Online Appendix

This Online Appendix accompanies the article *"How Central Bank Independence Shapes Monetary Policy Communication: A Large Language Model Application"* published in the *European Journal of Political Economy*. It details the construction and preprocessing of the BIS speech corpus, the large-language-model classification strategy, the full codebook, robustness checks, and all supplementary empirical results referenced in the main text. The replication files are available on OSF at https://osf.io/gpk7r/

Table of Contents

A BIS dataset processing

A.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/llm` 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.

A.2 Metadata extraction

The BIS dataset as downloaded from the BIS website[1] contains the text of the speech, the date on which it was given, the author, and a a non-standardised description string which contains

[1] https://www.bis.org/cbspeeches/index.htm

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.[2]

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 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 H.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.[3] 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

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

[3]Detecting 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 of the suggestions of the two approaches. Likely a small number of duplicates remain in the dataset.

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 H.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 H.4. Summary statistics of the selected metadata are presented in Table A.1.

Table A.1: Speech Metadata

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

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.

B 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 H.1 for our exact prompt. The sentence before and after each sentence is added as context. To illustrate, Table A.2 provides examples of sentences from our validation set and their classification including an explanation.

Table A.2: Classification examples

Classification	Example
Monetary Dominance	"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). *Explanation:* 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.
Fiscal dominance	"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 spend 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.) *Explanation:* 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.
Financial dominance	"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) *Explanation:* 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.
Monetary-fiscal coordination	"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). *Explanation:* This sentence refers to the monetary-fiscal interactions which is a key policy in the monetary-fiscal coordination.

Table A.2: 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). *Explanation:* This sentence shows that the bank wants coordinate with market participants to ensure financial stability.

Note: The classifications of the entire dataset can be accessed at
https://centralbanktalk.eu/data-speech

C Extended dataset description

C.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 Carribbean 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.

C.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.[4] 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 D.5 we show that results are robust to dropping observations with incompletely observed CBI indicators.

Table A.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. 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.

[4] The 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.

Table A.3: Dataset coverage

			Number of events since 1985		Largest event since 1985	
Country	Speeches	CBI coverage	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
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
LKA	68	1953	2	0	0.25	2023

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

			Number of events since 1985		Largest event since 1985	
Country	Speeches	CBI coverage	Increases	Decreases	Magnitude	Year
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
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

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

Country	Speeches	CBI coverage	Number of events since 1985		Largest event since 1985	
			Increases	Decreases	Magnitude	Year
GMB	1	1971	2	0	0.14	2018
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

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.

C.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.

D Event Study Robustness Checks

D.1 Relaxing the parallel trends assumption

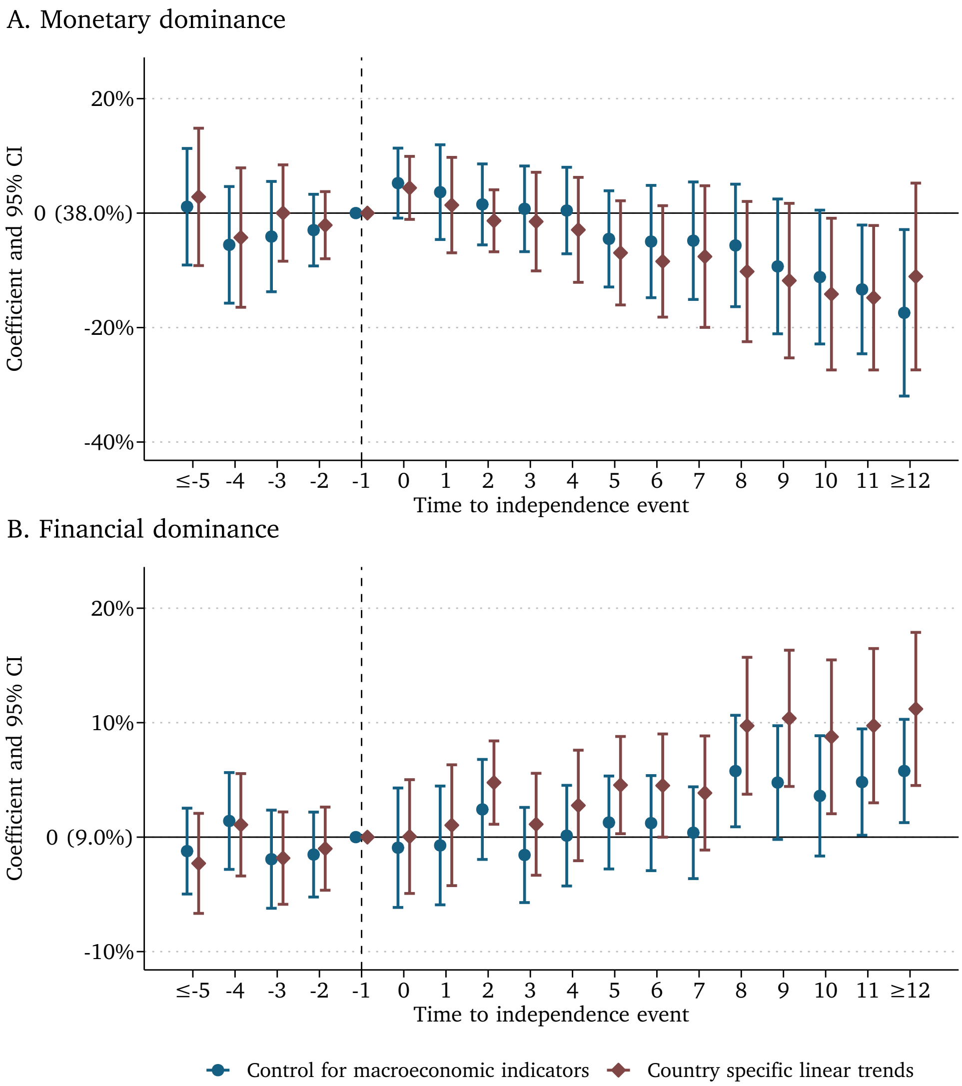
For our event study specification, to yield valid estimates of the causal effect of CBI on our communication measures, the treated and untreated units must follow parallel trends in the absence of treatment. While the generally small and statistically insignificant coefficient estimates on the pre-treatment 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:

$$\psi^m_{ict} = \sum_{k=-5}^{-2} \beta_k D^k_{ct} + \sum_{k=0}^{12} \beta_k D^k_{ct} + \mu_c + \theta_t + \mathbf{x}'_{ict}\boldsymbol{\gamma} + \epsilon_{ict} \tag{A.1}$$

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 A.1.

