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

Lauren Leek & Simeon Bischl

November 25, 2024

Motivation

In a nutshell

- Central bank communication is a core monetary policy and accountability for central banks, yet what shapes it?
- What? We develop and test a theory of how central bank independence shapes communication regarding monetary policy. We argue that increases in CBI alter the pressures a central bank faces and amends the reputation costs of not addressing this resulting in increases in communication about nancial constraints on monetary policy.
- How? We manually validate and ne-tune a LLM to develop a novel dataset regarding constraints in the monetary policy. Which we use as DV in staggered DiD and 2SLS regressions.
- Contributions:
  - A theory on what shapes it
  - Applied LLM research building on previous work
  - Long time-scope, all CBs

A theory of central bank independence and communication

- Communication as (i) monetary policy tool using it to shape market expectations and enhance predictability to anticipate pressures by scal and nancial policies and (ii) enhance democratic accountability and reputation (Blinder (2022); Casiraghi (2022))
- Independence can alter this relationship with monetary policy, yet how?
- Delegation to a more independent institutions changes policy pressures and changes the reputation costs needed which results in unexpected outcomes in monetary policy communication

Policy Pressure and Reputation

Mechanism	E ect
Policy Pressure (1)	In ation pressure decreases (e.g., Garringa (2020))
Policy Pressure (2)	Financial pressure increases (e.g., Aklin (2021); Masciandaro (2018))
Reputation (1)	More freedom to talk beyond its core task
Reputation (2)	More pro-active rather than reactive communication

Combined expectations

Heterogeneous e ects

Work ow: data and measurement

Data and measurement

- We construct indices of monetary, scal and nancial dominance based on speeches (18827 of 99 CBs covering the period 1997 to 2023). Map of coverage Examples classi cation Index construction
- We use a LLM to add interpretation and mimic human coding (prompt based, more parameters and higher interpretability). LLM validation metrics LLM confusion matrices LLM netuning experiments Hyperparameter optimisation
- Besides manual (human) validation, we also o er external validation with mainstream economic models. External validation
- Other data sources: VDem, IMF, Romelli's CBI indicator, scal and nancial crisis indicators Map of treatment Event distribution

Indicator over time

Note: Lines indicate a 365 day moving average (symmetric window) of the relative shares of dominance and coordination in the speeches given in the time window.

Empirical strategy: staggered DiD

- Main staggered two-way event speci cation:

$$Y_{ict} = \sum_{k=-5}^{k=-2} {}_{k}D_{ict}^{k} + \sum_{k=0}^{k=12} {}_{k}D_{ict}^{k} + {}_{c} + {}_{t} + x_{ict}^{0} + {}_{ict} \quad (1)$$

  - SE clustered at treatment level (country)
  - We bin the rst and the last event indicator
  - TWFE and Gardner (to deal with heterogeneous treatment e ects over time)

- Augmented for subgroups:

$$Y_{ict} = \sum_{k=-5}^{k=-2} {}_{k}D_{ct}^{k} + \sum_{k=0}^{k=12} {}_{k}D_{ct}^{k} + \sum_{j=2}^{j=J} \sum_{k=0}^{k=12} {}_{jk}D_{ct}^{k}S_{ct}^{j} + {}_{c} + {}_{t} + {}_{ict} \quad (2)$$

  - We aggregate the dynamic treatment e ects into a single average post treatment e ect by taking a unweighted and weighted average of all estimated lag coe cients.
  - SEs are calculated via the covariance estimates of the individual coe cients.

Results: main staggered DiD

Note: The event-study plots show the beta coe cients as estimated by the two-way xed e ects model (1). In addition, we report a equivalently speci ed heterogeneity robust estimators using the two-stage procedure of [1]

Results: heterogeneous e ects

	Financial Dominance		Fiscal Dominance		Monetary Dominance	
	Coe cient	Std. Error	Coe cient	Std. Error	Coe cient	Std. Error
Baseline						
Full sample	0.0368*	0.0194	0.0077	0.0106	-0.0943**	0.0403
Economic Development						
Non advanced	0.0073	0.0308	0.0095	0.0123	-0.0348*	0.0186
Advanced	0.0558***	0.0159	-0.0009	0.0138	-0.1456**	0.0595
Political System						
Autocracy	0.0311	0.0198	0.0069	0.0108	-0.0855**	0.0415
Democracy	0.0391**	0.0187	0.0074	0.0126	-0.0978**	0.0458
Supervision Capabilities						
Low	0.0386**	0.0185	0.0073	0.0100	-0.0857*	0.0512
Medium	0.0377*	0.0224	0.0046	0.0105	-0.0950***	0.0336
High	0.0078	0.0213	0.0057	0.0135	-0.1735**	0.0703

Note: Stars indicate signi cance levels: * p < 0:1; ** p < 0:05; *** p < 0:01. The table presents our main event study using the two-way xed e ects speci cation (1) strati ed by central bank and country characteristics. Heterogeneous e ects are estimated by interacting category dummies with the lags as in regression model (2). The shown coe cients are aggregations of the estimated dynamic treatment e ects for each subgroup. The rst row "baseline" reports the sample average e ect.

