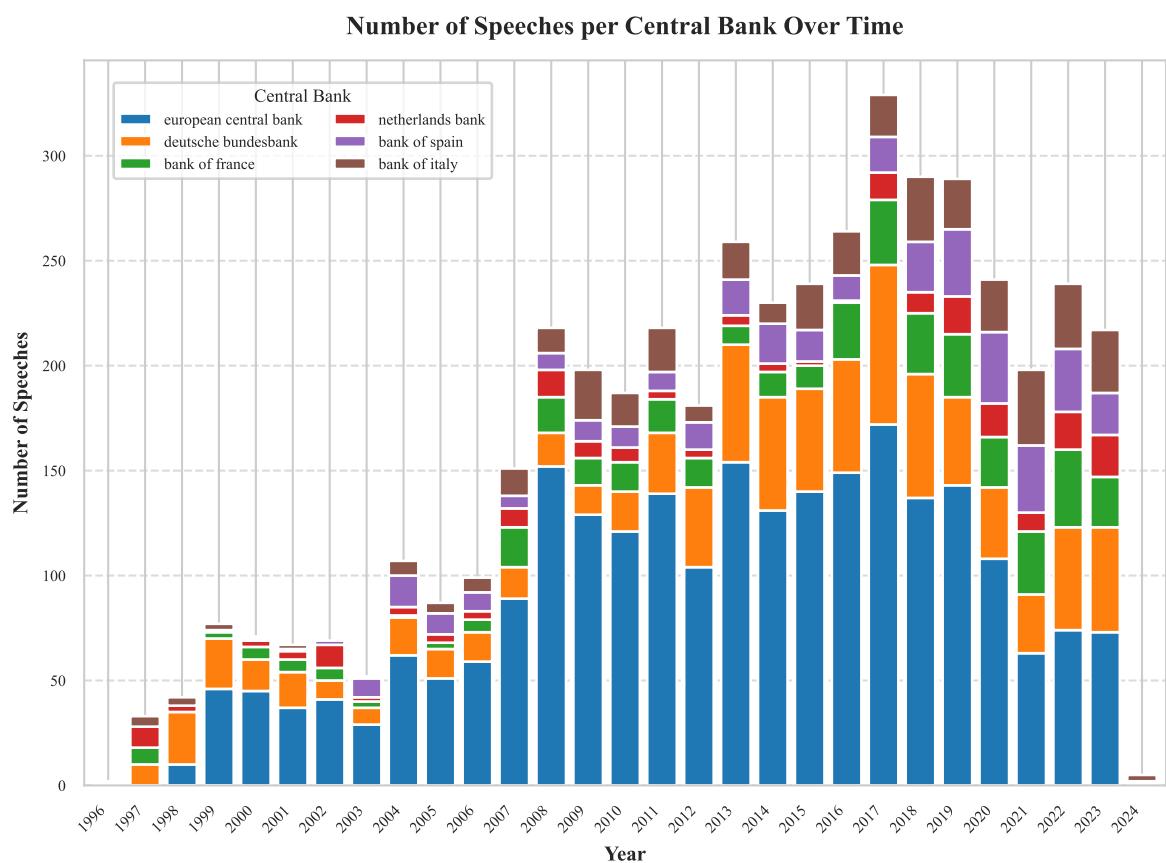


Figure C1.1: Frequency of speeches per central bank and over time

C1.2 Additional data

Table C1.3 shows the data sources and timespan of the variables used in the analysis.

Table C1.4 shows summary statistics of the main variables used after pre-processing.

After that, I describe the Google trends data, the eurobarometer data, media salience data and I provide a ranking of the countries based on the key variables.

Variable name	Data source	Timespan
Speeches topic proportions	Own calculations (raw speeches from BIS)	01/1997–01/2024
Inflation	Worldbank data	01/1997–01/2024
Trust in ECB	Eurobarometer survey	01/2004–01/2023
Real GDP Growth	OECD data	01/2000–01/2024
Unemployment	Worldbank data	01/1997–01/2024
Public salience	Google Trends data	01/2004–01/2024
Media salience	Financial Times	01/2000–01/2024

Table C1.3: Variables names, data source and timespan

Table C1.4 shows summary statistics of the main variables including the count, mean, standard deviation, minimum, maximum values and percentiles of the main quarterly data set used in the cross-sectional timeseries analysis. This dataset has already been aggregated to quarterly central bank observations by simply taking the mean of 3 months of the variables. The dataset only contains missing values in the Trust in the ECB and Google Trends measures since the data is only available in later years.

Table C1.4: Summary Statistics

(a) Economic Indicators

	unemployment	gdp	real growth	inflation	government debt	pgdp
count	553	553	553	553	553	553
mean	8.98	0.01	95.27		82.36	
std	4.19	0.03	12.52		27.17	
min	2.98		-0.11	69.76		35.77
25%	6.73		0.01	85.60		60.39
50%	8.46		0.02	97.00		75.28
75%	10.31		0.03	103.02		101.80
max	26.09		0.08	128.90		154.92

(b) Topics

	Monetary Policy	Economic Analysis	Financial Markets	Banking Regulation
count	553	553	553	553
mean	0.15	0.11	0.12	0.11
std	0.11	0.08	0.07	0.10
min	0.00	0.00	0.00	0.00
25%	0.08	0.07	0.07	0.06
50%	0.12	0.09	0.10	0.08
75%	0.18	0.13	0.14	0.11
max	1.00	0.62	0.54	1.00

(c) Topics (continued)

	Digital Finance	International Econ	Crisis Mgmt	Climate
count	553	553	553	553
mean	0.09	0.06	0.10	0.05
std	0.09	0.06	0.10	0.09
min	0.00	0.00	0.00	0.00
25%	0.05	0.04	0.06	0.02
50%	0.07	0.05	0.08	0.02
75%	0.11	0.07	0.10	0.03
max	1.00	0.56	1.00	1.00

(d) Topics (continued) and Public Sentiment

	Payments	National Econ	trust	ECB	google trends
count	553	553	448	451	
mean	0.07	0.14	1.46	11.29	
std	0.07	0.18	0.14	7.77	
min	0.00	0.00	1.12	0.00	
25%	0.04	0.05	1.35	7.00	
50%	0.06	0.07	1.45	10.33	
75%	0.08	0.14	1.55	14.83	
max	0.69	1.00	1.83	60.00	

C1.2.1 Google trends data

Next, Google trends allows you to specify the country, timespan and search terms. The guidelines and best practices of Hölzl et al. (2025) were employed when working with the Google Trends data. The five countries of the NCBs were used and the average of these five countries was used for the ECB. The longest possible timespan available was used, i.e., starting in January 2004 until January 2024. Given the simplicity of the task, i.e., seeing how salient the ECB was among the public, the search terms tested were limited to the ECB and the European Central Bank. Figures C1.2 and C1.3 below show for Germany and France that using ‘ECB’ as search term shows the most variation and frequency needed to determine the salience. Similar trends hold for other countries.

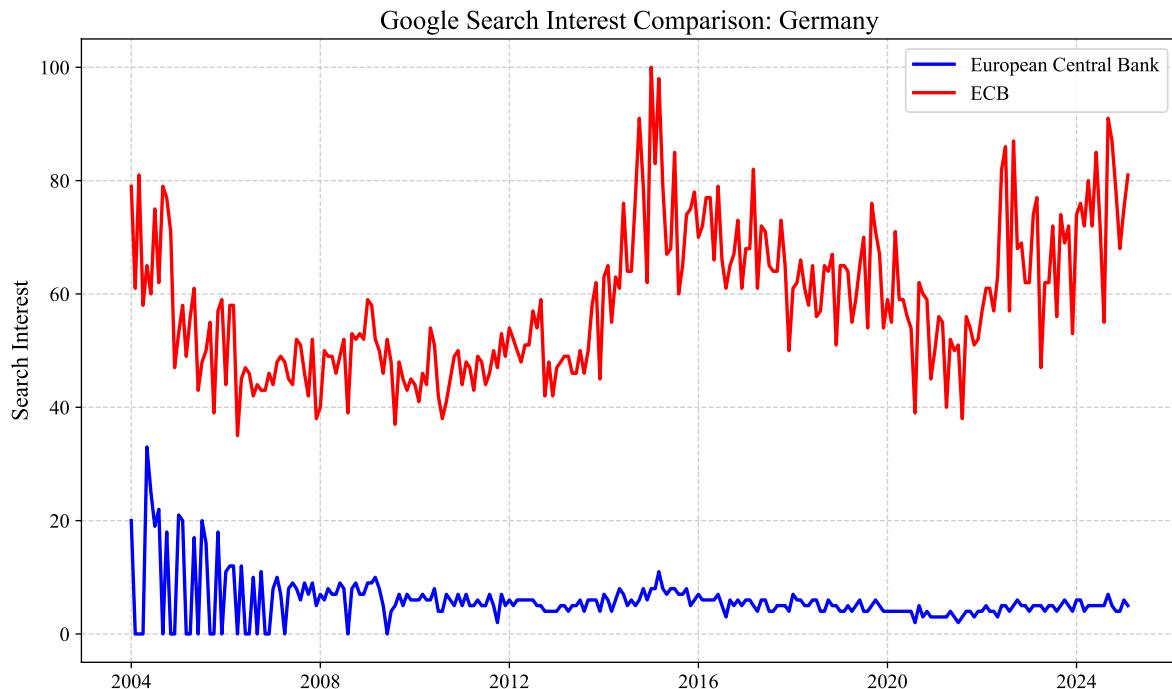


Figure C1.2: Google search interest for Germany

Note: This figure displays the temporal evolution of Google search interest for Germany, contrasting two search queries (“European Central Bank” and “ECB”) from 2004-2024.

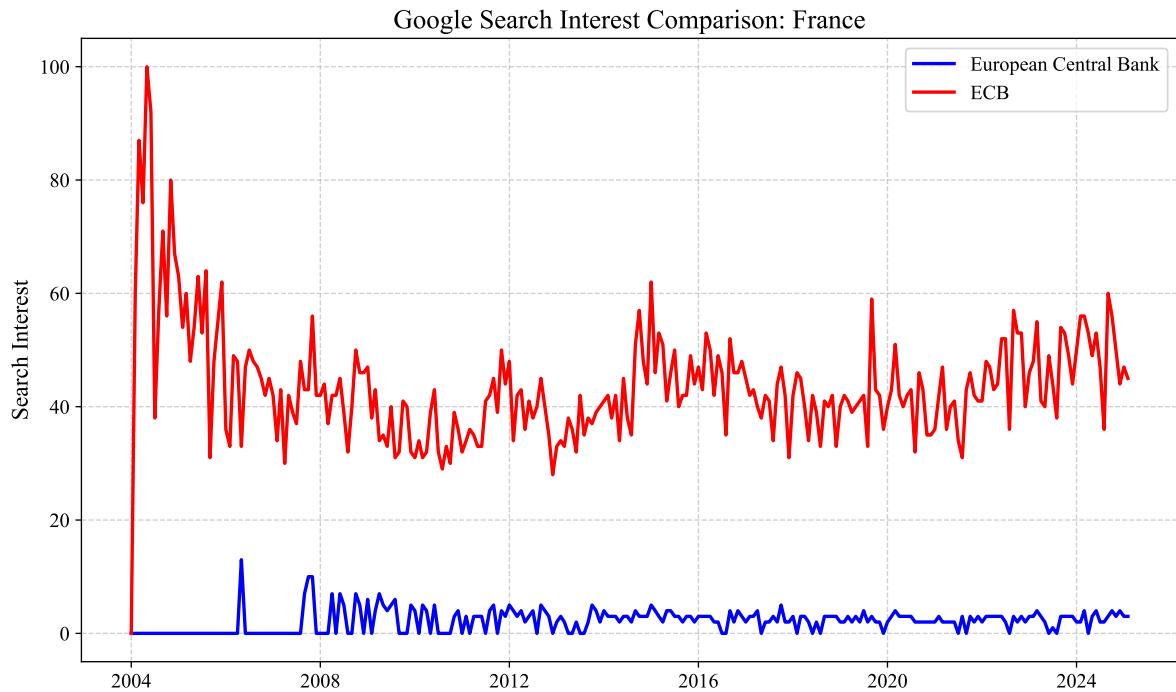


Figure C1.3: Google search interest for France

Note: This figure displays the temporal evolution of Google search interest for France, contrasting two search queries (“European Central Bank” and “ECB”) from 2004-2024.

C1.2.2 Eurobarometer data

For the data preparation for the public opinion proxy, I begin by loading each Standard Eurobarometer survey from 2004 and generating consistent identifier variables (e.g., Eurobarometer and study IDs, respondent IDs, and interview dates). I then harmonised country identifiers (e.g., “nation”), created and labelled weighting variables (national, European and total), and cleaned or recoded core opinion measures. Next, I standardised, labelled and harmonised all variables. I then re-labelled any divergent codes, merged and collapsed data at the year–country level and created an “EU average” row by calculating mean values of key variables across all European Union member states. Figure C1.4 shows the trust in the ECB variable and, as a benchmark, the trust in the EU in general variable.

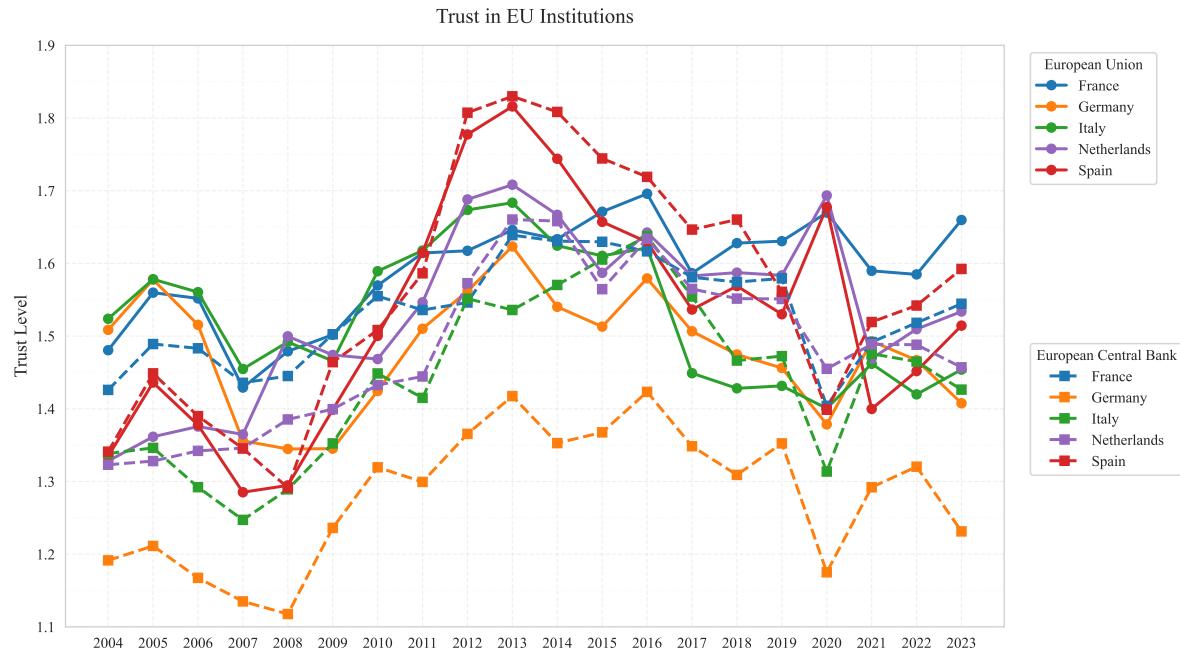


Figure C1.4: Trust in EU institutions

Note: This figure shows citizens' trust in European Union institutions retrieved from the standard Eurobarometer from 2004 to 2023. Trust in the EU is used as a benchmark next to trust in the European Central Bank and plotted for the five NCB countries (France, Germany, Italy, Spain, and the Netherlands). Trust in the EU is depicted using solid lines with circular markers, whereas trust in the ECB is shown with dashed lines accompanied by square markers.

C1.2.3 Media salience

Figure C1.5 shows that the eurozone crisis is the most salient period for the eurozone with regards to mentions of the European Central Bank in articles by the FT.

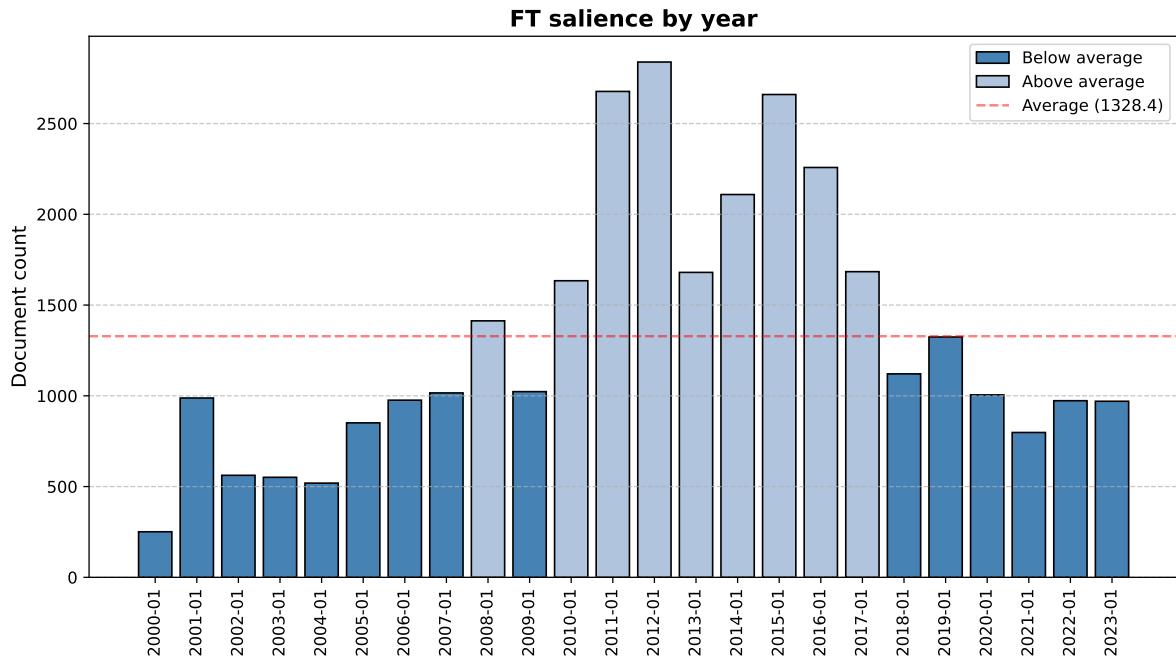


Figure C1.5: FT salience

Note: Number of Financial Times articles which mention the European Central Bank over time.

C1.2.4 Ranking countries

Table C1.2.4 shows the averages of the pressure proxies between countries and how they rank vis-a-vis each other.

Country	Mean salience	Mean trust in ECB	Mean Debt/GDP
Germany	0.614 (2)	0.524 (5)	0.545 (1)
Spain	0.663 (3)	0.693 (1)	0.733 (5)
France	0.584 (1)	0.554 (4)	0.683 (3)
Italy	0.634 (4)	0.584 (3)	0.703 (4)
Netherlands	0.703 (5)	0.604 (2)	0.634 (2)

Table C1.5: Country heterogeneity in terms of variables

Note: This table displays the mean salience, mean trust in the ECB and mean debt/GDP ratio by country. The ranking of the countries is added in the brackets.

C2 Topic modeling

Below, the pipeline of text pre-processing and inference of the topic model is discussed. This is followed with extra information on the topic modeling, how the optimal number of topics was determined including some descriptive statistics and figures to further interpret the topics.

C2.1 Speeches corpus pre-processing

All the text pre-processing follows standard methodology, (e.g., see Gentzkow et al. (2019)) and includes the following steps:

1. Deleting regular expressions
 - Page numbers
 - Page headers
 - New page characters
 - Footnotes
 - URLs
 - Subsequent whitespace characters
2. Creating a corpus
3. Tokenising, deleting stopwords, capital letters and removing words which skew the results unnecessarily (e.g., names of central bank governors and very common words).
4. Remove words with characters smaller than 3 and larger than 20.

C2.2 BERTopic model steps and choices made

The topic modeling pipeline begins by preparing the raw text data and computing document embeddings using a pre-trained SentenceTransformer model. Specifically, I use the all-MiniLM-L6-v2 which is a lightweight and efficient sentence transformer developed by Microsoft Research Wang et al. (2020). Despite its compressed architecture of only 6 layers (compared to larger models with 12 or 24 layers), it achieves very good performance (e.g., see <https://huggingface.co/spaces/mteb/leaderboard>). The model has been distilled from larger BERT architectures and offers multilingual abilities and high-quality 384-dimensional embeddings. This model generates dense vector representations for each document while leveraging GPU acceleration if available. The embeddings are normalised and processed in batches, ensuring efficiency and stability in handling large corpora. The resulting embeddings serve as the numerical foundation for subsequent clustering and topic extraction.

Next, the model incorporates a refined vectorisation step using a CountVectorizer configured with an extended set of stop words, a minimum document frequency, a maximum document frequency threshold and n-gram ranges covering bigrams to four-grams. Doing this helps to focus on meaningful multi-word expressions and reduces noise from common stopwords. The pipeline then reduces dimensionality. Specifically, I use Uniform Manifold Approximation and Projection (UMAP) of McInnes et al. (2020) as a dimensionality reduction step before clustering. UMAP preserves both the local and global structure of the high-dimensional embeddings while reducing them to a more manageable dimensional space. I configure UMAP with 5 components (n components=5) and a cosine distance metric, which is appropriate for comparing normalised document embeddings. The parameters (n neighbors=15, min dist=0.01) are selected to balance between preserving local relationships while maintaining a clear separation between distinct topic clusters.

The core topic modeling is performed using BERTopic Grootendorst (2022), which integrates embedding generation and vectorisation to automatically cluster similar documents into topics. An iterative refinement process is applied by first fitting the model to extract preliminary topics and then reducing outliers using both a clustered Term Frequency-Inverse Document Frequency (c-TF-IDF) strategy and probability thresholds. The c-TF-IDF aggregates documents belonging to the same topic into a single class-based document which makes it particularly effective for topic modeling as it emphasises terms that are distinctive for entire topics rather than individual documents. I use c-TF-IDF with a threshold of 0.2, which helps consolidate semantically similar topics and remove noise from the topic distribution. This two-step outlier removal minimises the influence of anomalous data points and yields more coherent topics. Further, the model is updated with the new topic assignments and the vectoriser, ensuring consistency across iterations. Metadata such as timestamps is also extracted to explore temporal trends in topic distribution, and the results, including detailed topic information and visualisations like an intertopic distance map, are saved for additional topic clustering.

C2.3 Topic creation

Figures C2.6 and C2.7 present wordclouds of the topics, Figure C2.8 provides a static representation of the intertopic distance map which aided in grouping the topics, Figure C2.9 provides a visualisation of the regrouping per topic and Tables C2.6 to C2.8 present the highest c-TF-IDF words. These figures are jointly used to determine 10 larger topics manually. As additional validation in line with Grootendorst (2022), ChatGPT was prompted to also regroup the topics within the 10 larger topics.



Figure C2.6: Wordcloud of the topics 1-22

Note: The wordclouds were generated for each topic to visually represent the frequency and relevance of key terms derived from the BERTopic model.



Figure C2.7: Wordcloud of the topics topics 23-44

Note: The wordclouds were generated for each topic to visually represent the frequency and relevance of key terms derived from the BERTopic model.

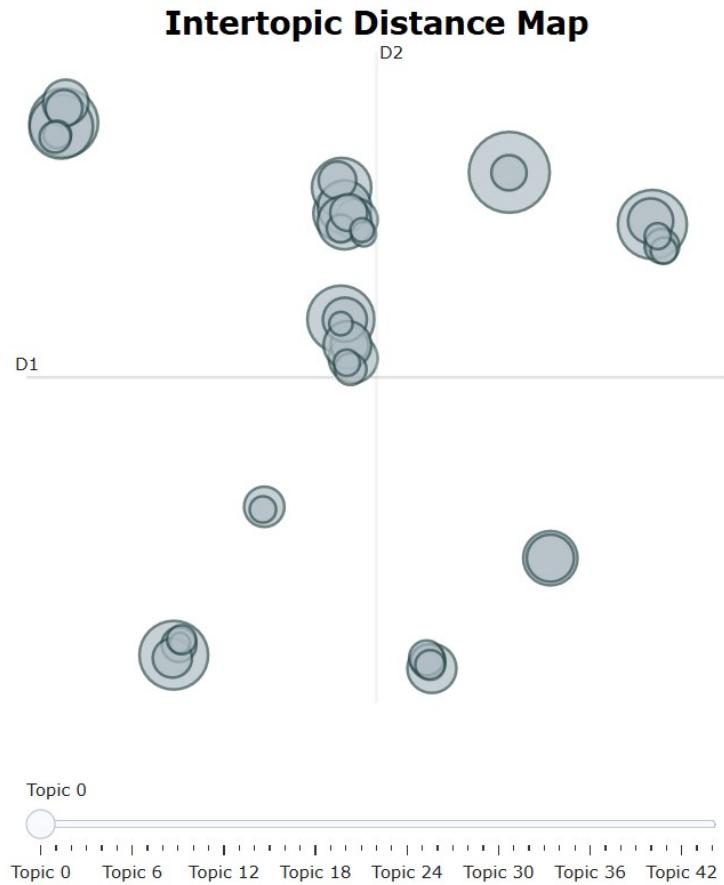


Figure C2.8: Static representation of the 2D intertopic distance map

Note: Within the BERTopic framework, the intertopic distance map is an interactive 2D visualisation that plots each topic as a point in a reduced-dimensional space derived from UMAP. The positioning of these topic points reflects their semantic relatedness, so that topics with overlapping keywords and thematic content appear closer together, while more distinct topics remain farther apart. I use this as an additional visual tool to help identify clusters that may be merged.

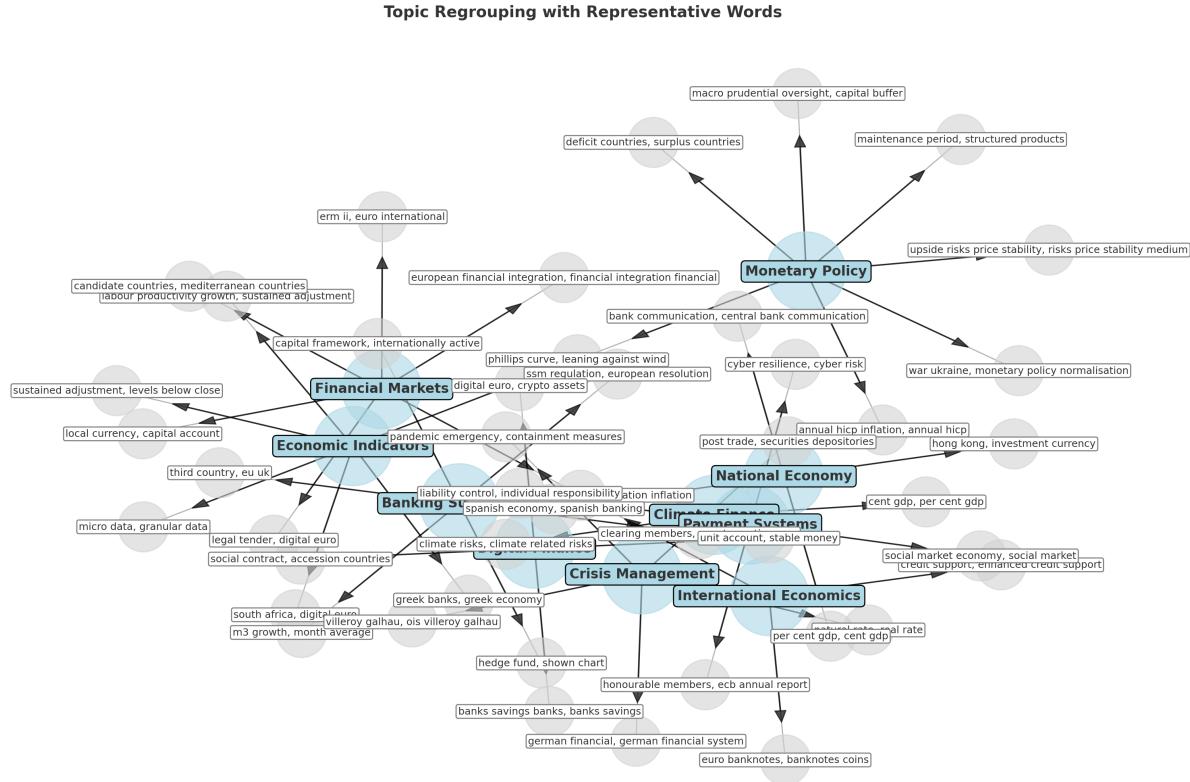


Figure C2.9: Topic regrouping map

Note: This figure shows the 10 overarching topics and the representative key words of the topics that have been aggregated into these overarching topics, i.e., the BERTopic output.

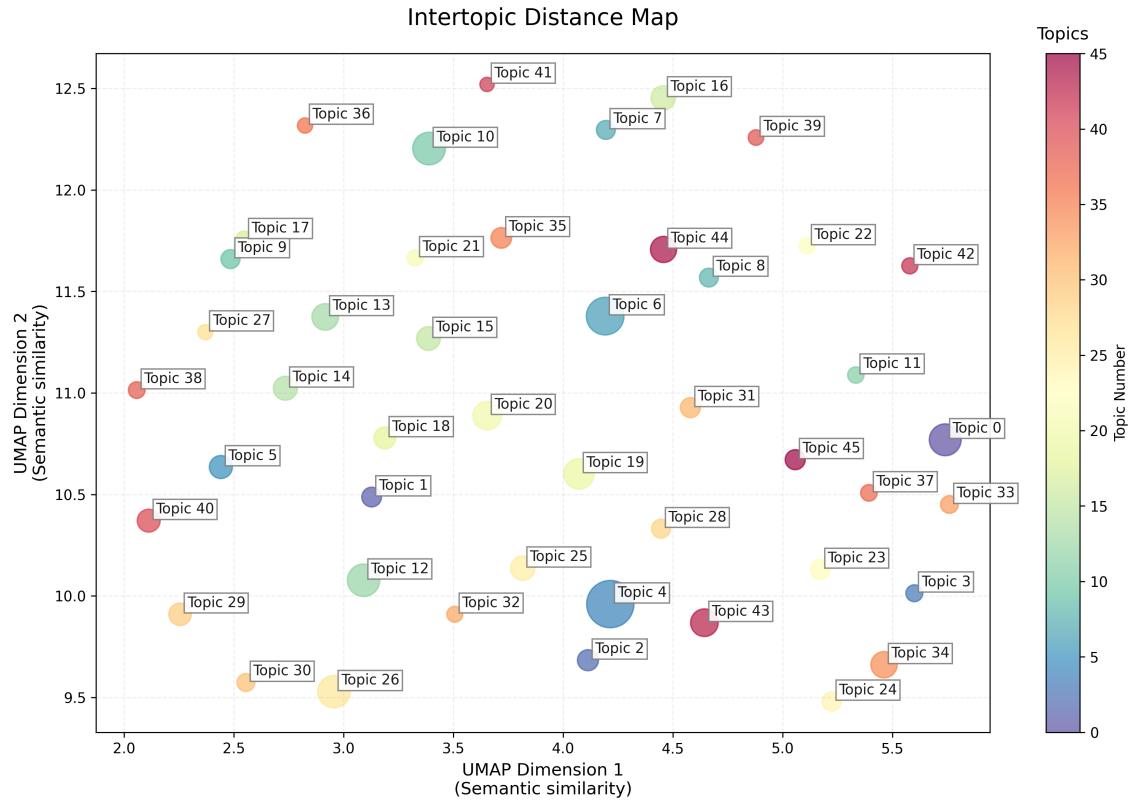


Figure C2.10: UMAP intertopic distance map

Note: The visualisation employs UMAP McInnes et al. (2020) to project high-dimensional topic embeddings onto a two-dimensional space. The apparent uniform distribution of topics stems from UMAP's preservation of both local and global topological structure while optimising the lower-dimensional representation. Inter-topic distances reflect semantic similarities: closer topics share more semantic content, while distant topics are more distinct. Topic sizes (represented by circle diameter) indicate relative frequency in the corpus.

