INTERNATIONAL MONETARY FUND

AI Meets Fiscal Policy:

Mapping Government Spending Actions Across 64 Countries

Shuvam Das, Davide Furceri, Nikhil Patel, and Adrian Peralta-Alva

WP/26/43

2026
MAR

WORKING PAPER

 WP/26/43

IMF Working Paper
Fiscal Affairs Department and Research Department

AI Meets Fiscal Policy: Mapping Government Spending Actions Across 64 Countries
Prepared by Shuvam Das, Davide Furceri, Nikhil Patel, and Adrian Peralta-Alva*

Authorized for distribution by Emine Boz and Davide Furceri
March 2026

ABSTRACT: We build the first global quarterly narrative database of discretionary government spending actions by applying a fixed GPT–4.1 prompt to *Economist Intelligence Unit* (EIU) *Country Reports*. The resulting series identifies exogenous spending shocks—expansions and contractions—for an unbalanced panel of 64 countries from 1952:Q1 to 2023:Q4. We validate the database by (i) replicating expert narrative coding in Romer and Romer (2019), (ii) showing that the identified shocks predict subsequent movements in measured government spending, and (iii) establishing alignment with action-based consolidation measures in Adler et al. (2024). Using country-by-country proxy SVARs that treat the narrative indicator as an internal instrument, we estimate cumulative government spending multipliers. The median multiplier is about 0.7 at horizons up to two years, with substantial heterogeneity across countries and states. Multipliers are larger in countries that are less open to trade, under fixed exchangerate regimes, during downturns, and at the zero lower bound. Political conditions also matter: multipliers are smaller when broad economic policy uncertainty and fiscal policy-specific uncertainty is high, but weak political support is associated with larger conditional multipliers.

JEL Classification Numbers:	C45, C82, C32, E62, E32, H50
Keywords:	fiscal multipliers; government spending shocks; narrative identification; large language models; text-as-data; proxy SVAR; cross-country database.
Author's E-Mail Address:	sdas7@imf.org; dfurceri@imf.org; npatel@imf.org; aperalta-alva@imf.org

* All authors are with the International Monetary Fund. Corresponding author: Adrian Peralta-Alva. We thank Alan Auerbach, Christiane Baumeister, Emine Boz, James Cloyne, Giancarlo Corsetti, Domenico Giannone, Oscar Jorda, Ivan Petrella, Giorgio Primiceri and Sarah Zubairy for helpful comments, Alejandro Acevedo Guillot for excellent research assistance, and Sultan Orazbayev and Max Yarmolinsky for help with the LLM implementation. This paper is part of a Macroeconomic Research in Low-Income Countries project supported by the UK's Foreign Commonwealth and Development Office (FCDO). The views expressed here are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

AI Meets Fiscal Policy: Mapping Government Spending Actions Across 64 Countries

Shuvam Das Davide Furceri Nikhil Patel Adrian Peralta-Alva*

March 4, 2026

Abstract

We build the first global quarterly narrative database of discretionary government spending actions by applying a fixed GPT–4.1 prompt to *Economist Intelligence Unit* (EIU) *Country Reports*. The resulting unbalanced panel covers 64 countries from 1952–2023 and identifies exogenous spending shocks. We validate the measure by replicating expert narrative coding in Romer and Romer (2019), showing that the actions predict subsequent movements in measured government spending, and documenting close alignment with action-based consolidation series in Adler et al. (2024). Using country-specific VARs that treat the narrative indicator as an internal instrument, we derive the first set of comparable cumulative government spending multipliers. The median multiplier is 0.7 at horizons up to two years, with substantial heterogeneity across countries and over time. Pooled estimates imply larger multipliers in less open economies, under fixed exchange-rate regimes, and in downturns. Multipliers are smaller when uncertainty is high and larger when political support is stronger.

JEL Classification Numbers: C45, C82, C32, E62, E32, H50

Keywords: fiscal multipliers; government spending shocks; narrative identification; large language models; text-as-data; proxy SVAR; cross-country database.

---

*All authors are with the International Monetary Fund. Corresponding author: Adrian Peralta-Alva. We thank Alan Auerbach, Christiane Baumeister, Emine Boz, James Cloyne, Giancarlo Corsetti, Domenico Giannone, Oscar Jorda, Ivan Petrella, Giorgio Primiceri and Sarah Zubairy for helpful comments, Alejandro Acevedo Guillot for excellent research assistance, and Sultan Orazbayev and Max Yarmolinsky for help with the LLM implementation. This paper is part of a Macroeconomic Research in Low-Income Countries project supported by the UK's Foreign Commonwealth and Development Office (FCDO). The views expressed here are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

Contents

1 Introduction

What is the government spending multiplier for a given country? This question has gained significant policy relevance over the past two decades, as many countries implemented large stimulus packages to counter the effects of major recessions such as the Global Financial Crisis (GFC) and the COVID-19 pandemic. Despite extensive research on this topic, existing literature has provided answers primarily for specific, mostly advanced economies (notably the United States) or has offered average multiplier estimates for groups of countries.

The main reason for this important gap in the literature is that estimating government spending multipliers requires identifying exogenous measures of government spending, which is inherently challenging. A prominent approach to identifying such measures is the narrative method, pioneered by Romer and Romer (2010), which involves analyzing historical documents to classify fiscal actions based on their stated motivations. By distinguishing measures taken for long-run or inherited deficit reasons from those implemented in response to current economic conditions, the narrative approach isolates exogenous policy actions from endogenous responses. This methodology has been central to the modern literature on fiscal transmission, but it is labor-intensive.

Not surprisingly, datasets of exogenous fiscal measures identified using the narrative approach are only available for a few specific countries (e.g., Romer and Romer (2010) for the United States and (Cloyne et al., 2024) for the United Kingdom), or for a relatively small group of countries (e.g., Guajardo et al. (2014) for OECD economies; Adler et al. (2024) for a set of OECD economies and other countries). However, these cross-country datasets are only available at an annual frequency and cover a relatively short period (since the 1980s), making time-series analysis challenging.

To fill this gap, this paper builds a new dataset of narrative exogenous government spending shocks—both expansionary and contractionary—for an unbalanced panel of 64 countries on a quarterly basis starting from 1952Q1. Then, we use this data to estimate country specific government spending multipliers. Finally, we examine the relationship between the size of the multiplier and slow-moving factors (e.g., exchange rate regimes, trade openness), as well as higher-frequency factors such as uncertainty measures. To the best of our knowledge, this is the first effort to develop a historical dataset of exogenous government spending shocks at quarterly frequency for a large and diverse group of developed and developing countries and covering almost the entire period post-World War II.

To construct these shocks, we follow the narrative approach outlined in Romer and Romer (2023) and implement it using AI large language models (LLMs). Specifically, we employ a fixed GPT–4.1 prompt to identify fiscal actions based on the narrative real-time records of fiscal measures discussed in the Economist Intelligence Unit (EIU) reports, classifying them as endogenous or exogenous according to their stated motivations. To ensure consistency across countries and over time, we use a fixed prompt specification and coding rubric—without any task-specific training or fine-tuning. This procedure is implemented in a secure, non-adaptive environment: prompts are prespecified rather than tuned on outcomes, and the model is employed strictly off the shelf. For transparency and replication, we retain the verbatim excerpts that trigger each classification and release them alongside the coded series.

To address potential concerns about the accuracy, reliability, and consistency of our dataset, we evaluate our approach to identify spending measures using multiple methods. First, we assess whether our off-the-shelf large language model can accurately replicate expert human coding of fiscal stance and motivations. Specifically, we compare the AI-narrative approach with the fiscal stance and motivation of fiscal actions identified by Romer and Romer (2019), based on the narrative accounts of the EIU reports covering episodes of financial distress across OECD countries from 1980 to 2017. Our results show that the AI matches expert coding with an accuracy exceeding 93 percent. Second, we show that the identified government spending shocks are correlated with, and can predict, changes in actual government spending data. Third, we provide internal timing and orthogonality diagnostics: at quarterly frequency, the identified shocks behave like innovations to the fiscal stance rather than systematic responses to realized macroeconomic outcomes. Fourth, we compare our expected or identified shocks with those reported by Adler et al. (2024). We show that years in which Adler et al. (2024) report sizable expenditure-based consolidations almost always coincide with at least one consolidation quarter in our database, and rarely feature only expansionary actions. In addition, larger consolidation packages in Adler et al.

(2024) are associated with more consolidation-dominant years in our narrative series. We interpret these results as evidence that, despite differences in frequency, coverage, and coding design, both sources capture the same underlying phenomenon—discretionary shifts in the government's expenditure stance aimed at medium-term fiscal adjustment.

There are, however, important differences between our dataset and existing action-based cross-country measures. The quarterly frequency of our dataset is crucial not only to provide sufficiently long time series for estimating country-specific multipliers but also for causal identification: it reduces the scope for predictable within-year policy responses and avoids distortions from temporal aggregation. In addition, our dataset covers both expansionary and contractionary spending shocks, whereas the other cross-country action-based datasets focus on consolidations.

Using the resulting cross-country dataset of identified narrative spending actions, we estimate country-specific government spending multipliers at various horizons. For each country, we fit a small Bayesian VAR that includes a narrative spending proxy—taking values -1 for contraction, $+1$ for expansion, and 0 for neutral or no action—ordered first, followed by government spending and real GDP. Cumulative spending multipliers are computed as the ratio of the cumulative GDP response to the cumulative government spending response. A limitation of our narrative data is that, unlike Adler et al. (2024), we do not observe dollar-equivalent sizes of the fiscal measures, but only their qualitative direction. To address this, we adopt the internal-instrument approach of Plagborg-Møller and Wolf (2021) and Li et al. (2024). In this framework, the contemporaneous response of government spending pins down the scale of the structural innovation, rendering the qualitative nature of the proxy non-critical for identification.

For the median economy in the sample, the estimated multiplier is 0.74 at the 1-year horizon and 0.67 at the 2-year horizon, values that fall squarely within the range reported in the existing literature (Ramey, 2019). While the responses of spending and output differ across countries, median multipliers at the 1- and 2-year horizons are similar for advanced and developing economies. At the same time, heterogeneity within each group is substantial, with the inter-quartile range of multipliers spanning approximately 0 to 1.5.