Empirical strategy: IV approach

Dominance	First Stage	2SLS E ect on Dominances		
Dependent Variables:	CBI	Monetary	Financial	Fiscal
Model:	(1)	(2)	(3)	(4)
Variables				
CBI		-2.344	0.4862	-0.0255
		(0.8162)	(0.1972)	(0.1021)
Inverse distance weighted World CBI $_{-1}$	0.4342			
	(0.1448)			
Neighbours Electoral Democracy Index $_{-1}$	0.8257			
	(0.3622)			
Neighbours Liberal Democracy Index $_{-1}$	-0.6284			
	(0.3092)			
Independence Judiciary	0.0027			
	(0.0385)			
In ation rate	0.0455	0.6852	-0.2894	-0.0675
	(0.0509)	(0.1742)	(0.0801)	(0.0396)
Unemployment rate	0.0008	-0.0176	0.0079	0.0020
	(0.0005)	(0.0084)	(0.0018)	(0.0011)
Other Covariates				
Country FE	X	X	X	X
Fit statistics				
Observations	12,271	12,271	12,271	12,271
R^2	0.97262	0.11170	0.06832	0.04233
Within R^2	0.15976	-0.04104	0.00159	0.00151

Clustered (Country FE) standard-errors in parentheses
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Note: In order to check whether the independence event is driven by endogenous factors (e.g., nancial pressure).

Mechanisms: policy pressure or accountability?

- Policy Pressure:in ation pressure is lower: less communication about in ation in the context of monetary dominance In ation pressure
- Policy Pressure: nancial pressure increases due to less supervision involvement Financial pressure
- Accountability: more freedom to talk beyond it's core task Freedom to talk
- Accountability: more pro-active rather than re-active communication Pro vs re-active

Alternative explanations

- Global increase in nancial dominance due to nancial crisis — Financial crisis event study
- Driven by euro area countries (due to SSM and large sample size) — Euro area dropped
- First time independence (epistemic community) — First independence event

Robustness checks

- Deviations from the parallel trends assumption — Linear trends and controls
- Allowing for multiple independence changes of varying intensities — Treatment variations
- Evaluating alternative heterogeneity-robust estimators — Estimation models DiD
- Testing the impact of sample variations — Sample variations
- Placebo randomization tests — Placebo randomization aggregated — Placebo randomization dynamic

Conclusion

- How does CBI shape monetary policy communication?
- When independence increases this results in a substitution e ect from a focus on central banks portraying themselves to be unconstrained in their monetary policy conduct to increased levels of nancial constraints due to changing policy pressures and changing reputational concerns.
- These e ects are robust and show heterogeneous e ects across di erent political regimes and economic development.
- Broader implications include that independence from scal authorities does not imply independence from nancial markets (also before the nancial crisis) and more independence does not imply that central banks are more focused on their narrow conduct of price stability through monetary policy.

Appendix

Speeches coverage

Main slide

Classi cation examples

Type	De nition	Example
Monetary Dominance	The central bank prioritizes to maintain price stability, and its monetary policy is not subordinated to scal policy or to nancial stability considerations.	"Furthermore, monetary policy implementation in line with the market e ciency principle would need to remain without prejudice to our primary mandate of safeguarding price stability." (Retrieved from: The European Central Bank, 14-06-2021)
Fiscal Dominance	The central bank accommodates its monetary policy to scal considerations, and its decisions are subordinated to meet the demands of scal policy.	"Moreover, although most of the resources administered by the BIS are invested in nancial 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 nancing markets in local currency." (Retrieved from the Central Bank of Argentina, 09-07-2008)
Financial Dominance	The central bank accommodates its monetary policy to - nancial considerations, and its decisions are subordinated to respond to the needs of nancial markets.	"It is thus signi cant that our exible and abundant provision of liquidity contained market participants' concerns over liquidity nancing." (Retrieved from the Bank of Japan, 04-07-2002)

Main slide

Index construction

- We calculate a relative share $_i^m$ 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.