Table C2.6: Topics 1 to 15: most important terms with c-TF-IDF scores

Topic	Documents	Top Terms (c-TF-IDF Score)
0	310	annual hicp inflation (0.014); annual hicp (0.014); upside risks price stability (0.013); according eurostat (0.013); today meeting (0.012); risks price stability medium (0.012); eurostat flash estimate (0.012); eurostat flash (0.012)
1	223	spanish economy (0.089); spanish banking (0.044); spanish banking system (0.025); financial education (0.017); pension system (0.016); spanish financial (0.015); financial position (0.014); budgetary policy (0.014)
2	224	digital euro (0.129); crypto assets (0.092); bank digital (0.056); central bank digital (0.056); instant payments (0.055); digital currency (0.050); crypto asset (0.049); payment solutions (0.049)
3	192	per cent gdp (0.038); cent gdp (0.038); italian economy (0.037); italian banking (0.025); net borrowing (0.022); per cent total (0.020); cent total (0.020); cooperative banks (0.020)
4	226	labour productivity growth (0.025); sustained adjustment (0.011); employment rate (0.010); labour utilisation (0.009); productivity developments (0.009); lisbon agenda (0.009); inflation aim (0.009); convergence inflation (0.008)
5	154	climate risks (0.075); climate related risks (0.072); sustainable finance (0.071); climate risk (0.054); paris agreement (0.053); net zero (0.052); greening financial (0.051); environmental risks (0.050)
6	140	ssm regulation (0.042); european resolution (0.024); national competent (0.023); supervisory information (0.023); joint supervisory (0.022); national competent authorities (0.021); supervisory powers (0.021); supervisory teams (0.020)
7	167	capital framework (0.054); internationally active (0.039); basel ii framework (0.032); ii framework (0.032); trading book (0.031); active banks (0.030); risk sensitivity (0.029); internationally active banks (0.027)
8	135	pandemic emergency (0.032); containment measures (0.026); impact pandemic (0.025); pandemic emergency purchase programme (0.025); pandemic emergency purchase (0.025); emergency purchase programme (0.025); emergency purchase (0.025); programme pepp (0.023)
9	114	villeroy galhau (0.054); ois villeroy galhau (0.045); franc ois villeroy galhau (0.045); ois villeroy (0.045); franc ois villeroy (0.045); franc ois (0.045); galhau governor banque france (0.043); galhau governor banque (0.043)
10	85	european financial integration (0.061); financial integration financial (0.030); post trading (0.029); accession countries (0.028); financial integration euro (0.026); standard deviation (0.025); financial integration euro area (0.025); corporate bond market (0.024)
11	213	erm ii (0.024); euro international (0.019); inflation differentials (0.016); convergence criteria (0.015); monetary growth (0.014); international euro (0.014); real convergence (0.013); euro adoption (0.011)
12	113	phillips curve (0.032); leaning against wind (0.018); monetary policy frameworks (0.015); relationship between monetary (0.014); price misalignments (0.013); asset price misalignments (0.013); dsge models (0.013); short term interest rate (0.012)
13	97	sustained adjustment (0.024); levels below close (0.024); sustained adjustment path (0.018); adjustment path (0.018); sustained adjustment path inflation (0.018); adjustment path inflation (0.017); path inflation (0.017); term premium (0.017)
14	98	honourable members (0.028); ecb annual report (0.024); ecb annual (0.024); committee economic monetary affairs (0.015); committee economic monetary (0.015); committee economic (0.015); economic monetary developments (0.014); madam chair (0.013)

Table C2.7: Topics 16 to 30: most important terms with c-TF-IDF scores

Topic	Documents	Top Terms (c-TF-IDF Score)
15	106	liability control (0.044); individual responsibility (0.043); principle individual (0.037); between liability control (0.032); between liability (0.032); maastricht framework (0.028); balance between liability (0.027); balance between liability control (0.027)
16	67	war ukraine (0.042); monetary policy normalisation (0.029); russia invasion (0.027); above target (0.026); energy costs (0.026); russia invasion ukraine (0.025); supply bottlenecks (0.025); incoming data (0.022)
17	59	upside risks price stability (0.045); risks price stability medium (0.035); price stability medium longer (0.023); stability medium longer (0.023); indirect taxes (0.022); annual hicp inflation (0.021); annual hicp (0.021); stability medium longer term (0.020)
18	92	international liquidity (0.030); globalisation inflation (0.026); exchange rate flexibility (0.024); rate flexibility (0.024); capital account (0.023); capital controls (0.022); reserve currency (0.021); global savings (0.020)
19	108	maintenance period (0.036); structured products (0.029); term money market (0.028); short term money (0.024); short term money market (0.022); term money market rates (0.021); originate distribute (0.019); open market operations (0.018)
20	48	micro data (0.107); granular data (0.092); machine learning (0.080); financial statistics (0.069); data collection (0.059); official statistics (0.057); household finance (0.052); data gaps (0.046)
21	67	macro prudential oversight (0.039); capital buffer (0.036); borrower based (0.035); macroprudential authorities (0.031); countercyclical capital buffer (0.029); prudential measures (0.023); warnings recommendations (0.023); risk warnings (0.023)
22	57	south africa (0.057); digital euro (0.053); digital currency (0.052); central bank digital currency (0.043); bank digital currency (0.043); central bank digital (0.039); bank digital (0.039); german banking (0.037)
23	77	euro banknotes (0.137); banknotes coins (0.116); euro banknotes coins (0.099); cash changeover (0.078); euro cash (0.048); introduction euro banknotes coins (0.045); introduction euro banknotes (0.044); january 2002 (0.039)
24	137	deficit countries (0.022); surplus countries (0.021); fiscal governance (0.018); greece ireland (0.013); framework monetary union (0.013); fiscal macroeconomic (0.012); governance euro (0.012); stable monetary union (0.012)
25	38	third country (0.140); eu uk (0.133); uk eu (0.102); hard brexit (0.065); third countries (0.053); march 2019 (0.028); eu capital (0.028); failing likely (0.027)
26	64	m3 growth (0.041); month average (0.028); monetary growth (0.027); marginal lending (0.026); euro banknotes (0.025); marginal lending facility (0.025); lending facility (0.024); average annual growth (0.023)
27	37	greek banks (0.143); greek economy (0.120); bank greece (0.069); capital controls (0.049); mr draghi (0.043); adjustment programme (0.042); emergency liquidity assistance (0.035); greek debt (0.035)
28	57	social contract (0.072); accession countries (0.071); civil society (0.061); real convergence (0.026); treaty rome (0.019); erm ii (0.019); european social (0.018); public goods (0.017)
29	33	candidate countries (0.157); mediterranean countries (0.132); accession countries (0.102); eu accession (0.084); eu central (0.047); erm ii (0.041); real convergence (0.039); czech republic (0.036)

Table C2.8: Topics 31 to 45: most important terms with c-TF-IDF scores

Topic	Documents	Top Terms (c-TF-IDF Score)
30	65	local currency (0.047); capital account (0.028); financial reform (0.025); financial safety (0.023); capital flow (0.022); international financial architecture (0.021); financial sector reforms (0.020); multilateral surveillance (0.019)
31	36	cent gdp (0.039); per cent gdp (0.039); spanish economy (0.037); italian banking (0.032); around per cent (0.027); around per (0.026); asian countries (0.023); external deficit (0.022)
32	56	german financial (0.077); german financial system (0.058); german banking (0.049); residential real estate (0.040); residential real (0.040); banks savings (0.038); banks savings banks (0.037); capital buffer (0.033)
33	33	legal tender (0.132); digital euro (0.120); euro banknotes (0.119); cash payments (0.080); payment methods (0.073); paper money (0.070); euro cash (0.064); security features (0.056)
34	33	hedge fund (0.075); shown chart (0.067); chart slide (0.062); chart left (0.061); household sector (0.057); credit cycle (0.044); commercial property (0.042); sources risk (0.042)
35	73	credit support (0.028); enhanced credit support (0.024); enhanced credit (0.024); credit easing (0.022); non conventional (0.017); rate full allotment (0.017); fixed rate full allotment (0.016); fixed rate full (0.016)
36	58	banks savings banks (0.056); banks savings (0.055); savings bank (0.032); bank sector (0.030); cost efficiency (0.029); border mergers (0.028); cross border mergers (0.028); return equity (0.022)
37	46	natural rate (0.093); real rate (0.063); natural rate interest (0.053); equilibrium real (0.041); real equilibrium (0.030); equilibrium rate (0.030); equilibrium interest (0.028); equilibrium real rate (0.028)
38	29	clearing members (0.200); counterparties ccps (0.077); otc derivatives markets (0.076); central counterparties ccps (0.076); third country (0.068); resolution planning (0.065); trade repositories (0.064); centrally cleared (0.062)
39	32	social market economy (0.074); social market (0.073); labour force participation (0.044); force participation (0.044); global value (0.042); german economic (0.041); global value chains (0.040); retirement age (0.034)
40	42	unit account (0.053); stable money (0.049); hans tietmeyer (0.049); gold standard (0.041); ludwig erhard (0.038); woods system (0.032); bretton woods system (0.032); nation state (0.029)
41	27	post trade (0.186); securities depositories (0.168); target2 securities (0.163); collateral management (0.135); central securities (0.113); clearing settlement systems (0.094); financial market infrastructure (0.091); securities clearing settlement (0.080)
42	26	hong kong (0.110); investment currency (0.076); chinese economy (0.065); bank china (0.063); reserve currency (0.062); direct debits (0.060); emerging asia (0.040); global finance (0.035)
43	20	cyber resilience (0.754); cyber risk (0.310); cyber security (0.302); cyber risks (0.268); financial infrastructures (0.265); cyber attacks (0.143); financial entities (0.080); financial ecosystem (0.061)
44	40	bank communication (0.082); central bank communication (0.082); communication monetary (0.029); labour utilisation (0.028); communication monetary policy (0.027); communication policy (0.026); employment protection (0.022); path policy (0.021)

C3 Sequence analysis robustness checks

Below, the robustness tests of the sequence analysis are presented. First, first-movers graphs are discussed. Next, different time-windows for the Markov chains are discussed.

C3.1 First-mover graphs

Figures C3.11 and C3.12 present the first-movers graph for the crisis management and climate topics as a means of further illustrating the first-mover dynamics. Figure C3.11 shows that the ECB clearly dominates but not per se initiates crisis management. There is also high correspondence in the peaks pointing towards the ECB clearly responding to the NCBs. Further, in line with the radically different views of crisis management, France and Germany seem to be discussing it more extensively than other countries. Figure C3.12 shows that climate is a less prominent topic and only emerges in the last decade. With regards to first-movers it shows that the Netherlands was an initial driver in the early years while other countries followed. Only very recently under ECB president Lagarde was climate really prominent in the ECB.

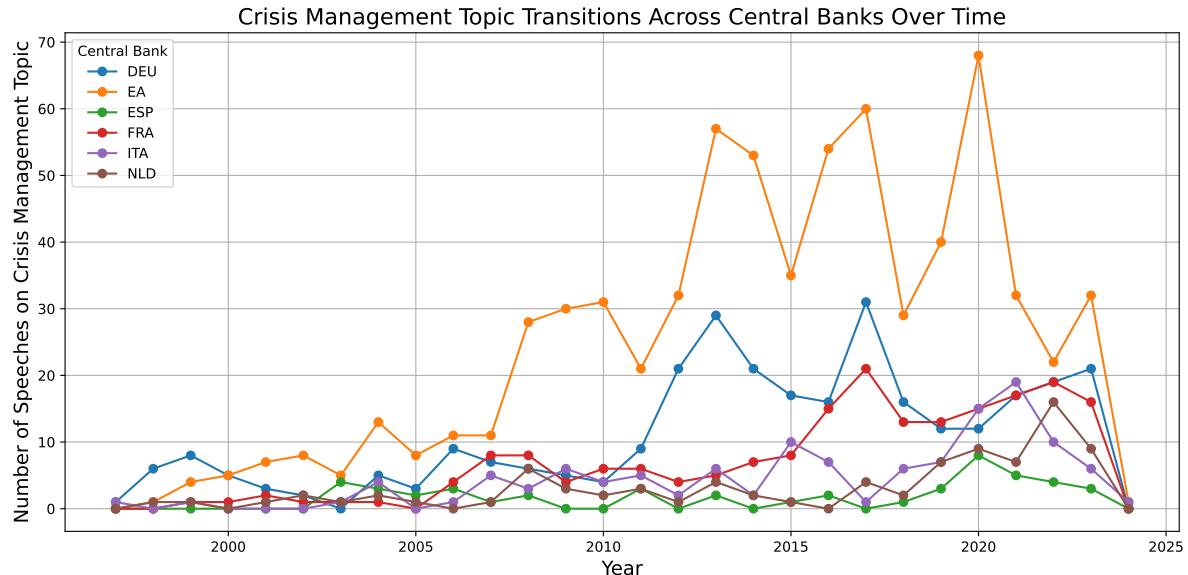


Figure C3.11: First-mover graph crisis management

Note: This figure displays the number of speeches given per central bank on the topic of crisis management. The speeches are counted when at least 25% of the speech concerns this topic.

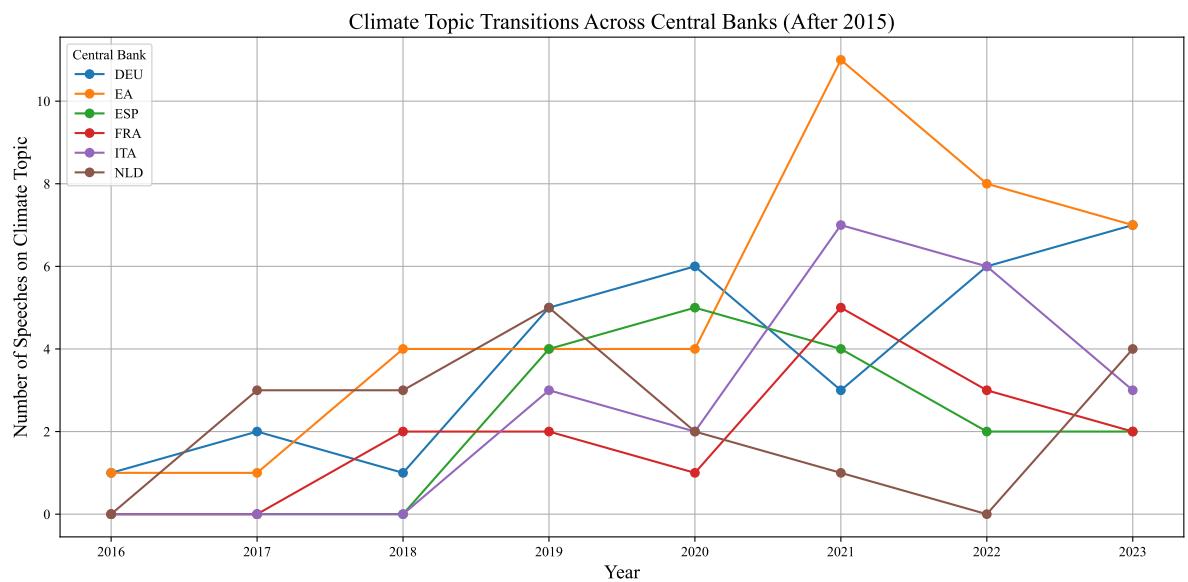


Figure C3.12: First-mover graph climate

Note: This figure displays the number of speeches given per central bank on the topic of climate. The speeches are counted when at least 25% of the speech concerns this topic.

C3.2 Markov windows robustness checks

In order to assess the sensitivity of the results to the specification of the Markov windows, I conduct a series of robustness tests. The main idea is to verify that the empirical findings are not driven by an arbitrary choice of window parameters. Moreover, the different patterns in varying timewindows can also be substantially interpreted.

Let w^* denote the baseline window length used in the analysis (e.g., $w^* = 4$ quarters). I re-estimate the main models using alternative window lengths $w \in \{w^* - \Delta, w^*, w^* + \Delta\}$. For each alternative specification, the Markov windows are recalculated by reassigning the state probabilities using the modified window length. Comparing the parameter estimates and their statistical significance across different window configurations shows consistency in the sign, magnitude and inference of the coefficients indicating that the main results of the study are robust to variations in the Markov window specification. Figures C3.13 to C3.22 summarises the findings from these alternative specifications for the between central bank transitions. Overall, the robustness tests confirm that the baseline results are not sensitive to the choice of window length.

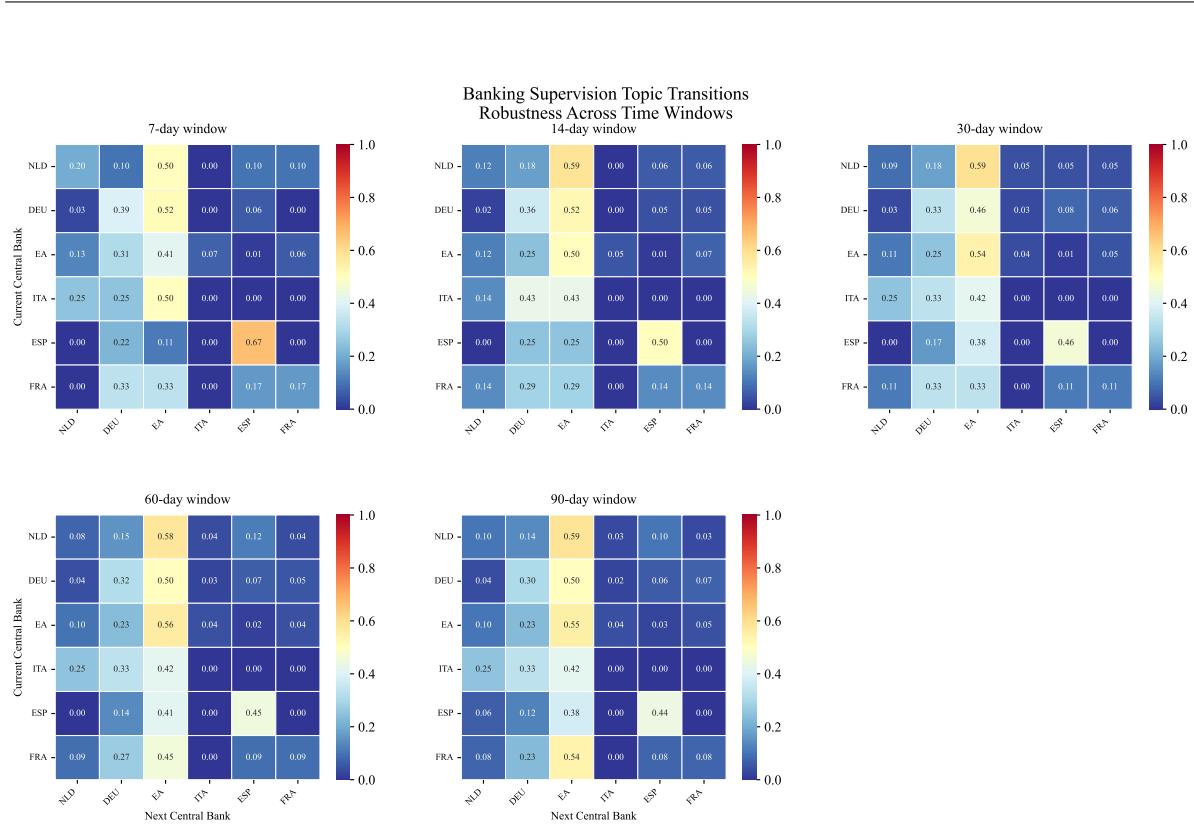


Figure C3.13: Matrix window robustness: banking supervision

Note: This figure shows the heatmaps for banking supervision transitions between central banks, displaying how one central bank (row) is followed by another central bank (column) within a 7, 14, 30, 60 and 90-day window, according to a first-order Markov process. Each cell represents the estimated transition probability, normalised so that each row sums to one, with darker (blue) shades indicating lower probabilities and warmer (yellow–red) shades indicating higher probabilities.

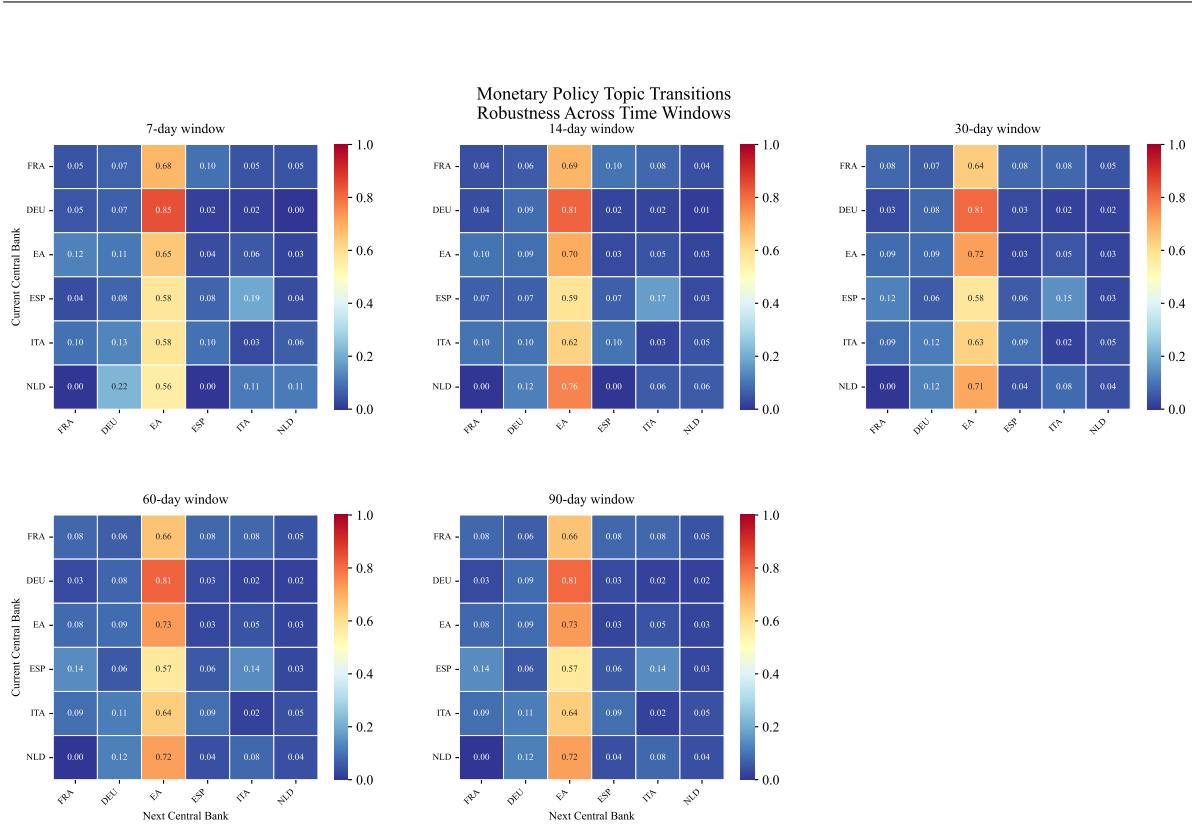


Figure C3.14: Matrix window robustness: monetary policy

Note: This figure shows the heatmaps for monetary policy transitions between central banks, displaying how one central bank (row) is followed by another central bank (column) within a 7, 14, 30, 60 and 90-day window, according to a first-order Markov process. Each cell represents the estimated transition probability, normalised so that each row sums to one, with darker (blue) shades indicating lower probabilities and warmer (yellow-red) shades indicating higher probabilities.

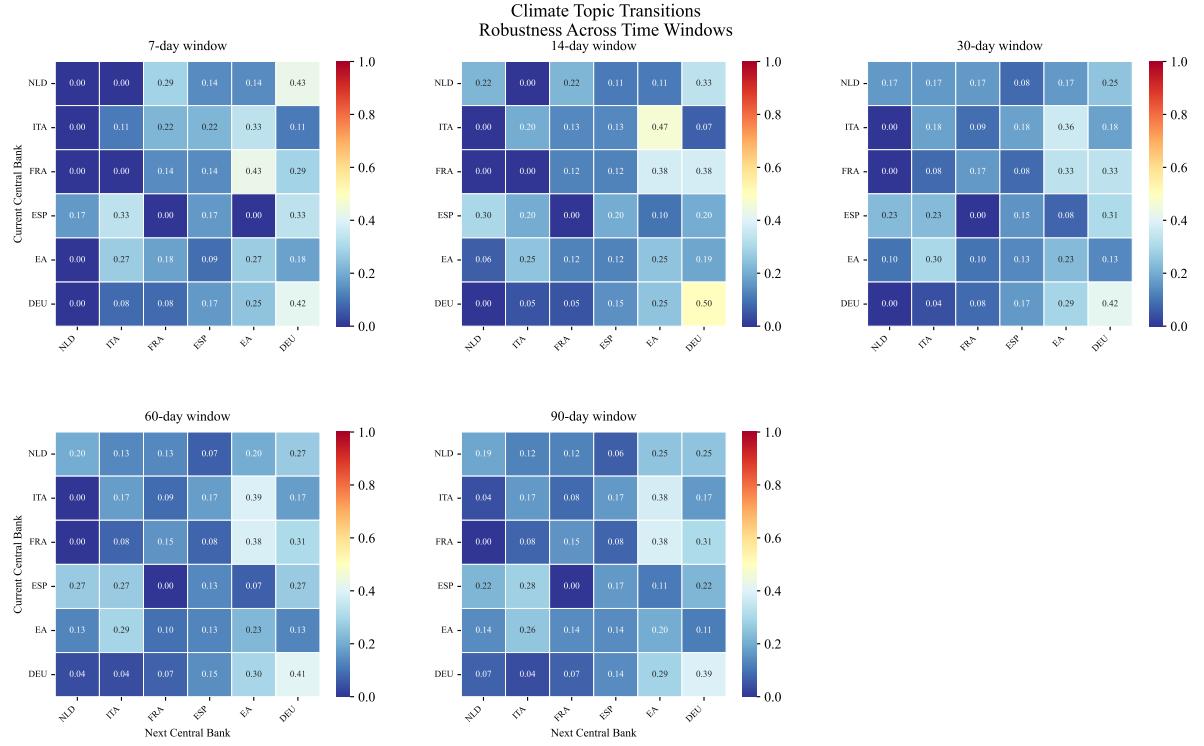


Figure C3.15: Matrix window robustness: climate

Note: This figure shows the heatmaps for climate transitions between central banks, displaying how one central bank (row) is followed by another central bank (column) within a 7, 14, 30, 60 and 90-day window, according to a first-order Markov process. Each cell represents the estimated transition probability, normalised so that each row sums to one, with darker (blue) shades indicating lower probabilities and warmer (yellow–red) shades indicating higher probabilities.

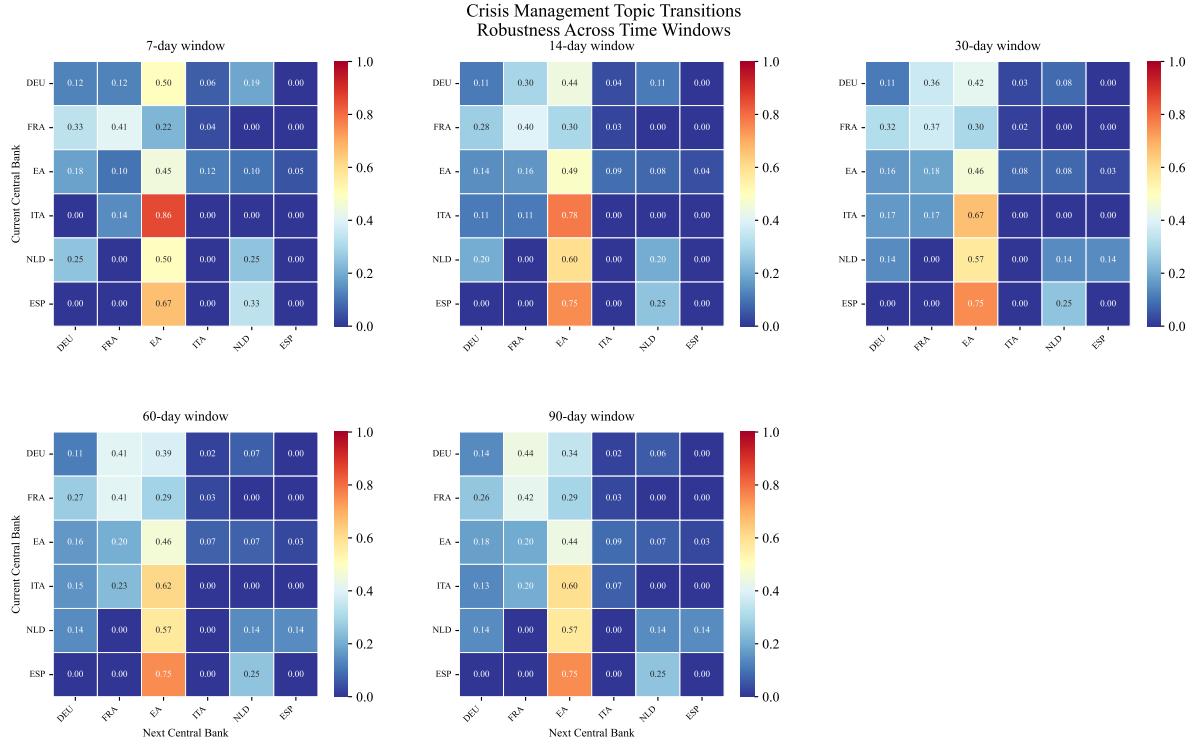


Figure C3.16: Matrix window robustness: crisis management

Note: This figure shows the heatmaps for crisis management transitions between central banks, displaying how one central bank (row) is followed by another central bank (column) within a 7, 14, 30, 60 and 90-day window, according to a first-order Markov process. Each cell represents the estimated transition probability, normalised so that each row sums to one, with darker (blue) shades indicating lower probabilities and warmer (yellow–red) shades indicating higher probabilities.

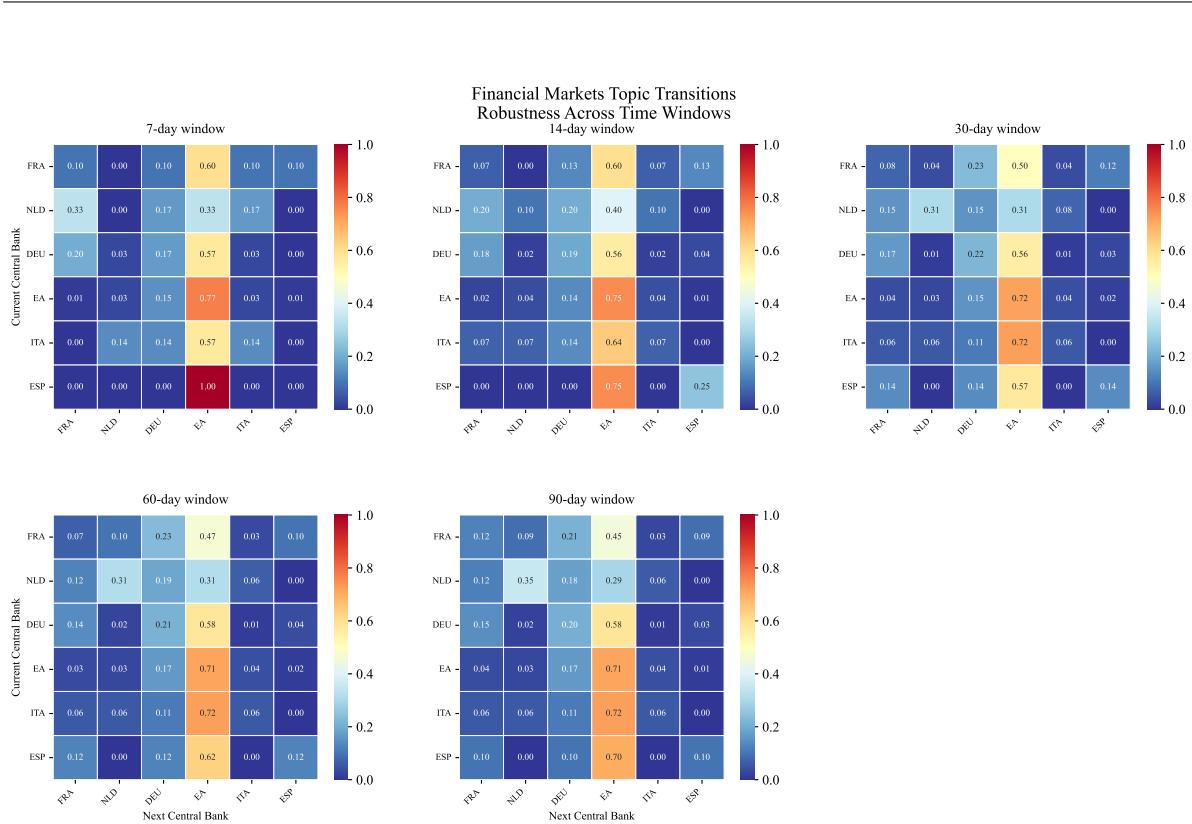


Figure C3.17: Matrix window robustness: financial markets

Note: This figure shows the heatmaps for financial markets transitions between central banks, displaying how one central bank (row) is followed by another central bank (column) within a 7, 14, 30, 60 and 90-day window, according to a first-order Markov process. Each cell represents the estimated transition probability, normalised so that each row sums to one, with darker (blue) shades indicating lower probabilities and warmer (yellow–red) shades indicating higher probabilities.