Finally, we examine how multipliers vary with (i) slow-moving structural characteristics,

(ii) the sign of spending shocks, and (iii) time-varying macroeconomic and political conditions. The results corroborate well-established findings in the literature. First, along the canonical dimensions emphasized by Ilzetzki et al. (2013), multipliers are smaller in more open economies and larger under fixed exchange-rate arrangements. Second, multipliers are state dependent and vary with the business cycle: spending multipliers are larger when output growth is below its country-specific average and when monetary policy is constrained at the effective lower bound (e.g., Auerbach and Gorodnichenko, 2011; Ramey and Zubairy, 2018). The results also point to two additional findings. First, for both overall uncertainty and fiscal policy uncertainty (based on the World Uncertainty Index (WUI) and the fiscal policy-specific (FUI) component from Ahir et al. (2022)), global fiscal multipliers are systematically lower in high-uncertainty quarters at one- and two-year horizons. Second, conditional multipliers are larger when political support is favorable. The evidence indicates that the support dependence operates primarily through implementation: narrative spending shocks translate into a significantly stronger realized government-spending response when political support is high, with correspondingly larger output responses. Moreover, election proximity is associated with spending announcements that translate little into changes in actual spending.

This paper contributes to three main strands of the literature. First, it advances research on the construction of exogenous fiscal measures by providing the first global, quarterly narrative database of exogenous government-spending shocks covering a broad set of advanced and developing economies and spanning more than 70 years. Second, we extend the literature on the use of AI and language processing techniques to identify macroeconomic policy actions (see, for example, (Aruoba and Drechsel, 2026) for the United States (U.S.) monetary policy). Third, we contribute to research on fiscal spending multipliers by offering comparable country-specific multiplier estimates and new evidence on how business-cycle conditions, monetary-policy constraints, political support, and distinct notions of uncertainty shape fiscal transmission.

The remainder of the paper is organized as follows. Section 2 describes the narrative sources and the macroeconomic and fiscal data. Section 3 details the GPT–4.1 classification procedure, the construction of the quarterly shock series, and the validation exercises. Section 4 outlines the empirical framework used to estimate impulse responses and spending multipliers and presents benchmark results. Section 5 studies heterogeneity with respect to standard macroeconomic characteristics and the sign of shocks. Section 6 examines time-varying state dependence with respect to business-cycle conditions. Section 7 examines state-dependence with respect to alternative measures of economic and policy uncertainty. Section 8 studies how political support conditions fiscal transmission. Section 9 concludes.

2 Construction of the AI-narrative database

This section discusses the methodology used to construct the narrative database and how it adheres to the "requirements" outlined by Romer and Romer (2023): (i) a reliable narrative source; (ii) a clear idea of the information sought in the source; (iii) a dispassionate and consistent approach to the source; and (iv) careful documentation of the narrative evidence.

2.1 Data Source

Our primary narrative source is the *Economist Intelligence Unit* (EIU) *Country Reports.* The EIU produces standardized country reports for a broad set of advanced, emerging, and low-income economies, with coverage for many countries extending back to the 1950s. Reports are typically released quarterly; in later years some countries are covered at monthly frequency. For consistency, we construct a quarterly archive with one report per country–quarter. When multiple reports exist within a quarter, we retain the chronologically last comprehensive report issued within a given quarter[1] because overlapping coverage provides limited incremental information on fiscal policy actions and their stated motivations. The reports discuss political and macroeconomic developments and describe fiscal measures and the motivations emphasized at the time, which makes them well suited for narrative identification.

[1]We use the quarterly report corresponding to March, June, September, and December. Between 2000 and 2007, and for countries covered at monthly frequency, the EIU typically issued two types of publications within a quarter: a short “Updater” (brief update, typically single-digit pages) and a longer “Main Report” (comprehensive analysis, typically double-digit pages). In such cases, we retain the “Main Report” for that quarter, regardless of the specific month in which it appears. For example, if a quarter contains a July “Updater,” an August “Main Report,” and a September “Updater,” we select the August “Main Report.”

The EIU reports follow a centralized editorial workflow that promotes comparability across countries and time. Country analysts prepare an initial draft drawing on local sources. The draft is then reviewed by specialists and edited for clarity and internal consistency prior to publication. This standardization is valuable for our application because it facilitates systematic extraction of fiscal actions and stated motives in a comparable format across country–quarters.

Relative to OECD and many IMF country reports, the EIU reports provide (i) higher-frequency coverage for many economies and (ii) long historical time series for a broad cross section of countries, including emerging and low-income economies. These features are useful for empirical work at business-cycle frequency. In addition, the structured and relatively stable report format facilitates extracting narrative information on fiscal actions and the motives emphasized by contemporaneous sources, which is especially valuable in settings where alternative high-frequency narrative records are limited.

While the full EIU archive potentially covers more than 140 economies, our empirical analysis focuses on a subset of 64 advanced and developing economies for which we could also collect consistent quarterly macroeconomic and fiscal data (see Section 2.5 for a description of the sources).

2.2 Information to look for

Our goal is to identify "exogenous" discretionary government spending actions. Following Romer and Romer (2010), we treat a fiscal action as exogenous when the stated motive is unrelated to contemporaneous macroeconomic conditions. In their U.S. narrative study, Romer and Romer emphasize two motive categories that satisfy this criterion: (i) measures taken to address an inherited fiscal imbalance; and (ii) measures reflecting long–run objectives for the size or composition of the public sector (e.g., growth, fairness, institutional design). Both motivations are orthogonal to those related to stabilize short–run fluctuations and therefore appropriate to identify the causal effects of government spending on economic activity.

2.3 Dispassionate and consistent approach to the source

We use a fixed GPT–4.1 prompt to identify fiscal actions and classify them as endogenous or exogenous. To maintain consistency, the prompt specification and coding rubric remain invariant across countries and over time, with the model used off-the-shelf—without any task-specific training or fine-tuning. This approach prevents sample-specific tailoring and look-ahead bias, while enabling replication by other researchers (see next section on validity).

The prompt also records whether identified measures are implemented in the current quarter or announced for future quarters. We retain these fields for auditing and robustness checks. However, the baseline econometric series used in the VAR relies only on the discrete sign classification (expansion, contraction, or no action).

Below, we present the prompt, the coding criteria, and the methodology used to construct our shocks from the extracted and coded text.

Prompt.

> *Task. "Read the Economist Intelligence Unit country report excerpt for country C and quarter Q. Identify exogenous government spending actions in Q based on the stated motivations in the text, and summarize their net directional impact on spending."*
>
> *Definition of exogenous action. "Classify a spending action as exogenous only if its stated motivation is clearly unrelated to short-term macroeconomic developments. Typical non-cyclical motives include medium-term consolidation targets, long-run ideological or structural objectives, or compliance with a law, treaty, or supranational rule. If the text states that the action is undertaken to counter recession/slowing growth, reduce unemployment, cool overheating, or respond to contemporaneous inflation, interest-rate moves, exchange-rate pressure, or financing stress, treat the action as endogenous (do not include it among exogenous actions).*
>
> *Acknowledging current conditions does not by itself imply endogeneity if the stated motive is explicitly non-cyclical."*

Outputs. "Return:

1. *EXOG_NET_SPENDING ∈ {EXP_, CON_, NEUTRAL, UNCLEAR} indicating the net directional impact on spending of the identified exogenous actions in Q; optionally append small/large when the text provides a clear directional intensity cue.*
2. *MOTIVATION: brief phrase citing the stated motive(s) for the exogenous action(s).*
3. *COMPONENT (if specified): spending component(s) referenced in the text.*
4. *INTENSITY_ORDINAL ∈ {-10, . . . , 10} (optional): directional intensity (-10 very contractionary, +10 very expansionary).*
5. *CONFIDENCE (optional): brief self-assessment of classification confidence.*

Notes. Base the classification strictly on the provided text; do not infer unstated motives. If net direction is not supported by the text, return NEUTRAL or UNCLEAR."

Coding criteria. An action is coded as exogenous only when two conditions are simultaneously satisfied: (i) the stated motive is unrelated to near-term macroeconomic conditions; and (ii) the narrative does not cite contemporaneous growth, inflation, unemployment, interest-rate movements, exchange-rate pressure, or financing stress as the rationale. Actions that fail to meet either condition are classified as endogenous and excluded from the exogenous series. Notably, acknowledgment of prevailing economic conditions does not negate exogeneity when the motive is explicitly non-cyclical.

Construction of the shock. Report-level outputs are aggregated into a country–quarter proxy $z_{i,t}$ of exogenous spending actions:

$$z_{i,t} \in \{-1, 0, +1\}, \qquad z_{i,t} = \begin{cases} +1, & \text{if } \texttt{EXOG\_NET\_SPENDING} = \texttt{EXP\_}, \\ -1, & \text{if } \texttt{EXOG\_NET\_SPENDING} = \texttt{CON\_}, \\ 0, & \text{if } \texttt{EXOG\_NET\_SPENDING} \in \{\texttt{NEUTRAL}, \texttt{UNCLEAR}\}. \end{cases}$$

Thus $z_{i,t}$ represents the net directional impact of exogenous actions on spending for country i in quarter t. The mapping is conservative: whenever the stated motive is not explicitly non-cyclical—or when the report links the action to contemporaneous growth, unemployment, inflation, interest rates, exchange-rate pressure, or financing stress—the prompt classifies the action as non-exogenous, and it does not contribute to the net effect. If the direction cannot be determined from the text, the model returns UNCLEAR, which is mapped to $z_{i,t} = 0$.

Reports describe discretionary policy actions as they are announced, legislated, and implemented, but it is challenging to achieve clean separation between announcement ("news") and implementation timing (e.g., Mertens and Ravn (2013)). Indeed, our database does not perform an announcement/implementation split at the classification stage. Hence, we interpret $z_{i,t}$ as a qualitative proxy for discretionary spending actions as recorded in real time, which may reflect a mixture of announcement and implementation elements. In the econometric analysis, $z_{i,t}$ is used as an instrument for innovations to realized government spending. Accordingly, our estimated impulse responses and multipliers capture the average effects of the composite policy-action shock component that is correlated with this narrative proxy, rather than a pure "surprise" implementation shock. Disentangling anticipated and implemented components systematically in a broad cross-country setting is an important topic for future work.

2.4 Documentation: narrative evidence

In this section, we present the narrative evidence. First, we provide detailed examples of the identified measures and explain how these episodes are classified based on the narrative evidence. Then, we highlight key stylized facts regarding the distribution of measures by type (expansions versus contractions), by country groups, and their evolution over time.

Examples of Narrative Episodes.

Example 1 (long-run ideas; expansion). Ghana, 2010Q3 Action: Parliament passed a controversial US$10bn housing deal with a South Korean construction company, STX Korea, in early August. Stance: Expansion; Motivation: stated aim is to expand housing supply and infrastructure capacity. Timing: implemented. Confidence: 95 percent.

Example 2 (inherited deficit; contraction). Ecuador, 1982Q4 Action: The authorities remove petrol and imported-wheat subsidies as part of a fiscal consolidation; the motive is to contain budgetary pressures and improve sustainability. This is an exogenous contraction (subsidies/transfers). Stance: Contraction; Motivation: Stated aim is to contain budgetary pressures. Timing: implemented. Confidence: 90 percent.