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



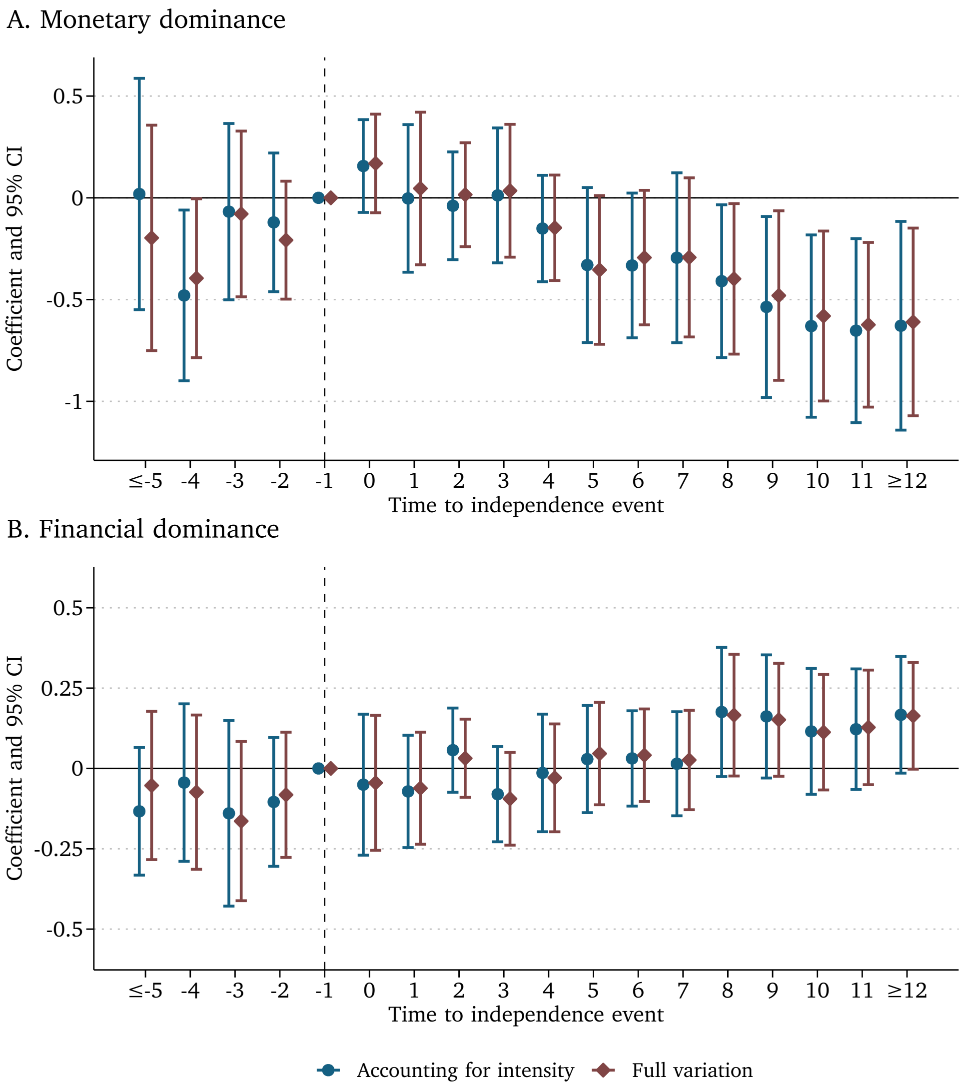
Note: The event-study plots show the beta coefficients as estimated by the two-way fixed effects model (equation (2) in the main text) with either macroeconomic controls or country specific linear trends added. Dynamic treatment effects are estimated relative to the year before the CBI change. The number 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 4.4 and the coefficient of changes in inflation on monetary dominance in our IV specification (equation (5) in the main text), 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.

D.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 4.1 and discussed in Schmidheiny and Siegloch (2023),

Figure A.2: Event study with continuous treatments



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

our main event study model with binned endpoints can handle multiple treatments of varying intensities. Figure A.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.

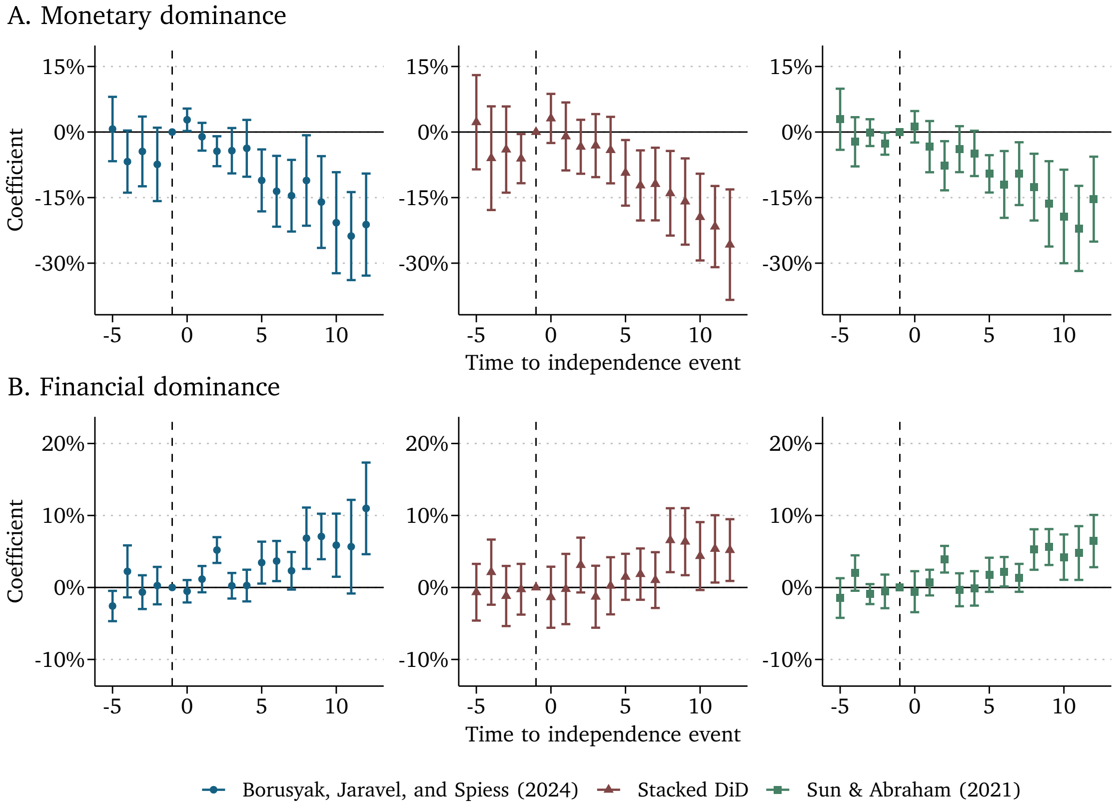
D.3 Heterogeneity robust estimators

Î̂n the main text, we show estimates of our event 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 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 A.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 A.4 reports aggregated estimates. The table also includes Callaway and Sant'Anna (2021)'s estimate.[5]

[5] The 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 A.3: Alternative event study estimators



Note: The columns show heterogeneity robust event study estimators which correspond to same setup as in the TWFE model (equation (2) in the main text). The vertical bars represent the 95% confidence intervals for the estimated coefficients.