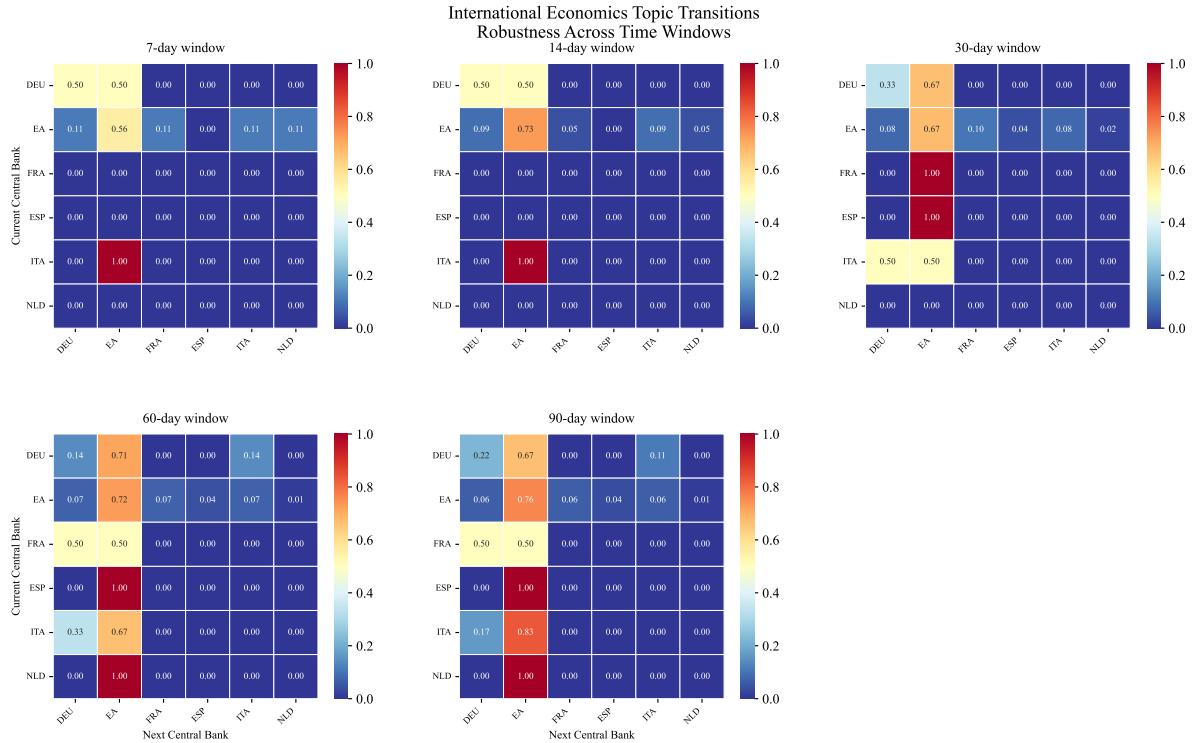


Figure C3.18: Matrix window robustness: international economics

Note: This figure shows the heatmaps for international economics transitions between central banks, displaying how one central bank (row) is followed by another central bank (column) within a 7, 14, 30, 60 and 90-day window, according to a first-order Markov process. Each cell represents the estimated transition probability, normalised so that each row sums to one, with darker (blue) shades indicating lower probabilities and warmer (yellow–red) shades indicating higher probabilities.

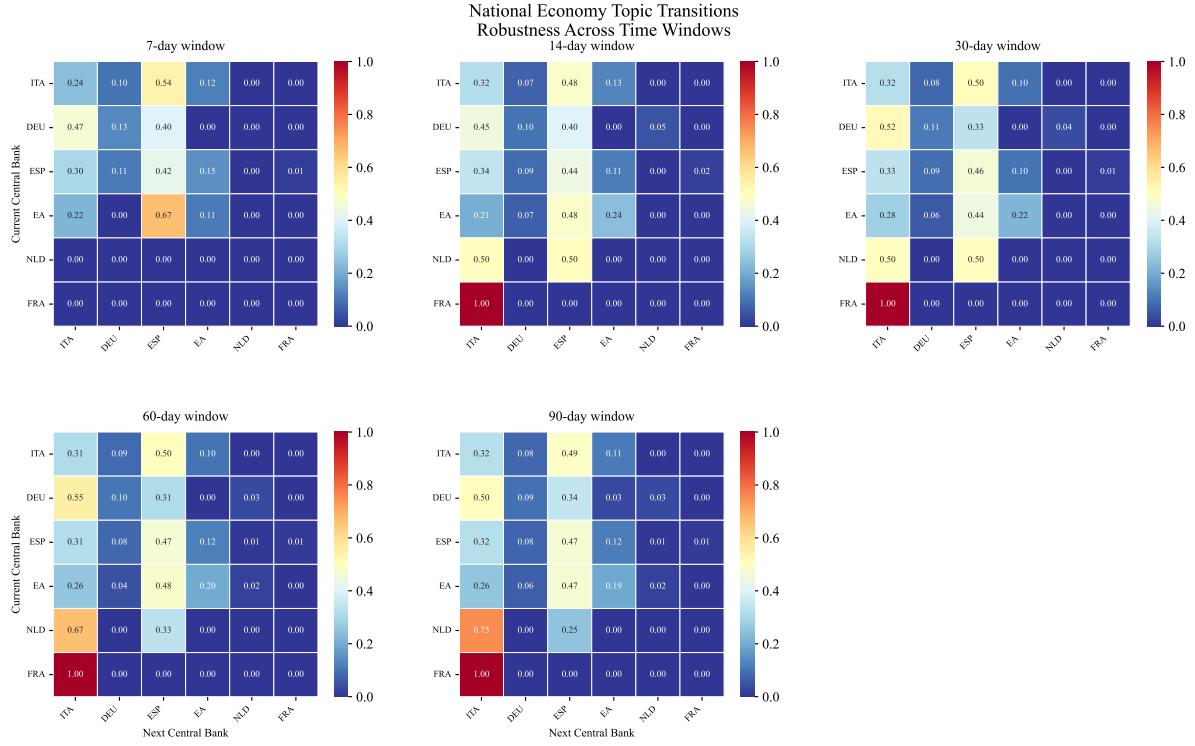


Figure C3.19: Matrix window robustness: national economy

Note: This figure shows the heatmaps for national economy transitions between central banks, displaying how one central bank (row) is followed by another central bank (column) within a 7, 14, 30, 60 and 90-day window, according to a first-order Markov process. Each cell represents the estimated transition probability, normalised so that each row sums to one, with darker (blue) shades indicating lower probabilities and warmer (yellow–red) shades indicating higher probabilities.

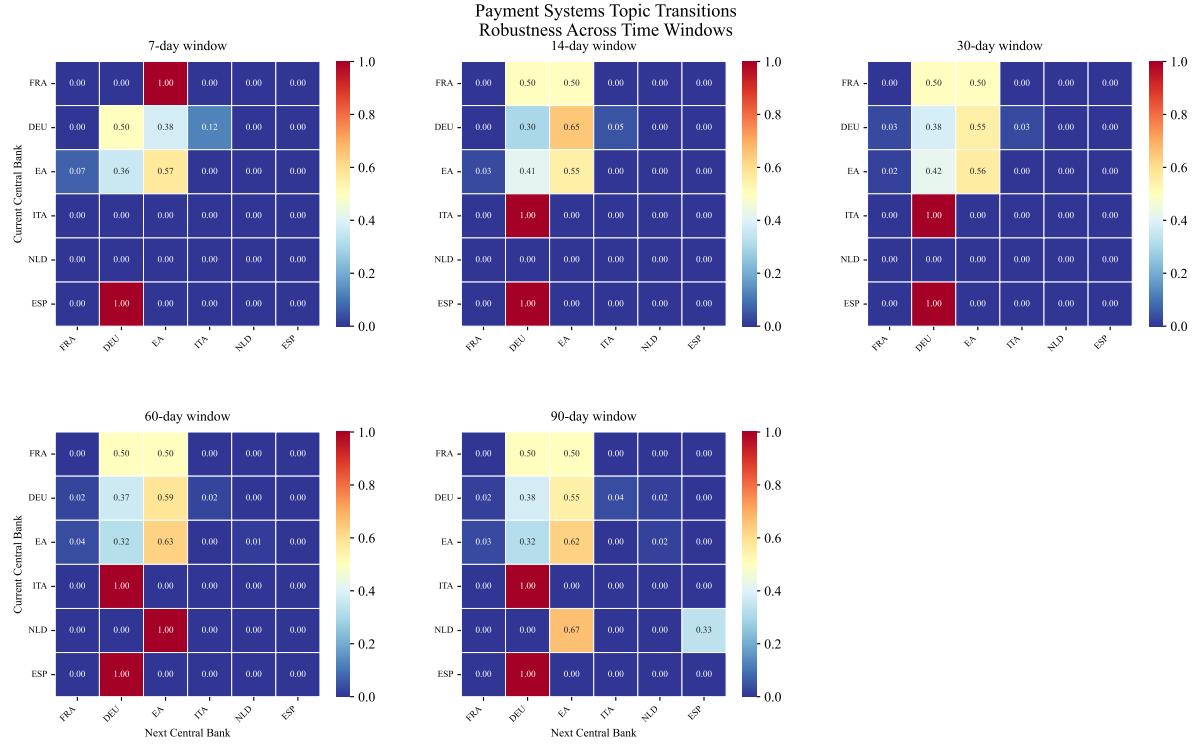


Figure C3.20: Matrix window robustness: payment systems

Note: This figure shows the heatmaps for payment system transitions between central banks, displaying how one central bank (row) is followed by another central bank (column) within a 7, 14, 30, 60 and 90-day window, according to a first-order Markov process. Each cell represents the estimated transition probability, normalised so that each row sums to one, with darker (blue) shades indicating lower probabilities and warmer (yellow–red) shades indicating higher probabilities.

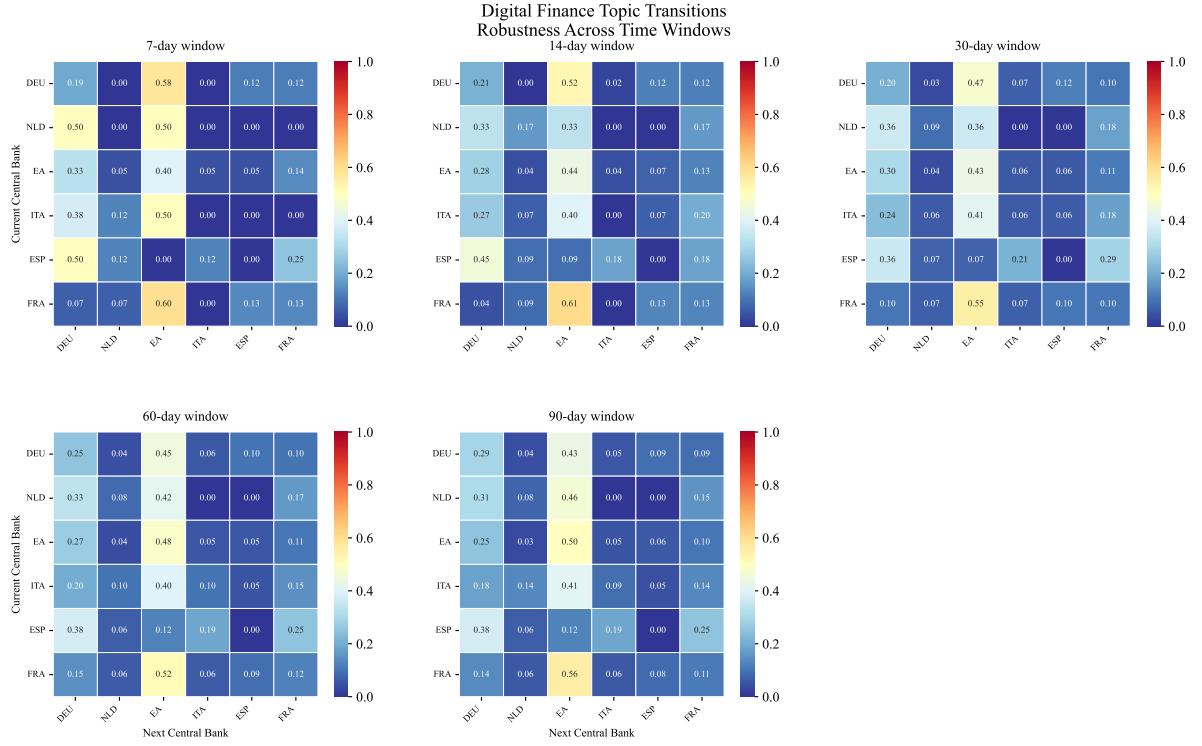


Figure C3.21: Matrix window robustness: digital finance

Note: This figure shows the heatmaps for digital finance transitions between central banks, displaying how one central bank (row) is followed by another central bank (column) within a 7, 14, 30, 60 and 90-day window, according to a first-order Markov process. Each cell represents the estimated transition probability, normalised so that each row sums to one, with darker (blue) shades indicating lower probabilities and warmer (yellow–red) shades indicating higher probabilities.

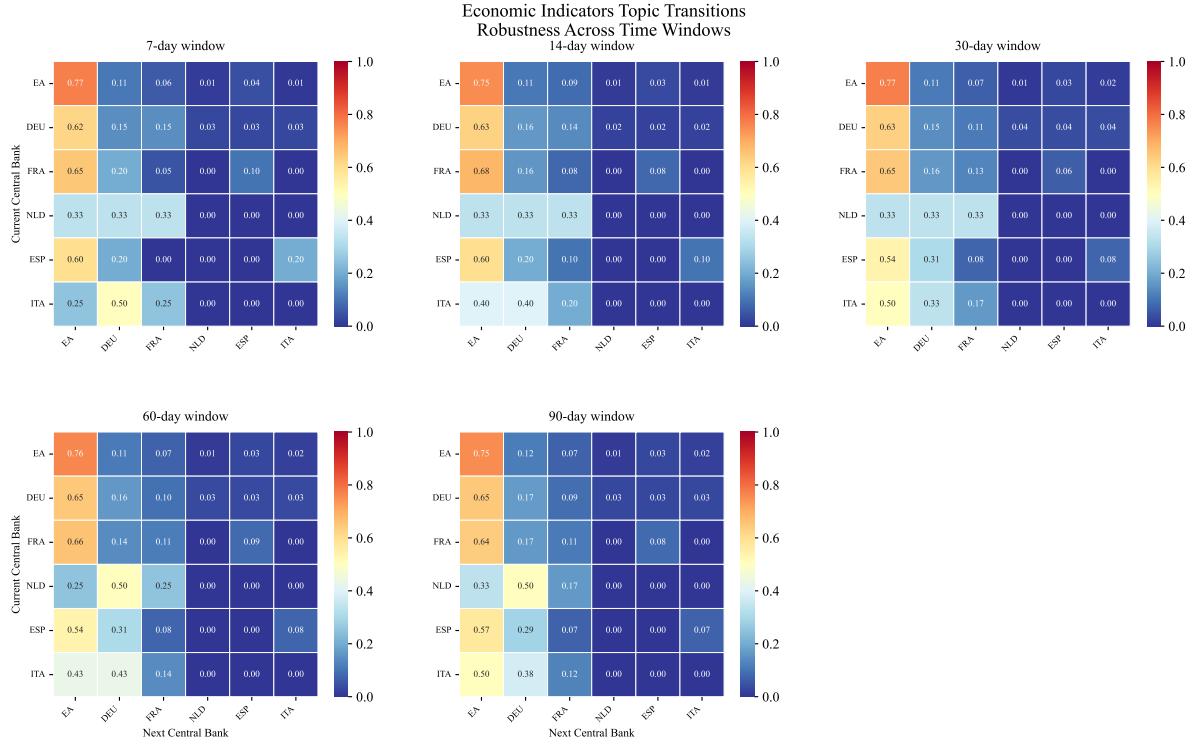


Figure C3.22: Matrix window robustness: economic indicators

Note: This figure shows the heatmaps for economic indicator transitions between central banks, displaying how one central bank (row) is followed by another central bank (column) within a 7, 14, 30, 60 and 90-day window, according to a first-order Markov process. Each cell represents the estimated transition probability, normalised so that each row sums to one, with darker (blue) shades indicating lower probabilities and warmer (yellow–red) shades indicating higher probabilities.

C4 Timeseries analysis robustness checks

Below the robustness tests of the two-way coefficient plots, the three way interaction margin plots, the half-yearly three-way interaction grid and several tables of the figures presented in the main text are discussed.

C4.1 Two-way interactions

Figure C4.23 below shows the pre and post crisis two-way interaction coefficient plots. The results show differences pre and post-crisis, however, they do not point systematically into one direction and can all be interpreted when accounting for the changed context. Figure C4.24 shows the coefficient plot estimated with leads instead of lags. This figure shows that the responsiveness also goes the other way, i.e., the NCBs responds to the ECB. Nevertheless, the results of the leads are less strong. Figure C4.25 below shows the coefficient plot with half-year instead of quarterly lags. The results go into a similar direction as the quarterly results showing robustness.

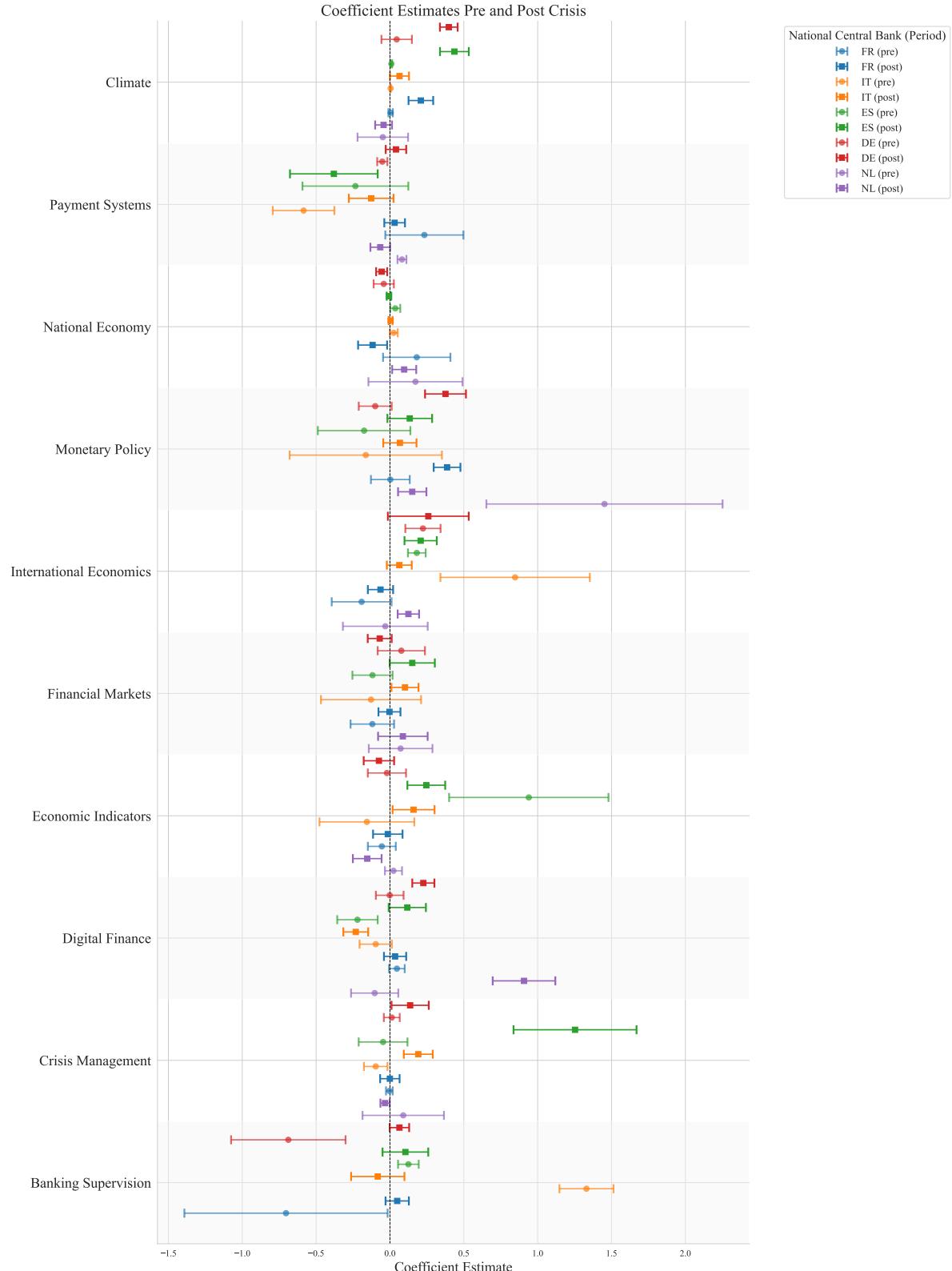


Figure C4.23: Coefficient plot pre and post-crisis

Note: Coefficient plot following Model B. The sample is split in pre and post-crisis (with 2008 as the cut-off date). 95% confidence intervals are included.

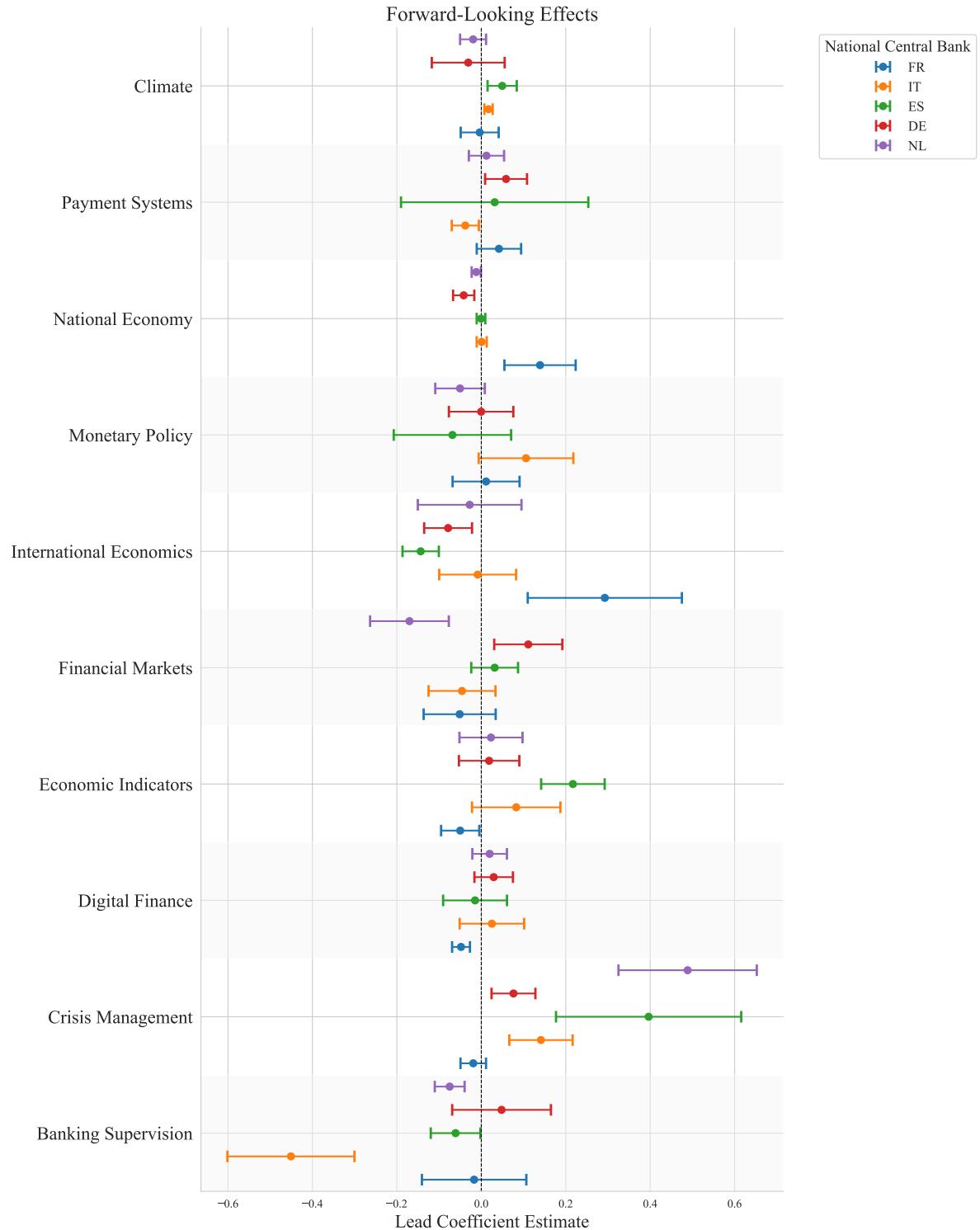


Figure C4.24: Coefficient plot with leads

Note: Coefficient plot following Model B. The lags are replaced by leads. 95% confidence intervals are included.

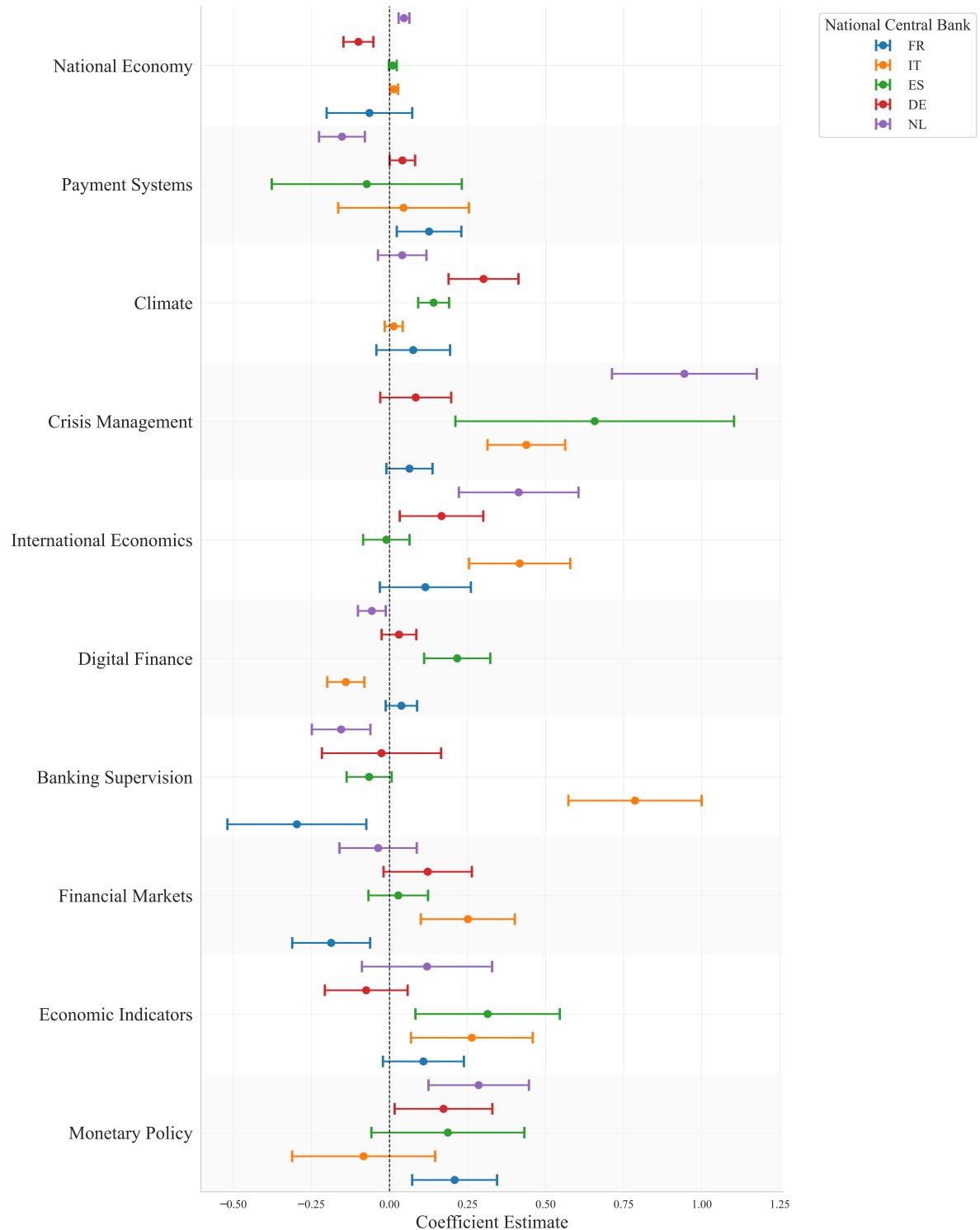


Figure C4.25: Coefficient plot half-year lags

Note: Coefficient plot following Model B. Half year lags are used instead of 3 month lags. 95% confidence intervals are included.

C4.2 Three-way interaction margin plots

For further exploration of the three-way interactions, I added all other three-way interaction margin plots following Model C presented in the main text. Rather than using binary indicators, the continuous indicators are now used of the proxies of salience, trust in the ECB, debt-GDP ratio and inflation. The results predominately go in the same direction for most countries showing some heterogenous effects between countries in line with expectations discussed in the text.