Example 3 (no action). United Kingdom, 2005Q4 Action: The report discusses technical adjustments and postponements made in the previous quarters, and future intentions. However, there are no explicit spending changes or new actions identified by the report. The AI assigns the exogenous effect on government spending to be unclear.

Table 1 compiles these cases along with additional concise episodes that map the two motive families across instruments and countries. Many more examples are available in the appendix, where we provide a 5 percent random sample of the complete database, including brief summaries of the narrative information.

Table 1: Selected Exogenous Spending Actions

Country–Quarter	**Sign**	**EIU cue (concise)**
Bangladesh 1980Q2	+1	"five-year plan envisages Tk 300–500 bn in public programs"
Honduras 1978Q1	+1	"US$450 mn over five years in electric-power projects"
Côte d'Ivoire 1970Q2	+1	"special capital budget; Canadian-supported electrification"
Ghana 2010Q3	+1	"Parliament approves large public-housing program (∼US$10 bn)"
Indonesia 1985Q1	−1	"real development spending cut by about 7%"
Ecuador 1982Q4	−1	"removal of petrol and imported-wheat subsidies"

Notes: Episodes are coded from *EIU Country Reports* using the fixed, pre-specified GPT 4.1 prompt described in Section 2.3. "Sign" is the net directional effect on spending in the report quarter (+1 expansion, −1 contraction). The "EIU cue" reproduces a short fragment from the relevant report indicating the stated motive/measure.

Stylized facts. Out of 16,029 country-quarter observations processed, we identify 8,636 discretionary spending episodes, corresponding to 53.9% of the sample. Among these episodes, 4,374 are classified as expansions and 4,262 as contractions, implying that expansions account for 50.6% of all non-zero observations.

Table 2: Coverage of Narrative Data at a Glance (nonzero quarters only)

Countries in universe	64
Countries with nonzero quarters (this listing)	64
Sample period (unbalanced; country start dates vary)	1952:Q1–2023:Q4
Nonzero quarters ($z_{i,t} = \pm 1$)	8,636
Expansions / Contractions (%)	50.6 / 49.4

Notes: "Nonzero quarters" are country–quarters in which the narrative spending–action proxy takes values in $\{-1, +1\}$. A value of +1 denotes an exogenous net *expansion* in discretionary government spending and −1 denotes an exogenous net *contraction*, as coded from *EIU Country Reports* using the fixed GPT–4.1 prompt and exogeneity screen described in Section 2.3. Quarters classified as NEUTRAL or UNCLEAR are mapped to zero and are excluded from the nonzero count. The panel is unbalanced; "1952:Q1–2023:Q4" denotes the union of available country–quarter observations across countries.

Coverage. Table 2 provides an overview of the database's coverage. Three key features stand out from Table 2. First, the database spans long time periods and offers broad cross-sectional coverage. Second, the incidence of expansions and contractions among nonzero quarters is nearly balanced (50.6% vs. 49.4%), which is useful for identification strategies that rely on sign variation rather than magnitudes. Third, low-income countries (LICs) account for 1,920 of the 8,636 nonzero quarters (22.2%), highlighting the added value of extending narrative coverage beyond advanced and emerging economies.

Cross-country heterogeneity Table 3 reports the counts of nonzero observations and the share of expansions by IMF income group. The data reveal significant variation in the share of expansions and contractions across income groups. Advanced economies exhibit a higher proportion of spending contractions (54.2%) compared to expansions (45.8%), possibly reflecting the more widespread application of fiscal rules in these countries. In emerging markets, the share of expansion episodes (49.5%) is nearly equal to that of contractions. In low-income countries (LICs), the majority of measures (60.7%) are expansionary.

Table 3: Exogenous Spending Episodes by Country Group (nonzero quarters)

Group	Nonzero quarters	Expansion share (%)
Advanced	3,129	45.8
Emerging	3,587	49.5
LIC	1,920	60.7

Notes: Country groups (Advanced, Emerging, LIC) follow the IMF income classification used in WEO October 2025. "Nonzero quarters" are country–quarters in which the narrative spending–action proxy takes values in $\{-1, +1\}$. "Expansion share" is the fraction of expansionary quarters among nonzero quarters within the group.

Motivation behind spending actions To shed light on what governments are actually doing when they undertake exogenous fiscal actions, we analyze a random subsample of narrative episodes and classify the associated measures using a simple text-based dictionary. A clear asymmetry emerges. Expansionary actions are overwhelmingly associated with increases in public investment, particularly infrastructure and capital projects, which account for nearly 60 percent of expansion episodes. By contrast, contractionary actions are dominated by explicit consolidation measures—such as spending restraint and deficit reduction—which represent over two-thirds of contraction episodes. Other categories, including social spending, privatization, and external financing, play a secondary role and appear in both expansions and contractions. These patterns suggest that fiscal expansions and consolidations are not mirror images of each other: expansions are primarily investment-driven, while contractions largely reflect deliberate deficit-reduction efforts.

2.5 Quarterly macroeconomic and fiscal series

In addition to the narrative data, we assemble standard macroeconomic time series at a quarterly frequency for use in the VAR analysis. For each country, we obtain real GDP (to measure output responses) and real government expenditure from a mix of sources such as the EIU, Haver, CEIC, and central banks. Detailed information about data sources is provided in Table 4. We express variables in log levels for the VARs. The fiscal shock series constructed from the narrative data will be one of the variables in the VAR (entered as described in the empirical approach). Given the focus on identifying exogenous shocks, we will zero out the narrative fiscal impulse in quarters where the action was judged to be endogenous to current economic conditions (so that only exogenous shock values remain in the series). The final panel dataset thus combines: (i) the narrative fiscal impulse series (with exogenous shocks), and (ii) macroeconomic series for output and other variables, for a large sample of countries and time periods.

All results going forward are based on the intersection of the narrative series and available quarterly macro data.

Table 4: Data coverage and sources by countries

Country	First Quarter	Sources	Country	First Quarter	Sources
Argentina	2004 Q1	EIU	Indonesia	1999 Q2	EIU
Australia	1993 Q1	EIU	Ireland	2002 Q1	HA
Austria	2001 Q1	HA	Italy	1999 Q1	HA
Bangladesh	2018 Q1	EIU	Japan	1993 Q1	EIU
Belgium	1995 Q1	HA	Jamaica	2015 Q1	EIU
Bolivia	2002 Q1	EIU	Kenya	2009 Q1	EIU
Botswana	2007 Q2	CEIC	Korea	2000 Q1	EIU
Brazil	1997 Q1	EIU	Lithuania	1999 Q1	HA
Bulgaria	2003 Q1	EIU	Mexico	1993 Q1	EIU
Burundi	2005 Q1	HA	Moldova	2016 Q1	EIU
Canada	1993 Q1	EIU	Netherlands	1999 Q1	HA
Chile	2003 Q1	EIU	Nicaragua	2016 Q1	EIU
China	1998 Q1	EIU	Nigeria	2010 Q1	EIU
Colombia	2000 Q1	EIU	Norway	2002 Q1	EIU
Côte d'Ivoire	2015 Q1	EIU	Paraguay	2003 Q1	EIU
Costa Rica	2006 Q1	EIU	Peru	2006 Q1	EIU
Croatia	2000 Q1	EIU	Poland	1999 Q1	EIU
Czech Republic	1996 Q1	EIU	Portugal	1999 Q1	HA
Denmark	1999 Q1	EIU	Rwanda	2021 Q1	EIU
Dominican Republic	2000 Q1	EIU	Romania	2004 Q3	EIU
Ecuador	2012 Q1	EIU	Slovak Republic	1999 Q1	HA
Finland	1999 Q1	HA	Slovenia	1999 Q1	HA
France	1980 Q1	HA	South Africa	1990 Q1	HA, SARB
Georgia	2006 Q1	EIU	Spain	1995 Q1	HA
Germany	2002 Q1	HA	Sweden	1995 Q1	EIU
Ghana	2008 Q1	EIU	Switzerland	2007 Q1	EIU
Greece	1999 Q1	HA	Tanzania	2010 Q1	EIU
Guatemala	2001 Q1	EIU	United Kingdom	1997 Q2	EIU
Honduras	2014 Q1	EIU	United States	1952 Q1	BEA
Hungary	2004 Q1	EIU	Uruguay	1999 Q1	EIU
India	1997 Q2	HA	Vietnam	2008 Q1	EIU

Notes: HA = Haver Analytics; EIU = Economist Intelligence Unit; BEA = Bureau of Economic Analysis; SARB = South African Reserve Bank.

3 Validation

This section presents four complementary assessments of our AI-assisted narrative approach. The exercises are designed to speak separately to (i) the model's ability to interpret narrative sources in the manner of expert coders and (ii) empirical validation of the spending-action proxy.

First, we ask whether an off-the-shelf LLM can reproduce Romer and Romer (2019) (RR19) report-level classifications of fiscal stance and stated motivations. We consider this useful because it holds fixed the narrative source, the set of reports, and the coding rubric. It thus tests the ability of the LLM model to read EIU reports as well as extracting and evaluating fiscal narratives adequately. Second, we assess *replicability* by re-running the fixed prompt on the same inputs and quantifying the stability of the model's outputs across runs. Third, we evaluate whether the *production* proxy for exogenous discretionary spending actions (Section 2.3) maps into realized fiscal aggregates by testing whether movements in the proxy predict subsequent movements in measured government spending. Fourth, we provide an *external alignment* check by comparing our quarterly sign proxy to the action-based annual consolidation measures reported by Adler et al. (2024). We discuss each exercise in turn.

3.1 Benchmarking narrative interpretation: comparison to Romer and Romer (2019)

Romer and Romer (2019)—hereafter RR19—study fiscal policy responses in episodes of financial distress across 31 OECD countries over 1980–2017 using narrative evidence from *EIU Country Reports*. For each episode, they examine a fixed sequence of reports (typically nine; one episode includes eleven) and, for each report, answer four questions:

1. What is the current and/or prospective stance of fiscal policy?
2. What is the motivation given for the fiscal developments?
3. Does the EIU mention the debt-to-GDP ratio as a concern or as an underlying motivation?

4. Is there anything else of note relevant to fiscal policy actions?

RR19 provides a stringent benchmark for the narrative *interpretation* component of our approach. Their classifications are produced by expert readers, based on the same narrative source we use (EIU reports), and the coding rubric is described transparently in their paper. Moreover, RR19's emphasis on stated motivations speaks directly to the key conceptual step in the narrative method: extracting the intent of policy actions from contemporaneous text.