D.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 A.4 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 D.3 for more details).

Table A.4: Single coefficient estimates

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

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 of the main text.

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 heterogeneous treatment effects (Goodman-Bacon, 2021).

D.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 C) 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.

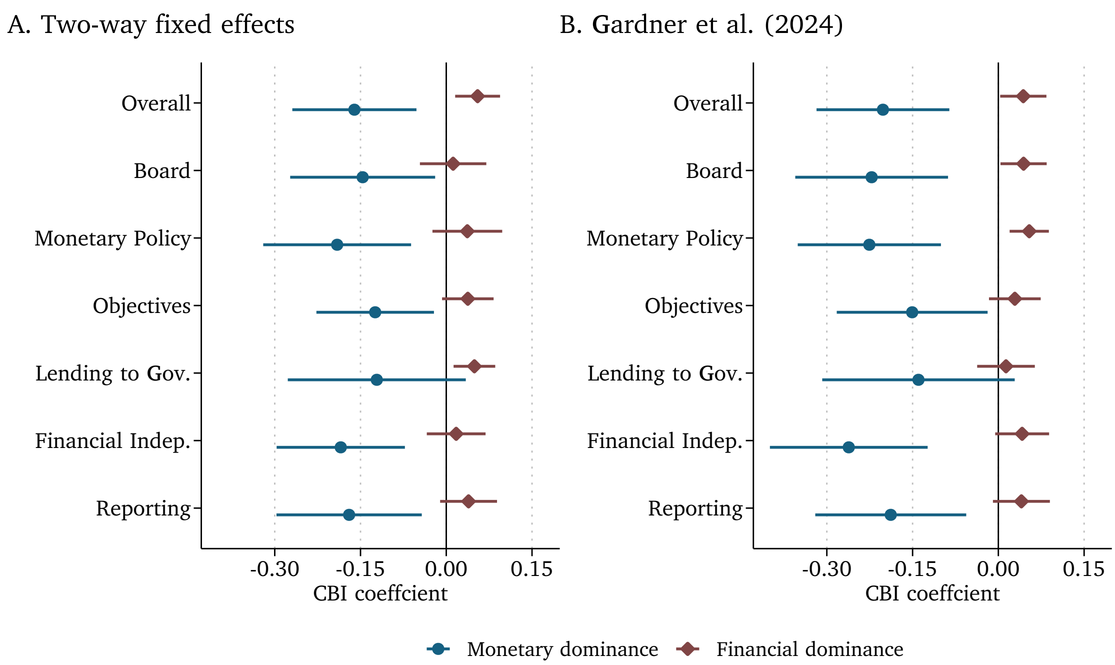
Second, we account for anticipation of the independence changes and alternatively estimate a specification where all independence changes are shifted to occur one year before the independence change is recorded in Romelli (2024). In the event study plots, effects are, therefore, reported relative to 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 A.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 9). Relying only on never treated units confirms the findings from the other specifications regarding monetary and financial dominance. 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 A.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 A.4: Effect of CBI sub-indicator changes on dominances



Note: Panel A plots the 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.

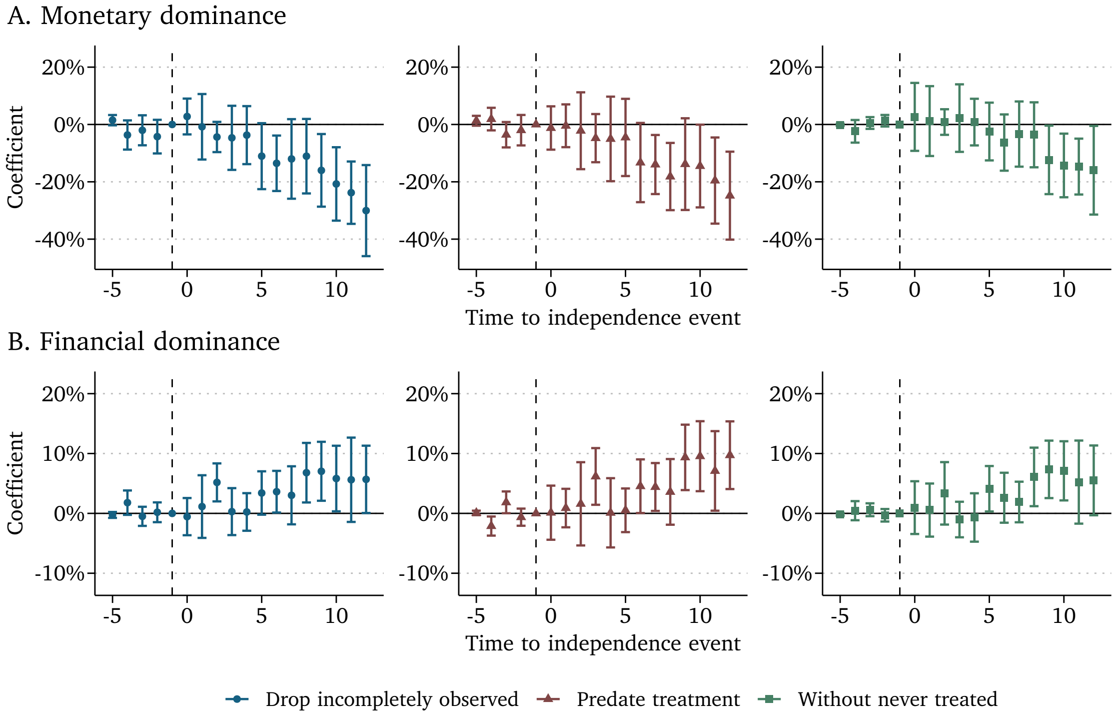
We find similar coefficient estimates across most dimensions of CBI, explained by a relatively high correlation among the CBI sub-indicators. Based on our definition of an independence event using the overall CBI indicator, we observe that typically (in the median case) an independence event entails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

D.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 pressures. In section 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 the BIS dataset as one of 'Academic', 'General Central Banking', 'Financial Market' and 'Political' (See section A.2). Below, in Figure A.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 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 ob-

Figure A.5: Robustness sample construction



Note: The panels show the three variations in the estimation sample discussed in Appendix D.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.

served effects are explained by different audiences, and in particular not by an increase in speeches directed at financial markets.