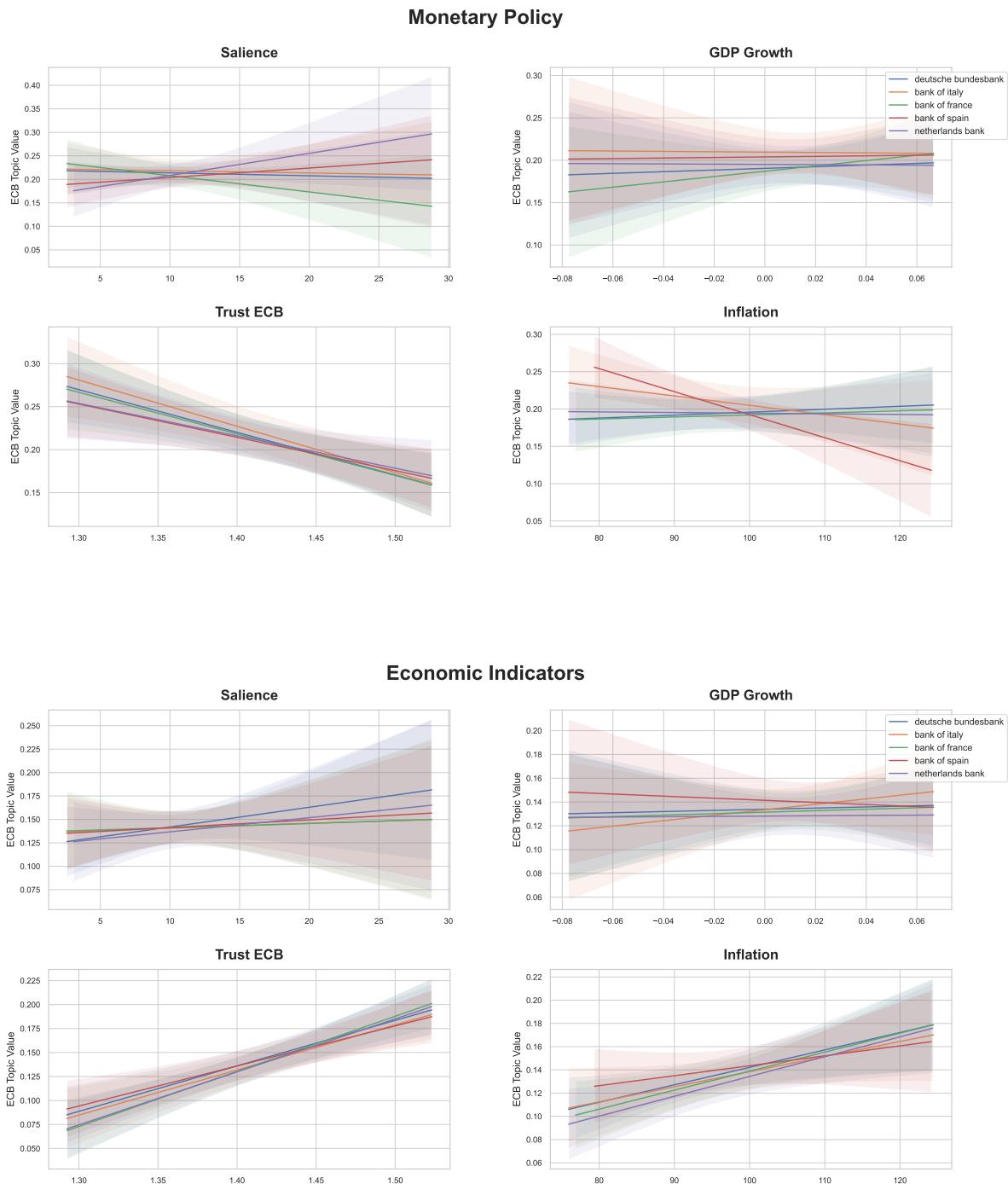


Figure C4.26: Margin plots: monetary policy and economic indicators

Note: Three-way interactions of monetary policy and economic indicators following Model C. 95% confidence intervals are included.

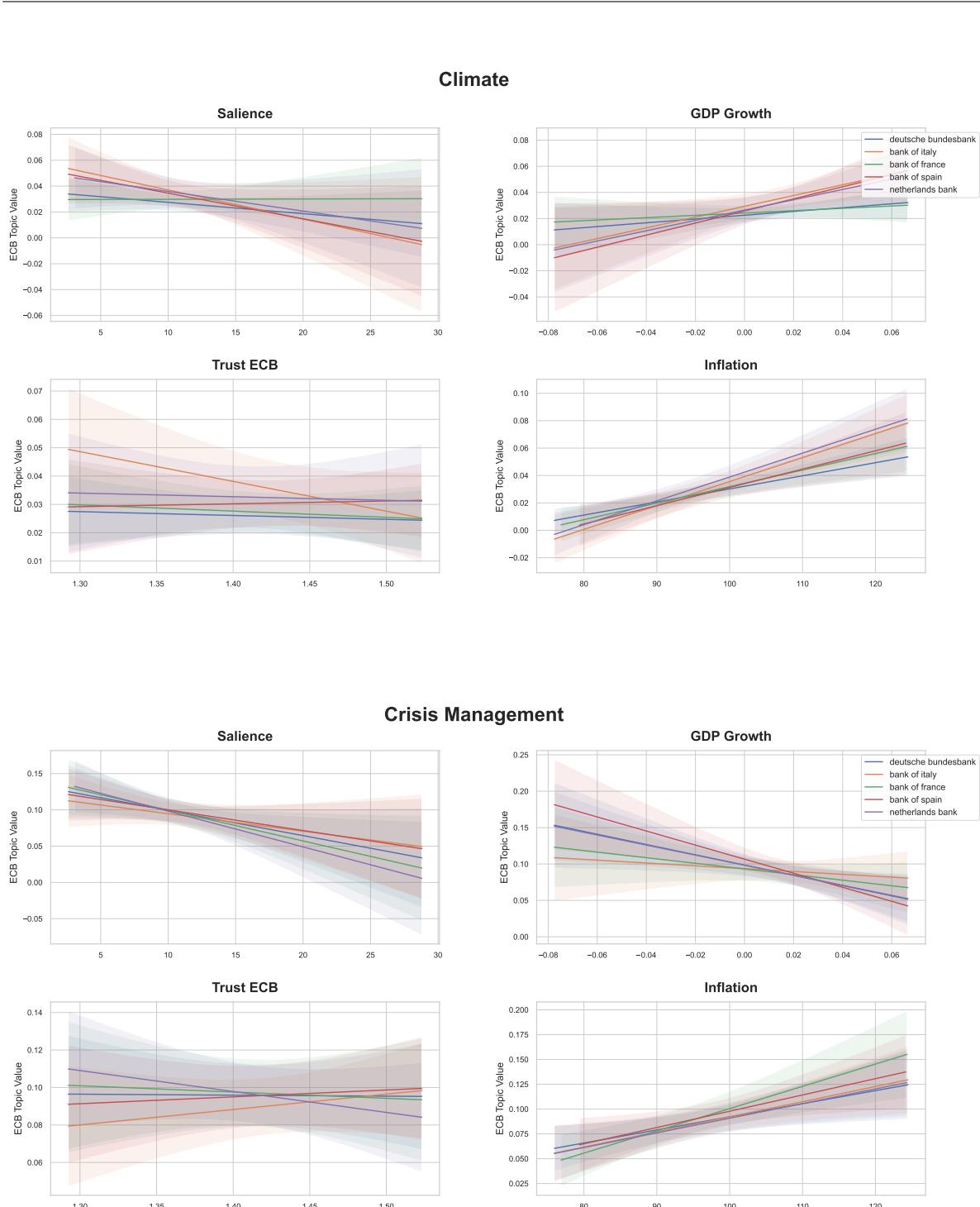
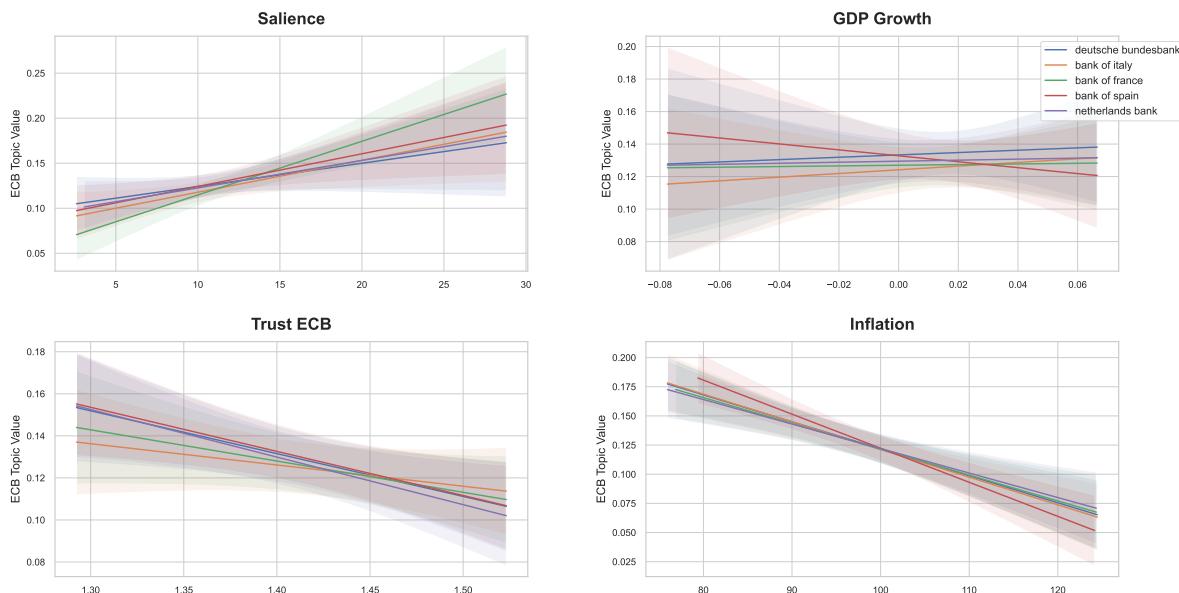


Figure C4.27: Margin plots: climate and crisis management

Note: Three-way interactions of climate and crisis management following Model C. 95% confidence intervals are included.

Financial Markets



Banking Supervision

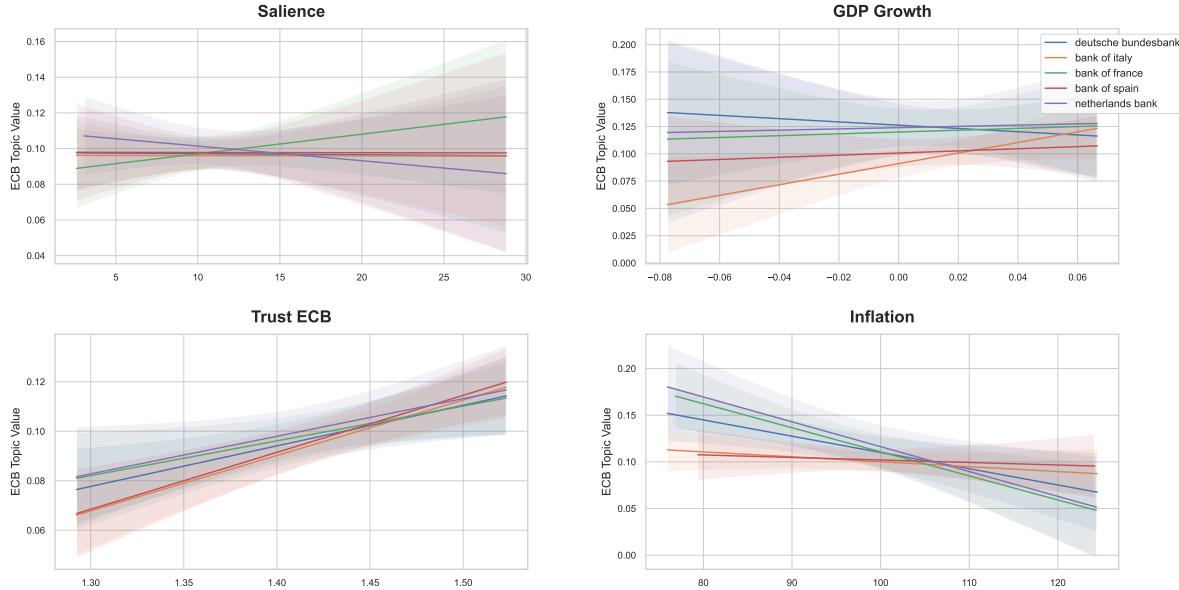


Figure C4.28: Margin plots: financial markets and banking supervision

Note: Three-way interactions of financial markets and banking supervision following Model C. 95% confidence intervals are included.

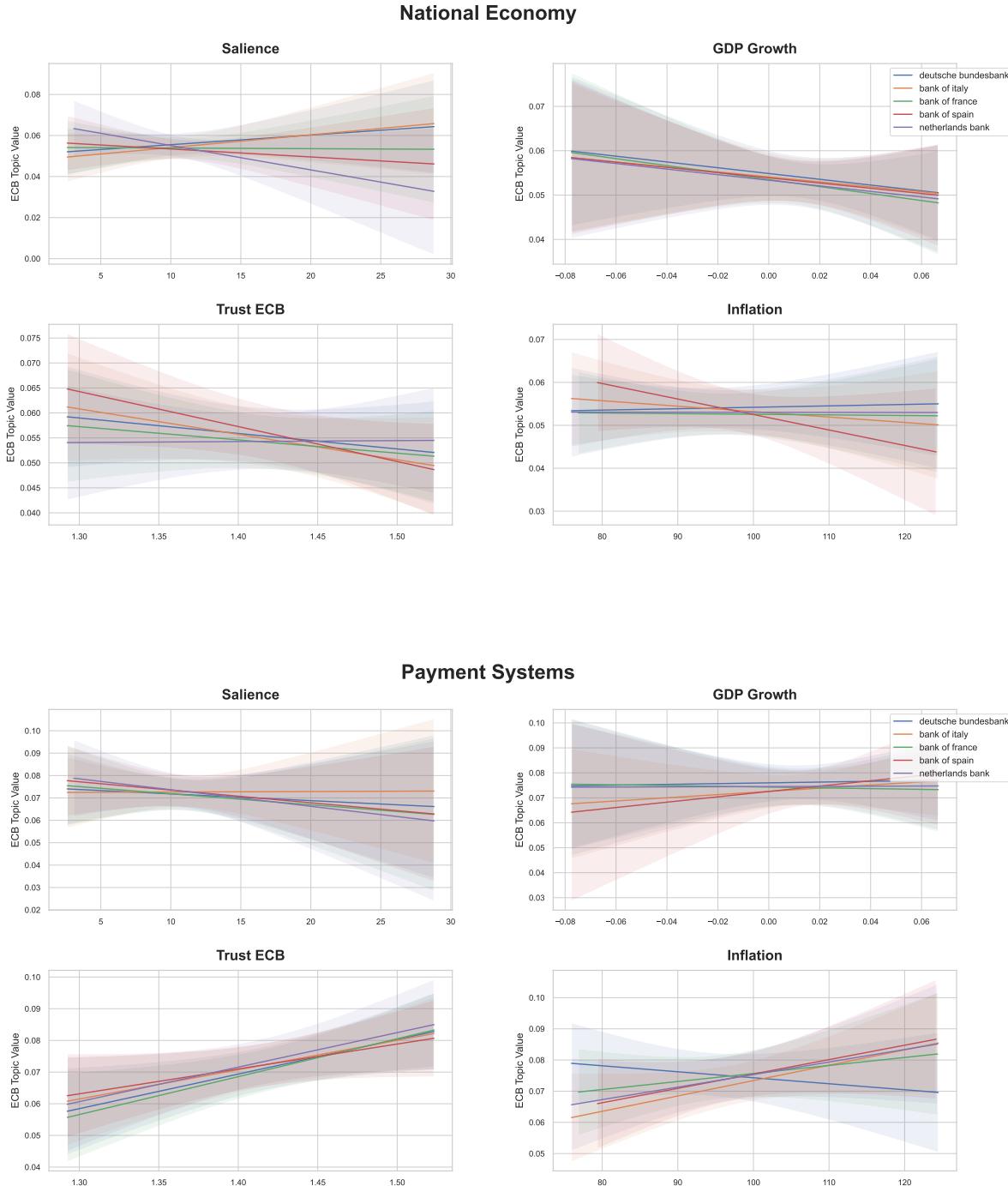
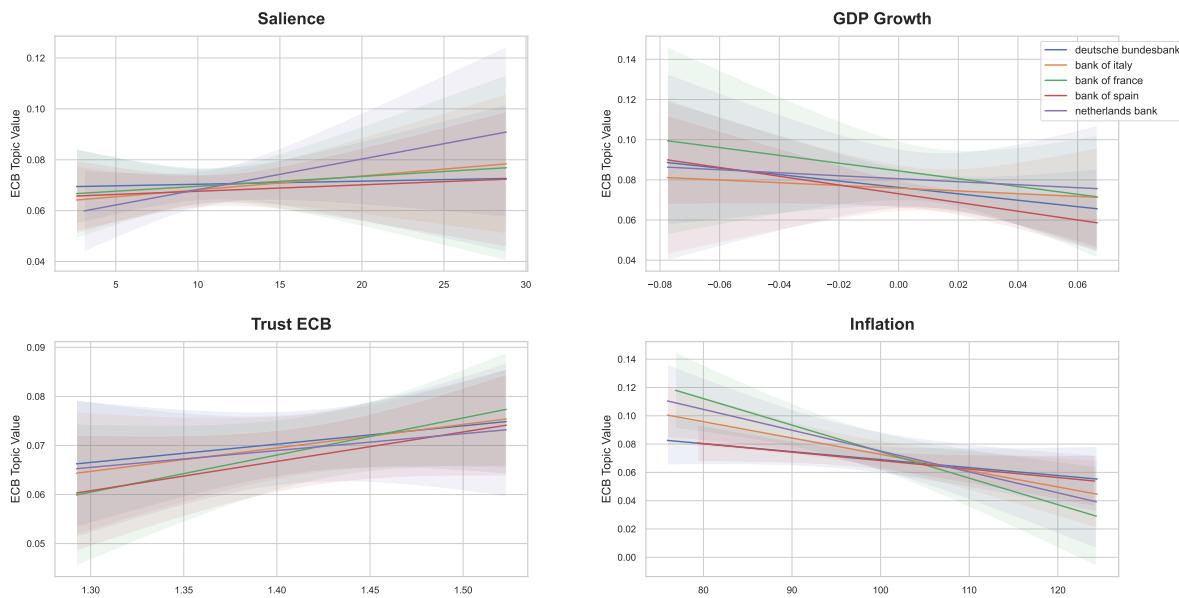


Figure C4.29: Margin plots: national economics and payment systems

Note: Three-way interactions of national economics and payment systems following Model C. 95% confidence intervals are included.

International Economics



Digital Finance

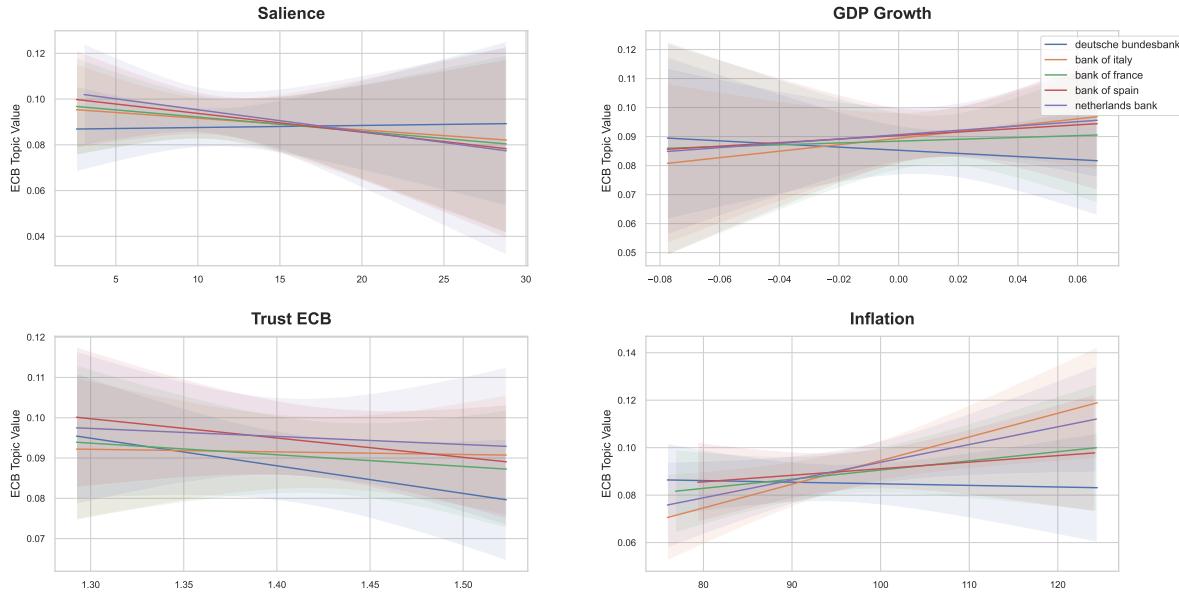


Figure C4.30: Margin plots: international economics and digital finance

Note: Three-way interactions of international economics and digital finance following Model C. 95% confidence intervals are included.

C4.3 Three-way interaction grid robustness tests

Figure C4.31 below shows the three-way interaction grid with half-yearly lags instead of quarterly lags as specified in the main text. Again, the results seem to go in a similar direction, which shows robustness.

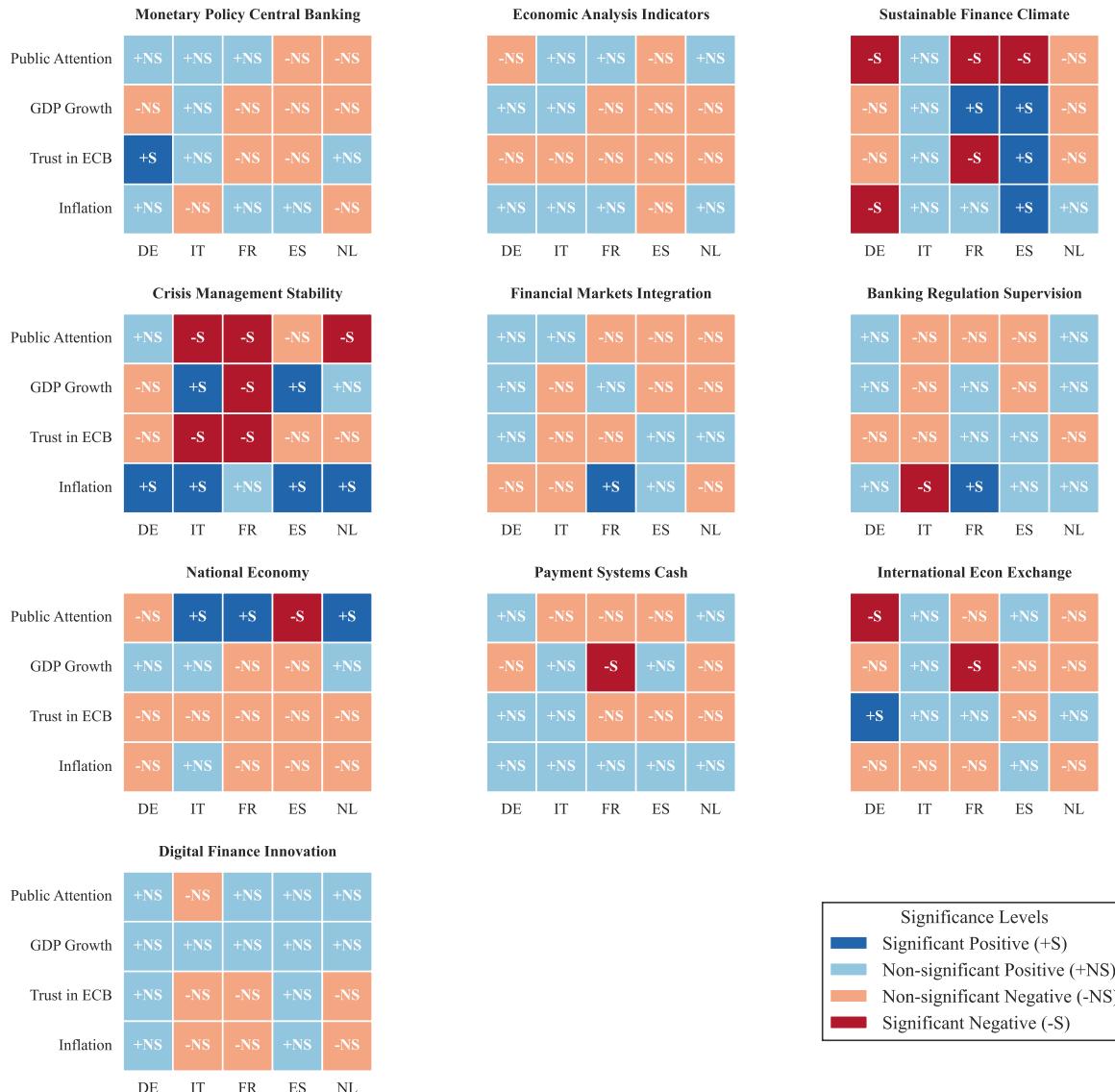


Figure C4.31: Interaction grid of half yearly three-way interactions

Note: This interaction grid follows the regressions as depicted in C3 but then for half yearly lags. The proxies of salience, trust in the ECB, GDP growth and debt-GDP are binary and represent a lower and higher within country average. For significance, a p-value of 0.05 is employed. Panel clustered standard errors are employed.

C5 Heterogeneous effects

This section provides extra information on the heterogeneous effects discussed in the main text.

C5.1 Salience: Eurocrisis

I further examine whether, under very salient conditions, the responsiveness dynamics change. I subset the dataset for only the euro crisis years (2010-2015) and re-estimate the transitions between central banks conditional on the topic. Figure C5.32 shows the results.

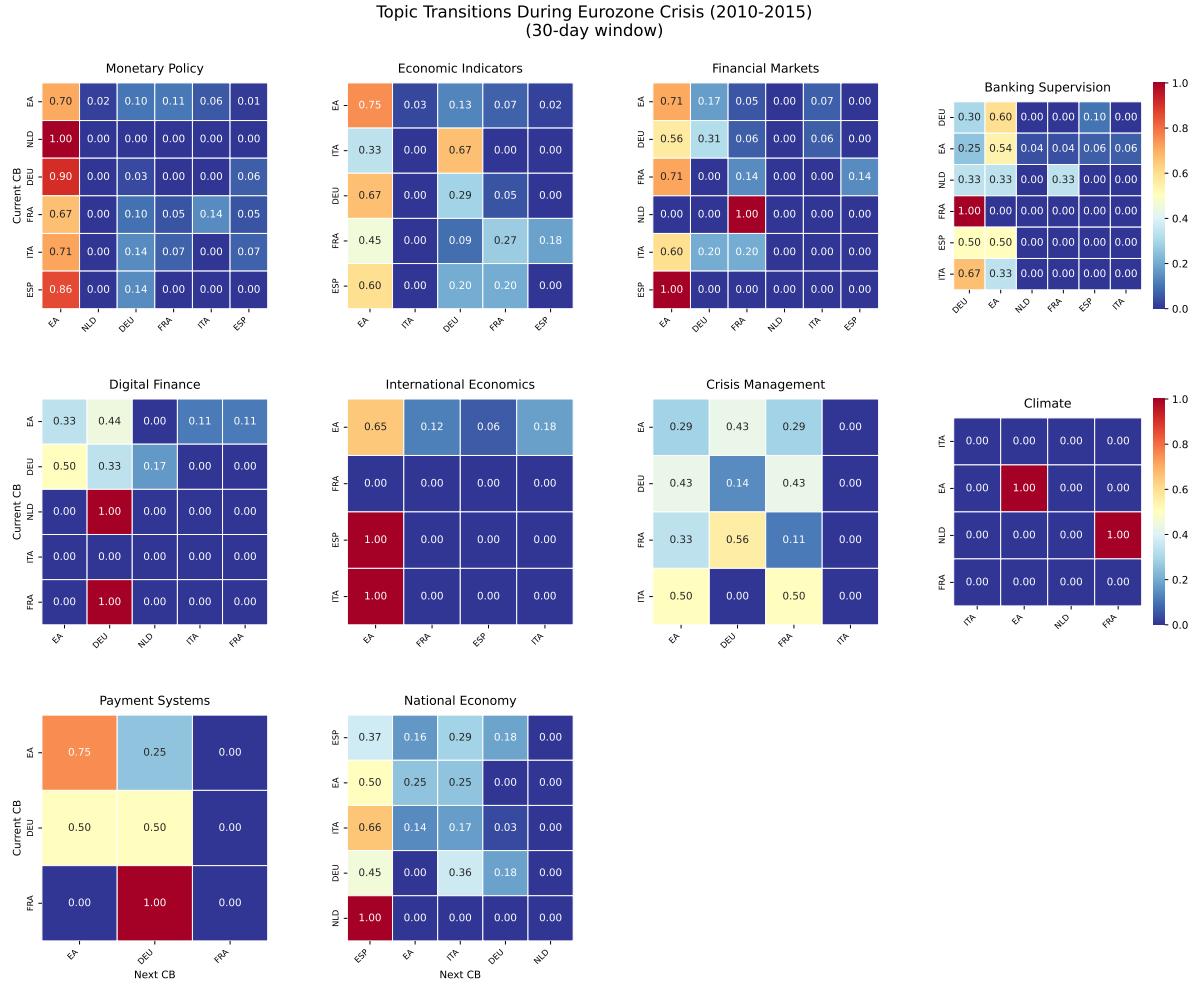


Figure C5.32: Markov transition matrix between central banks during the eurocrisis years

Note: This figure shows ten heatmaps, each corresponding to a distinct topic, depicting how one central bank (row) is followed by another central bank (column) within a 30-day window, according to a first-order Markov process. Each cell represents the estimated transition probability, normalised so that each row sums to one, with darker (blue) shades indicating lower probabilities and warmer (yellow-red) shades indicating higher probabilities.

C5.2 Coalitions

To test the effect of coalitions, I define the topic proportion for each NCB as $\text{NCB}_{b,q,t}(\text{C5.0})$, where b indexes a central bank, q indexes the quarter, and t indexes the topic. Averages are then used of the Dovish coalition (France, Italy, Spain) and the Hawkish one (Germany

and the Netherlands):

$$\text{CoalitionTalk}_{\text{dovish},q,t} = \frac{1}{3} \left[\text{NCB}_{\text{FR},q,t} + \text{NCB}_{\text{IT},q,t} + \text{NCB}_{\text{ES},q,t} \right].$$

Hawkish coalition (Germany, Netherlands), average talk:

$$\text{CoalitionTalk}_{\text{hawkish},q,t} = \frac{1}{2} \left[\text{NCB}_{\text{DE},q,t} + \text{NCB}_{\text{NL},q,t} \right].$$

I then use these indices in an augmented version of Model C of the timeseries analysis to test the issue-responsiveness replacing NCB lags with the coalition talk lags. Figure C5.33 shows the results. It can be seen that quite a few more results are significant and positive compared to the full interaction grid (Figure 3.7). In most topics, there is no real difference between hawkish and dovish coalitions in terms of responsiveness. Only with higher or lower GDP growth, the ECB seems to be more responsive to Hawkish countries while with regards to other pressures it seems to be balanced.

	+NS	+NS	-NS	-NS
Banking Supervision (Dovish)	+NS	-NS	-NS	+S
Banking Supervision (Hawkish)	+NS	+S	-NS	+S
Climate (Dovish)	+S	+S	-NS	-S
Climate (Hawkish)	-NS	+NS	-NS	-S
Crisis Management (Dovish)	-NS	+S	+NS	-NS
Crisis Management (Hawkish)	-NS	-NS	+NS	+NS
Digital Finance (Dovish)	+NS	+NS	-NS	+NS
Digital Finance (Hawkish)	-NS	-NS	+NS	+NS
Economic Indicators (Dovish)	-NS	+S	+NS	+NS
Economic Indicators (Hawkish)	+NS	+NS	+NS	+NS
Financial Markets (Dovish)	+NS	+NS	+NS	+NS
Financial Markets (Hawkish)	-NS	-NS	+NS	-NS
International Economics (Dovish)	+NS	-NS	+S	-NS
International Economics (Hawkish)	-NS	-NS	+NS	+S
Monetary Policy (Dovish)	-NS	+NS	-NS	+NS
Monetary Policy (Hawkish)	-NS	+NS	-NS	+NS
National Economy (Dovish)	+NS	+NS	-S	-NS
National Economy (Hawkish)	+S	-NS	+S	-NS
Payment Systems (Dovish)	-NS	+NS	+NS	-NS
Payment Systems (Hawkish)	+NS	+NS	+NS	-NS

GDP Growth Inflation Salience Trust ECB

Figure C5.33: Coalition interaction grid

Note: The coefficients are retrieved from the regressions of Model C, whereby the NCB lag is replaced with the coalitiontalk lag. The proxies of salience, trust in the ECB, GDP growth and debt-GDP are binary and represent a lower and higher within country average. For significance, a p-value of 0.05 is employed.

C5.3 Sintra effects

I test the Sintra effect by creating a Sintra dummy variable:

$$\text{Sintra}_q = \begin{cases} 1, & \text{if quarter } q \text{ includes or follows the Sintra conference date,} \\ 0, & \text{otherwise.} \end{cases}$$

I add this dummy variable to regression Model B and also create an interaction term of the dummy with the NCB lag to examine how responsiveness changes following the conference every year in early July. Table C5.9 below shows the results.