Accordingly, our first exercise is a matched-sample benchmark that holds the underlying documents and rubric fixed. We design a fixed prompt that mirrors RR19's four questions and applies it to the same set of EIU reports used by RR19 for the same 19 countries (200 reports in total). We evaluate accuracy for outputs corresponding to questions (1) and (2).[2]

For fiscal stance, the model assigns one of four labels: *Expansionary*, *Contractionary*, *Neutral*, or *Unclear*. For motivations, we adopt RR19's taxonomy and allow multiple motive categories to be recorded for a given report:

1. Financial rescue (bank bailouts or financial-crisis-driven stimulus);
2. Countercyclical stimulus (stimulus during a downturn);
3. Politics (stimulus taken to win an election);
4. Market access (austerity due to rising borrowing costs);
5. Conditionality (program requirements associated with official support);
6. Policymaker ideas (domestic policy preferences);
7. EU fiscal rules (adherence to supranational rules);
8. Countercyclical austerity (actions to prevent overheating);
9. Other.

[2]This RR19 benchmark is intentionally distinct from the construction of our quarterly exogenous spending-action proxy $z_{i,t}$ in Section 2.3. The production proxy uses a different fixed prompt and coding rule tailored to identifying *exogenous discretionary government spending actions* and aggregating them to a quarterly sign. The RR19 exercise instead isolates whether the model can implement an expert narrative-reading task when the documents and rubric are held constant.

RR19 classify motives (1)–(3) as pertaining to expansionary actions and motives (4)–(8) as pertaining to contractionary actions; motivation categories are not mutually exclusive, and multiple motives may be recorded for a given fiscal action.

Independence of stance vs. motive validation Panels (a) and (b) evaluate two different outputs from the model on the same RR19 reports, but they are scored separately. Panel (a) assesses whether the model's net stance for that report matches RR19's net classification. Panel (b) assesses whether the model detects the same motivation category that RR19 associates with that report—regardless of whether the net stance in Panel (a) matches. Consequently, entries such as Turkey (2001M7) can show a stance disagreement in Panel (a) alongside a motivation match in Panel (b): the model can correctly recognize "financial rescue" language even if our conservative aggregation rule yields a different net stance in a mixed-action passage.

The table below reproduces Table 3 from Romer and Romer (2019). Tick marks (✓) indicate instances where the AI-based classification agrees with the Romer and Romer assessment, while crosses (✗) denote disagreement.

Romer and Romer 2019, TABLE 3
Size and Motivations for Fiscal Expansions in Episodes of High Financial Distress

a. Size[a]

(Date expansion is first mentioned in parentheses)

Small	Large
U.S. (1992Q1)[b] ✓	
Norway (1991Q4)[c] ✓	
Korea (1999Q1)[d] ✓	Japan (1998Q1) ✓
Austria (2009M1) ✓	Turkey (2001M7) ✗
France (2009M1) ✓	U.S. (2008M7) ✓
Italy (2009M1)[b] ✓	Iceland (2009M1)✗
Norway (2009M1)✓	U.K. (2009M1)✓
Portugal (2009M1)✓	
Spain (2009M1)✓	
Sweden (2009M1)[c]✓	
Denmark (2009M7)✓	

b. Motivation

(Date motivation is first mentioned in parentheses)

Financial Rescue[e]	Countercyclical	Politics
	U.S. (1992Q1)[f] ✓	U.S. (1992Q1)[f] ✓
Norway (1992Q1) ✓	Norway (1992Q1) ✓	Norway (1993Q1)[f] ✓
Japan (1998Q1) ✓	Japan (1998Q1) ✓	Japan (1998Q1)[f] ✓
Korea (1999Q1)[d] ✓	Korea (1999Q1)✓	Korea (1999Q3)[f] ✗
Turkey (2001M7) ✓		
U.S. (2009M1) ✓	U.S. (2008M7) ✓	
Iceland (2009M1) ✓		
U.K. (2009M1) ✓	U.K. (2009M1)[f] ✓	
Austria (2009M1) ✓	Austria (2009M1) ✓	
France (2009M1)[f] ✓	France (2009M1)[f] ✓	
Italy (2009M1)[f] ✓	Italy (2009M1)[f] ✓	
Norway (2009M1)[f] ✗	Norway (2009M1) ✓	
Portugal (2009M1) ✓	Portugal (2009M1) ✓	Portugal (2009M1) ✓
Spain (2009M1) ✓	Spain (2009M1) ✓	
Sweden (2009M1) ✓	Sweden (2009M1) ✓	
Denmark (2009M7) ✓	Denmark (2009M7)✓	

Note: This table is based on Table 3 of Romer and Romer (2019), with additional annotations comparing the AI-based classifications to those in RR19. Check marks (✓) and crosses (✗) indicate whether the AI-based assessment agrees or disagrees with the corresponding classification in RR19. Panel (a) reports agreement in the direction of the fiscal action in each country–quarter episode. Panel (b) reports agreement in the underlying motivation for fiscal expansion, where a match indicates that the AI assessment identifies the same motivation category as RR19 for the given episode. Motivation categories are not mutually exclusive. Agreement in Panel (b) is evaluated cell-by-cell and is not conditional on agreement in Panel (a). Source: Authors' calculations based on Romer and Romer (2019).

Romer and Romer 2019, TABLE 4

Size and Motivations for Fiscal Austerity in Episodes of High Financial Distress

a. Size[a]

(Date austerity is first mentioned in parentheses)

Small	Large
U.S. (1991Q1)[b] ✗	Finland (1993Q3) ✓
Norway (1994Q3) ✓	Sweden (1994Q2) ✓
Korea (1998Q1)[c] ✓	Mexico (1996Q3) ✓
U.S. (2011M7) ✓	Turkey (2002M1) ✓
France (2010M7)[b] ✓	Iceland (2009M7) ✓
Denmark (2010M7) ✓	U.K. (2010M7) ✓
	Austria (2010M7) ✓
	Italy (2010M7)[b] ✓
	Portugal (2010M7) ✓
	Spain (2010M1) ✓
	Greece (2010M1) ✓
	Hungary (2009M7) ✓
	Ireland (2010M1) ✓

b.

Motivation

(Date motivation is first mentioned in parentheses)

Market Access	Conditionality	Ideas	EU Rules	Countercyclical
		Norway (1994Q3) ✓		Norway (1994Q3) [d] ✓
Finland (1994Q3)[d] ✗		Finland (1993Q3) ✓	Finland (1995Q1) ✓	
Sweden (1995Q1) ✓		Sweden (1994Q2) ✓	Sweden (1994Q3) ✓	
Mexico (1996Q3) ✓	Mexico (1996Q3) ✓			
Korea (1998Q1) ✓	Korea (1998Q1) ✓			
Turkey (2002M1) ✓	Turkey (2002M1) ✓		Turkey (2005M1) ✓	
		U.S. (2011M7) ✓		
Iceland (2010M1) ✓	Iceland (2009M7) ✓	Iceland (2010M1)[d] ✓		
U.K. (2010M7)[d] ✓			U.K. (2010M7)[d] ✓	
		Austria (2010M7) ✓	Austria (2010M7) ✓	
		France (2010M7) ✓	France (2010M7) ✓	
Italy (2010M7) ✓			Italy (2011M1) [d] ✓	
Portugal (2010M7) ✓	Portugal (2011M7) ✓	Portugal (2012M1)[d] ✓	Portugal (2010M7) ✓	
Spain (2010M1) ✓		Spain (2010M1)[d] ✓	Spain (2011M1)[d] ✓	
		Denmark (2011M1) ✓	Denmark (2011M1) ✓	
Greece (2010M1) ✓	Greece (2010M7) ✓		Greece (2010M1) ✓	
Hungary (2009M7) ✓	Hungary (2009M7) ✓	Hungary (2009M7)[d] ✓	Hungary (2009M7) ✓	
Ireland (2010M1) ✓	Ireland (2011M1) ✓			

Note: This table is based on Table 4 of Romer and Romer (2019), with additional annotations comparing the AI-based classifications to those in RR19. Check marks (✓) and crosses (✗) indicate whether the AI-based assessment agrees or disagrees with the corresponding classification in RR19. Panel (a) reports agreement in the direction of the fiscal action in each country–quarter episode. Panel (b) reports agreement in the underlying motivation for fiscal consolidation, where a match indicates that the AI assessment identifies the same motivation category as RR19 for the given episode. Motivation categories are not mutually exclusive. Agreement in Panel (b) is evaluated cell-by-cell and is not conditional on agreement in Panel (a). Source: Authors' calculations based on Romer and Romer (2019).

Each row in RR19's Table 3 corresponds to a specific EIU report that RR19 identify as the first mention of either an expansion (panel a) or a particular motivation (panel b). In the top panel, episodes are grouped by the size of the expansion ("Small" or "Large"), but our validation is performed at the report level: for each entry, we locate the corresponding report and assess whether the AI classifies it as expansionary, contractionary, neutral, or unclear—irrespective of size. In the lower panel, we take the report and quarter in which RR19 first identify a given motivation and check whether the AI assigns the same motive in that instance. Hence, each tick or cross in the table represents agreement or disagreement between the AI's classification for that specific EIU report and RR19's human-coded judgment. The same validation approach applies to RR19's Table 4, which mirrors Table 3 but for contractionary episodes and their associated motivations.

We therefore check whether AI:

- identifies an expansionary (contractionary) fiscal stance in the quarters that RR classify as expansionary (contractionary); and
- identifies the same motivation in the quarter that RR associate with that action.

Overall, we find that the AI output aligns closely with RR19's identification of fiscal stance and motivation. For expansions, the AI matches the stance in 14 out of 16 cases (87.5%) and the motivation in 32 out of 34 cases (94.1%).[3] For contractions, the AI matches the direction in 18 out of 19 cases (95 %) and the motivations in 44 out of 45 cases (97.8%).

Moreover, it is important to emphasize that the apparent "disagreements" do not stem from failures to interpret the text, but rather arise in reports describing both expansionary and contractionary measures simultaneously. In such cases, the model generally extracts the individual actions and their signs accurately. However, because magnitudes are unobserved and our classification approach is deliberately conservative, the net fiscal stance assigned by the model may differ from RR19's ex post judgment. For instance, in Turkey and Iceland, both sources identify concurrent increases and cuts; RR19 code these as a net expansion, whereas our rule assigns a contraction or labels the stance as Unclear in the absence of

[3] Four RR19 expansion episodes (Finland 1993Q1, Sweden 1993Q2, Greece 2009M7, and Ireland 2009M7) are excluded from the baseline stance comparison because RR19 classify these as net expansions primarily due to large financial-rescue measures, while the accompanying EIU text describes concurrent conventional fiscal contraction or ambiguous net effects. Our stance rule is deliberately conservative in mixed-action passages and does not infer relative magnitudes when they are not explicitly stated. Including these four boundary cases reduces the expansion stance match rate from 14/16 (87.5%) to 14/20 (70%), but the underlying disagreement reflects differences in aggregation of mixed fiscal actions rather than failure to detect expansionary or contractionary measures.

textual evidence on relative magnitudes. Norway presents a boundary case hinging on classification—specifically, whether financial-sector support is treated as fiscal spending—on which reasonable coders may differ.