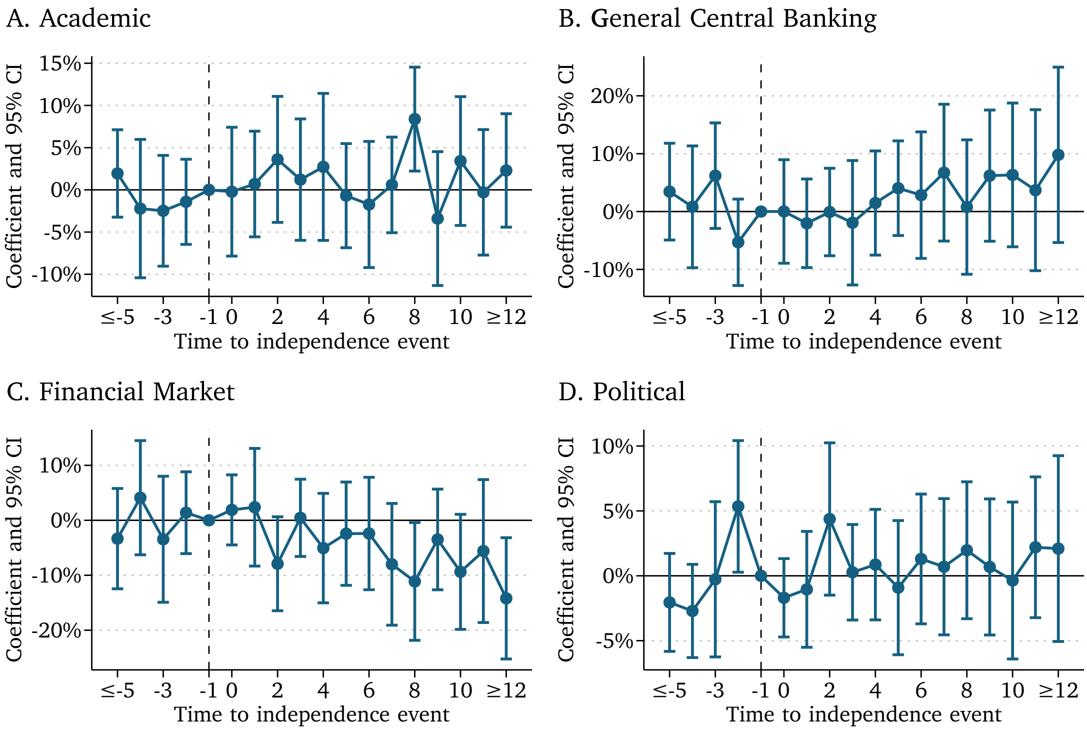
In addition, we estimate heterogenous effects by audience (as we do in the main text with equation (4)). Results are reported in Table A.5. Point estimates have the same signs and are of a similar magnitude throughout. With the exception of political audiences, we also find significant effects in all 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 our analysis.

D.7 Alternative CBI dataset

In this section, we repeat our estimation using the CBI dataset of Garriga (2025). 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 (2025), 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 (2025) covering 14 more

Figure A.6: The effect of CBI on audiences



Note: The panels show our event study design (equation (2) in the main text) 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.

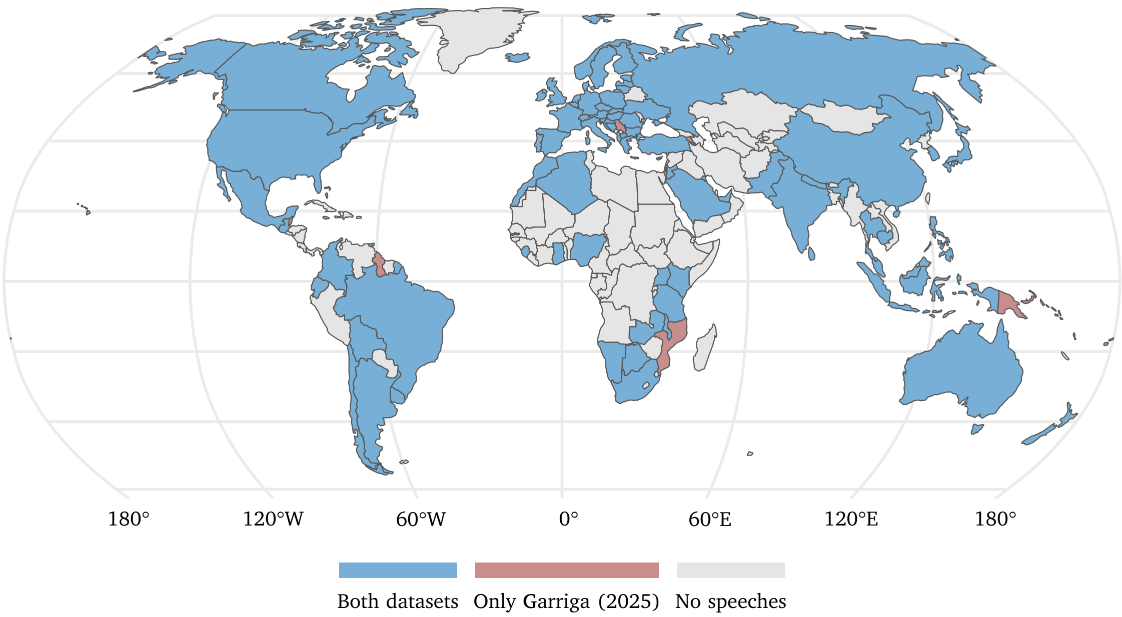
countries from our speeches dataset that together account for 557 speeches (see note below Figure A.7). Only one central bank, Macao, is tracked by Romelli (2022) but missing in Garriga (2025), with a total of 29 speeches.

Table A.5: Effect of CBI by audience

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

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 (4)

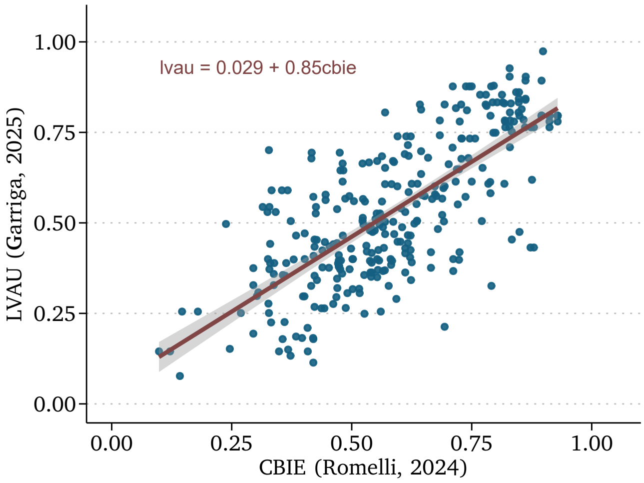
Figure A.7: Dataset coverage and number of independence changes



Note: This map compares the coverage of the dataset used for our empirical analysis (Romelli, 2024) and Garriga (2025). Countries shown in red are only available in Garriga (2025). 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 Fijii, 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 A.8.