Topic	FR	IT	ES	DE	NL
Monetary Policy	-0.336 (0.280)	-0.332 (0.238)	0.039 (0.209)	0.188 (0.212)	0.045 (0.250)
Economic Indicators	-0.221 (0.149)	0.344 (0.860)	-1.534 (0.959)	-0.275 (0.199)	-0.203 (0.292)
Financial Markets	-0.143 (0.091)	-0.742** (0.310)	0.144 (0.480)	-0.355** (0.147)	0.270 (0.196)
Banking Supervision	0.200 (0.248)	-1.183*** (0.404)	-0.039 (0.115)	0.030 (0.161)	0.055 (0.074)
Digital Finance	0.029 (0.227)	-0.098 (0.142)	-0.183 (0.129)	-0.001 (0.163)	-0.013 (0.064)
International Economics	-0.016 (0.183)	-0.266 (0.284)	0.160 (0.584)	-0.635** (0.297)	0.129 (0.230)
Crisis Management	0.045 (0.096)	-0.128 (0.263)	-0.655 (1.322)	0.127 (0.173)	-0.297 (0.380)
Climate	-0.078 (0.403)	0.040* (0.023)	-0.230* (0.120)	0.451 (0.522)	-0.032 (0.096)
Payment Systems	-0.053 (0.076)	0.320 (0.270)	1.121** (0.525)	0.061 (0.064)	-0.158 (0.386)
National Economy	-0.512** (0.240)	0.053 (0.035)	0.008 (0.036)	-0.012 (0.044)	-0.692** (0.310)

Table C5.9: Impact of NCB Topic \times Sintra interaction on ECB responsiveness

Note: Coefficients and clustered standard errors for the interaction term $NCB\ Topic \times Sintra$. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

C6 Alternative explanations

This section provides extra information on the alternative explanations discussed in the main text.

C6.1 Between topic transitions

The transition between topics discussed across all central banks is defined as follows: let $\mathcal{T} = \{1, 2, \dots, K\}$ be the set of possible topics. At discrete time t , the observed topic is $T_t \in \mathcal{T}$. I model (T_t) as a discrete-time, first-order Markov chain. Whereby I define

$$\text{count}\{t : T_t = i \text{ and } T_{t+1} = j\}$$

as the number of times I observe a transition from topic i to topic j in the data. I denote this by

$$n_{ij} = \text{count}\{t : T_t = i \text{ and } T_{t+1} = j\},$$

where $i, j \in \mathcal{T}$. I then define the transition probability

$$P_{ij} = \frac{n_{ij}}{\sum_{k \in \mathcal{T}} n_{ik}},$$

which ensures that the rows of $P = [P_{ij}]$ sum to 1, making it a row-stochastic matrix. Therefore, P_{ij} represents the probability of moving from topic i to topic j .

To examine direct responsiveness (within a 30-day window) in discussions between various topics across all central banks, I look at the topic sequences presented in the Markov transition matrix in Figure C6.34. Notably, some topics have high self-persistence, for instance, the topic of monetary policy, with self-transition probabilities nearing 0.33, indicating a strong likelihood that discussions between central banks remain within this topic once initiated. Overall, most transitions between topics generally occur with low probability, pointing towards the volatile nature of speech topics. However, there

are a few exceptions revealing logical connections among certain topics. For instance, international economics often transitions to financial markets, showing the inter-linkage between international issues and financial issues. Moreover, many topics transition into monetary policy, which is logical given that it remains the primary mandate, thus, often discussed in relation or follow-up to other topics. To test whether these findings are not dependent on the window length, Figure C3.2 shows the same transition matrices for 7, 14, 60 and 90 day windows. From these robustness checks, it can be seen that the longer the window length, the more monetary policy responds to other topics. This indicates that monetary policy is still a consistently core topic that is discussed in response to various other topics.

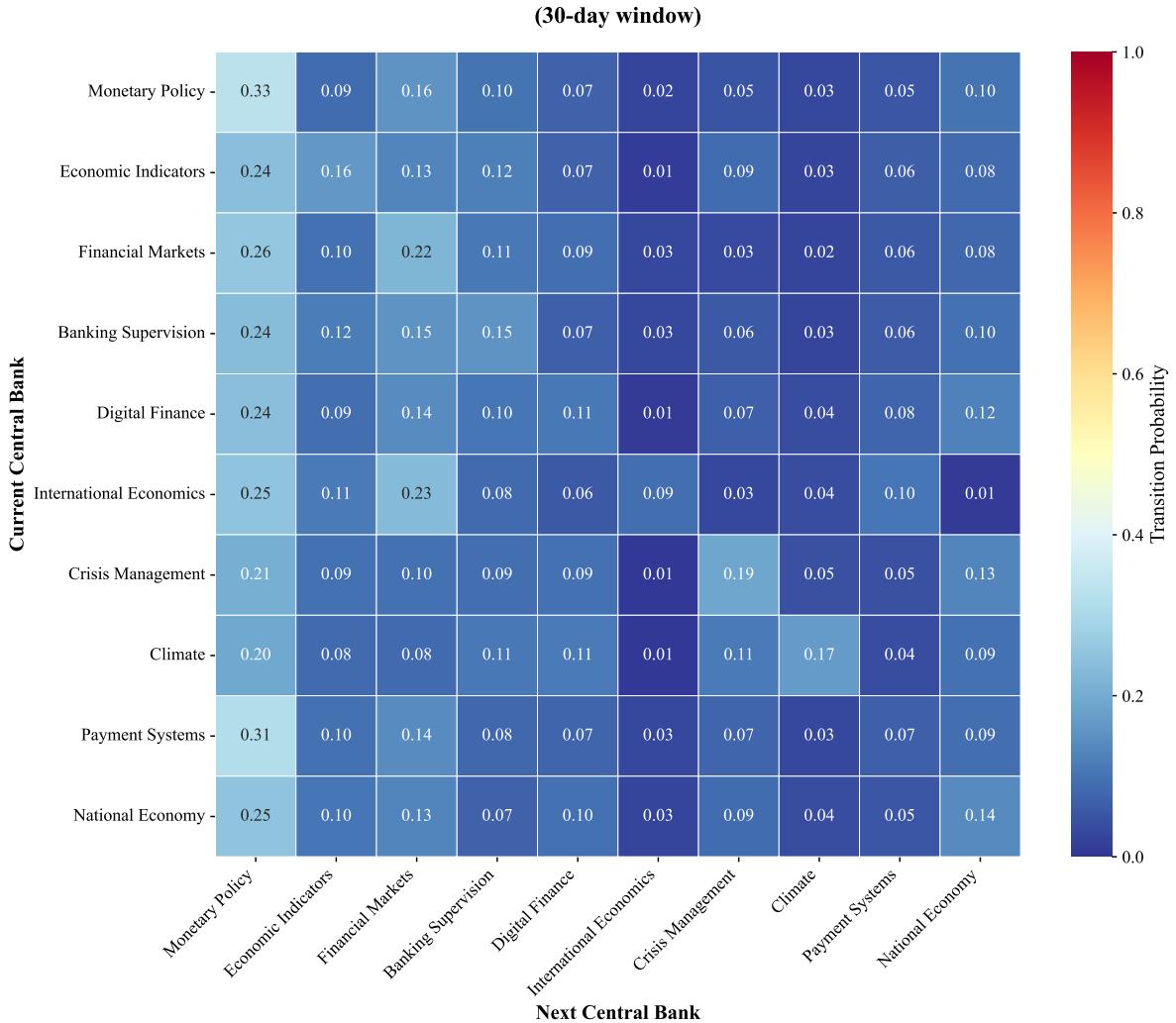


Figure C6.34: Markov Transition Matrix of topic transitions

Note: This heatmap shows the row-stochastic transition matrix of topics, estimated via a first-order Markov process. Each cell (i, j) reports the estimated probability of moving from topic i to topic j in consecutive time periods, with row sums equal to one.

In order to assess the sensitivity of the results to the specification of the Markov windows, I conduct a series of robustness tests. The main idea is to verify that the empirical findings are not driven by an arbitrary choice of window parameters. Moreover, the different patterns in varying timewindows can also be substantially interpreted.

Let w^* denote the baseline window length used in the analysis (e.g., $w^* = 4$ quarters).

I re-estimate the main models using alternative window lengths $w \in \{w^* - \Delta, w^*, w^* + \Delta\}$. For each alternative specification, the Markov windows are recalculated by reassigning the state probabilities using the modified window length. Comparing the parameter estimates and their statistical significance across different window configurations shows consistency in the sign, magnitude and inference of the coefficients indicating that the main results of the study are robust to variations in the Markov window specification. Figure C6.35 shows the different window lengths of between topics. Overall, the robustness tests confirm that the baseline results are not sensitive to the choice of window length.

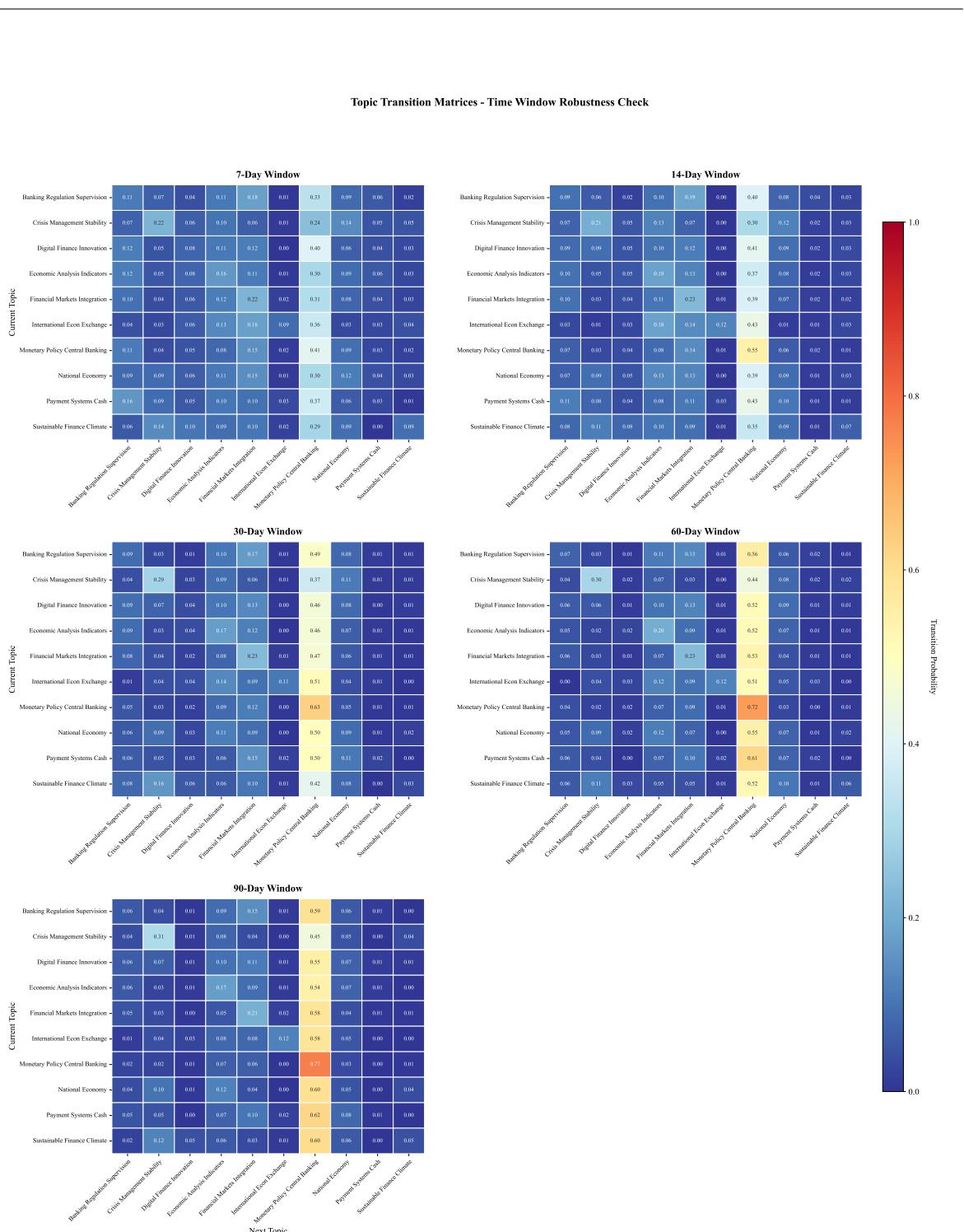


Figure C6.35: Topic transition matrix time window robustness

Note: This heatmap shows the row-stochastic transition matrix of topics, estimated via a first-order Markov process as outlined in the text. Each cell (i, j) reports the estimated probability of moving from topic i to topic j in consecutive time periods, with row sums equal to one. Darker cells correspond to higher transition probabilities. These estimates capture both the persistence of particular topics (on the diagonal) and the likelihood of switching between distinct topics.

C6.2 Within-topic responsiveness

By measuring textual similarity within a single topic, I can examine whether the views within a topic converge and diverge over time. To measure within-topic alignment, I first identify speeches where the topic probability exceeds a threshold of 0.3. This threshold ensures that the analysis is only conducted on topic-relevant speeches, thereby, eliminating noise. Then, for each speech s , I calculate sentence embeddings using the ‘all-mpnet-base-v2’ model. This model was chosen for its state-of-the-art performance in semantic text similarity tasks and its ability to capture nuanced differences in monetary policy communication. These embeddings are then aggregated by central bank and quarter:

$$\text{Emb}_{b,t} = \frac{1}{|\mathcal{S}_{b,t}|} \sum_{s \in \mathcal{S}_{b,t}} \text{Emb}(s),$$

where $\mathcal{S}_{b,t}$ represents the set of speeches by bank b in quarter t that exceed the topic probability threshold, and $\text{Emb}(s)$ is the sentence embedding for speech s .

Following this, I create a measure of communication alignment by computing, for each quarter t , the cosine distance between the ECB’s current communication and the average position of the NCBs from the previous quarter:

$$\text{Dist}_{\text{ECB},t} = \cos \left(\text{Emb}_{\text{ECB},t}, \frac{1}{|B|} \sum_{b \in B} \text{Emb}_{b,t-1} \right)$$

where B is the set of NCBs (Bundesbank, Bank of France, Netherlands Bank, Bank of Italy, and Bank of Spain), and $\cos(\cdot, \cdot)$ denotes the cosine distance. Higher values indicate greater divergence in communication content. Figures C6.36 and C6.37 show these cosine distances for the topics of monetary policy and climate. Using cosine distance allows to focus on the similarity in content and style rather than differences in scale, making it suitable for comparing textual embeddings.

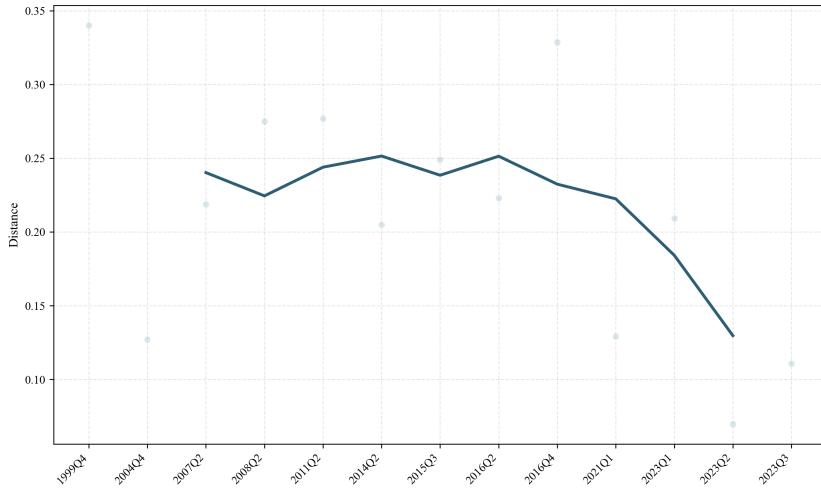


Figure C6.36: Intra-topic distance monetary policy

Note: This figure shows the temporal dynamics of communication alignment between the ECB and NBCs on the topic of monetary policy. The semantic distance between the ECB's current communication and the aggregate position of NCBs from the previous quarter is calculated using sentence embeddings (all-mpnet-base-v2). The analysis focuses on speeches where the topic probability exceeds 30%. The dark blue line represents a smoothed trend (4-quarter rolling average) of the cosine distances, where higher values indicate greater divergence in communication content.

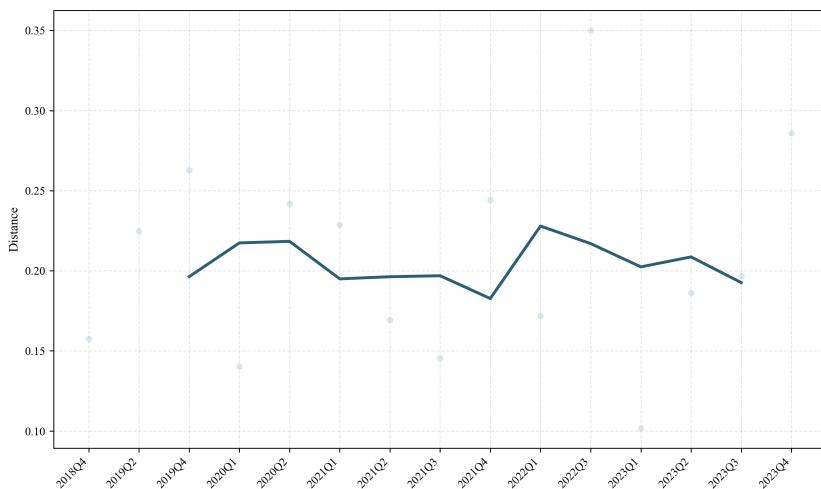


Figure C6.37: Intra-topic distance climate

Note: This figure shows the temporal dynamics of communication alignment between the ECB and NBCs on the topic of climate. The semantic distance between the ECB's current communication and the aggregate position of NCBs from the previous quarter is calculated using sentence embeddings (all-mpnet-base-v2). The analysis focuses on speeches where the topic probability exceeds 30%. The dark blue line represents a smoothed trend (4-quarter rolling average) of the cosine distances, where higher values indicate greater divergence in communication content.

C6.3 Permutation-based validation of transition patterns

I use permutations to ask the question: “If the null hypothesis were true and the topic sequence was completely random, how often would I see a pattern at least as extreme as the one we actually observed?” In other words, I am testing whether the observed sequencing of central bank speech topics over time exhibits a meaningful structure - or whether the patterns could plausibly arise just by chance.

More specifically, to determine whether the observed topic-pressure transition matrices could plausibly arise from random sequencing, I implement a speech-level permutation test that treats every speech as the unit of randomisation. For each topic, speeches whose topic share exceeds 25% are first dichotomised into a low-pressure versus high-pressure state, using the topic-specific median as the cut-off. Speeches are then chronologically ordered within each national central bank, and the four first-order Markov transition probabilities - Low→Low, Low→High, High→Low, High→High - are computed.¹¹ To build an empirical null distribution, the entire pipeline is rerun 100 times after independently shuffling the topic-intensity column across all speeches. This permutation preserves (i) the marginal distribution of topic intensities, (ii) the number of speeches per NCB per day, and (iii) country-specific seasonality, yet eliminates any sequential structure. The 100 transition matrices yield a reference distribution whose mean and standard deviation are displayed as orange bars with 95% whiskers in Figure C6.38; two-tailed p -values are calculated as the proportion of permuted probabilities that are at least as far from their mean as the observed value.

Figure C6.38 juxtaposes the observed transition probabilities (blue) with the permutation mean (orange). Across almost all topics, within-band continuations are markedly higher than the null expectation, whereas cross-band jumps are suppressed. For example,

¹¹I restrict the window to $\Delta t \leq 30$ days, but Appendix A.4 shows that results are robust to 7, 14, 60, and 90-day windows.

the Low→Low probability in economic indicators is roughly twice its permuted benchmark (0.85 vs 0.40), and climate exhibits a near-certain Low→Low repeat (1.00 vs 0.45 under the null). A global χ^2 test comparing the full observed matrix with the permutation mean rejects the null of randomness at $p < 0.001$ (see Appendix A.3), indicating that the sequence of speeches is far from a “noise machine”.

The magnitude of this stickiness varies systematically by issue area. Core macroeconomic domains - monetary policy, economic indicators, and financial markets - exhibit the strongest excess probabilities, suggesting coordinated communication in topics with direct market repercussions. By contrast, digital finance and banking supervision display a mixed pattern: cohesion spikes in high-pressure moments but approximates randomness during tranquil periods. Finally, the national economy panel aligns almost perfectly with the null envelope, corroborating earlier evidence that the ECB avoids sustained engagement in NCB-specific domestic debates. Together, these findings demonstrate that the observed transition structure is both statistically and substantively distinct from random sequencing, and that its intensity is conditioned by the strategic salience of individual topics.

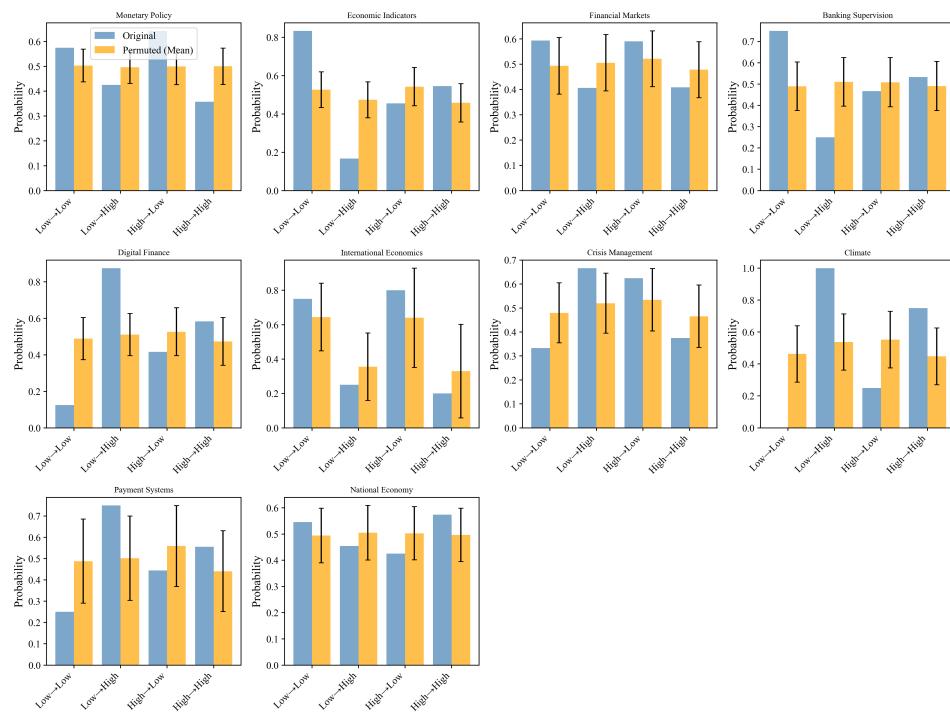


Figure C6.38: Permutations of runs

Appendix Chapter 4

D1 Replication files on Github

We provide a repository with replication codes on Github. It contains all the codes to pre-process the dataset, run ChatGPT on two million sentences, and finally produce our indices and aggregated results. Moreover, we provide our manually classified validation sample

`inputdata/validation_sample_all.xlsx` and the codes to conduct prompt engineering experiments, fine-tune GPT-3.5, and assess the classification quality of various GPT models and Gemini Pro against this validation set. The following contains the descriptions of the included files and their function. The same instructions can be found inside the `README.md`.

We share a yearly aggregation of our indices of dominance and coordination `dominance_coordination_dataset.csv`. This file is sufficient to produce all charts inside the Appendix and main part of the chapter. Importantly, we don't include any speeches or sentence-level results. The output files are more than a gigabyte in size and too large for this repository. To rerun the full analysis, the speech data would need to be scraped with the python code here. We do, however, provide the sentence-level classification of our prompt engineering results, validation exercise, and model comparisons. These are stored as Pandas DataFrames in `.pkl` format inside the `outputdata` folder.

D1.1 Instructions to run codes

- To rerun any of our analyses, an API key for GPT and/or Gemini needs to be set inside the `llm_functions.py` file. Also note that these LLMs, even at a temperature set to zero, are non-deterministic. Exact results vary with each run, although with

GPT-3.5, usually 97%-99% of sentences are identically classified across two runs.

In addition, changes to the model on OpenAI’s/Google’s side can impact results.

- To run R codes, the working directory should be set to the root of the project.
- Python codes expect to be run from the folder they are in.
- Validation, prompt engineering, and model comparison codes are self-contained and can be run with the inputs provided inside this repository, provided that an API key is set.

D1.2 Included files

The `codes` folder contains the following files:

- `0_text_preprocessing.py`: This file runs the preprocessing steps described in the Appendix.
- `1_gpt_main_analysis.py`: This code consists of the code required to run the full dataset. It requires the output produced by `0_text_preprocessing.py`.
- `2_validation_and_robustness.py`: This file contains the code for the robustness checks, prompt engineering results, and different GPT versions. It requires only our validation set as input `validation_sample_all.xlsx`.
- `3_fine_tuning_and_few_shot.py`: This file constructs a training dataset from our validation set, trains a fine-tuned GPT 3.5 model, and evaluates it with the remaining sample. Moreover, it contains code to run Gemini Pro using (i) the same prompts as GPT and (ii) a few-shot prompting strategy.
- `llm_functions.py`: Functions that are shared by the python codes are in this file. Most notably, it contains the function that takes a dataframe as input and calls

either the Gemini or GPT API with our prompt design. This function allows for parallel API queries to maximize rate limits.

- `merge_datasets.R`: This R code calculates our relative indicator of dominance and coordination. It requires the outputs saved by `1_gpt_main_analysis.py`. It also sketches how our shared dataset `dominance_coordination_dataset.csv` is produced (without including the third-party data sources).
- `run_all_charts.R`: Produces all of the charts.

D1.3 Replication of Charts

All our charts can be replicated with the R codes inside the `codes/figures` folder. Run `run_all_charts.R` to produce all charts. The R files read from the GPT results provided inside the `outputdata` and the yearly aggregation of the full dataset `dominance_coordination_dataset.csv`. No access to GPT is required to produce the charts. These are the files to produce the charts:

- `bin_scatter.R`
- `correlation.R`
- `crisis.R`
- `levels_over_time.R`
- `sentence_count_charts.R`
- `stability.R`
- `temperature_charts.R`

Common functions and settings to change the size of the charts are inside `functions_and_settings.R`.

D1.4 Prompts

The instructions part of our prompts are stored in the `prompts` folder. The sentences/excerpts are automatically appended to the prompt. We use a `.yaml` format to store the prompts. Our final instructions for level 1, level 2, and level 3 are in the `11`, `12`, `13` subfolders. To change the prompts either modify the prompt file or modify the python code to load a different prompt.

D2 Database Construction

This dataset of central bank official speeches is scraped from the Bank of International Settlement website. In mid-2023, we downloaded a total of 18,081 speeches, press conferences, interviews and lectures issued by 118 institutions, mainly central banks, over the period from 1997 to 2023. See Table D2.1 for the contained central banks, the number of speeches scraped and the first and last year speech obtained per institution. The speeches are downloaded in PDF format and converted to text files. The PDF format is not entirely standardised, which can result in the core text being interspersed with page numbers, footnotes, tables, charts, and literature references, thereby disrupting the flow of the main text. To clean the text we apply several pre-processing steps (see the next section D2.1).

Table D2.1: List of central banks, time periods and number of speeches

Central Bank	First Year	Last Year	Speeches Count
Bank of France	1997	2023	377
Bank of Finland	1997	2023	173
Sveriges Riksbank	1997	2023	481
Reserve Bank of India	1997	2023	865
People's Bank of China	1997	2023	142
Bank of Italy	1997	2023	370

[]List of central banks, time periods and number of speeches (*continued*)

Central Bank	First Year	Last Year	Speeches Count
European Monetary Institute	1997	1997	2
South African Reserve Bank	1997	2023	393
Deutsche Bundesbank	1997	2023	802
Federal Reserve Bank of New York	1997	2023	411
Reserve Bank of Australia	1997	2023	528
Reserve Bank of New Zealand	1997	2023	190
Czech National Bank	1997	2023	55
De Nederlandsche Bank	1997	2023	205
Bank of Canada	1997	2023	552
Banque Nationale Suisse	1997	1997	3
Central Bank of the Republic of Turkey	1997	2022	99
Hong Kong Monetary Authority	1997	2023	249
Bank for International Settlements	1997	2015	17
Federal Reserve Bank of Kansas City	1997	2011	19
Bank of Greece	1997	2023	157
Swiss National Bank	1997	2023	383
International Monetary Fund	1997	2014	4
Central Bank of Ireland	1998	2023	318
Central Bank of Iceland	1998	2023	90
Central Bank of the Republic of Austria	1998	2023	81
European Central Bank	1998	2023	2377
Bank of Korea	1998	2023	89
Central Bank of Norway	1999	2023	284
Central Bank of Brazil	1999	2015	11
Bank Indonesia	1999	2023	63
Monetary Authority of Singapore	1999	2023	290
National Bank of Belgium	1999	2022	38
Bank of Latvia	1999	2018	11
Bank of Namibia	1999	2022	32
Bank of Thailand	2000	2023	221
Croatian National Bank	2000	2023	10
Federal Reserve Bank of San Francisco	2000	2012	3

[]List of central banks, time periods and number of speeches (*continued*)

Central Bank	First Year	Last Year	Speeches Count
National Bank of North Macedonia	2001	2023	92
Bank of Israel	2001	2022	105
Central Bank of Bosnia and Herzegovina	2001	2020	15
Central Bank of Trinidad and Tobago	2001	2018	101
Central Bank of Chile	2001	2023	125
Bank of Estonia	2001	2023	21
State Bank of Pakistan	2001	2022	132
Central Bank of Malta	2001	2022	58
Bank of Poland	2002	2007	20
Central Bank of Malaysia	2002	2023	487
National Bank of Denmark	2002	2023	101
Bank of Spain	2002	2023	329
Central Bank of Luxembourg	2002	2015	38
Bank of Mauritius	2002	2023	160
Bank of Zambia	2003	2023	156
Bank of Botswana	2003	2022	46
Bank of Papua New Guinea	2003	2022	60
National Bank of Slovakia	2003	2019	5
Bank of Mexico	2003	2023	98
Central Bank of Nigeria	2003	2023	35
Reserve Bank of Malawi	2003	2017	25
Bank of Portugal	2003	2021	60
Eastern Caribbean Central Bank	2003	2019	17
Central Bank of the Bahamas	2004	2019	13
Bank of Albania	2004	2023	290
Central Bank of Barbados	2004	2023	91
Bank of Sierra Leone	2004	2011	12
Central Bank of Sri Lanka	2004	2019	67
Central Bank of the Philippines	2004	2023	521
Central Bank of Argentina	2004	2021	34
Reserve Bank of Fiji	2004	2021	132
National Bank of Serbia	2004	2023	114

[]List of central banks, time periods and number of speeches (*continued*)

Central Bank	First Year	Last Year	Speeches Count
Bank of Jamaica	2005	2023	25
National Bank of Romania	2005	2023	71
Bank of Uganda	2005	2019	152
Bank of Ghana	2005	2022	56
Bulgarian National Bank	2006	2023	44
Saudi Arabian Monetary Agency	2006	2014	28
Central Bank of Hungary	2006	2022	11
Central Bank of the United Arab Emirates	2006	2019	10
Bank of Mozambique	2006	2010	5
Reserve Bank of Vanuatu	2006	2013	2
Central Bank of Solomon Islands	2006	2019	18
Monetary Authority of Macao	2007	2014	27
Central Bank of Bahrain	2007	2017	45
Central Bank of Kenya	2007	2023	182
Central Bank of Colombia	2007	2009	3
Central Bank of Samoa	2007	2014	6
Bank of Algeria	2009	2022	8
National Bank of Cambodia	2009	2022	3
Central Bank of Aruba	2009	2009	1
Federal Reserve Bank of Boston	2010	2011	5
Central Bank of Jordan	2010	2010	1
Central Bank of Bolivia	2010	2010	1
Central Bank of Cyprus	2011	2023	17
Federal Reserve Bank of Philadelphia	2011	2015	32
Federal Reserve Bank of Dallas	2011	2015	19
Central Bank of Belize	2011	2011	1
Central Bank of Curaçao and Sint Maarten	2011	2022	14
Federal Reserve Bank of Minneapolis	2011	2014	23
Federal Reserve Bank of Chicago	2011	2014	10
Bank of Guatemala	2012	2012	1
Central Bank of Uruguay	2012	2012	1
Bank of Guyana	2012	2012	2

[]List of central banks, time periods and number of speeches (*continued*)

Central Bank	First Year	Last Year	Speeches Count
Central Bank of Nepal	2013	2019	14
Bank of Tanzania	2013	2013	1
Bank of Slovenia	2014	2023	14
Bank of Russia	2014	2022	43
Federal Reserve Bank of Atlanta	2015	2015	2
Federal Reserve Bank of Richmond	2015	2015	2
Bank of Lithuania	2015	2022	34
Central Bank of Kuwait	2017	2022	5
Central Bank of Seychelles	2017	2023	23
Maldives Monetary Authority	2018	2021	5
Central Bank of the Republic of Kosovo	2018	2023	28
National Bank of Ukraine	2019	2023	33
Bank of Morocco	2019	2021	6

D2.1 Pre-processing steps

All the text pre-processing steps are undertaken in Python and available in the replication files (see `text_processing.py` in the codes folder). The pre-processing steps follow the following steps:

1. Regular expressions

We examined the extracted text for recurring patterns of:

- Page numbers
- Page headers
- New page characters
- Footnotes
- URLs
- Subsequent whitespace characters

, which we remove using appropriate regular expressions.