By contrast, we identify only two genuine errors made by the model: (i) in Korea, the model missed an explicit political-motive cue ("with an eye to elections...") and consequently misclassified the stance; and (ii) in the United States entry in Table 4, although the model correctly listed the measures, it inferred the overall stance from remarks about rising deficits and mis-signed the net effect. A near-miss occurs in Finland, where the model failed to map language about rising long-term rates and financing pressure to a Market Access motive; this affects the motive label but not the stance sign.

Overall, these findings suggest that most apparent disagreements reflect differences in aggregation/judgment in mixed-action passages rather than inability to detect the underlying fiscal actions.

3.2 Replicability

To assess replicability, we re-ran the prompt 51 times on a subset of 20 reports using the same prompt used in section 3.1. The fiscal stance's sign remained consistent for every report.

Along with the original Romer and Romer questions, we also extracted other fields such as intensity and the confidence in the prompt. Intensity is an integer between -10 to +10 which measures the size and direction of the fiscal stance, and confidence refers to how confident the AI is in its own assessment of all fields. These are counterparts of the fields described in Section 2.3.

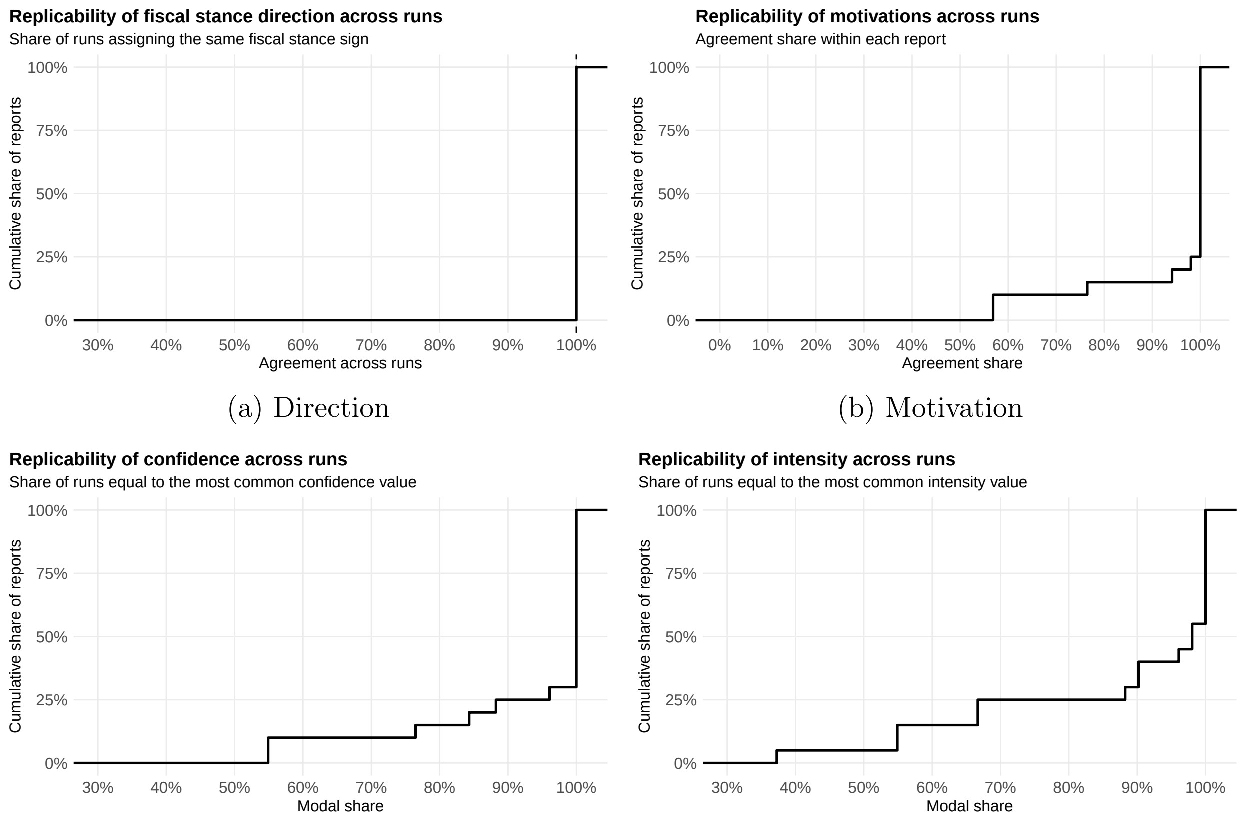
We then examine how replicable these four fields are across the 51 runs. For each report and each field, we summarize replicability using the modal share, defined as the fraction of runs in which the most frequently assigned value appears. This measure equals one when the same value is returned in every run, and declines as the model's outputs become more dispersed. We present the distribution of modal shares across reports using empirical cumulative distribution functions (ECDFs).

The results reveal a clear ranking in replicability across fields. The direction of the fiscal stance is perfectly stable: for all 20 reports, the modal share equals one, indicating that every run assigns the same sign. Motivation exhibits slightly lower, but still high, replicability across runs. Dominant motivation is identified consistently across runs, with occasional variation reflecting multiple plausible motives emphasized in the underlying narrative. By contrast, intensity and confidence exhibit lower replicability. While many reports still display high modal shares—indicating that a single value dominates across runs—several cases show non-trivial dispersion, with alternative values appearing intermittently.

This pattern is not unexpected. Direction is a coarse, binary classification closely tied to the qualitative narrative emphasized in the reports. Motivation, while still grounded in narrative evidence, is a multi-category and often multi-faceted construct: fiscal actions may plausibly serve several objectives simultaneously (e.g., stabilization and redistribution), and small differences in textual emphasis across runs can affect which motive is selected as primary. Intensity and confidence, in turn, require finer quantitative judgments that are inherently more sensitive to phrasing, emphasis, and ambiguity in the underlying text. As a result, small differences in interpretation across runs are more likely to affect these fields without altering the overall direction of fiscal policy.

Taken together, these results suggest that the AI-based classification is highly robust along the dimension most relevant for narrative identification—fiscal policy direction—while exhibiting greater but interpretable variation in secondary, more granular fields.

Figure 1: Replicability of results across 51 runs: Cumulative Distributions



(a) Direction (b) Motivation

(c) Confidence (d) Intensity

Notes: The figure plots ECDFs of the "modal share" across 20 randomly selected reports re-run 51 times under an identical fixed prompt. For each report and field (Direction/Motivation/Confidence/Intensity), the modal share is the fraction of runs in which the most frequently returned value occurs. A modal share of 1 indicates perfect replicability for that report/field. Source: Authors' calculations.

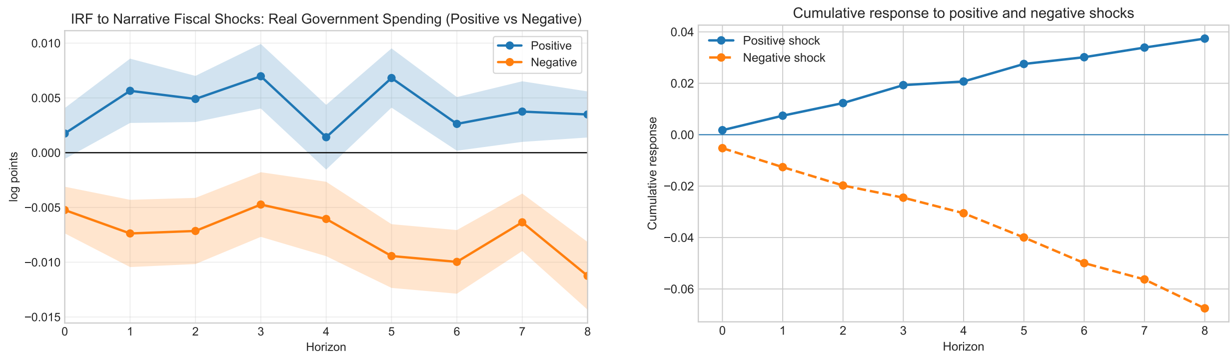
3.3 AI vs. government spending data

We examine whether the fiscal actions identified are correlated with and can predict changes in actual government spending data. To this end, we estimate the following equation of government expenditure :

$$\begin{aligned} \ln G_{i,t+h} = {} & \beta_h^{+}\mathbf{1}\{z_{i,t}=+1\} + \beta_h^{-}\mathbf{1}\{z_{i,t}=-1\} \\ & + \sum_{k=1}^{4}\Big(\phi_{h,k}\ln G_{i,t-k} + \psi_{h,k}\ln Y_{i,t-k} + \gamma_{h,k} z_{i,t-k}\Big) \\ & + \alpha_i + \lambda_t + \varepsilon_{i,t+h}, \qquad h = 0,\ldots,8. \end{aligned} \tag{1}$$

At impact ($h = 0$), contractions (negative shocks) reduce $\log G$ by roughly -0.005 (log

Figure 2: Local-projection responses of real government spending to narrative spending shocks (expansions vs. contractions)



(a) IRFs from LP

(b) Cumulative effect on G

Notes: Panel 2a shows local-projection impulse responses of G to expansionary ($z_{i,t} = +1$) and contractionary ($z_{i,t} = -1$) narrative shocks estimated from Equation 1, including country and time fixed effects and the lag controls listed in the regression. Panel 2b reports the associated cumulative responses. Shaded areas denote 68% confidence intervals with standard errors clustered at the country level. Horizons are quarters. Source: Authors' calculations.

points; 0.5% level), while expansions (+1) rise one quarter later with $\beta^+ \approx 0.002$. Cumulatively, contractions reach about -0.06 by $h \approx 8$, while expansions reach about 0.04 over the same period.

3.4 External Validation against Adler et al. 2024

This subsection evaluates the external validity of the quarterly narrative fiscal stance series against an independent action-based measure of fiscal consolidation. The external source is the annual dataset constructed by Adler et al. (2024), which reports, for a set of 31 advanced and emerging economies, the announced size of discretionary fiscal consolidation packages, decomposed into tax measures and expenditure measures, expressed as a percent of GDP. The coverage begins in 1978 for advanced economies and in 1989 for Latin American and Caribbean economies, and extends through 2020. The definition of consolidation in Adler et al. (2024) is explicit: only measures primarily motivated by medium-term deficit reduction and debt sustainability are included, while measures taken in response to short-run cyclical weakness are excluded. The expenditure component of these packages, denoted here by $spend_{c,y}$, is the size (in percent of GDP) of spending-based consolidation in country c and year y. Positive values of $spend_{c,y}$ indicate expenditure cuts or restraint (fiscal tightening).

Negative values indicate net loosening of expenditure in that year.