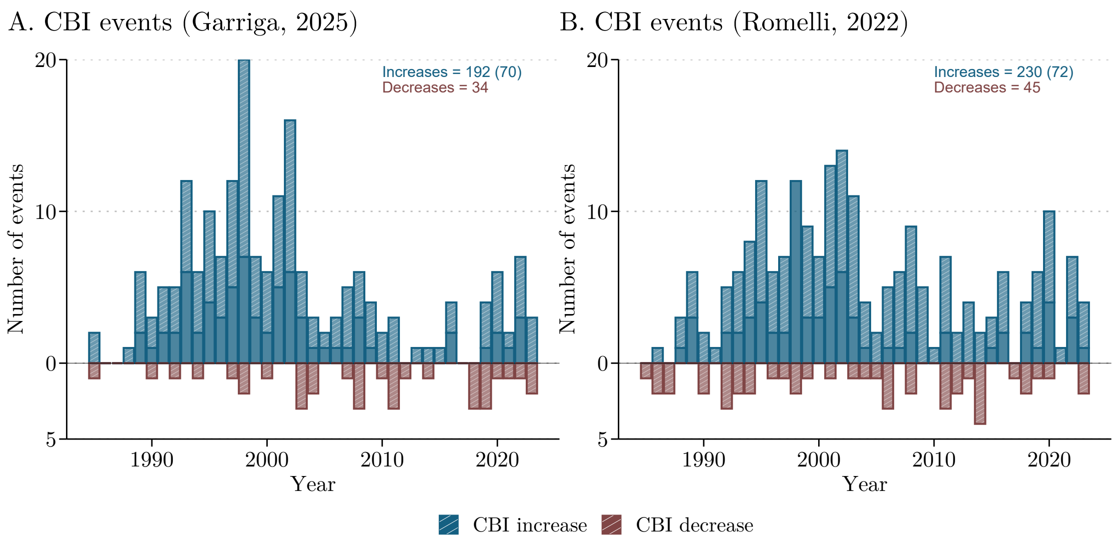
Figure A.8: Correlation of CBI scores



Note: The plot shows all unique CBI × country × 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 4.1 of the main text to construct independence events from Garriga (2025) 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 A.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 A.9: Comparison of CBI events



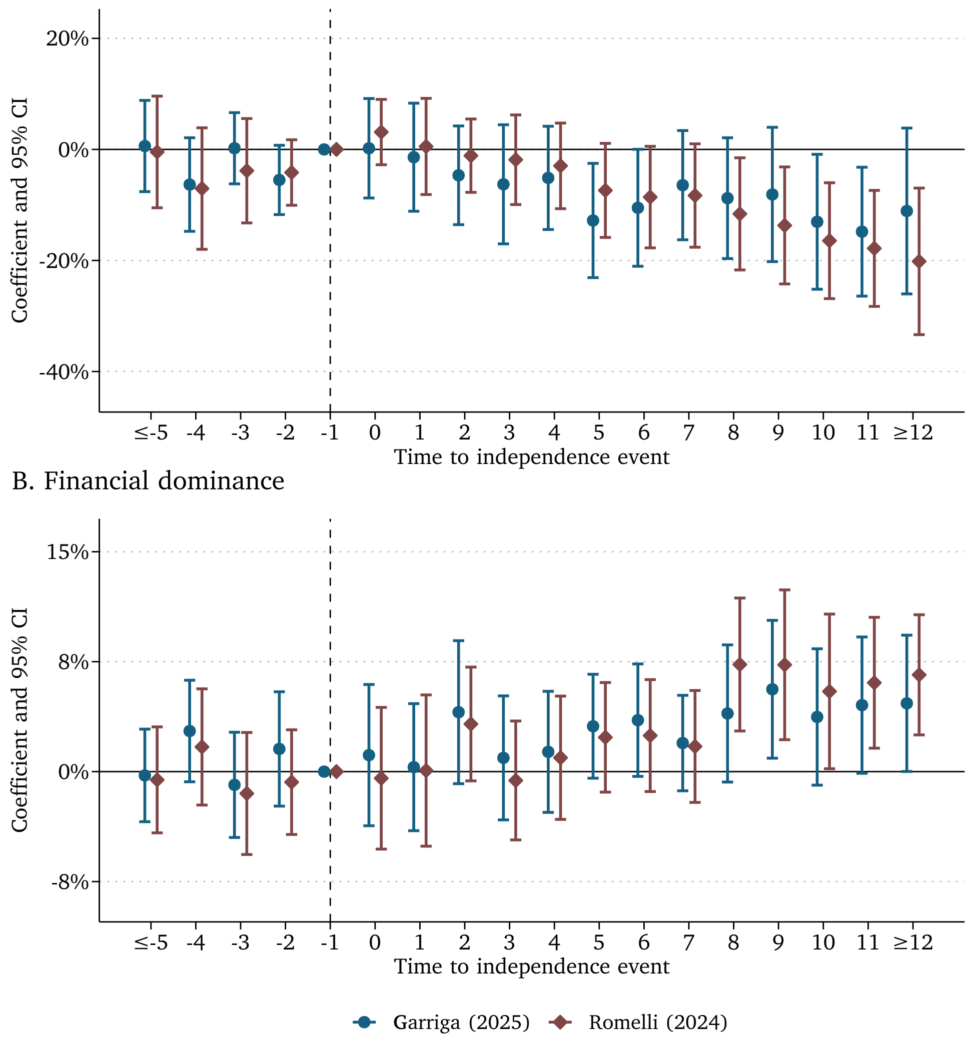
Note: Both panels display the number of CBI events over time, categorised into increases (blue bars) and decreases (red bars). Events excluded from the event study design are shaded. In the top right, the total number of events is shown with the number of relevant increases for the event study in brackets.

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 (2025), 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 (equation (2) in the main text) using both datasets. The results are shown in Figure A.10. We observe largely similar estimates both in terms of pre-trends and the effects of independence. Using Garriga (2025)'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. $H_0 : \beta_0 = \beta_1 = ... = \beta_{12} = 0$, for both financial ($p = 0.0325$) and monetary dominance ($p = 0.0151$). Our main results are therefore robust to the alternative CBI indicator.