2. Conversion to sentence level

We convert the entire corpus to the sentence level using the *Punkt* sentence tokenizer from the NLTK python package. We also tried the sentence extraction from spacy's *en_core_web_lg* model, which we found to produce similar results while being much slower.

3. Sentence level heuristics

After segmenting the corpus into individual sentences, the dataset still contains entries that do not constitute genuine sentences of the primary text. Instead, these entries include tables, annotations, or binary data erroneously recognized as text during the PDF conversion process. To address this issue, we implement conservative rules aimed at filtering out clearly irrelevant entries:

-
- Remove sentences with less than 2/3 ASCII characters
 - Remove sentences that consists of less than 6 tokens or more than 200 tokens
 - Remove sentences with less than 20 characters

After pre-processing, we obtain a dataset that consists of 2,034,313 sentences. A small share of sentences do not contain relevant text, i.e., some chart annotations, references etc. remain. We are reluctant to more aggressively delete sentences as GPT will identify irrelevant sentences anyways. After pre-processing, we run the three levels of GPT classification, construct our relative indicator and aggregate on the central bank-year level as described in section 4.3.5 of the main text.

D2.2 Variables in the Database

Our shared database contains our index of the relative shares of monetary-dominance, fiscal-dominance, monetary-fiscal coordination, monetary-financial coordination and financial dominance as well as additional country-year level data on macroeconomic and political indicators. Table D2.2 describes the variables and where applicable their data sources will be provided.

Table D2.2: Variables in the database

Variable name	Description
Metadata	
central_bank	Name of the central bank
country	The ISO 3166-1 3 letter country code of the central bank's country
year	The year of the observation
currency_code	The ISO 4217 3 letter currency code of the central bank's currency
advanced	Takes value 1 if the country is a advanced economy according to the IMF classification of advanced, emerging and developing countries.

Variables in the database (*continued*)

Variable name	Description
number_of_speeches	The number of speeches that were classified by the central bank in a given year
number_of_sentences	The total number of sentences that were classified by the central bank in a given year
Level 1	
topic_monetary	Share of sentences that address monetary issues (e.g., inflation, price stability, primary mandate, interest rate)
topic_fiscal	Share of sentences that address fiscal issues (e.g., sovereign debt, budget balance, fiscal governance, taxes, pensions)
topic_financial	Share of sentences that address financial issues (e.g., banking supervision, financial instability, credit risks)
topic_climate	Share of sentences that address climate issues (e.g., environmental issues, CO2, climate change, sustainable development goals)
topic_macro	Share of sentences that address macroeconomic issues (e.g., GDP, economic growth, unemployment, productivity, economic outlook)
topic_international	Share of sentences that address international economics issues (e.g., trade, exchange rates, capital mobility, tariffs)
topic_other	Share of sentences that address topics of monetary, fiscal, financial, climate, macro or international economics.
Level 2	
share_normative	Share of sentences that contain a value judgement.
share_descriptive	Share of sentences that are descriptive and do not provide a value judgement.
normative_!topic!	For each of the topics (see Level 1) the dataset contains a variables which describes the share of sentences belonging to the topic that are classified as normative
Level 3	
monetary_dominance	Share of excerpts that classify as “monetary dominance”, i.e., if the excerpt clearly and explicitly says that the central bank subordinates fiscal or financial policies to the central bank’s monetary policy objective of price stability

Variables in the database (*continued*)

Variable name	Description
fiscal_dominance	Share of excerpts that classify as “fiscal dominance”, i.e., if the excerpt clearly and explicitly says that the central bank subordinates itself to fiscal authorities, that is, where the speaker says that monetary policy is primarily driven by fiscal considerations rather than maintaining price stability
financial_dominance	Share of excerpts that classify as “financial dominance”, i.e., if the excerpt clearly and explicitly says that the central bank subordinates to financial markets or the financial regulation authorities, that is, where the speaker says that monetary policy is primarily driven by financial stability considerations rather than maintaining price stability
monetary_financial_coordination	Share of excerpts that classify as “monetary-financial coordination”, i.e., if the excerpt suggests that the central bank and financial regulators should cooperate, this is, where the speaker says that monetary policy and financial regulation are best coordinated to achieve the right policy mix
monetary_fiscal_coordination	Share of excerpts that classify as “monetary-fiscal coordination”, i.e., if the excerpt suggests that fiscal authorities and the central bank should cooperate, this is, where the speaker says that monetary and fiscal policy are best coordinated to achieve the right policy mix
none	Share of excerpts that classify as “none”, i.e., if there is no reference to monetary, financial or fiscal developments or if the excerpt describes monetary, financial or fiscal developments, that is, if the speaker does not make any normative reference to monetary, financial or fiscal policy.

Macroeconomic variables

inflation	HICP inflation, IMF WEO variable code: PCPI
gdp_real_ppp_capita	GDP per capita measured in purchasing power parity (PPP) USD, IMF WEO variable code: NGDPRPPP PC
spread	Government bond spread measured as difference of 10Y government bond yield of the country to the German government bond yield. Data from Bloomberg, and in some cases where Bloomberg does not contain the relevant bond yield Refinitiv/Datastream

Political Variables

Variables in the database (*continued*)

Variable name	Description
democracy_ind	Dummy variable based on the v2x_regime variable from the VDEM dataset indicating a democracy. Democracy (1) includes electoral and liberal democracy; autocracy (0) includes closed and electoral autocracy
polarization_ind	Dummy variable based on the variable v2cacamps_mean from the VDEM dataset indicating high polarization, whereby 0 and 1 map to low polarization (0) and 2 and 3 map to high polarization (1).

D3 Validation

D3.1 Graphical coding overview

The coding of dominance and coordination is conducted in three steps. First, we classify the topic of the sentence as either 'monetary', 'fiscal', 'financial', 'climate', 'macro', 'international', or 'other'. The second step, level 2, determines whether the sentences is normative or descriptive. Level 3 then categorizes sentences as a form of dominance or coordination, also entailing the most frequent option of none. For level 3, the sentence before and after each sentence is added as context. The main text provides more details on why certain choices were made regarding the classification. See Figure D3.1 below for an overview of the coding scheme.

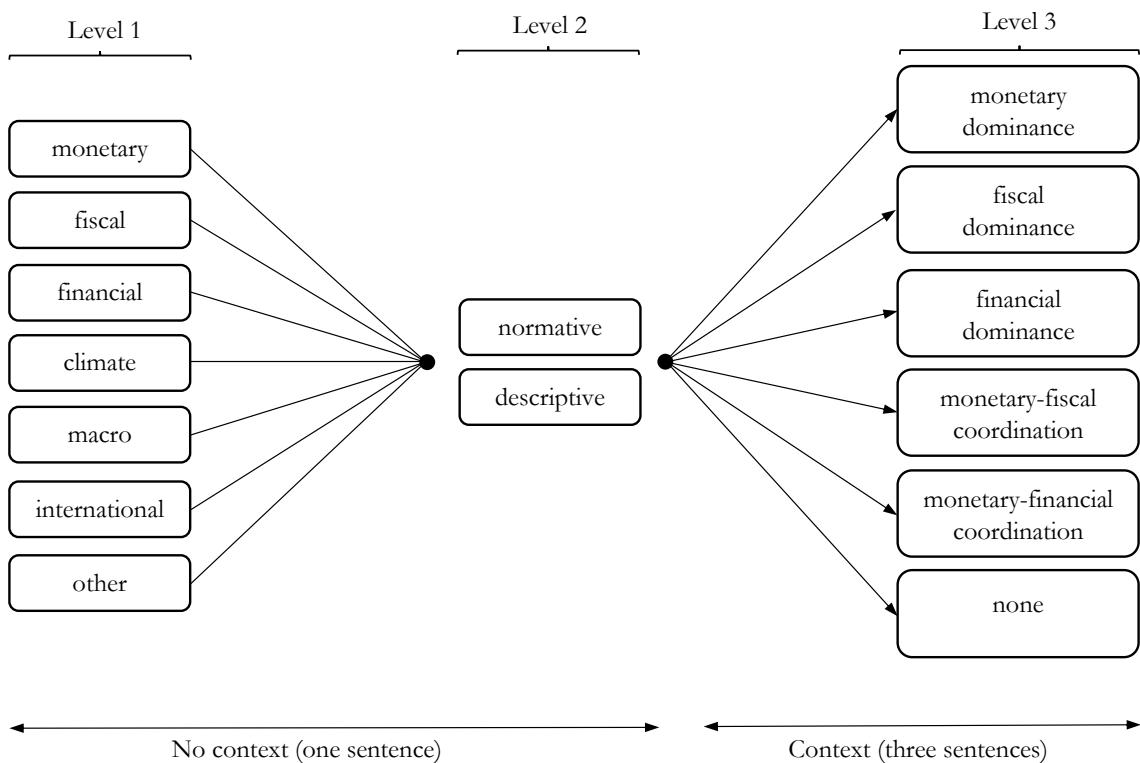


Figure D3.1: Three-level coding classification overview.

D3.2 Ambiguous sentence coding guidelines

We recognize that there are sentences for which the coding can be ambiguous. Therefore, these are some extra guidelines for possibly ambiguous cases.

1. Level 1:

- References to digital are classified as ‘other’
- Topics that refer to payment systems are classified as ’other’
- When a sentence discusses both monetary and fiscal topics, or both monetary and financial topics, it should be categorized under fiscal or financial, rather than under monetary.
- References to exchange rates should be classified as ’international’
- References to forecast and general developments in the economy should be classified as ’macro’

2. Level 2:

- Classify as normative when words such as “should” or “our opinion” indicate a value judgement or a course of action.
- Also implicit value judgements highlighting the importance of a certain topic can be classified as normative.
- Any sentences that are a neutral recital of facts should be classified as descriptive.

3. Level 3:

- If the excerpt alone does not provide sufficient context to identify the type of dominance or coordination, or if the context is assumed rather than explicitly

provided, it should only be classified as a form of dominance or coordination if the context is widely acknowledged as common knowledge.

- To differentiate between dominance and coordination, one should examine whether there are signs of a hierarchical relationship or an emphasis on interactions without a power hierarchy.

D3.3 Classification Examples

The following Tables D3.3, D3.4, D3.5 provide examples of sentences from our validation set and their classification

Table D3.3: Level 1 classification examples

Classification	Example and explanation
Fiscal	<p>Some countries, such as the United States and Ireland, also implemented extensive and costly government measures to support the financial sector. (Retrieved from: The Sveriges Riksbank, 18-10-2011).</p> <p><i>Explanation:</i> the sentence mentions both fiscal and financial but the focus is on governments taking action.</p>
Monetary	<p>Among other things, this outcome complicates our ability to assess the present stance of monetary policy. (Retrieved from the Federal Reserve Bank of New York, 31-10-2006.)</p> <p><i>Explanation:</i> The sentence describes an implication for their monetary policy.</p>
Financial	<p>Recent events have demonstrated the important role that banks play as liquidity providers and the potential for broader market turbulence when banks have difficulty performing this role. (Retrieved from the Board of Governors of the Federal Reserve System, 07-03-2008)</p> <p><i>Explanation:</i> This sentence mentions how the role of bank's as liquidity providers is important for the economy.</p>

[

]Classification examples (*continued*)

Classification	Example and explanation
Climate	<p>In the longer run, the only way to address the climate crisis and to safeguard Europe's energy security is by accelerating the energy transition. (Retrieved from the Bundesbank, 15-11-2022).</p> <p><i>Explanation:</i> The sentence mentions the importance of addressing the climate crisis.</p>
International	<p>However, experience shows that it is always correct to allow the exchange rate to change based on market forces. (Retrieved from the Bank of Israel, 23-04-2013).</p> <p><i>Explanation:</i> Sentences referring to exchange rate are classified as international.</p>
Macro	<p>It does seem likely that productivity calculated for the entire economy using GDP data weakened in the second quarter. (Retrieved from the Board of Governors of the Federal Reserve System, 24-07-1998)</p> <p><i>Explanation:</i> The sentence provides information on the macroeconomic outlook.</p>
Other	<p>The ECB's Legal Committee is a genuine example of the identity of the Eurosystem and the ESCB., (Retrieved from the European Central Bank, 24-01-2019)</p> <p><i>Explanation:</i> The sentence talks about a legal committee which cannot be linked to any of the other categories thus by default will be 'other'.</p>

Table D3.4: Level 2 classification examples

Classification	Example and explanation
Normative	<p>Economic development calls for clear and predictable rules, and institutions that assure they will be enforced. (Retrieved from: The Bank of Mexico, 28-04-2003).</p> <p><i>Explanation:</i> The sentence clearly entails a value judgement since the bank argues that rules enforced by institutions are necessary.</p>

Classification	Example and explanation
Descriptive	<p>Such a large increase in income has only happened on a few occasions over the past decades. (Retrieved from the Sveriges Riksbank, 22-05-2002.)</p> <p><i>Explanation:</i> The sentence relates macroeconomic conditions to the previous periods.</p>

Table D3.5: Level 3 classification examples

Classification	Example
Monetary dominance	<p>Furthermore, monetary policy implementation in line with the market efficiency principle would need to remain without prejudice to our primary mandate of safeguarding price stability. (Retrieved from: The European Central Bank, 14-06-2021).</p> <p><i>Explanation:</i> The topic concerns a monetary topic and they emphasize their primary mandate of price stability being above other priorities. Therefore, this sentence can be classified as monetary dominance.</p>
Fiscal dominance	<p>Moreover, although most of the resources administered by the BIS are invested in financial assets of top quality at international level and their exposure to the various risks are managed conservatively, a greater portion of such funds could be spent toward the direct purchase of debt denominated in local currencies of emerging countries or to the use of them as collateral of certain bond issuance of countries with limited depth of their financing markets in local currency. (Retrieved from the Central Bank of Argentina, 09-07-2008.)</p> <p><i>Explanation:</i> This sentence refers to funds being spent towards the direct purchase of debt (=monetary financing) instead of considering pure price stability considerations, thus we consider this sentence to be fiscal dominance.</p>

Level 3 classification examples (*continued*)

Classification	Example
Financial dominance	<p>It is thus significant that our flexible and abundant provision of liquidity contained market participants' concerns over liquidity financing. (Retrieved from the Bank of Japan, 04-07-2002)</p> <p><i>Explanation:</i> This sentence states that monetary policy is accommodating financial markets by providing liquidity, thus showing that financial markets are a consideration for the bank in conducting their monetary policy.</p>
Monetary-fiscal coordination	<p>Since restarting our strategy review, we have introduced a new work stream on monetary-fiscal interactions precisely to address such questions. (Retrieved from the European Central Bank, 30-09-2020).</p> <p><i>Explanation:</i> This sentence refers to the monetary-fiscal interactions which is a key policy in the monetary-fiscal coordination.</p>
Monetary-financial coordination	<p>If market participants are willing to continue to work together, then we can safely achieve the transitions needed to create a better and more robust system that will help to ensure our ongoing financial stability. (Retrieved from the Board of Governors of the Federal Reserve System, 07-11-2017).</p> <p><i>Explanation:</i> This sentence shows that the bank wants coordinate with market participants to ensure financial stability.</p>

D3.4 Coder reliability scores

The random sample of 1000 sentences was coded by the 3 authors who each have at least a MSc degree in Economics/Political Economy. Before starting the coding process, the authors went through the codebook and the examples provided above together, which have in turn been developed inductively based on another random sample of 100 sentences. All three human coders subsequently independently coded the same first 400 sentences. Table D3.6 shows a matrix of the agreement between the human coders (L,M,S) and GPT (C). The complete dataset was extended to 1000 sentences by each author coding an additional 200 sentences.

A. Level 1				B. Level 2				C. Level 3						
	L	M	S	C	L	M	S	C	L	M	S	C		
L	1.00	0.87	0.88	0.85	L	1.00	0.81	0.88	0.85	L	1.00	0.84	0.87	0.62
M	0.87	1.00	0.81	0.77	M	0.81	1.00	0.78	0.73	M	0.84	1.00	0.82	0.60
S	0.88	0.81	1.00	0.78	S	0.88	0.78	1.00	0.76	S	0.87	0.82	1.00	0.61
C	0.85	0.77	0.78	1.00	C	0.85	0.73	0.76	1.00	C	0.62	0.60	0.61	1.00

Note: Cells of the matrix show the share of classifications that agree between the pair of coders.

Table D3.6: Coder overlap matrices

	Level 1	Level 2	Level 3
A. Krippendorff's α			
Humans	0.81	0.56	0.62
One human replaced by	0.76	0.48	0.36
GPT			
Humans + GPT	0.77	0.50	0.42
B. Cohen's κ			
Humans	0.81	0.56	0.62
One human replaced by	0.76	0.48	0.38
GPT			
Humans + GPT	0.78	0.50	0.44

Table D3.7: Coder reliability scores

We calculate Krippendorff's alpha that measures how well different coders agree when coding data at different levels of measurement. We find alphas of 0.81, 0.56 and 0.62 for level 1, level 2 and level 3 respectively among the human coders. In addition, we report Cohen's Kappa. Table D3.7 provides coder reliability scores with GPT added as forth coder and the average reliability score of all possible combinations of replacing one human coder with GPT.

D3.5 Validation metrics

In the following we calculate validation metrics of our GPT classifier that we compare against the "ground truth" of our human classification (see Table 3 of the main text). We differentiate between the full sample of 1000 sentences and a agreement sample,

which consists of the sentences that all 3 human coders have classified identical. We rely mostly on F1 scores to assess our models. The F1 score is calculated as the harmonic mean of precision and recall, where the F1 score is between 0 (worst) and 1 (best). This score is especially useful when dealing with unbalanced classifications as is the case in our dataset. We report two different kinds of F1 scores to adapt to the multiple label classification task: F1 weighted and F1 macro. F1 macro calculates the F1 score for each class separately and then takes the average of the scores. Thus, each class is given equal weight, regardless of its size. The weighted F1 scores weighs the category-specific F1 scores by the share of each category in the dataset. Accuracy refers to the share of correctly classified sentences and balanced accuracy is the mean of sensitivity and specificity averaged across categories. In addition, we report macro averages of Precision and Recall. We prefer macro averages as these are sensitive to changes in prediction quality in the less frequent categories.

D4 Additional prompt engineering results

Figure D4.2 corresponds to Figure 4.4 inside the main text, pointing out the trade-off between the number of sentences classified in one prompt regarding accuracy and the number of tokens used. Level 1 (topic) and level 2 (normative) appear to be more sensitive to the number of sentences included inside the prompt. We find highest accuracy for 10 sentences per prompt.

Figure D4.3 is the equivalent of Figure 4.3 in the main text, depicting how accuracy and stability of the classification vary with the temperature setting of GPT-3.5 for Level 1 (topic) and Level 2 (normative) classifications. As we found for level 3, accuracy is less stable with higher temperature and declines on average with higher temperature.

In Table D4.8, we present accuracy metrics of our final prompt with different messages. Leaving the default message "You are a helpful assistant" performs virtually identical to providing GPT-3.5 with a central bank context by assigning it the role of a research assistant at a central bank. Finally, a very elaborate system message giving GPT-3.5 the persona of a expert on central bank communication does, if anything, slightly lower the prediction accuracy.

System Message	Accuracy	F1 (weighted)	F1 (macro)	Recall (macro)
You are a helpful assistant.	0.62	0.66	0.36	0.44
You are a research assistant at a central bank.	0.62	0.66	0.37	0.45
You are a distinguished expert on central bank communication. Through your thorough studies, having read countless speeches and other central bank documents, you are familiar with the language central bankers use and know how to interpret their statements. This expertise enables you to understand nuanced differences in central bank communications and accurately decode the sometimes hard to grasp messages conveyed inside their communication.	0.60	0.66	0.35	0.43

Table D4.8: Validation metrics of system messages.

	Minimal Instructions	Final	Detailed instructions
Accuracy	0.50	0.62	0.47
F1 (weighted)	0.57	0.66	0.55
F1 (macro)	0.30	0.36	0.29
Precision (macro)	0.29	0.34	0.31
Recall (macro)	0.44	0.44	0.40
Tokens used	106379	137193	16099
Prompt length	672	2433	3559

Table D4.9: Performance metrics of different prompt configurations

Note: The mid column indicates the validation metrics of our final prompt. The other prompts can be found in the Appendix.

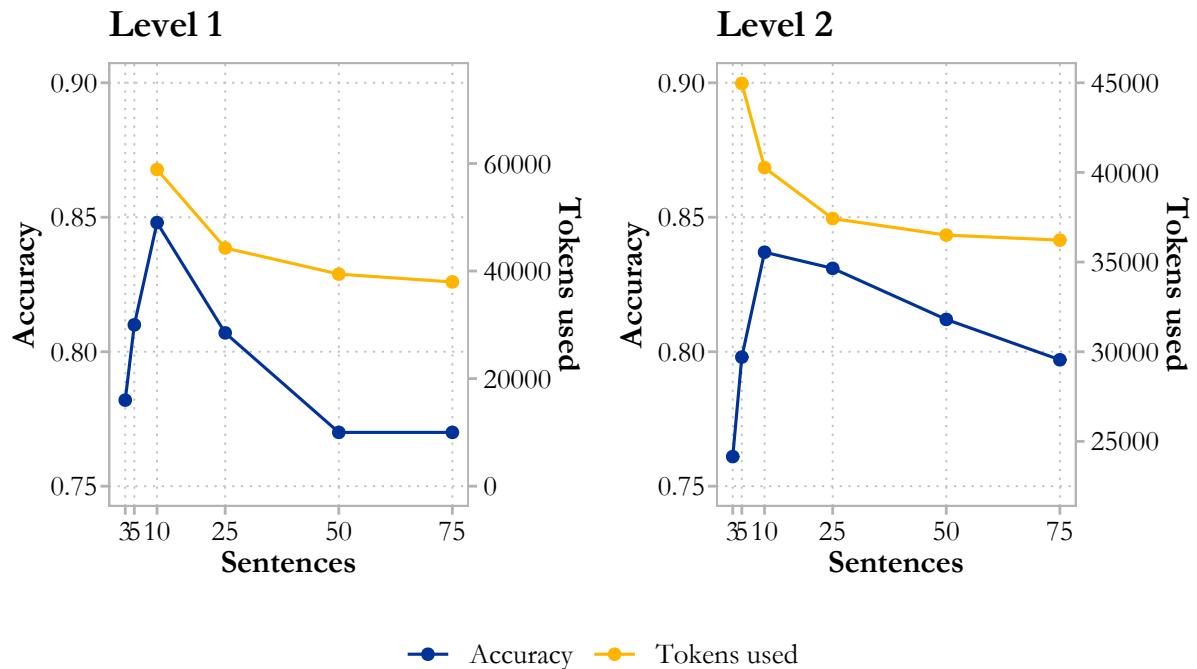
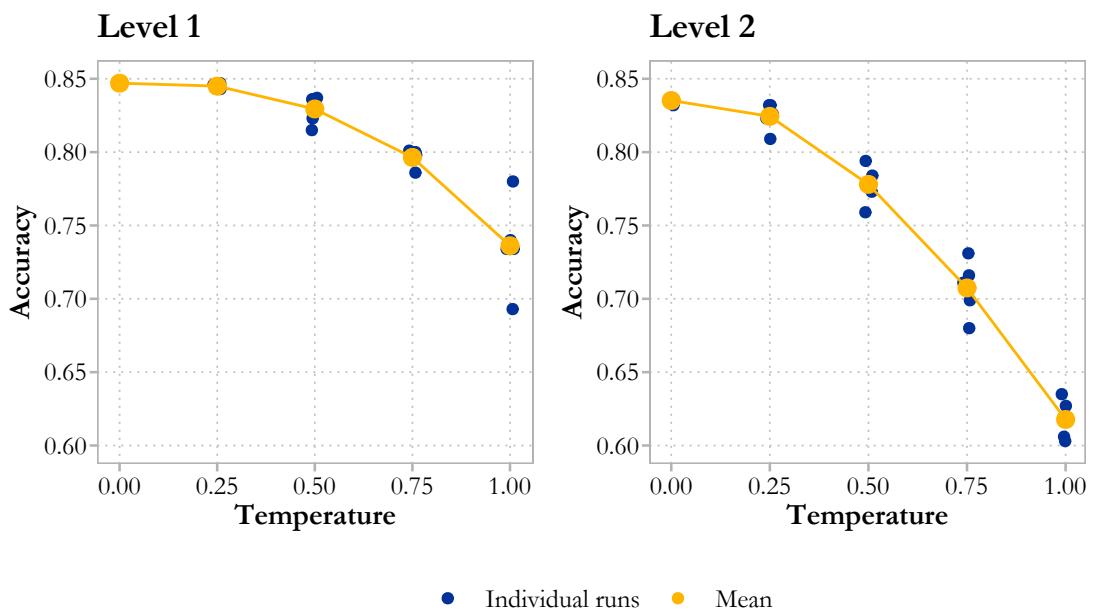


Figure D4.2: Relationship between sentence count, accuracy and token usage

Figure D4.3: Variation in accuracy with different temperature settings



D5 Additional variation in indices

Figures D5.5, D5.6, D5.7, D5.8 present further correlations of our dominance and co-ordination indicators vis-a-vis government bond spreads, polarization and purchasing power parity GDP per capita. Government bond spreads are measured as the difference of the 10 year government bond yield to the bond yield of the German government bond. The polarization indicator is constructed using the "v2cacamps_mean" variable from the VDEM dataset, whereby 0 and 1 map to low polarization and 2 and 3 map to high polarization. GDP figures are taken from the IMF's World Economic Outlook. Figure D5.4 shows the share of normative sentences over time for each of the country groups as defined in Table 6 of the main text.