Our narrative database documents, at quarterly frequency, whether the government in a given country–quarter announces or implements a discretionary spending action. Each nonzero action is classified as either a *discretionary consolidation in government spending* (coded -1) or a *discretionary expansion in government spending* (coded $+1$).[4] Let $indicator_{c,y,q} \in \{-1, 0, +1\}$ denote this quarterly classification for country c, calendar year y, and quarter $q \in \{1, 2, 3, 4\}$.

Relative to Adler et al. (2024), our database has clear advantages and one clear limitation. First, it is quarterly (rather than annual) and has substantially broader country coverage, enabling sharper timing tests and wider cross-country analysis. Second, we document below that our quarterly sign indicator is not predictable based on standard lagged macro controls . By contrast, at the annual frequency the announced package magnitudes in Adler et al. (2024) display significant predictability in simple forecasting regressions. The main limitation of our measure is intensity: the proxy records direction but not size. For our identification, however, this is not first order—the proxy is used as an internal instrument and the contemporaneous response of spending pins the scale (Plagborg-Møller and Wolf, 2021; Li et al., 2024).

To compare our quarterly narrative series to the annual consolidation data in Adler et al. (2024), we use a set of proxies. Concretely, for each country c and year y, we construct:

$$\begin{aligned}
consol_quarters_{c,y} &= \sum_{q=1}^{4} \mathbf{1}\{indicator_{c,y,q} = -1\}, \\
expand_quarters_{c,y} &= \sum_{q=1}^{4} \mathbf{1}\{indicator_{c,y,q} = +1\}, \\
any_consol_{c,y} &= \mathbf{1}\{consol_quarters_{c,y} \geq 1\}, \\
any_expand_{c,y} &= \mathbf{1}\{expand_quarters_{c,y} \geq 1\}.
\end{aligned}$$

In words: $any\_consol_{c,y} = 1$ if at least one quarter in year y contains a discretionary spending

[4]Quarters without an identified discretionary spending action are recorded in the narrative database as 0 and are treated as "no decisive discretionary spending action." A value of -1 reflects cuts, freezes, or restraint of spending that are described by contemporaneous sources as deliberate fiscal tightening; a value of $+1$ reflects discretionary spending increases such as wage hikes, transfers, or program expansions that are described as active policy choices.

consolidation; $any\_expand_{c,y} = 1$ if at least one quarter contains a discretionary expansion.

We also form a (signed) annual stance index,

$$net\_stance_{c,y} = \frac{consol\_quarters_{c,y} - expand\_quarters_{c,y}}{4}, \tag{2}$$

which lies in $[-1, +1]$. By construction, $net\_stance_{c,y} = +1$ if all four quarters in year y are coded as consolidation (-1) and none are coded as expansion; $net\_stance_{c,y} = -1$ if all four quarters are coded as expansion ($+1$) and none are coded as consolidation. Intermediate values reflect a mixture of consolidation and expansion across quarters in the same calendar year. This variable is used below as an annual measure of the relative prevalence of consolidation versus expansion actions in our high-frequency narrative series.

Finally, we merge the two datasets at the country–year level using ISO 3 country codes and calendar year.

Test 1: Extensive-margin alignment. We first ask whether years that Adler et al. (2024) classify as featuring large spending-based consolidations correspond to years in which our quarterly narrative series ever records a consolidation quarter. Define

$$AE\_consol_{c,y}(\tau) = \mathbf{1}\{spend_{c,y} \geq \tau\},$$

where τ is a threshold in percent of GDP (we consider $\tau \in \{0.5, 1.0, 2.0\}$). For each τ, we compute: (i) the "Hit rate," defined as the share of (c, y) with $AE\_consol_{c,y}(\tau) = 1$ such that $any\_consol_{c,y} = 1$ in our data; and (ii) the "Expansion rate," defined as the share of those same (c, y) such that $any\_expand_{c,y} = 1$.

Panel A of Table 5 summarizes these extensive-margin results for the full merged sample. For $\tau = 0.5$ (a spending-based consolidation of at least 0.5 percent of GDP), there are 138 country–years that Adler et al. (2024) classify as consolidation years. In 96.4 percent of these cases, our quarterly database records at least one consolidation quarter ($any\_consol_{c,y} = 1$). For larger consolidation packages ($\tau = 1.0$ and $\tau = 2.0$), the hit rate remains above 96 percent and reaches 100 percent for $\tau = 2.0$. At the same time, between 23 and 25 percent of these "large consolidation" years also contain at least one expansion quarter in our quarterly series

($any\_expand_{c,y} = 1$). This reflects within-year mixtures: governments often implement multiple discretionary spending actions of different signs within the same calendar year, while their dataset reports the net announced consolidation package at the annual level.

Panel B of Table 5 repeats the exercise separately for advanced economies and for Latin America / Caribbean economies, using the sample-group indicators in Adler et al. (2024). For the Latin America / Caribbean group, the hit rate is 100 percent for all thresholds, with 25 to 30 percent of these consolidations containing at least one quarter classified as expansions in our database. For advanced economies, hit rates remain above 95 percent, with a slightly lower expansion rate. Taken together, these results indicate that when Adler et al. (2024) report a sizable expenditure-based consolidation package for a given country–year, our quarterly narrative classification almost always records at least one discretionary consolidation quarter in that year, and rarely records only expansionary actions.

We also consider the converse case, namely years in which $spend_{c,y} \leq -\tau$, i.e. years that Adler et al. (2024) classify as featuring a net loosening of expenditure (negative spending-based consolidation). Such observations are rare in the intersection sample (one country–year for $\tau = 0.5$ and zero for $\tau \geq 1.0$), but in that case our quarterly database classifies the same year as expansionary ($any\_expand_{c,y} = 1$) and never as consolidation ($any\_consol_{c,y} = 0$).

Table 5: Extensive-Margin Validation against Spending-Based Consolidation Episodes

Threshold τ	Full sample					
	N_{big}	Hit rate ($any\_consol=1$)	Expansion rate ($any\_expand=1$)	N_{loose}	Exp. hit rate ($any\_expand=1$)	Consolidation rate ($any\_consol=1$)
0.5	138	0.964	0.246	1	1.000	0.000
1.0	65	0.969	0.231	0	. . .	. . .
2.0	16	1.000	0.250	0	. . .	. . .

	Panel B. By country group					
	Advanced / AEs			LAC		
Threshold τ	Hit rate	Expansion rate	N_{big}	Hit rate	Expansion rate	N_{big}
0.5	0.955	0.234	111	1.000	0.296	27
1.0	0.962	0.222	54	1.000	0.273	11
2.0	1.000	0.250	12	1.000	0.250	4

Notes: $spend_{c,y}$ is the expenditure-based consolidation size in Adler et al. (2024), measured in percent of GDP for country c and year y, with positive values indicating discretionary expenditure cuts and restraint. "Threshold τ" defines a large consolidation year as $spend_{c,y} \geq \tau$. N_{big} counts such country–years. "Hit rate" is the share of those country–years in which the quarterly narrative database reports at least one consolidation quarter ($any\_consol_{c,y} = 1$). "Expansion rate" is the share of those same country–years in which the quarterly narrative database reports at least one expansion quarter ($any\_expand_{c,y} = 1$). "N_{loose}" counts country–years with $spend_{c,y} \leq -\tau$, interpreted as net loosening of expenditure; "Exp. hit rate" and "Consolidation" are defined analogously for those years. "Advanced / AEs" and "LAC" correspond to the advanced economy and Latin America/Caribbean subsamples in Adler et al. (2024), respectively. Hit rates for loosening episodes are omitted when $N_{\text{loose}} = 0$. *Source:* Authors' coding/classification using the merged country–year panel described in the text.

Test 2: Probability of being identified as a consolidation year. We next estimate a country–fixed-effects linear probability model that relates $any\_consol_{c,y}$ and $any\_expand_{c,y}$ to the probability that Adler et al. (2024) classify year y in country c as a sizable consolidation year:

$$\mathbf{1}\{spend_{c,y} \geq 0.5\} = \alpha_c + \beta_1\, any\_consol_{c,y} + \beta_2\, any\_expand_{c,y} + \varepsilon_{c,y}, \tag{3}$$

where α_c are country fixed effects and standard errors are clustered at the country level. The dependent variable is $AE\_consol_{c,y}(0.5) = \mathbf{1}\{spend_{c,y} \geq 0.5\}$, i.e. an indicator for whether Adler et al. (2024) record an expenditure-based consolidation package of at least 0.5 percent of GDP in that country–year.

Table 6 (column 1) reports the estimates. The coefficient on $any\_consol_{c,y}$ is 0.126 with a clustered standard error of 0.019. Thus, conditional on country fixed effects, observing at least one consolidation quarter in our quarterly narrative database is associated with an 12.6 percentage point higher probability that Adler et al. (2024) classify that same country–year as featuring a sizable expenditure-based consolidation package. The coefficient on $any\_expand_{c,y}$ is -0.073 (s.e. 0.02), implying that observing an expansionary quarter in our classification is associated with a 7.3 percentage point *lower* probability that the external dataset codes that year as a sizable consolidation. Both coefficients are statistically significant with standard errors clustered by country. This indicates that the sign of discretionary spending actions in our quarterly narrative series is informative about whether the external source classifies the year as a consolidation episode.

Test 3: Intensity alignment. Finally, we study whether the *intensity* of consolidation in Adler et al. (2024)—measured by $spend_{c,y}$ in percent of GDP—co-moves with the annual stance index $net\_stance_{c,y}$ defined in (2). We estimate the country–fixed-effects regression

$$spend_{c,y} = \gamma_c + \delta\, net\_stance_{c,y} + u_{c,y}, \tag{4}$$

again clustering standard errors by country. Results (Table 6, column 2) indicate that $\delta = 0.263$ with a clustered standard error of 0.044. Since $net\_stance_{c,y}$ ranges from -1 (all observed actions in the year are expansionary) to $+1$ (all observed actions are consolidation), moving from a purely expansionary year ($net\_stance = -1$) to a purely consolidation year ($net\_stance = +1$) is associated, on average within a given country, with an increase of approximately $2 \times 0.263 \approx 0.5$ percentage points of GDP in $spend_{c,y}$. In other words, years that our quarterly narrative database classifies as dominated by consolidation actions correspond to larger expenditure-based consolidation packages in the external quantitative dataset, conditional on country fixed effects.