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

A. Monetary dominance

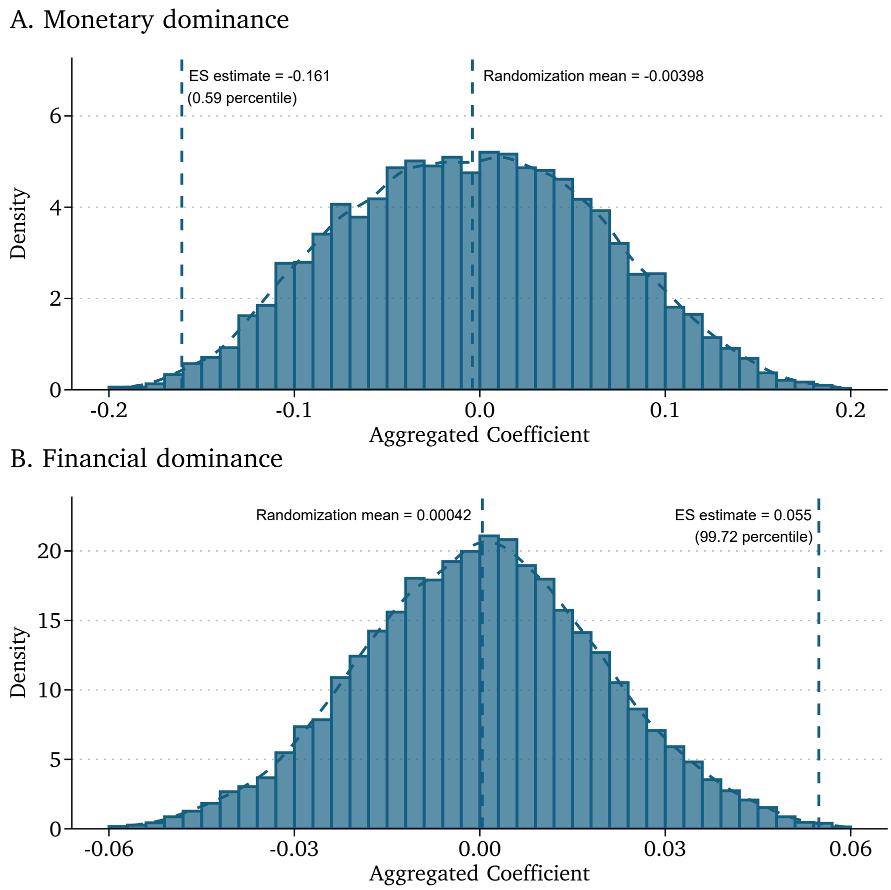


Note: The event-study plots show the beta coefficients as estimated by the two-way fixed effects model (equation (2) in the main text) for the datasets of Garriga (2025) and Romelli (2024). Dynamic treatment effects are estimated relative to the year before the CBI change. The vertical bars represent the 95% confidence intervals for the estimated coefficients.

D.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 a 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 A.11 shows the distribution of the aggregated event-study coefficient with 10,000 replications.

Figure A.11: Treatment randomisation placebo (aggregated)

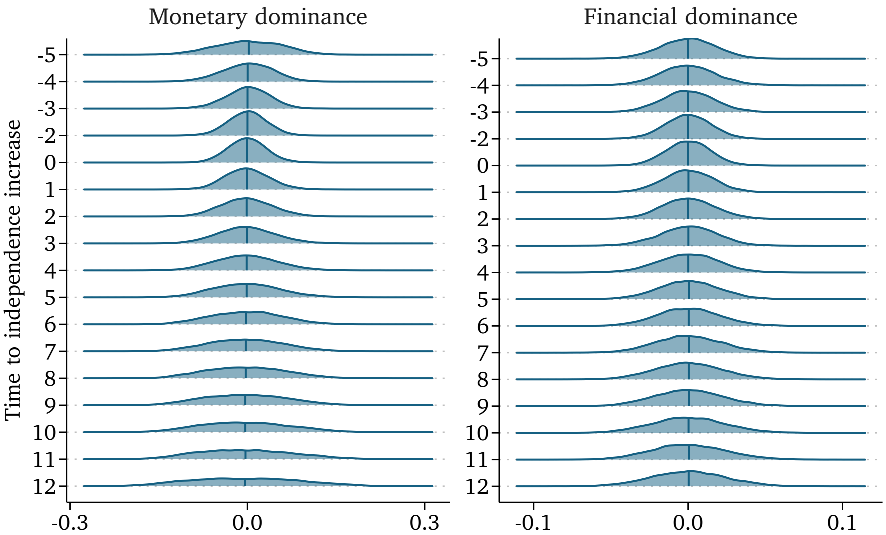


Note: The three panels illustrate the aggregated coefficients from our event study specification, based on 10,000 randomisations of the treatment countries and years. Vertical lines mark the means of the randomisation exercise and our estimated coefficients from the baseline specification

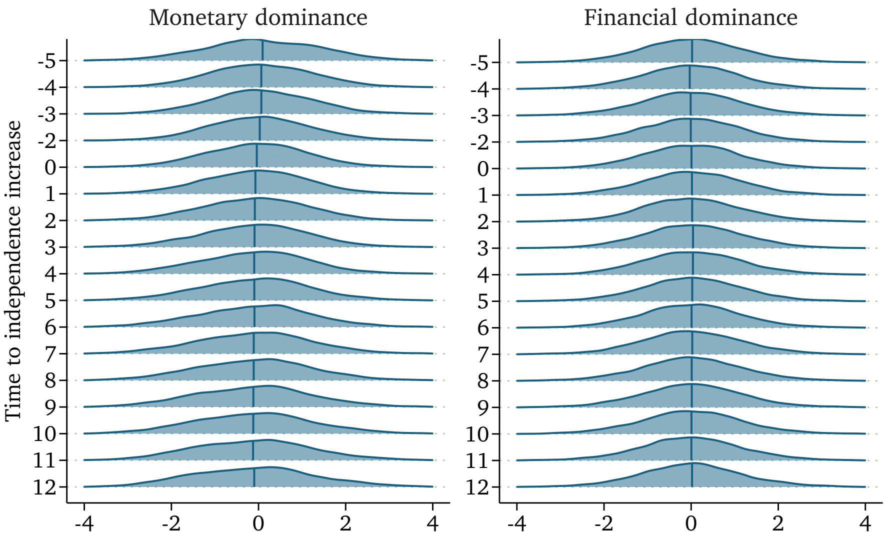
The distribution's mean is practically zero in line with the zero effect expectation of the placebo test. Our estimated coefficient fall into the the 0.59% and 99.72 % percentiles of the placebo distributions of monetary, financial and fiscal dominance. Further, Chart A.12 plots the distributions of the coefficient estimate and the t-statistic for the dynamic effects estimated with our main two-way fixed effects model (equation (2) in the main text). 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. Coefficients towards the edges of the event study window display a higher variance, yet the means are narrowly spread around zero.

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

A. Event study coefficient



B. t-statistic



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.

E 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 H.1. These categories correspond to the human-annotated sample of Leek et al. (2024), which consists of 1,000 sentences (see also Appendix B 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 A.6 for the options we consider.