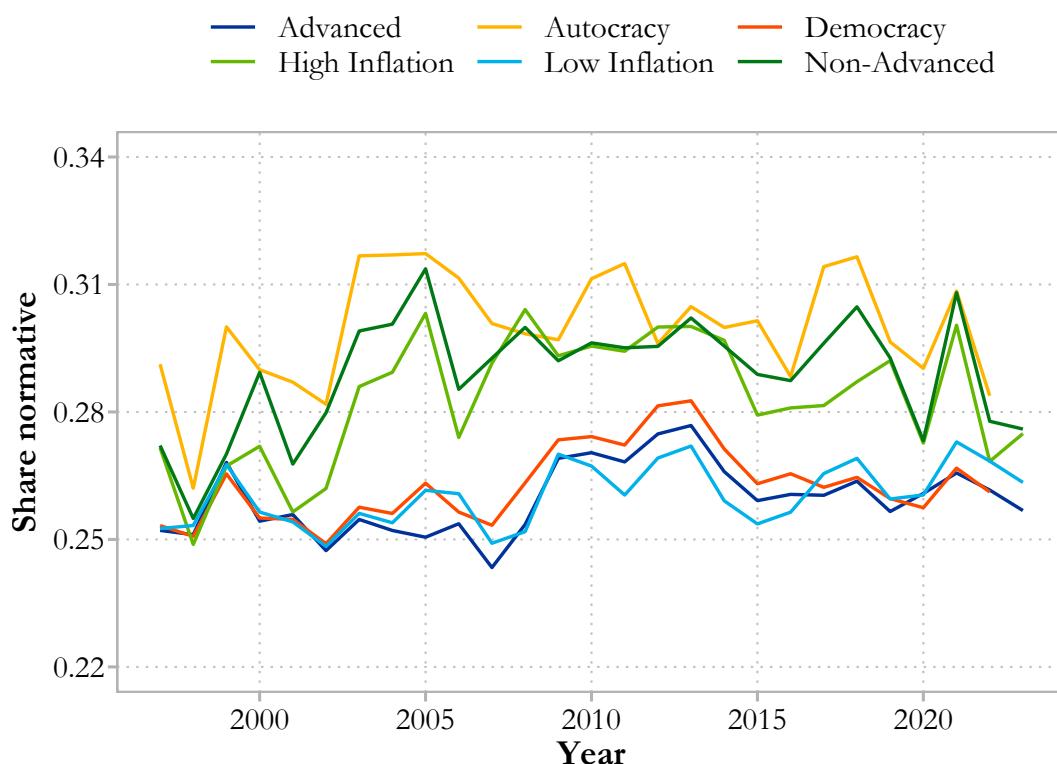
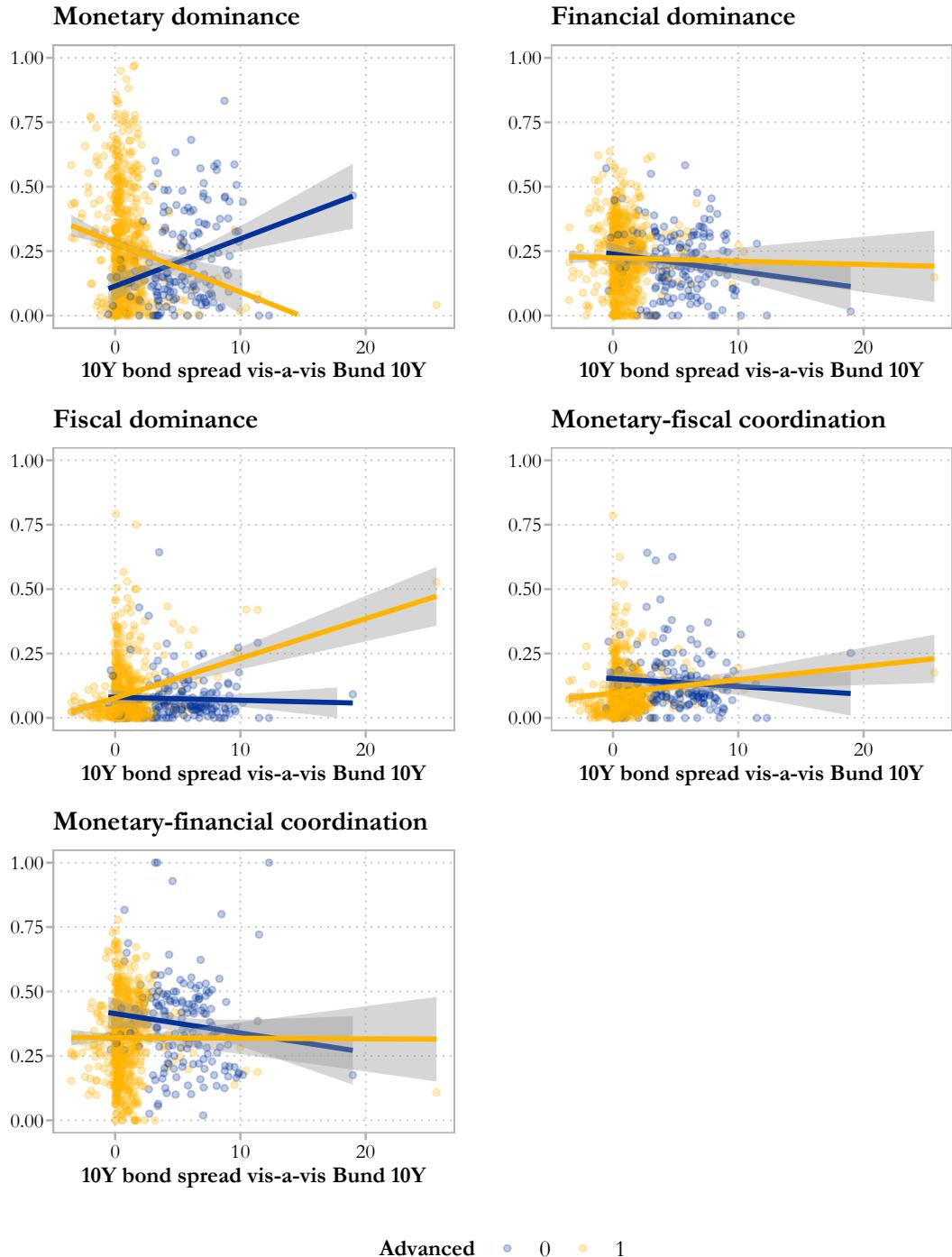
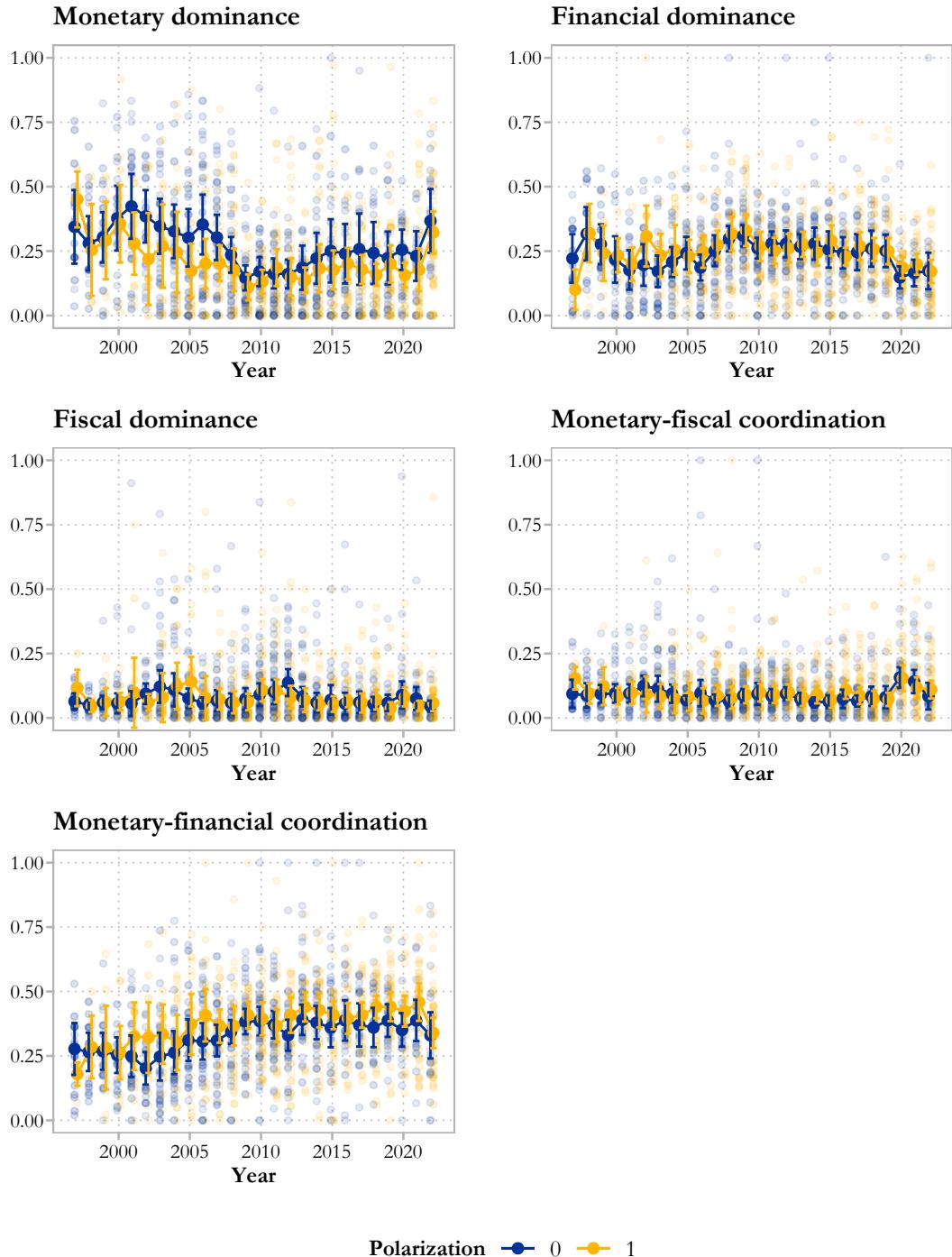


Figure D5.4: Level 2 classification by country group.



Note: Dots show central bank-year observations. The solid line indicates indicates a pooled linear fit by country group. The shaded region around the regression line is the 95% confidence band.

Figure D5.5: Scatter chart of correlation of bond spreads with coordination and dominances in advanced and non-advanced countries.



Note: Scattered line chart of the means of high and low polarization countries over time. Small dots show central bank-year observations. Observations with more speeches are less opaque. The solid line indicates the average of the category-average weighted by the number of speeches. Error bars indicate the 95% confidence interval of the weighted mean.

Figure D5.6: High and low polarization countries over time.

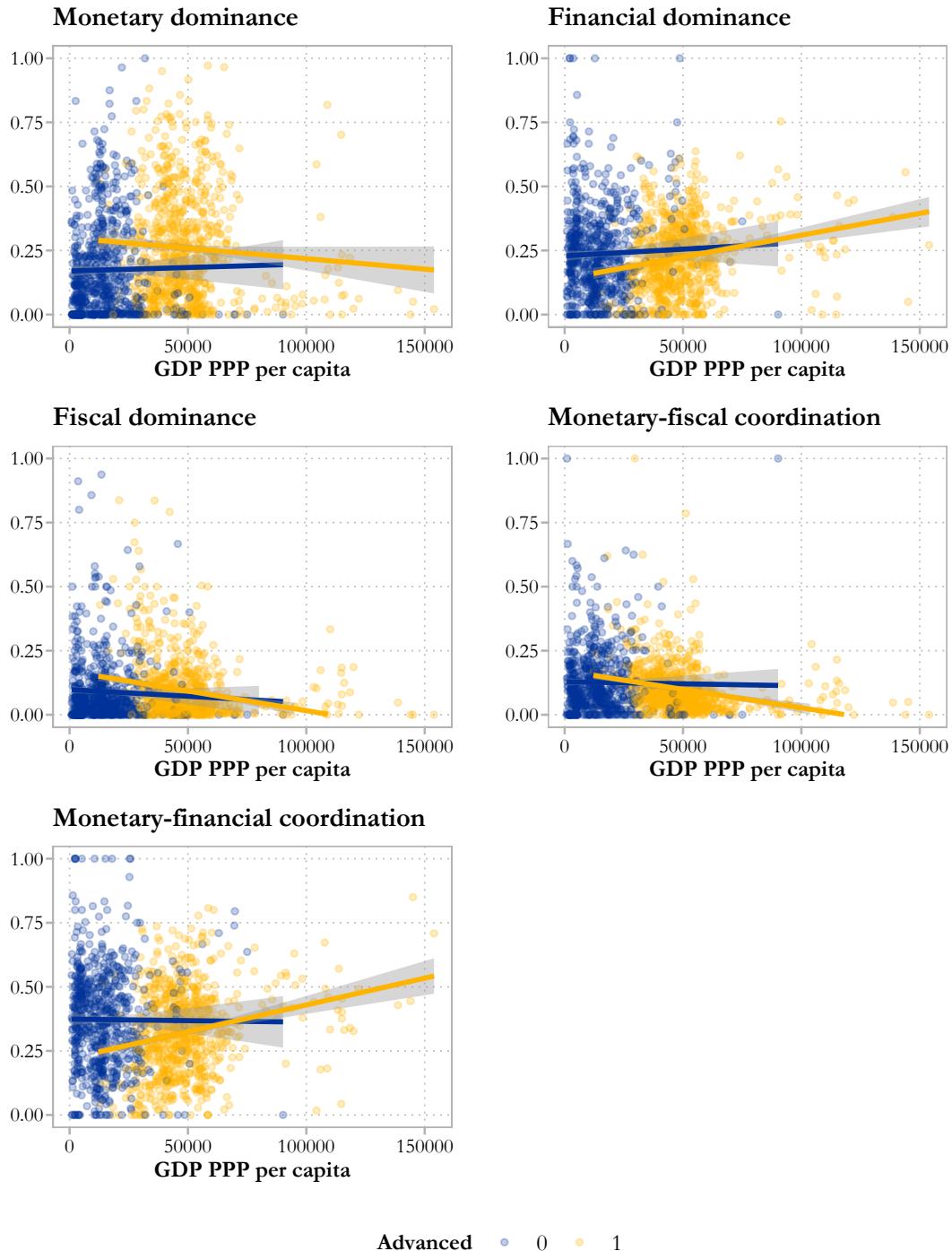


Figure D5.7: Correlation of GDP per capita (PPP) with policy-linkages by economic development

Note: Scatter chart of correlation of GDP per capita (PPP) with coordination and dominances in advanced and non-advanced countries. Dots show central bank-year observations. The solid line indicates a linear fit by country group. The shaded region around the regression line is the 95% confidence band

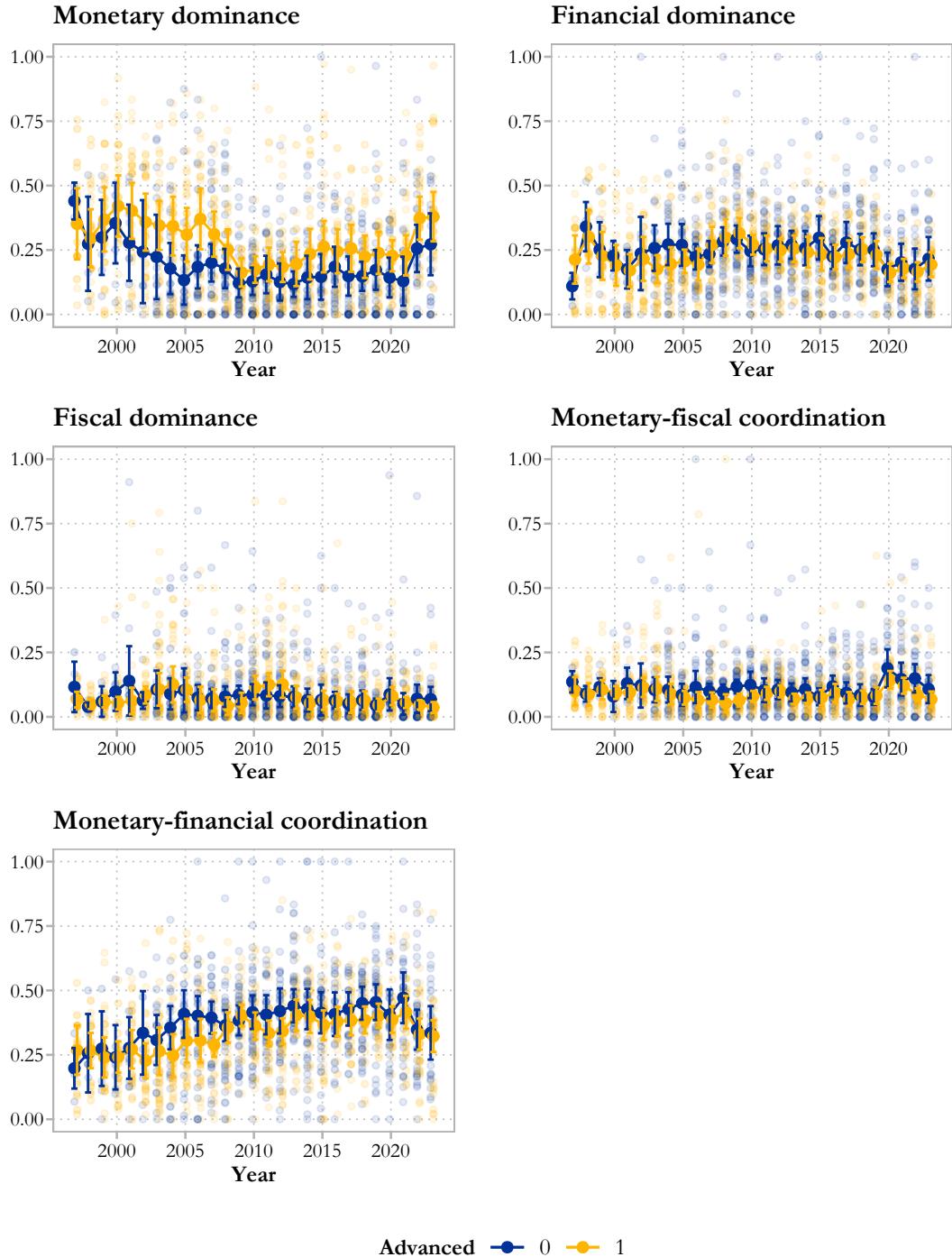


Figure D5.8: Means of advanced and emerging and developing economies over time.

Note: Scattered line chart of the means of advanced and emerging and developing economies over time. Small dots show central-bank year observations. Observations with more speeches are less opaque. The solid line indicates the average of the category-average weighted by the number of speeches. Error bars indicate the 95% confidence interval of the weighted mean.

D6 Prompt variations

D6.1 Minimal Prompt

Classify excerpts from a central bank speech as one the following categories:

- "monetary dominance", "monetary-financial coordination", "monetary-fiscal coordination", "financial dominance" or "fiscal dominance" if the speaker suggests the presence thereof.
- Rely on your knowledge of what those categories mean. Classify in one of the coordination categories if there is no clear hierarchy. When there is no indication of a dominance or coordination relation with regard to fiscal authorities or financial markets classify as "none".

Classify each of the excerpts individually. Reply only with the number of the excerpt and the assigned label.

These are the excerpts:

1. <Excerpt 1>
2. <Excerpt 2>

...

D6.2 Extended Prompt

Classify excerpts from a central banker speech as one of the following categories:

- Classify the excerpts as either financial, monetary, or fiscal dominance when there is a hierarchy between the central bank and another actor. Monetary dominance if the central banker is explicitly or implicitly prescribing a policy to others. Financial or

fiscal dominance if the central banker explicitly or implicitly suggests that monetary policy will accommodate other policies.

- Classify it as fiscal-monetary coordination or monetary-financial coordination when there is not a clear hierarchy and there is a request to coordinate or cooperate.

More information:

- “none” if the central bank official:
 - when the excerpt is ambiguous to the extent that the context cannot be implied
 - this is also the default category
- ”monetary dominance” if the central bank official:
 - emphasizes price stability above other objectives (e.g., inflation target is more important than financial stability or public debt sustainability)
 - suggests that governments should pursue prudent fiscal policy (e.g., fiscal consolidation, reduce public debt/deficit)
 - suggests to raise interest rates despite negative consequences on growth and employment
 - suggests a rules-based fiscal framework or new fiscal institutions
- “monetary-financial coordination” if the central bank official:
 - suggests to work together to create more efficient or better functioning financial markets (e.g., more market transparency)
 - suggests to facilitate liquidity of markets

-
- suggests improving the deposit guarantee funds together with financial institutions
 - “monetary-fiscal coordination” if the central bank official:
 - suggest to introduce a policy-mix of monetary and fiscal policy or tighter cooperation regarding fiscal and monetary policy
 - suggests to provide governments with additional liquidity through loan facilities, accept greater range of securities as collateral for the loans or loans with longer-than-usual maturities.
 - suggests in liaison with governments to use a combination of monetary easing, fiscal expansion, and targeted credit support.
 - suggests to ensure effective transmission of monetary policy to public and private spending.
 - ”financial dominance” if the central bank official:
 - suggests that the central bank should support the financial markets (e.g., provide liquidity support, recapitalize or bail out banks) regardless of inflation
 - suggests not to tighten the monetary policy stance (e.g. not to raise interest rates) if this threatens financial stability
 - suggests that financial market stress is making monetary policy more difficult
 - is concerned about negative feedback effects from the financial markets (e.g. doom loop, contagion)
 - suggests that the central bank will inject liquidity via non-bank financial intermediaries or repo markets, will bail out or recapitalise banks
 - ”fiscal dominance” if the central bank official:

-
- is accommodating to government policies (e.g., suggests to bring down sovereign bond spreads to help reach sustainable growth).
 - suggests monetary financing (e.g., financing the government deficit)
 - suggests to intervene in the sovereign bond market to stabilise it prioritizes government debt sustainability over price stability (e.g. a situation where fiscal policy is not sustainable forcing the central bank to prioritise the government's solvency above its own objectives)

Classify each of the excerpts individually. Reply only with the number of the excerpt and the assigned label.

These are the excerpts:

1. <Excerpt 1>

2. <Excerpt 2>

...

Bibliography

Abadie, A., Athey, S., Imbens, G., and Wooldridge, J. (2023).

When Should You Adjust Standard Errors for Clustering?

The Quarterly Journal of Economics, 138(1):1–35.

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2020).

Sampling-Based versus Design-Based Uncertainty in Regression Analysis.

Econometrica, 88(1):265–296.

Acemoglu, D., Johnson, S., Querubin, P., and Robinson, J. A. (2008).

When Does Policy Reform Work? The Case of Central Bank Independence.

Brookings Papers on Economic Activity, 39(1 (Spring)):351–429.

Acerbi, A. and Stubbersfield, J. M. (2023).

Large language models show human-like content biases in transmission chain experiments.

Proceedings of the National Academy of Sciences, 120(44):e2313790120.

Publisher: Proceedings of the National Academy of Sciences.

Adolph, C. (2013).

Bankers, Bureaucrats, and Central Bank Politics.

Cambridge University Press.

Afonso, A. and Toffano, P. (2013).

Fiscal Regimes in the EU.

- Ahir, H., Dell’Ariccia, G., Furceri, D., Papageorgiou, C., and Qi, H. (2023).
Financial Stress and Economic Activity: Evidence from a New Worldwide Index.
Technical report.
- Ahmed, R., Aizenman, J., and Jinjarak, Y. (2021).
Inflation and Exchange Rate Targeting Challenges Under Fiscal Dominance.
Journal of Macroeconomics, 67:103281.
- Ainsley, C. (2021).
Decentralized central banks: Political ideology and the Federal Reserve System of regional banks.
Governance, 34(2):277–294.
- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019).
Optuna: A Next-generation Hyperparameter Optimization Framework.
arXiv:1907.10902 [cs, stat].
- Aklin, M. and Kern, A. (2021).
The Side Effects of Central Bank Independence.
American Journal of Political Science, 65(4):971–987.
- Aklin, M., Kern, A., Seddon, J., and Leng, N. (2022).
Central Bank Independence in Autocracies.
- Alesina, A. and Summers, L. (1993).
Central Bank Independence and Macroeconomic Performance: Some Comparative Evidence.
Journal of Money, Credit and Banking, 25(2):151–62.
Publisher: Blackwell Publishing.
- Alexandrova, P., Rasmussen, A., and Toshkov, D. (2020).

- Agenda responsiveness in the European Council: public priorities, policy problems and political attention.
West European Politics, 39(4):605–627.
- Algan, Y., Papaioannou, E., Passari, E., and Guriev, S. (2018).
The European Trust Crisis and the Rise of Populism.
- Alizadeh, M., Kubli, M., Samei, Z., Dehghani, S., Zahedivafa, M., Bermeo, J. D., Krobeynikova, M., and Gilardi, F. (2024).
Open-Source LLMs for Text Annotation: A Practical Guide for Model Setting and Fine-Tuning.
arXiv:2307.02179 [cs].
- Anderes, M., Rathke, A., Streicher, S., and Sturm, J.-E. (2021).
The role of ECB communication in guiding markets.
Public Choice, 186(3-4):351–383.
- Armelius, H., Bertsch, C., Hull, I., and Zhang, X. (2020).
Spread the Word: International spillovers from central bank communication.
Journal of International Money and Finance, 103:102116.
- Aruoba, B. and Drechsel, T. (2022).
Identifying Monetary Policy Shocks: A Natural Language Approach.
- Baerg, N. and Lowe, W. (2020).
A textual Taylor rule: Estimating central bank preferences combining topic and scaling methods.
Political Science Research and Methods, 8(1):106–122.
- Baker, A., Hindmoor, A., and McDaniel, S. (2024).

- Performing central bank independence: The Bank of England's communicative financial stability strategy.
Regulation & Governance, 18(3):1000–1017.
- Barro, R. and Gordon, D. (1983a).
Rules, Discretion and Reputation in a Model of Monetary Policy.
Journal of Political Economy, 91(4):589–610.
- Barro, R. J. and Gordon, D. B. (1983b).
A Positive Theory of Monetary Policy in a Natural Rate Model.
Journal of Political Economy, 91(4):589–610.
- Battaglia, L., Christensen, T., Hansen, S., and Sacher, S. (2024).
Inference for Regression with Variables Generated by AI or Machine Learning.
arXiv:2402.15585 [econ].
- Baumgartner, F. R. and Jones, B. D. (1993).
Agendas and Instability in American Politics.
University of Chicago Press.
- Baumgartner, F. R. and Jones, B. D. (2009).
Agendas and Instability in American Politics, Second Edition.
University of Chicago Press, Chicago, IL.
- Beck, N. (1987).
Elections and the Fed: Is There a Political Monetary Cycle?
American Journal of Political Science, 31(1):194–216.
- Beck, N. and Katz, J. N. (2006).
What to Do (and Not to Do) with Time-Series Cross-Section.
American Political Science Review, 100(4):676–677.
Publisher: Cambridge University Press.

- Belke, A. and Potrafke, N. (2012).
Does government ideology matter in monetary policy? A panel data analysis for OECD countries.
Journal of International Money and Finance, 31(5):1126–1139.
- Bennani, H. and Neuenkirch, M. (2017).
The (home) bias of European central bankers: new evidence based on speeches.
Applied Economics, 49(11):1114–1131.
- Benoit, K. (2020).
Text as Data: An Overview.
In *The SAGE Handbook of Research Methods in Political Science and International Relations*, pages 461–497. SAGE Publications Ltd, 1 Oliver’s Yard, 55 City Road London EC1Y 1SP.
- Berger, W. and Kißmer, F. (2013).
Central bank independence and financial stability: A tale of perfect harmony?
European Journal of Political Economy, 31:109–118.
- Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B. (2011).
Algorithms for Hyper-Parameter Optimization.
In *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc.
- Bernanke, B. (2010).
Central Bank Independence, Transparency, and Accountability.
- Bertsch, C., Hull, I., Lumsdaine, R. L., and Zhang, X. (2024).
Central Bank Mandates and Monetary Policy Stances: through the Lens of Federal Reserve Speeches.

- Besley, T. and Case, A. (2003).
Political Institutions and Policy Choices: Evidence from the United States.
Journal of Economic Literature, 41(1):7–73.
- Bestvater, S. E. and Monroe, B. L. (2023).
Sentiment is Not Stance: Target-Aware Opinion Classification for Political Text Analysis.
Political Analysis, 31(2):235–256.
- Betz, T., Cook, S. J., and Hollenbach, F. M. (2018).
On the Use and Abuse of Spatial Instruments.
Political Analysis, 26(4):474–479.
- Bhat, S. and Dugal, I. (2025).
India’s new central bank chief joins Modi’s effort to regain lost growth.
Reuters.
- Bianchi, F. and Ilut, C. (2017).
Monetary/Fiscal policy mix and agents’ beliefs.
Review of Economic Dynamics, 26:113–139.
- Bird, S., Klein, E., and Loper, E. (2009).
Natural Language Processing with Python.
- BIS (2016).
Macroprudential Policy.
- Blauberger, M., Heindlmaier, A., Kramer, D., Martinsen, D. S., Sampson Thierry, J., Schenk, A., and Werner, B. (2018).
ECJ Judges read the morning papers. Explaining the turnaround of European citizenship jurisprudence.

- Journal of European Public Policy*, 25(10):1422–1441.
- Blei, D., Ng, A., and Jordan, M. (2003).
Latent Dirichlet Allocation.
Journal of Machine Learning Research, 3:993–1022.
- Blinder, A., Goodhart, C., Hildebrand, P., Lipton, D., and Wyplosz, C. (2001).
How do Central Banks Talk?
Geneva Reports on the World Economy, 3.
- Blinder, A. S. (2004).
The Quiet Revolution: Central Banking Goes Modern.
Yale University Press.
- Blinder, A. S., Ehrmann, M., de Haan, J., and Jansen, D.-J. (2022).
Central Bank Communication with the General Public: Promise or False Hope?
- Blinder, A. S., Ehrmann, M., Fratzscher, M., de Haan, J., and Jansen, D.-J. (2008).
Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence.
Journal of Economic Literature, 46(4):910–945.
- Blinder, A. S., Ehrmann, M., Haan, J. d., and Jansen, D.-J. (2024).
Central Bank Communication with the General Public: Promise or False Hope?
Journal of Economic Literature, 62(2):425–457.
- Bloomberg, I. O. D. (2024).
Argentina takes steps to speed up Milei’s dollarisation strategy.
Section: economy.
- Board, F. S. (2024).
Post-2008 Financial Crisis Reforms.

Bodea, C. and Hicks, R. (2015).

Price Stability and Central Bank Independence: Discipline, Credibility, and Democratic Institutions.

International Organization, 69(1):35–61.

Bodea, C. and Kerner, A. (2022).

Fear of inflation and gender representation in central banking.

European Journal of Political Economy, 74:102192.

Boelaert, J., Coavoux, S., Ollion, E., Petev, I. D., and Präg, P. (2024).

Machine Bias. Generative Large Language Models Have a Worldview of Their Own.

Bohl, M. T., Kanelis, D., and Siklos, P. (2023).

Central bank mandates: How differences can influence the content and tone of central bank communication.

Journal of International Money and Finance, 130(C).

Publisher: Elsevier.

Boneva, L., Ferrucci, G., and Mongelli, F. P. (2022).

Climate change and central banks: what role for monetary policy?

Climate Policy, 22(6):770–787.

Bordo, M. D. (2007).

A Brief History of Central Banks.

Economic Commentary, (12/1/2007).

Publisher: Federal Reserve Bank of Cleveland.

Borusyak, K., Jaravel, X., and Spiess, J. (2024).

Revisiting Event Study Designs: Robust and Efficient Estimation.

arXiv:2108.12419 [econ].

- Brand, C., Buncic, D., and Turunen, J. (2010).
The Impact of Ecb Monetary Policy Decisions and Communication on the Yield Curve.
Journal of the European Economic Association, 8(6):1266–1298.
Publisher: Oxford University Press.
- Braun, B. (2016).
Speaking to the people? Money, trust, and central bank legitimacy in the age of quantitative easing.
Review of International Political Economy, 23(6):1064–1092.
- Braun, B. and Dusterhoft, M. (2024).
Noisy politics, quiet technocrats? Central banking in contentious times.
- Broz, J. L. (2002).
Political System Transparency and Monetary Commitment Regimes.
International Organization, 56(4):861–887.
- Brunnermeier, M. K. (2015).
Financial Dominance.
Banca d’Italia.
- Brunnermeier, M. K., James, H., and Landau, J.-P. (2016).
The Euro and the battle of ideas.
In *The euro and the battle of ideas*. Princeton University Press.
- Bursian, D. and Faia, E. (2018).
Trust in the Monetary Authority.
Journal of Monetary Economics, 98.
- Busuioc, E. M. and Lodge, M. (2016).
Citation for: The Reputational Basis of Public Accountability.

Wiley Online Library.

Busuioc, M. and Rimkutė, D. (2020).

Meeting expectations in the EU regulatory state? Regulatory communications amid conflicting institutional demands.

Journal of European Public Policy, 27(4):547–568.

Publisher: Routledge _eprint: <https://doi.org/10.1080/13501763.2019.1603248>.

Callaway, B., Goodman-Bacon, A., and Sant'Anna, P. (2024).

Difference-in-Differences with a Continuous Treatment.

Callaway, B. and Sant'Anna, P. H. C. (2021).

Difference-in-Differences with multiple time periods.

Journal of Econometrics, 225(2):200–230.

Calvo, D., Crisanto, J. C., Hohl, S., and Gutiérrez, O. P. (2018).

Financial supervisory architecture: what has changed after the crisis?

Campello, R. J. G. B., Moulavi, D., and Sander, J. (2013).

Density-Based Clustering Based on Hierarchical Density Estimates.

In Pei, J., Tseng, V. S., Cao, L., Motoda, H., and Xu, G., editors, *Advances in Knowledge Discovery and Data Mining*, pages 160–172, Berlin, Heidelberg. Springer.

Canes-Wrone, B. and Shotts, K. W. (2004).

The Conditional Nature of Presidential Responsiveness to Public Opinion.

American Journal of Political Science, 48(4):690–706.

Casiraghi, M. and Perez, L. P. (2022).

Central Bank Communications. Monetary and Capital Markets Department Technical Assistance Handbook.

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019).

The Effect of Minimum Wages on Low-Wage Jobs*.

The Quarterly Journal of Economics, 134(3):1405–1454.

Chappell, H. W., McGregor, R. R., and Vermilyea, T. A. (2008).

Regional economic conditions and monetary policy.

European Journal of Political Economy, 24(2):283–293.

Chen, X., Leeper, E. M., and Leith, C. (2022).

Strategic interactions in U.S. monetary and fiscal policies.

Quantitative Economics, 13(2):593–628.

Christelis, D., Georgarakos, D., Jappelli, T., and van Rooij, M. (2020).

Trust in the Central Bank and Inflation Expectation.

SSRN Electronic Journal.

Christiano, L. J., Eichenbaum, M. S., and Trabandt, M. (2018).

On DSGE Models.

Journal of Economic Perspectives, 32(3):113–140.

Conti-Brown, P. (2017).

The Power and Independence of the Federal Reserve.

Princeton University Press, Princeton, NJ.

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Altman, D., Bernhard, M., Agnes Cornell, M. F., Gastaldi, L., Gjerløw, H., Glynn, A., Hicken, A., Lührmann, A., Maerz, S. F., Marquardt, K. L., McMann, K., Mechkova, V., Paxton, P., Pemstein, D., Römer, J., Seim, B., Sigman, R., Skaaning, S.-E., Staton, J., Sundström, A., Tzelgov, E., Uberti, L., Wang, Y.-t., Wig, T., and Ziblatt, D. (2024).

'V-Dem Codebook v11' Varieties of Democracy (V-Dem) Project.

- Cox, G. W. and McCubbins, M. D. (2000).
Political structure and economic policy: The institutional determinants of policy outcomes.
Publisher: Cambridge University Press.
- Crisanto, J. C., Hohl, S., and Calvo, D. (2018).
Financial supervisory architecture since the Global Crisis: Supervisory models remain diverse, but more powers for central banks.
- Cross, J. P. and Greene, D. (2020).
Talk is not cheap: Policy agendas, information processing, and the unusually proportional nature of European Central Bank communications policy responses.
- Cross, J. P., Greene, D., Umansky, N., and Calò, S. (2024).
Speaking in unison? Explaining the role of agenda-setter constellations in the ECB policy agenda using a network-based approach.
Journal of European Public Policy, 31(11):3676–3702.
- Crowe, C. and Meade, E. E. (2008).
Central bank independence and transparency: Evolution and effectiveness.
European Journal of Political Economy, 24(4):763–777.
- Cukierman, A., Webb, S. B., and Neyapti, B. (1992).
Measuring the Independence of Central Banks and Its Effect on Policy Outcomes.
The World Bank Economic Review, 6(3):353–398.
- Davig, T. and Leeper, E. M. (2011).
Monetary–fiscal policy interactions and fiscal stimulus.
European Economic Review, 55(2):211–227.
- de Chaisemartin, C. and D'Haultfœuille, X. (2024).