Table 6: Country Fixed-Effects Regressions Linking the Quarterly Narrative Stance to Annual Consolidation Measures

	(1) $\mathbf{1}\{spend_{c,y} \geq 0.5\}$	(2) $spend_{c,y}$
$any\_consol_{c,y}$	0.126***	
	(0.019)	
$any\_expand_{c,y}$	-0.073***	
	(0.020)	
$net\_stance_{c,y}$		0.263***
		(0.044)
Country FEs	Yes	Yes
Obs.	1161	1161
R^2	0.125	0.145
Dep. var. mean	0.134	0.068
SEs	Clustered by country	Clustered by country

Notes: Column (1) reports a linear probability model with country fixed effects. The dependent variable is $\mathbf{1}\{spend_{c,y} \geq 0.5\}$, an indicator for whether Adler et al. (2024) report a spending-based consolidation package of at least 0.5 percent of GDP in country c, year y. The regressors are $any\_consol_{c,y}$ (indicator for at least one consolidation quarter in our quarterly narrative classification) and $any\_expand_{c,y}$ (indicator for at least one expansion quarter). Column (2) reports an OLS regression with country fixed effects. The dependent variable is $spend_{c,y}$, the size of expenditure-based consolidation in percent of GDP from Adler et al. (2024). The regressor is $net\_stance_{c,y}$, defined in (2) to lie in $[-1, +1]$. Standard errors, in parentheses, are clustered at the country level. *** indicates statistical significance at the 1 percent level. Source: Authors' calculations using the merged country–year panel described in the text.

In summary, our quarterly narrative series and the action-based annual consolidation dataset co-move along both the extensive margin and the intensive margin. When Adler et al. (2024) indicate that a country–year features a sizable spending-based consolidation package, our quarterly narrative series almost always records at least one discretionary consolidation quarter in that same country–year, and almost never records only expansionary actions. Moreover, within countries, the annual size of spending-based consolidation (in percent of GDP) in Adler et al. (2024) is increasing in the share of quarters our database codes as discretionary spending consolidations. We interpret these results as evidence that both sources capture the same underlying policy object—discretionary shifts in the government's expenditure stance aimed at medium-term fiscal adjustment—despite differences in frequency (quarterly versus annual), scope (global coverage versus a 31-country sample in Adler et al. (2024)), and coding design (narrative sign versus quantitative package size).

4 Computing country-by-country multipliers

This section presents new evidence on country-specific fiscal multipliers. It first discusses the use of the shocks identified in the previous sections, then presents the methodological approach to estimate country-specific multipliers, and finally shows the ranges of multipliers obtained.

4.1 Predictability of shocks

Although our fiscal shocks are intended to be exogenous with respect to their *motivation* (as established through textual analysis), credible causal interpretation also requires that they are not systematically related to lagged, observable macroeconomic conditions. A prominent concern in this regard is raised by Jordà and Taylor (2015), who show that narrative fiscal shocks from an earlier version of the Adler et al. (2024) compilation for advanced economies (i.e., Guajardo et al. (2014)) can be forecast using standard macroeconomic indicators such as the debt-to-GDP ratio, the output gap, and GDP growth.

To assess whether the quarterly government expenditure shocks in our database exhibit similar predictability, we estimate panel regressions of the form

$$z_{i,t} = \alpha_i + \beta_1 \,\mathrm{Debt/GDP}_{i,t-1} + \beta_2 \,\mathrm{OutputGap}_{i,t-1} + \beta_3 \,\mathrm{GDPGrowth}_{i,t-1} + \rho \, z_{i,t-1} + \varepsilon_{i,t}, \quad (5)$$

where $z_{i,t}$ denotes the quarterly narrative spending-action shock and α_i captures country fixed effects. We consider three samples: our full country sample, the subset of countries covered in Guajardo et al. (2014), and the subset of countries covered in Adler et al. (2024).

Table 7 reports the results. Across all three samples, the only robust predictor of $z_{i,t}$ is its own lag: the coefficient on the lagged expenditure shock is positive and statistically significant, whereas the coefficients on lagged debt-to-GDP, the output gap, and GDP growth are statistically indistinguishable from zero. Consistent with this, the regressions have very limited explanatory power, with $R^2 \leq 0.1$ in all cases.

Table 7: Predictability Results: New Expenditure Shocks

	(1) Full sample	(2) Guajardo et al. (2014) Sample	(3) Adler et al. (2024) Sample
Debt to GDP	0.000101	0.000240	0.0000893
(lagged)	(0.000178)	(0.000307)	(0.000324)
Output Gap	0.00755	0.0139	0.00827
(lagged)	(0.00464)	(0.00940)	(0.00765)
GDP growth	0.404	0.795	0.561
(lagged)	(0.505)	(0.692)	(0.581)
Expenditure shock	0.266***	0.308***	0.298***
(lagged)	(0.0140)	(0.0275)	(0.0228)
Constant	-0.0775***	-0.103**	-0.0915**
	(0.0237)	(0.0410)	(0.0412)
R^2	0.07	0.10	0.09

Notes: Dependent variable is the quarterly narrative spending-action shock $z_{i,t}$ (coded -1,0,+1 as described in Section 2). Regressors are one-quarter lagged debt-to-GDP, output gap, and real GDP growth, and the lagged shock $z_{i,t}$. All specifications include country fixed effects; standard errors are clustered by country. Columns correspond to (i) full sample, (ii) Guajardo et al. (2014) country sample, and (iii) Adler et al. (2024) country sample, using the intersection of available quarterly macro controls. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: Authors' calculations.

For comparison, Table 8 reports analogous predictability regressions for the annual narrative shocks in Adler et al. (2024). In both the primary balance and expenditure specifications, the lagged narrative shock is highly significant; moreover, lagged GDP growth also contributes to predicting the shocks. The resulting fit is substantially higher than in Table 7, with R^2 in the range of approximately 0.24–0.31. Taken together with the results in Table 7, these patterns indicate that our quarterly expenditure shocks contain less systematic information about lagged macroeconomic conditions than the annual narrative shocks from Adler et al. (2024).

Table 8: Predictability Results: Adler et al (2024) shocks

	(1) Primary Balance shock	(2) Expenditure shock
Debt to GDP	0.0000425	-0.000445
(lagged)	(0.000555)	(0.000361)
Output Gap	0.536	0.202
(lagged)	(0.955)	(0.536)
GDP Growth	-2.869**	-1.747***
(lagged)	(1.149)	(0.628)
Narrative Shock	0.461***	0.530***
(lagged)	(0.0526)	(0.0642)
Constant	0.214***	0.146***
	(0.0467)	(0.0288)
R^2	0.243	0.305

Notes: Dependent variable is the annual narrative consolidation shock from Adler et al. (2024). Regressors are one period lagged debt-to-GDP, output gap, and real GDP growth, and the lagged shock $z_{i,t}$. All specifications include country fixed effects; standard errors are clustered by country. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: Authors' calculations.

As an additional check, we evaluate whether the frequency at which shocks are measured affects their predictability. Specifically, we aggregate the quarterly expenditure shocks to the annual level by summing within year, yielding a measure ranging from -4 (all quarters contractionary) to $+4$ (all quarters expansionary). We then re-estimate the same predictability specification at the annual frequency, regressing the aggregated annual shock on lagged debt-to-GDP, the output gap, GDP growth, and the lagged shock.

Table 9 shows that, once aggregated, predictability becomes more pronounced. In the baseline and contractionary-only specifications, both lagged debt-to-GDP and lagged GDP growth are statistically significant predictors, in addition to the lagged shock; in the expansionary-only specification, predictability is driven primarily by GDP growth and the shock's own persistence. Overall, these results are consistent with the interpretation that the weak predictability at the quarterly horizon reflects the higher-frequency nature of the shock measure, whereas systematic information about macroeconomic conditions is more readily detectable when the same underlying actions are viewed at the annual frequency.

Table 9: Predictability Results: Annual Aggregation

	(1) Baseline	(2) Contractionary Only	(3) Expansionary Only
Debt to GDP	0.00545*	0.00493**	0.000462
(lagged)	(0.00292)	(0.00199)	(0.00156)
Output Gap	-0.00267	0.0113	-0.00991
(lagged)	(0.0225)	(0.0153)	(0.0120)
GDP Growth	6.030***	3.490***	2.839***
(lagged)	(1.876)	(1.277)	(1.003)
Expenditure shock	0.324***	0.311***	0.207***
(lagged)	(0.0295)	(0.0295)	(0.0309)
Constant	-0.727***	-1.009***	0.329***
	(0.192)	(0.135)	(0.104)
R^2	0.125	0.116	0.053

Notes: Dependent variable is the narrative spending-action shock aggregated to annual frequency by summing. Regressors are one-quarter lagged debt-to-GDP, output gap, and real GDP growth, and the lagged shock $z_{i,t}$ (all annual). All specifications include country fixed effects; standard errors are clustered by country. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: Authors' calculations.

4.2 Econometric Identification Strategy

Using the narrative series as an observable proxy for unexpected spending actions, we estimate country-specific structural VARs and identify government-spending shocks via a recursive ("internal instrument") ordering that places the proxy first (Plagborg-Møller and Wolf, 2021). With this ordering, the first orthogonalized innovation (the first Cholesky shock) is interpreted as the fiscal shock: it may affect all remaining variables contemporaneously, whereas the remaining innovations are constrained to have no impact effect on the proxy. In population, this implementation delivers the same *relative* impulse-response estimand as IV local projections that use the same narrative proxy and, under standard invertibility conditions, coincides with the proxy-SVAR estimand (Plagborg-Møller and Wolf, 2021; Li et al., 2024; Stock and Watson, 2018).

In addition to this proxy-based recursive identification, we incorporate economically motivated *admissibility conditions* on the implied spending response and the resulting multipliers. The purpose is purely inferential: to avoid drawing conclusions from posterior draws in which the proxy innovation is effectively uninformative about realized spending (e.g., near-zero spending responses) and hence produces mechanically extreme multipliers. Operationally, these conditions are implemented as an accept–reject filter (i.e., posterior truncation) applied to impulse responses computed under the proxy-first ordering.[5]

System, ordering, and lag length. For each country we estimate a VAR that always includes the narrative proxy, real government spending, and real GDP, and may include additional macro variables when available. Let

$$x_t = \begin{pmatrix} z_t \\ g_t \\ y_t \end{pmatrix},$$

where z_t is the narrative spending series, $g_t \equiv \log G_t$ is real government spending, $y_t \equiv \log Y_t$ is real GDP. We set the lag order to $p = 8$ and order the variables recursively with z_t first.

Reduced-form VAR. Let x_t denote the n-dimensional vector of observables, ordered as above with $n \geq 3$. We estimate a VAR(p),

$$x_t = c + \sum_{\ell=1}^{p} B_\ell x_{t-\ell} + u_t, \qquad \mathbb{E}[u_t] = 0, \qquad \Sigma_u \equiv \mathbb{E}[u_t u_t'] \succ 0, \tag{6}$$

and interpret (6) as a system of best linear predictors. For identification and impulse-response analysis, only the reduced-form parameters $\{B_\ell\}_{\ell=1}^{p}$ and Σ_u are required.