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

Table A.6: Hyper-parameter tuning options

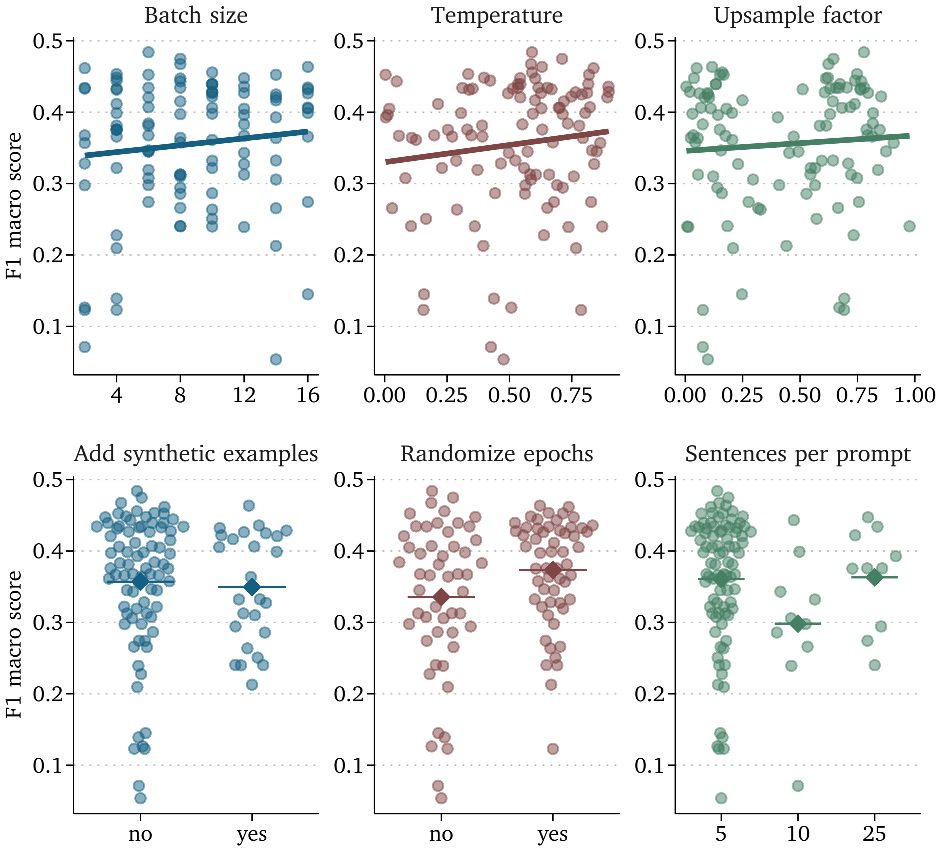
Parameter	Description	Possible values
Optimisation settings		
Epochs	Number of training cycles.	$[1, 10]$
Learning rate	Size of the steps taken in the model parameter space during optimisation.	$[0.0001, 0.01]$
Batch size	Number of training samples utilised in one iteration of model updating.	$\{2, 4, 6, \ldots, 16\}$
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.

of the trained models and their F1 macro score evaluated on the evaluation set are shown in Figure A.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 A.7 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 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 computational cots by substantially lowering the number of processed tokens. In Figure A.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.

Figure A.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.

F LLM validation

F.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 A.8 reports standard evaluation metrics for the Gemini and ChatGPT 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.

Table A.7: Final model parameters

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

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.

Table A.8: Validation metrics

	ChatGPT			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**

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.

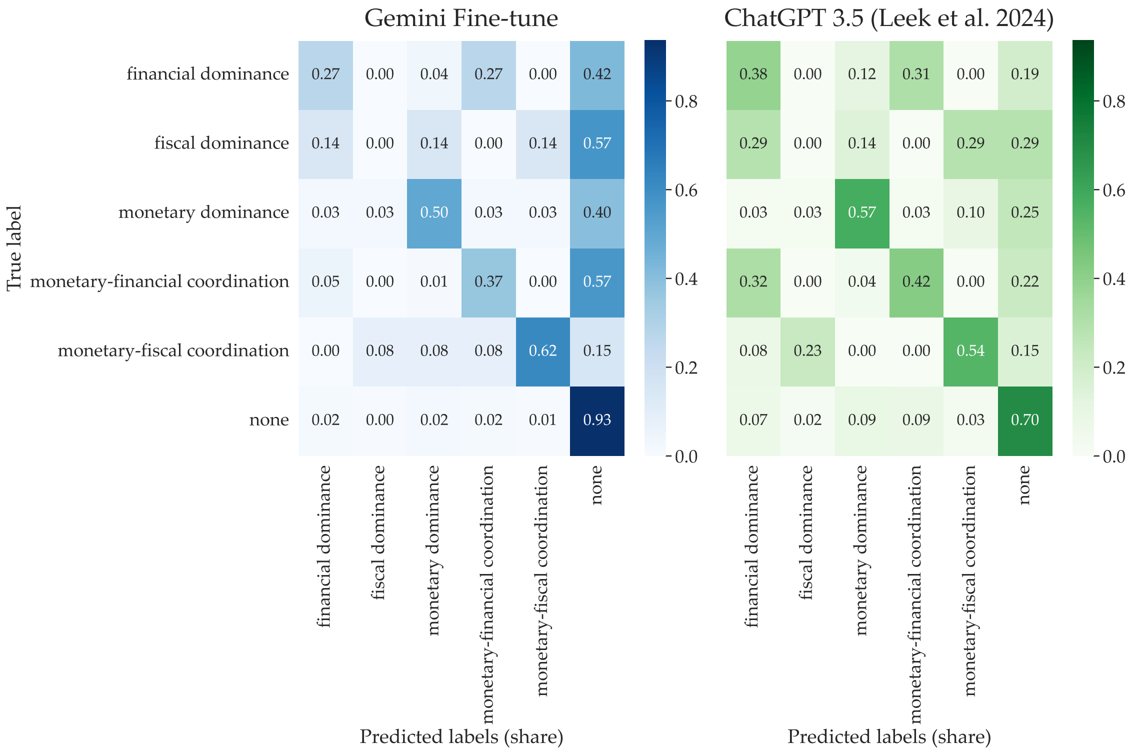
To further understand model model performance, we report the confusion matrices of our Gemini fine-tune and the zero-shot ChatGPT-3.5 used by Leek et al. (2024) (see Figure A.14).

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 A.8. Conversely, Gemini tends to be less sensitive in the categories of dominance, aligning more closely with human coders. Overall, Gemini provides better classifications with comparable ability to identify dominance and cooperation, but much fewer 'bad' mistakes, where a dominance or coordination category is erroneously assigned.