$$_i^m = \frac{\sum_{j \in J} 1(\text{Classification}_j = m)}{\sum_{j \in J} 1(\text{Classification}_j \in M)} \tag{3}$$

with $M = \{$Monetary dominance ; Fiscal dominance ; Financial dominance ; Monetary-fiscal coordination ; Monetary-financial coordination$\}$ and $j \in J$ indexing all sentences that belong to speech i. The proportions add up to 1, i.e, $\sum_{m \in M} {}_i^m = 1 \ \forall i$

Main slide

Validation metrics

	ChatGPT			Gemini Pro 1.0		
	gpt-3.5	gpt-3.5- ne-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 [2] except the bold ne-tune column, which shows the validation metrics of our Gemini 1.0 Pro ne-tune based on 300 sentences. The validation scores are calculated on the holdout sample of 700 sentences. Given that precision is higher than recall, we are more restrictive in assigning categories.

Main slide

Confusion matrices

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 classi er. 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 [2] is shown.

LLM netuning experiments

Note: The six panels plot hyperparameter settings against our main validation metric, the F1 macro score calculated out of sample. In total, 108 successful ne-tunes were trained. Each dot corresponds to a ne tune. The scatter plots in the rst row are meant to illustrate the correlation between the hyperparameter 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. It is important to note that the hyperparameter settings for each run were not randomly sampled but selected by the Baysian optimization algorithm implemented in Optuna. If a hyperparameter is sampled more frequently, the algorithm predicts that this parameter leads to a better model.

Hyper-parameter optimisation

Parameter	Description	Possible values
Optimization settings		
Epochs	Number of training cycles.	[1; 10]
Learning rate	Size of the steps taken in the model parameter space during optimization.	[0:0001; 0:01]
Batch size	Number of training samples utilized in one iteration of model updating.	f 2; 4; 6; : : : ; 16g
Dataset composition		
Synthetic sentences	Add AI generated sentences to training sample.	f True, Falseg
Upsample factor	A factor governing the degree of upsampling, where a factor of 1 would result in a fully balanced training set.	[0; 1]
Randomize epochs	Re-randomize sentences included in prompts for each epoch.	f True, Falseg
Prompt engineering		
Sentences per prompt	The number of sentences to be included in one prompt	f 5; 10; 25g
Temperature	Parameter controlling the variation in generation output.	[0; 0:9]
Format instructions	Include instructions on output format.	f True, Falseg

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

Main slide

External validation

Note: The line shows a 200 day moving average of the relative shares of monetary vs. scal dominance for the US. Shaded regions indicate time periods which [3] identi ed as scal dominance using a machine learning classi er trained on synthetic data generated from a DSGE model. The blue horizontal line is the sample average of the US outside of the shaded regions. Red lines indicate the the average in each of the four scal dominance periods.

Main slide

Map of Independence Events

Main slide

Events distribution

Note: Panel A describes all changes in independence identi ed by [4] from 1990-2023 in the countries contained in the speeches dataset. Panel B is the subset which we use for our event study and di erence in di erences speci cations where events are restricted to one per country

Main slide

Mechanisms: in ation pressure

Note: Each dot represents binned averages of central bank year observations for which at least one speech is included in our sample. 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 nancial 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 t. Shaded areas in red and blue represent the 95% con dence of tted piece-wise polynomials of order two. In Panel A, the vertical line and the grey shaded area represent in ation rates below 2%, which is the common de nition of stable prices.

Mechanisms: nancial pressure

Note: Both panels present event study estimates using our main event study speci cation (1), but with the adjustment that the regressions are performed on the country-year level instead of on individual speeches. The in ation variable is winsorized at the 99th percentile to limit the in uence of extreme outliers. The grey shaded area indicates the sup-t con dence band which delimits the range of event-time paths of the e ect that would still be consistent with a zero e ect given the con dence level of 95%.

Main slide

Mechanisms: freedom to talk

Note: Average cosine distance and average number of topic clusters are rst calculated for each central bank based on Gemini embedding. The indicators are evaluated incorporating all speeches within a symmetric one year window. The blue and red lines refer to the average of high and low independence central banks using a cuto value of 0.8 in the [4] dataset. To limit the in uence of outliers with few speeches, central banks are weighted by their number of speeches.

Main slide

Mechanisms: proactive versus reactive

Note: Solid lines indicate a 365 day moving average (symmetric window) of the relative shares of communication targeted towards past, future and current topics (left axis) from 1997 to 2023. As reference, the development of average CBI, weighted by the number of speeches per country in our sample, is shown as a dashed line for the same time period (right axis).

Alternative: nancial crisis

Note: Coe cients presented are aggregations of cohort time-speci c e ects, using the approach of [5]. In a given year, the average treatment e ect on the treated (ATT) is calculated as the observations weighted average of all cohorts-year e ects 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% con dence intervals are indicated by the shaded area above and below the line.