- Difference-in-Differences Estimators of Intertemporal Treatment Effects.
de Grauwe, P. (2020).
Economics of Monetary Union.
Oxford University Press.
- De Haan, J. and Sturm, J.-E. (2019).
Central bank communication.
In *The Oxford handbook of the economics of central banking*, page 231. Oxford University Press.
- Dettmers, T., Pagnoni, A., Holtzman, A., and Zettlemoyer, L. (2023).
QLoRA: Efficient Finetuning of Quantized LLMs.
arXiv:2305.14314 [cs].
- Deyris, J. (2023).
Too green to be true? Forging a climate consensus at the European Central Bank.
New Political Economy, 28(5):713–730.
- Diessner, S. (2023).
The political economy of monetary-fiscal coordination: central bank losses and the specter of central bankruptcy in Europe and Japan.
Review of International Political Economy, 0(0):1–23.
Publisher: Routledge eprint: <https://doi.org/10.1080/09692290.2023.2295373>.
- Diessner, S. and Lisi, G. (2020).
Masters of the ‘masters of the universe’? Monetary, fiscal and financial dominance in the Eurozone.
Socio-Economic Review, 18(2):315–335.
- Dincer, N. N. and Eichengreen, B. (2018).

- Central Bank Transparency and Independence: Updates and New Measures.
34th issue (March 2014) of the International Journal of Central Banking.
- Dreher, A., Sturm, J.-E., and de Haan, J. (2010).
When is a central bank governor replaced? Evidence based on a new data set.
Journal of Macroeconomics, 32(3):766–781.
- Dube, A., Girardi, D., Jorda, O., and Taylor, A. M. (2023).
A Local Projections Approach to Difference-in-Differences.
- ECB (2020).
The shadow of fiscal dominance: Misconceptions, perceptions and perspectives.
Technical report.
- ECB (2022).
Independence.
- Ehrmann, M. and Fratzscher, M. (2007).
The timing of central bank communication.
European Journal of Political Economy, 23(1):124–145.
- Ehrmann, M. and Fratzscher, M. (2010).
Politics and Monetary Policy.
The Review of Economics and Statistics, 93:941–960.
- Ehrmann, M., Holton, S., Kedan, D., and Phelan, G. (2021).
Monetary Policy Communication: Perspectives from Former Policy Makers at the ECB.
- Eichengreen, B. (2008).
Globalizing Capital: A History of the International Monetary System - Second Edition.
Princeton University Press, rev - revised, 2 edition.

Eichengreen, B., Hausmann, R., and Panizza, U. (2023).

Yet it Endures: The Persistence of Original Sin.

Open Economies Review, 34(1):1–42.

Eijffinger, S. C. W. and Hoeberichts, M. (2002).

Central Bank Accountability and Transparency: Theory and Some Evidence.

International Finance, 5(1):73–96.

_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-2362.00088>.

El-Erian, M. A. (2016).

The Only Game in Town: Central Banks, Instability, and Recovering from Another Collapse.

Elgie, R. (1998).

Democratic accountability and central bank independence: Historical and contemporary, national and European perspectives.

West European Politics, 21(3):53–76.

Publisher: Routledge _eprint: <https://doi.org/10.1080/01402389808425257>.

Epstein, D. and O'Halloran, S. (1999).

Delegating Powers: A Transaction Cost Politics Approach to Policy Making Under Separate Powers.

Ersel, H. and Özatay, F. (2008).

Fiscal Dominance and Inflation Targeting: Lessons from Turkey.

Emerging Markets Finance and Trade, 44(6):38–51.

Eulau, H. and Karps, P. D. (1977).

The Puzzle of Representation: Specifying Components of Responsiveness.

Legislative Studies Quarterly, 2(3):233.

- Evdokimova, T., Mohácsi, P. N., Ponomarenko, O., and Ribakova, E. (2023).
Central banks and policy communication: How emerging markets have outperformed
the Fed and ECB | PIIE.
Section: Working Papers 23-10.
- Faiz, A., Kaneda, S., Wang, R., Osi, R., Sharma, P., Chen, F., and Jiang, L. (2024).
LLMCarbon: Modeling the end-to-end Carbon Footprint of Large Language Models.
arXiv:2309.14393 [cs].
- Fanta, N. and Horvath, R. (2024).
Artificial intelligence and central bank communication: the case of the ECB.
Applied Economics Letters, 0(0):1–8.
Publisher: Routledge _eprint: <https://doi.org/10.1080/13504851.2024.2337318>.
- Farhi, E. and Tirole, J. (2012).
Collective Moral Hazard, Maturity Mismatch, and Systemic Bailouts.
American Economic Review, 102(1):60–93.
- Favero, C. A. and Monacelli, T. (2005).
Fiscal Policy Rules and Regime (In)Stability: Evidence from the U.S.
- Feldkircher, M., Hofmarcher, P., and Siklos, P. L. (2024).
One money, one voice? Evaluating ideological positions of euro area central banks.
European Journal of Political Economy, 85:102582.
- Fernández-Albertos, J. (2015).
The Politics of Central Bank Independence.
Annual Review of Political Science, 18(1):217–237.
- Ferrara, F. M. (2020).
The battle of ideas on the euro crisis: evidence from ECB inter-meeting speeches.

Ferris, S. (2008).

Electoral politics and monetary policy: does the Bank of Canada contribute to a political business cycle?

Public Choice, 135(3):449–468.

Fischer, S. (1995).

Central-Bank Independence Revisited.

The American Economic Review, 85(2):201–206.

Publisher: American Economic Association.

Fraga, A., Goldfajn, I., and Minella, A. (2003).

Inflation Targeting in Emerging Market Economies.

Freyaldenhoven, S., Hansen, C., Pérez Pérez, J., and Shapiro, J. M. (2021).

Visualization, Identification, and Estimation in the Linear Panel Event-Study Design.

Fromage, D., Dermine, P., Nicolaides, P., and Tuori, K. (2019).

ECB independence and accountability today: Towards a (necessary) redefinition?

Maastricht Journal of European and Comparative Law, 26(1):3–16.

Publisher: SAGE Publications Ltd.

Gabor, D. (2016).

The (impossible) repo trinity: the political economy of repo markets.

Review of International Political Economy, 23(6):967–1000.

Gambacorta, L., Kwon, B., Park, T., Patelli, P., and Zhu, S. (2024).

CB-LMs: language models for central banking.

Gardner, J., Thakral, N., To, L. T., and Yap, L. (2024).

Two-stage differences in differences.

arXiv:2207.05943 [econ].

Garriga, A. C. (2010).

Determinants of Central Bank Independence in Developing Countries: A Two-Level Theory.

Garriga, A. C. (2022).

Central banks and civil war termination.

Journal of Peace Research, 59(4):508–525.

Garriga, A. C. (2025a).

Revisiting Central Bank Independence in the World: An Extended Dataset.

International Studies Quarterly.

Garriga, A. C. (2025b).

Revisiting Central Bank Independence in the World: An Extended Dataset.

International Studies Quarterly.

Garriga, A. C. and Rodriguez, C. M. (2020).

More effective than we thought: Central bank independence and inflation in developing countries.

Economic Modelling, 85:87–105.

Garriga, A. C. and Rodriguez, C. M. (2023).

Central bank independence and inflation volatility in developing countries.

Economic Analysis and Policy, 78:1320–1341.

Gemini Team, n. (2023).

Gemini: A Family of Highly Capable Multimodal Models.

Technical report.

Genovese, F., Schneider, G., and Wassmann, P. (2016).

The Eurotower Strikes Back: Crises, Adjustments, and Europe's Austerity Protests.

Comparative Political Studies, 49(7):939–967.

Publisher: SAGE Publications Inc.

Gentzkow, M., Kelly, B., and Taddy, M. (2019).

Text as Data.

Journal of Economic Literature, 57(3):535–574.

Geraats, P. M. (2002).

Central Bank Transparency*.

The Economic Journal, 112(483):F532–F565.

Publisher: John Wiley & Sons, Ltd.

Giavazzi, F. and Pagano, M. (1988).

The advantage of tying one's hands: EMS discipline and Central Bank credibility.

European Economic Review, 32(5):1055–1075.

Gilad, S., Maor, M., and Bloom, P. B.-N. (2015).

Organizational Reputation, the Content of Public Allegations, and Regulatory Communication.

Journal of Public Administration Research and Theory, 25(2):451–478.

Gilardi, F., Alizadeh, M., and Kubli, M. (2023).

ChatGPT outperforms crowd workers for text-annotation tasks.

Proceedings of the National Academy of Sciences, 120(30):e2305016120.

Publisher: Proceedings of the National Academy of Sciences.

Gleditsch, K. S. and Ward, M. D. (1997).

Double Take: A Reexamination of Democracy and Autocracy in Modern Polities.

The Journal of Conflict Resolution, 41(3):361–383.

Gleditsch, K. S. and Ward, M. D. (2006).

Diffusion and the International Context of Democratization.
International Organization, 60(4):911–933.

Goodhart, C. and Lastra, R. (2018).
Populism and Central Bank Independence.
Open Economies Review, 29(1):49–68.

Goodman-Bacon, A. (2021).
Difference-in-differences with variation in treatment timing.
Journal of Econometrics, 225(2):254–277.

Green-Pedersen, C. and Walgrave, S. (2014).
Agenda Setting, Policies, and Political Systems: A Comparative Approach.
University of Chicago Press.

Grier, K. B. (1989).
On the Existence of a Political Monetary Cycle.
American Journal of Political Science, 33(2):376–389.

Grilli, V., Masciandaro, D., and Tabellini, G. (1991).
Political and monetary institutions and public financial policies in the industrial countries.
Economic Policy, 6(13):341–392.

Grimmer, J., Roberts, M. E., and Stewart, B. M. (2022).
Text as Data: A New Framework for Machine Learning and the Social Sciences.
Princeton University Press, Princeton (N.J.).

Grimmer, J. and Stewart, B. (2013).
Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.

- Political Analysis.*
- Grootendorst, M. (2022).
BERTopic: Neural topic modeling with a class-based TF-IDF procedure.
- Gros, D. and Shamsfakhr, F. (2021).
Financial Dominance: Not an Immediate Danger.
- Haan, J. and Amtenbrink, F. (2000).
Democratic accountability and central bank independence: A response to Elgie.
West European Politics - WEST EUR POLIT, 23:179–190.
- Haan, J. and Sturm, J.-E. (2019).
Central Bank Communication: How to Manage Expectations?
pages 231–262.
- Haan, J. D., Amtenbrink, F., and Waller, S. (2004).
The Transparency and Credibility of the European Central Bank.
JCMS: Journal of Common Market Studies, 42(4):775–794.
Publisher: John Wiley & Sons, Ltd.
- Haldane, A. (2020).
What has central bank independence ever done for us?
- Haldane, A., Macaulay, A., and McMahon, M. (2020).
The 3 E's of central bank communication with the public.
- Haldane, A. and McMahon, M. (2018).
Central Bank Communications and the General Public.
AEA Papers and Proceedings, 108:578–583.
- Hansen, A. L. and Kazinnik, S. (2023).

Can ChatGPT Decipher Fedspeak?

SSRN Electronic Journal.

Hansen, D. (2022).

The Economic Consequences of Banking Crises: The Role of Central Banks and Optimal Independence.

American Political Science Review, 116(2):453–469.

Hansen, S. and McMahon, M. (2016).

Shocking language: Understanding the macroeconomic effects of central bank communication.

Journal of International Economics, 99:S114–S133.

Hansson, M. (2021).

Evolution of topics in central bank speech communication.

arXiv:2109.10058 [econ, q-fin].

Harary, F., Norman, R. Z., and Cartwright, D. (1965).

Structural models: an introduction to the theory of directed graphs.

Wiley, New York.

Harms, P. and Knaze, J. (2021).

Bilateral de-jure exchange rate regimes and foreign direct investment: A gravity analysis.

Journal of International Money and Finance, 117:102438.

Hayo, B. and Méon, P.-G. (2013).

Behind closed doors: Revealing the ECB’s decision rule.

Journal of International Money and Finance, 37:135–160.

Hayo, B. and Neuenkirch, M. (2015).

- Central bank communication in the financial crisis: Evidence from a survey of financial market participants.
Journal of International Money and Finance, 59:166–181.
- Heinemann, F. and Kemper, J. (2021).
The ECB Under the Threat of Fiscal Dominance – The Individual Central Banker Dimension.
The Economists' Voice, 18(1):5–30.
- Hellwig, M. F. (2014).
Financial Stability, Monetary Policy, Banking Supervision, and Central Banking.
- Herzog, A. and Benoit, K. (2015).
The Most Unkindest Cuts: Speaker Selection and Expressed Government Dissent during Economic Crisis.
The Journal of Politics, 77(4):1157–1175.
- Hinterlang, N. and Hollmayr, J. (2022).
Classification of monetary and fiscal dominance regimes using machine learning techniques.
Journal of Macroeconomics, 74:103469.
- Hix, S., Hoyland, B., and Vivyan, N. (2007).
From Doves to Hawks: A Spatial Analysis of Voting in the Monetary Policy Committee of the Bank of England, 1997-2007.
- Hobolt, S. and Klemmensen, R. (2008).
Government Responsiveness and Political Competition in Comparative Perspective.
Comparative Political Studies, 41(3).
- Holmes, D. R. (2013).

- Economy of Words: Communicative Imperatives in Central Banks.*
University of Chicago Press, Chicago, IL.
- Horvath, R. and Katuscakova, D. (2016).
Transparency and trust: the case of the European Central Bank.
Applied Economics.
- Horvath, R. and Vaško, D. (2013).
Central Bank Transparency and Financial Stability: Measurement, Determinants and Effects.
- Huber, J. D. and Shipan, R., C. (2002).
Deliberate Discretion?
Cambridge University Press.
- Hwang, I. D., Lustenberger, T., and Rossi, E. (2023).
Central bank communication and public trust: The case of ECB speeches.
Journal of International Money and Finance, 137:102916.
- Hölzl, J., Keusch, F., and Sajons, C. (2025).
The (mis)use of Google Trends data in the social sciences - A systematic review, critique, and recommendations.
Social Science Research, 126:103099.
- IMF (2023).
World Economic Outlook Database - Groups and Aggregates.
- IMF (2024).
World Economic Outlook Database.
- Issing, O. (2014).
Communication and transparency – The example of the ECB.

- Journal of Economic Dynamics and Control*, 49:70–73.
- Issing, O. (2019).
The Long Journey of Central Bank Communication.
MIT Press.
- Jarociński, M. and Karadi, P. (2020).
Deconstructing Monetary Policy Surprises—The Role of Information Shocks.
American Economic Journal: Macroeconomics, 12(2):1–43.
- Jazeera, A. (2021).
Turkey: Erdogan central bank firing clears way for more rate cuts.
- Johnson, J. (2016).
Priests of Prosperity: How Central Bankers Transformed the Postcommunist World.
In *Priests of Prosperity*. Cornell University Press.
- Jones, B. D. and Baumgartner, F. R. (2004).
Representation and Agenda Setting.
Policy Studies Journal, 32(1):1–24.
- Jones, E. (2009).
Output Legitimacy and the Global Financial Crisis: Perceptions Matter.
JCMS: Journal of Common Market Studies, 47(5):1085–1105.
- Jones, E. (2020).
Why We Need More Diverse Central Bankers.
Survival, 62(6):113–121.
- Jácome, L. I. and Vázquez, F. (2008).
Is there any link between legal central bank independence and inflation? Evidence from Latin America and the Caribbean.

- European Journal of Political Economy*, 24(4):788–801.
- Kaltenthaler, K. (2005).
The Bundesbank and the formation of the ECB's monetary policy strategy.
German Politics, 14:297–314.
- Kindleberger, C. P. (1996).
Manias, Panics and Crashes: A History of Financial Crisis.
Wiley.
Google-Books-ID: Wr6Nc2juZOcC.
- King, M. (2005).
Epistemic Communities and the Diffusion of Ideas: Central Bank Reform in the United Kingdom.
West European Politics, 28(1):94–123.
- Kirshner, J. (2018).
The Inescapable Politics of Money.
In *Monetary Orders: Ambiguous Economics, Ubiquitous Politics*, pages 3–24. Cornell University Press.
- Klimm, F. (2022).
Quantifying the ‘end of history’ through a Bayesian Markov-chain approach.
Royal Society Open Science, 9(11):221131.
- Klomp, J. and De Haan, J. (2010).
Inflation and Central Bank Independence: A Meta-Regression Analysis.
Journal of Economic Surveys, 24(4):593–621.
- Kocherlakota, N. R. (2017).

- The Decentralized Central Bank: A Review Essay on "The Power and Independence of the Federal Reserve".
Journal of Economic Literature, 55(2):621–636.
- Koop, C. and Lodge, M. (2020).
British economic regulators in an age of politicisation: from the responsible to the responsive regulatory state?
Journal of European Public Policy, 27(11):1612–1635.
Publisher: Routledge _eprint: <https://doi.org/10.1080/13501763.2020.1817127>.
- Koop, C., Reh, C., and Bressanelli, E. (2022).
Citation for: Agenda-setting under pressure: Does domestic politics influence the European Commission?
European Journal of Political Research.
- Kydland, F. E. and Prescott, E. C. (1977).
Rules Rather than Discretion: The Inconsistency of Optimal Plans.
Journal of Political Economy, 85(3):473–491.
Publisher: University of Chicago Press.
- Laeven, L. and Valencia, F. (2018).
Systemic Banking Crises Revisited.
- Laskar, M. T. R., Bari, M. S., Rahman, M., Bhuiyan, M. A. H., Joty, S., and Huang, J. X. (2023).
A Systematic Study and Comprehensive Evaluation of ChatGPT on Benchmark Datasets.
arXiv:2305.18486 [cs].
- Lee, J. and Wessel, D. (2024).

What did the Fed do after Silicon Valley Bank and Signature Bank failed?
Technical report.

Leek, L. and Bischl, S. (2025).

How central bank independence shapes monetary policy communication: A Large Language Model application.

European Journal of Political Economy, 87:102668.

Publisher: Elsevier.

Leek, L., Bischl, S., and Freier, M. (2024).

Introducing Textual Measures of Central Bank Policy-Linkages Using ChatGPT.

Leeper, E. M. (1991).

Equilibria under ‘active’ and ‘passive’ monetary and fiscal policies.

Journal of Monetary Economics, 27(1):129–147.

Liedorp, F., Mosch, R., van der Cruijsen, C., and de Haan, J. (2013).

Transparency of Banking Supervisors.

IMF Economic Review, 61(2):310–335.

Lindé, J. (2018).

DSGE models: still useful in policy analysis?

Oxford Review of Economic Policy, 34(1-2):269–286.

Lohmann, S. (2003).

Why Do Institutions Matter? An Audience–Cost Theory of Institutional Commitment
- Lohmann - 2003 - Governance - Wiley Online Library.

Governance.

Mabbett, D. and Schelkle, W. (2019).

Independent or lonely? Central banking in crisis.

- Review of International Political Economy*, 26(3):436–460.
- Maddaloni, A., Mendicino, C., and Laeven, L. (2022).
Monetary and macroprudential policies: Trade-offs and interactions.
- Majone, G. (1996).
Temporal Consistency and Policy Credibility: Why Democracies Need Non-Majoritarian Institutions.
European University Institute, Working Paper RSC No 96/57.
- Majone, G. (2001).
Two Logics of Delegation: Agency and Fiduciary Relations in EU Governance.
European Union Politics, 2(1):103–122.
- Makochekanwa, A. (2008).
Impact of Budget Deficit on Inflation in Zimbabwe.
Economic Research Guardian.
- Malmendier, U., Nagel, S., and Yan, Z. (2021).
The making of hawks and doves.
Journal of Monetary Economics, 117(C):19–42.
Publisher: Elsevier.
- Malsin, J. (2024).
Disgraced Lebanese Central Banker Who Was a Fugitive Now Awaits His Fate - WSJ.
- Marozzi, A. (2021).
Beware of Fiscal Signalling. The Effects of the ECB’s Fiscal Communication in the Euro Area.
- Marsh, D. (2021).
Threats to central bank independence.

- Masciandaro, D. and Romelli, D. (2018).
Central bankers as supervisors: Do crises matter?
European Journal of Political Economy, 52:120–140.
- McInnes, L., Healy, J., and Melville, J. (2020).
UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.
- McNamara, K. (2002).
Rational Fictions: Central Bank Independence and the Social Logic of Delegation.
25(1).
- Meade, E. and Sheets, N. (2005).
Regional Influences on FOMC Voting Patterns.
Journal of Money, Credit and Banking, 37(4):661–77.
- Meijers, M. J., Schneider, C. J., and Zhelyazkova, A. (2019).
Dimensions of input responsiveness in the EU: actors, publics, venues.
Journal of European Public Policy, 26(11):1724–1736.
- Mengus, E., Plantin, G., and Barthelemy, J. (2021).
Large public debts need not imply fiscal dominance.
- Miller, D. L. (2023).
An Introductory Guide to Event Study Models.
Journal of Economic Perspectives, 37(2):203–230.
- Miller, S., Johnson, N., and Wherry, L. R. (2021).
Medicaid and Mortality: New Evidence From Linked Survey and Administrative Data*.
The Quarterly Journal of Economics, 136(3):1783–1829.
- Miranda-Agrippino, S. and Rey, H. (2020).
U.S. Monetary Policy and the Global Financial Cycle.

The Review of Economic Studies, 87(6):2754–2776.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., and Gebru, T. (2019).

Model Cards for Model Reporting.

In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, pages 220–229, New York, NY, USA. Association for Computing Machinery.

Moschella, M. (2024).

Unexpected Revolutionaries: How Central Banks Made and Unmade Economic Orthodoxy.

Cornell University Press.

Moschella, M. and Diodati, N. (2020).

Does politics drive conflict in central banks' committees? Lifting the veil on the European Central Bank consensus.

European Union Politics.

Moschella, M. and Pinto, L. (2019).

Central banks' communication as reputation management: How the Fed talks under uncertainty.

Public Administration, 97(3):513–529.

Moschella, M., Pinto, L., and Diodati, N. (2020).

Let's speak more? How the ECB responds to public contestation.

Journal of European Public Policy, 27:400–418.

Müller, M. and Braun, C. (2021).

Guiding or Following the Crowd? Strategic Communication as Reputational and Regulatory Strategy.

- Journal of Public Administration Research and Theory*, 31(4):670–686.
- Neuenkirch, M. (2013).
Central Bank Transparency and Financial Market Expectations: The Case of Emerging Markets.
- Neundorf, A. and Adams, J. (2018).
The Micro-Foundations of Party Competition and Issue Ownership: The Reciprocal Effects of Citizens' Issue Salience and Party Attachments.
British Journal of Political Science, 48(2):385–406.
- Oriola, H. (2023).
Political monetary cycles: An empirical study.
European Journal of Political Economy, 79:102437.
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., and Lowe, R. (2022).
Training language models to follow instructions with human feedback.
arXiv:2203.02155 [cs].
- O'Connor, C. and Joffe, H. (2020).
Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines.
International Journal of Qualitative Methods, 19:1609406919899220.
Publisher: SAGE Publications Inc.
- Parkin, M. and Bade, R. (1978).
Central Bank Laws and Monetary Policies: A Preliminary Investigation.

- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, (2011).
Scikit-learn: Machine Learning in Python.
Journal of Machine Learning Research, 12(85):2825–2830.
- Persson, T. and Tabellini, G. (2001).
Political institutions and policy outcomes: what are the stylized facts?
Publisher: CEPR discussion paper.
- Plosser, C. I. (2014).
Systematic Policy and Forward Guidance at the Fed.
Journal of Applied Corporate Finance, 26(2):15–18.
- Pond, A. (2018).
Financial Liberalization: Stable Autocracies and Constrained Democracies.
Comparative Political Studies, 51(1):105–135.
- Puppis, M., Maggetti, M., Gilardi, F., Biela, J., and Papadopoulos, Y. (2014).
The Political Communication of Independent Regulatory Agencies.
Swiss Political Science Review, 20(3):388–412.
- Qanas, J. and Sawyer, M. (2024).
'Independence' of Central Banks and the Political Economy of Monetary Policy.
Review of Political Economy, 36(2):565–580.
Publisher: Routledge _eprint: <https://doi.org/10.1080/09538259.2023.2189006>.
- Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., and Radev, D. R. (2010).
How to analyze political attention with minimal assumptions and costs.
American Journal of Political Science, 54(1):209–228.

- Publisher: Wiley Online Library.
- Rajan, R. G. (2010).
Fault Lines: How Hidden Fractures Still Threaten the World Economy.
Princeton University Press, rev - revised edition.
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., and Rand, D. G. (2014).
Structural Topic Models for Open-Ended Survey Responses.
American Journal of Political Science, 58(4):1064–1082.
- Rogoff, K. (1985).
The Optimal Degree of Commitment to an Intermediate Monetary Target.
Quarterly Journal of Economics, 100:1169–1189.
- Romelli, D. (2022).
The political economy of reforms in Central Bank design: evidence from a new dataset.
Economic Policy, 37(112):641–688.
- Romelli, D. (2024).
Trends in Central Bank Independence: A De-jure Perspective.
- Romer, C. D. and Romer, D. H. (2004).
A New Measure of Monetary Shocks: Derivation and Implications.
American Economic Review, 94(4):1055–1084.
- Roth, F., Gros, D., and Nowak-Lehmann D., F. (2022).
Crisis and Citizens' Trust in the European Central Bank: Panel Data Evidence for the Euro Area, 1999–2012.

- In Roth, F., editor, *Public Support for the Euro: Essays on Labor Productivity, Monetary Economics, and Political Economy, Vol. 2*, Contributions to Economics, pages 137–155. Springer International Publishing, Cham.
- Sabaté, M., Escario, R., and Gadea, M. D. (2015). Fighting fiscal dominance. The case of Spain, 1874-1998. *European Review of Economic History*, 19(1):23–43. Publisher: Oxford University Press.
- Sabaté, M., Gadea, M. D., and Escario, R. (2006). Does fiscal policy influence monetary policy? The case of Spain, 1874–1935. *Explorations in Economic History*, 43(2):309–331.
- Sargent, T. and Wallace, N. (1981). Some Unpleasant Monetarist Arithmetic. *Quarterly Review*, 5(3).
- Schelkle, W. (2023). Monetary re-insurance of fiscal states in Europe. *SM*, pages 29–52. Publisher: Societ editrice il Mulino Section: 1/2023.
- Schmidheiny, K. and Siegloch, S. (2023). On event studies and distributed-lags in two-way fixed effects models: Identification, equivalence, and generalization. *Journal of Applied Econometrics*, 38(5):695–713.
- Schnabel, I. (2020). The shadow of fiscal dominance: Misconceptions, perceptions and perspectives.
- Schneider, C. (2018).

- The Responsive Union: National Elections and European Governance*, volume 15.
Cambridge: Cambridge University Press.
- Publisher: Springer.
- Schonhardt-Bailey, C. (2013).
Deliberating American Monetary Policy.
MIT press.
- Schulz, D. F. (2017).
Too little, too late? : how central bankers' beliefs influence what they do.
Thesis, European University Institute.
- Shambaugh, G. E. and Shen, E. B. (2018).
A clear advantage: The benefits of transparency to crisis recovery.
European Journal of Political Economy, 55:391–416.
- Siderius, K. (2023).
An unexpected climate activist: central banks and the politics of the climate-neutral economy.
Journal of European Public Policy, 30(8):1588–1608.
- Smets, F. (2014).
Financial Stability and Monetary Policy: How Closely Interlinked?
35th issue (June 2014) of the International Journal of Central Banking.
- Soroka, S. and Wlezien, C. (2010).
Degrees of Democracy: Politics, Public Opinion and Policy.
Cambridge University Press.
- Soroka, S. N. and Wlezien, C. (2009).
Degrees of Democracy: Politics, Public Opinion, and Policy.

- Cambridge University Press, Cambridge.
- Soskice, D. (2007).
Macroeconomics and Varieties of Capitalism.
In Hancké, B., Rhodes, M., and Thatcher, M., editors, *Beyond Varieties of Capitalism: Conflict, Contradictions, and Complementarities in the European Economy*, page 0.
Oxford University Press.
- Stella, P. (2005).
Central Bank Financial Strength, Transparency, and Policy Credibility.
- Stiglitz, J. E. (2017).
Where Modern Macroeconomics Went Wrong.
- Stimson, J. (1991).
Public Opinion In America: Moods, Cycles, And Swings, Second Edition.
Westview Press.
- Strasser, G., Gaballo, G., Hoffmann, P., and Ehrmann, M. (2019).
Signalling a future path of interest rates: The international evidence on forward guidance.
- Strohecker, K. and George, L. (2024).
Nigeria targets diaspora cash, \$1 bln monthly remittances, cenbank chief says | Reuters.
- Strong, C. O. (2021).
Political influence, central bank independence and inflation in Africa: A comparative analysis.
European Journal of Political Economy, 69:102004.
- Sturm, J.-E., Bodea, C., De Haan, J., and Hicks, R. (2024).
Central bank independence, income inequality and poverty: What do the data say?

The Journal of Economic Inequality.

Sun, L. and Abraham, S. (2021).

Estimating dynamic treatment effects in event studies with heterogeneous treatment effects.

Journal of Econometrics, 225(2):175–199.

Swanson, E. T. (2021).

Measuring the effects of federal reserve forward guidance and asset purchases on financial markets.

Journal of Monetary Economics, 118:32–53.

Thatcher, M. and Sweet, A. (2010).

The Theory and Practice of Delegation to Non-Majoritarian Institutions.

West European Politics, 25.

Thatcher, M., Sweet, A. S., and Rangoni, B. (2023).

Reversing delegation? Politicization, de-delegation, and non-majoritarian institutions.

Governance, 36(1):5–22.

Publisher: John Wiley & Sons, Ltd.

Thiemann, M. (2019).

Is resilience enough? The macroprudential reform agenda and the lack of smoothing of the cycle.

Public Administration, 97(3):561–575.

Times, F. (2024).

Lebanon's former central banker detained on corruption allegations.

Tortola, P. D. (2020).

The Politicization of the European Central Bank: What Is It, and How to Study It?

- Journal of Common Market Studies*, 58(3):501–513.
- Tucker, P. (2018).
Unelected Power.
Princeton University Press.
- Törnberg, P. (2024).
Best Practices for Text Annotation with Large Language Models.
arXiv:2402.05129 [cs].
- Valencia, Fabian, L. L. (2018).
Systemic Banking Crises Revisited.
- van der Cruijsen, C. and Samarina, A. (2023).
Drivers of trust in the ECB during the pandemic.
Applied Economics, 55(13):1454–1476.
- van 't Klooster, J. (2022).
Technocratic Keynesianism: a paradigm shift without legislative change.
New Political Economy, 27(5):771–787.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017).
Attention Is All You Need.
- Vayid, I. (2013).
Central bank communications before, during and after the crisis: From open-market operations to open-mouth policy.
Working Paper 2013-41, Bank of Canada Working Paper.
- Veyrune, Kei Moriya, R. M. T. C. S. (2025).

From Text to Quantified Insights: A Large-Scale LLM Analysis of Central Bank Communication.

Wang, W., Wei, F., Dong, L., Bao, H., Yang, N., and Zhou, M. (2020).

MiniLM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., and Zhou, D. (2023).

Chain-of-Thought Prompting Elicits Reasoning in Large Language Models.

arXiv:2201.11903 [cs].

Widmann, T. and Wich, M. (2023).

Creating and Comparing Dictionary, Word Embedding, and Transformer-Based Models to Measure Discrete Emotions in German Political Text.

Political Analysis, 31(4):626–641.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q., and Rush, A. M. (2020).

HuggingFace’s Transformers: State-of-the-art Natural Language Processing.

arXiv:1910.03771 [cs].

Woodford, M. (2005).

Central Bank Communication and Policy Effectiveness.

Technical Report w11898, National Bureau of Economic Research, Cambridge, MA.

Xu, Paulo A. Medas, T. P. J. F.-Y. Y. K. G. (2017).

Fiscal Crises.