F.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

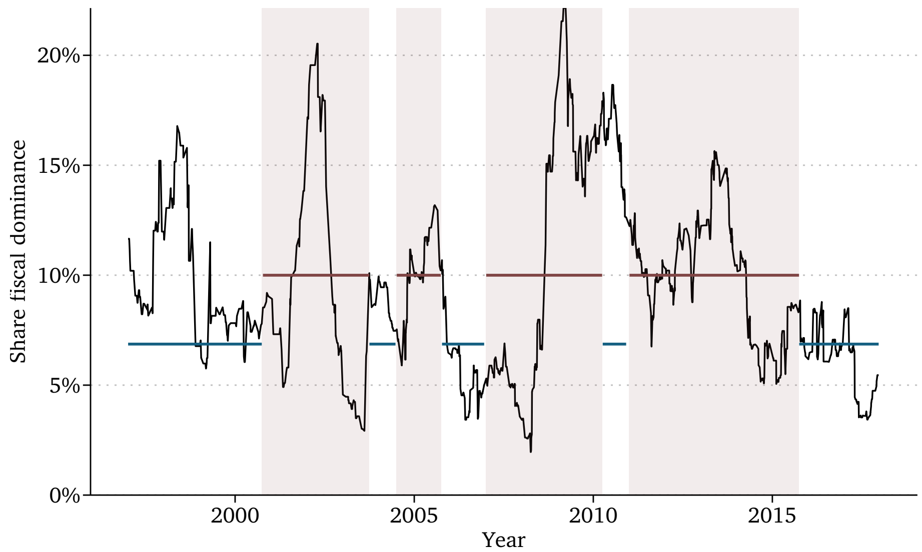
Figure A.14: Confusion Matrices of LLM classifiers



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 paper to classify our sample. On the right, the zero shot ChatGPT 3.5 model used by Leek et al. (2024) is shown.

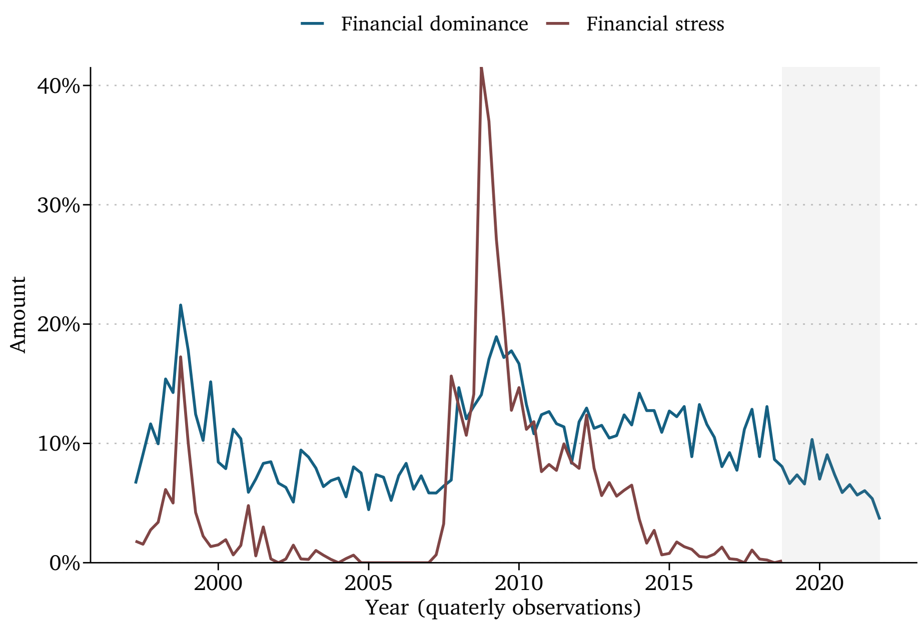
and a fiscal dominance regime, characterized through different paramerisations of the government spending and Taylor rules inside a DSGE model. In the second step, a tree-boosting machine learning classifier is trained to predict the regime based on the simulated macroeconomic data. Finally, the classifier is applied to predict the policy regime using actual macroeconomic data of the US from 1968 to 2017. Figure A.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 A.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 A.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 A.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 A.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.

G Instrumental variable design

G.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 $\widetilde{\psi}_{ict}^{m}$ of past speeches as the lagged dependent variable in the estimation equation (Equation (5) in the main text):

$$\psi_{ict}^{m} = \rho\,\widetilde{\psi}_{ict}^{m} + \beta_1\,\mathrm{CBI}_{ct} + \beta_2\,\Delta\pi_{ct} + \beta_3\,\Delta u_{ct} + \theta_t + \mu_c + \epsilon_{ict}, \tag{A.2}$$

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 A.9 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.

G.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:

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

Table A.9: Lagged dependant variable definition

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

Note: The table shows the coefficient of $\widehat{\text{CBI}}_{ct}$ in the second stage of the IV panel model (A.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.

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:

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

κ_t and λ_c denote year and country fixed effects. $\widetilde{\psi}^m_{ict}$ is the lagged dependent 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.

Table A.10: Instrumental variable regressions with interaction term

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

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 (A.3) and (A.4). Standard errors are clustered on the country level.

H Prompts

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

H.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.

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:

```
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.

H.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 Goveror 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:

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>
```

H.3 Audience classification

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

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:

```json
[
  {
    "identifier": "b060714",
    "audience": "academic"
  },
  {
    "identifier": "a151202",
    "audience": "financial_market"
  },
  {
    "identifier": "a170831",
    "audience": "central_bank"
  },
  {
    "identifier": "e180108",
    "audience": "financial_market"
  },
  {
    "identifier": "a170418",
    "audience": "political"
  }
]
```

H.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.

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:


```json
[
  {
    "identifier": "a981012",
    "latitude": 50.8465,
    "longitude": 4.3517
  },
  {
    "identifier": "c090907",
    "latitude": -34.6037,
    "longitude": -58.3692
  },
  {
    "identifier": "c050609",
    "latitude": 50.1109,
    "longitude": 8.6821
  }
]
```


References

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.

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.

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. arXiv:1907.10902 [cs, stat].

Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B. (2011). Algorithms for Hyper-Parameter Optimization. In *Advances in Neural Information Processing Systems*, volume 24.

Bird, S., Klein, E., and Loper, E. (2009). *Natural Language Processing with Python*.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting Event Study Designs: Robust and Efficient Estimation. arXiv:2108.12419 [econ].

Callaway, B. and Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

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.

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., Sundtröm, A., Tzelgov, E., Uberti, L., Wang, Y.-t., Wig, T., and Ziblatt, D. (2024). 'V-Dem Codebook v11' Varieties of Democracy (V-Dem) Project.

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.

Gardner, J., Thakral, N., To, L. T., and Yap, L. (2024). Two-stage differences in differences. arXiv:2207.05943 [econ].

Garriga, A. C. (2025). Revisiting Central Bank Independence in the World: An Extended Dataset. *International Studies Quarterly*.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

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.

Hinterlang, N. and Hollmayr, J. (2022). Classification of monetary and fiscal dominance regimes using machine learning techniques. *Journal of Macroeconomics*, 74:103469.

IMF (2024). World Economic Outlook Database.

Leek, L., Bischl, S., and Freier, M. (2024). Introducing Textual Measures of Central Bank Policy-Linkages Using ChatGPT.

Masciandaro, D. and Romelli, D. (2018). Central bankers as supervisors: Do crises matter? *European Journal of Political Economy*, 52:120–140.

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.

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.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.