Alternative: euro area and rst event

Note: The panels show event studies estimated using the two-way xed e ects equation (1). In the left column, we drop all the current euro area countries. In the middle panel, we add the ECB speeches to the control group. In the right panel, we modify the treatment indicator to turn on when independence changes for the rst 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.

Linear country-speci c trends & Covariates

Main slide

Alternative Event Study estimators

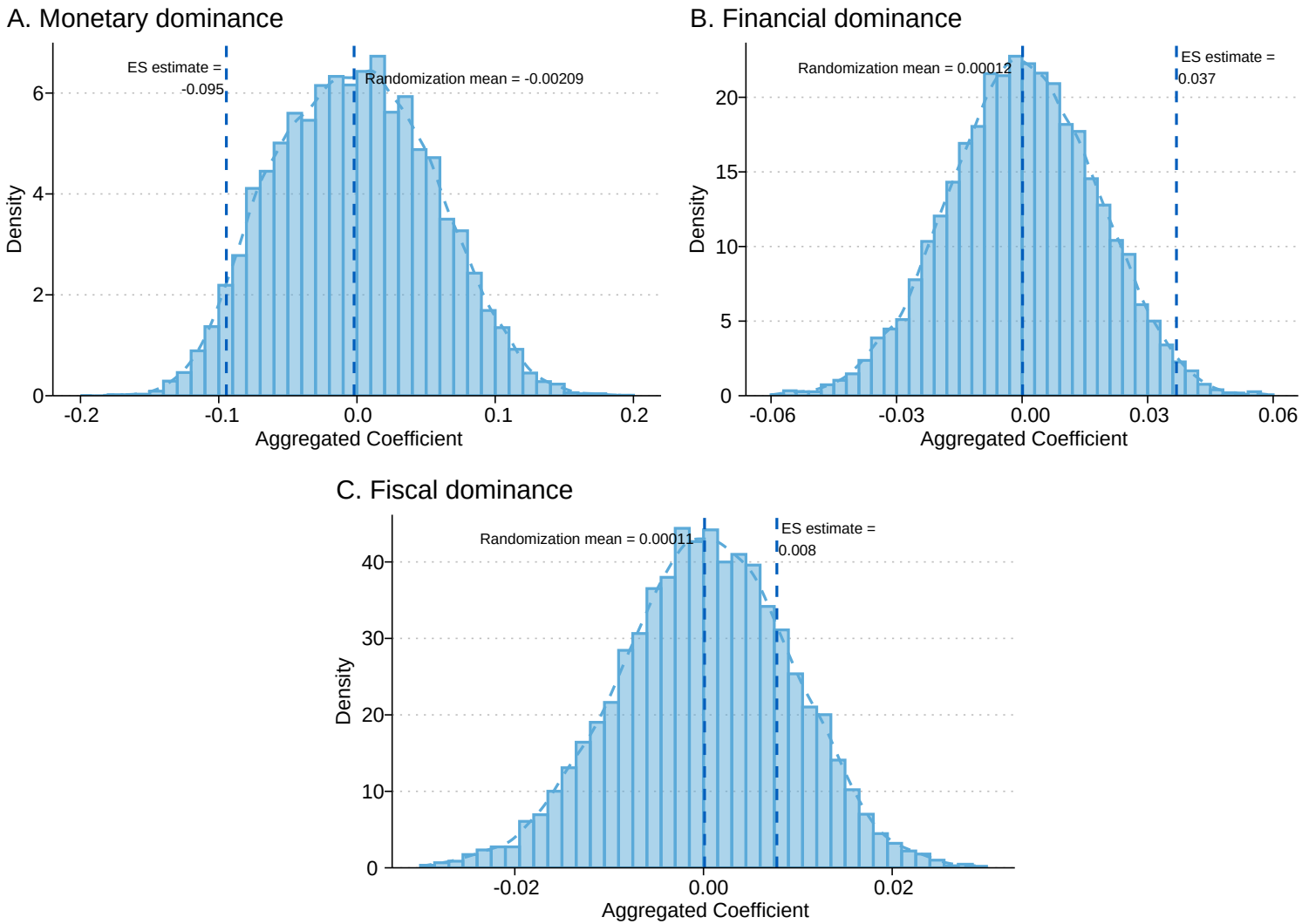
Continuous and multiple treatments

Main slide

Sample variations

Main slide

Placebo randomization aggregated



Note: The three panels illustrate the aggregated coefficients from our event study specification, based on 10,000 randomizations of the treatment countries and years.