Stacking observations for $t = p+1, \ldots, T$, let $T_{\text{eff}} = T - p$ be the effective sample size, Y the $T_{\text{eff}} \times n$ matrix of dependent variables, and X the $T_{\text{eff}} \times k$ regressor matrix containing a constant and p lags of x_t, with $k = 1 + np$. Let B be the $k \times n$ coefficient matrix stacking c and $B_1, \ldots, B_p$. Then

$$Y = XB + U, \tag{7}$$

[5]Appendix A provides full details on the estimation setup, priors, posterior computation (including the working likelihood used for computation), and the exact admissibility bounds and horizons used in baseline and robustness exercises. The admissibility conditions are chosen ex ante based on the existing multiplier literature and applied symmetrically across countries; they do not modify the underlying proxy-based identification of the shock.

where U stacks the reduced-form innovations $\{u_t\}$.

Proxy-first (Cholesky) identification. Let $\Sigma_u = PP'$ be the Cholesky factorization, with P lower triangular and positive diagonal. Define orthogonalized innovations ε_t by

$$u_t = P\varepsilon_t, \qquad \mathbb{E}[\varepsilon_t] = 0, \qquad \mathbb{E}[\varepsilon_t\varepsilon_t'] = I_n. \tag{8}$$

The identified fiscal shock is the first orthogonalized innovation ε_{1t}, i.e., the standardized innovation in the narrative-proxy equation. Because $P_{11} > 0$, a positive ε_{1t} corresponds to a positive impact innovation in z_t, which fixes the shock normalization relative to the proxy series.

Let $\Theta_h(B, \Sigma_u)$ denote the $n \times n$ matrix of orthogonalized impulse responses at horizon h, computed from the companion form implied by (6) and the impact matrix P. The response of variable j to the identified fiscal shock ε_{1t} at horizon h is

$$\theta_{j,h} = e_j'\Theta_h(B, \Sigma_u)\, e_1.$$

Under standard conditions, interpreting ε_{1t} as the government-spending shock yields the same population *relative* impulse-response estimand as IV local projections that use z_t as an instrument (Plagborg-Møller and Wolf, 2021; Li et al., 2024; Stock and Watson, 2018).

Our objects of interest—orthogonalized impulse responses and cumulative spending multipliers—are deterministic functionals of the reduced-form parameters (B, Σ_u) through the VAR moving-average representation and the Cholesky factorization $\Sigma_u = PP'$. In the baseline implementation, we draw (B, Σ_u) from a Bayesian reduced-form VAR with Minnesota priors, compute the corresponding impulse responses and multipliers, and retain a draw if and only if the admissibility inequalities hold. This accept–reject procedure is equivalent to conditioning (truncating) the posterior distribution on the set of admissible reduced-form parameters.[6] Robustness tests to alternative admissibility tests are conducted.

[6]Bayesian reduced-form VAR estimation with Minnesota (normal-inverse-Wishart) priors relies on a Gaussian working likelihood, which is potentially misspecified given that the narrative proxy is discrete. Appendix B compares the baseline approach to a frequentist counterpart that does not impose Gaussianity and finds the differences to be minor.

4.3 Impulse Responses and Spending Multipliers

Let $y_t = \log Y_t$ and $g_t = \log G_t$ denote real GDP and government spending, respectively. Denote by $\widehat{\mathrm{IRF}}_y(h)$ and $\widehat{\mathrm{IRF}}_g(h)$ the orthogonalized impulse responses (in log points) at horizon h to the proxy-identified shock ε_{1t} defined in (8). For small responses, the implied level changes satisfy

$$\Delta Y_{t+h} \approx \bar{Y}\,\widehat{\mathrm{IRF}}_y(h), \qquad \Delta G_{t+h} \approx \bar{G}\,\widehat{\mathrm{IRF}}_g(h),$$

where $\bar{Y}$ and $\bar{G}$ are sample averages (or pre-shock means) over the estimation window.

The cumulative spending multiplier at horizon h is the ratio of cumulative level responses,

$$\widehat{\mathcal{M}}_h = \frac{\sum_{j=0}^{h} \Delta Y_{t+j}}{\sum_{j=0}^{h} \Delta G_{t+j}} \approx \frac{\bar{Y}}{\bar{G}} \cdot \frac{\sum_{j=0}^{h} \widehat{\mathrm{IRF}}_y(j)}{\sum_{j=0}^{h} \widehat{\mathrm{IRF}}_g(j)} = \frac{1}{\bar{s}_G} \cdot \frac{\sum_{j=0}^{h} \widehat{\mathrm{IRF}}_y(j)}{\sum_{j=0}^{h} \widehat{\mathrm{IRF}}_g(j)}, \tag{9}$$

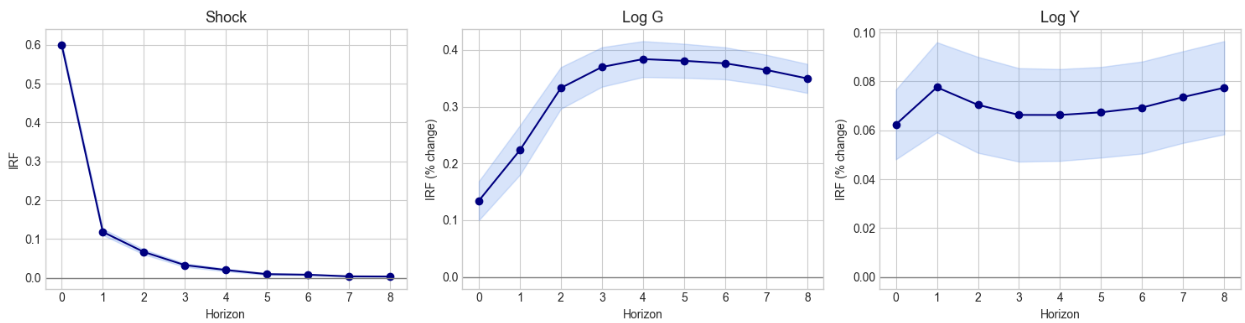
where $\bar{s}_G = (\frac{\bar{G}}{\bar{Y}})$ is the average spending to GDP ratio for a given country. Thus, with variables in logs, one obtains level multipliers by dividing the ratio of cumulative log responses by $\bar{s}_G$.

We summarize impulse responses and multipliers using posterior medians (and, when reported, posterior uncertainty bands) computed from the accepted posterior draws described in Section 4.2 and Appendix A.

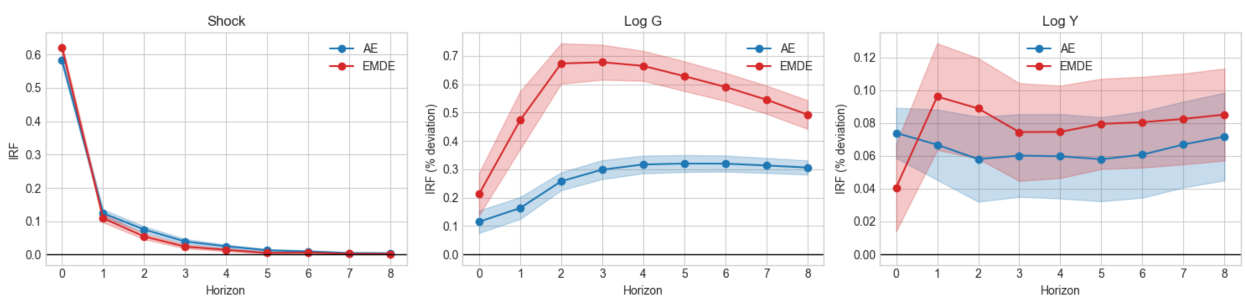
Panel (a) of Figure 3 shows the inverse variance-weighted impulse responses (dotted lines) for the full sample along with the associated ± 1 standard error bands (shaded areas) for each horizon (quarter) from zero to eight. A positive spending shock triggers a sharp and immediate increase in government expenditure, peaking at approximately 0.4 percent around the one-year horizon and remaining elevated over the two-year period. The output response is also positive across all horizons, with effects ranging between 0.06 and 0.08 percent. Results are robust to alternative admissibility bounds on cumulative multipliers (see Appendix A).

These responses, however, mask significant differences across countries. This is evident in Figure 3, Panel (b), which shows the responses of spending and GDP separately for advanced and developing economies. The figures indicate that spending shocks are associated with larger government spending responses in the typical developing economy compared to the typical advanced economy. In contrast, although GDP responses are also larger in the typical developing economies, the differences are not statistically significant (as indicated by the overlapping bands).

Figure 3: IRFs of a shock to the fiscal variable in the baseline VAR.



(a) Full sample of countries



(b) Subsamples of AE and EMDE countries

Note: Solid lines report inverse-variance–weighted averages of country-specific impulse responses. For each horizon, country-level impulse responses are summarized by their posterior medians and pooled across countries within each group using weighted least squares with a constant only, where weights are given by the inverse of the estimated variance of the country-specific response. Shaded areas denote ±1 standard error bands computed from the weighted regression. Horizons are in quarters. Source: Authors' calculations.

Turning to the spending multipliers, Table 10 reports the 1- and 2-year inverse variance-weighted average [7] in the full sample, as well as for the median advanced and developing economies. The multipliers for the full sample at the 1-year (0.74) and 2-year (0.67) horizons are consistent with the range commonly reported in the literature for individual countries

[7]For each country i and horizon h, let $\widehat{IRF}_i^X(h)$ denote the estimated impulse response of variable $X \in \{Y, G\}$ to an expenditure shock. Let $I\tilde{R}F_i^X(h)$ denote the within-country median of the estimated responses , and let $\widehat{V}_i^X(h)$ denote the corresponding estimated variance.

Group-level impulse responses are constructed horizon-by-horizon as inverse-variance-weighted averages as well as panels of countries (e.g., Ramey (2019)).

Despite differences in the responses of spending and output to fiscal shocks, the 1- and 2-year multipliers for the average advanced and developing economies are quite similar (0.70 and 0.61 for the advanced economy, and 0.80 and 0.81 for the developing economy). This is because the share of government spending to GDP—which is inversely used to rescale the ratio of the two responses—is much larger in the median advanced economy (39 percent) than in the median developing economy (17 percent).

Table 10: Inverse Variance- Weighted Multipliers Across Samples

Sample	1-year multiplier	2-year multiplier
Full Sample	0.74	0.67
Advanced Economies	0.70	0.61
EMDEs	0.80	0.81

Notes: For each country, we estimate impulse responses of output and government spending to an expenditure shock. At each horizon, we construct group-level impulse responses by aggregating the country-specific median responses using inverse-variance weights (Figure 3). The reported multiplier at horizon h is then calculated as the cumulative ratio of these pooled impulse responses, scaled by the group-level median government expenditure share. "Advanced economies" and "EMDEs" follow the paper's country classification.

At the same time, while the spending multipliers for the median and average economies are similar, there is significant heterogeneity in their magnitudes within each group of countries, as evidenced in Figure 4. In particular, the interquartile range of multipliers for each group varies approximately between 0 and 1.5.

of the country-level medians:

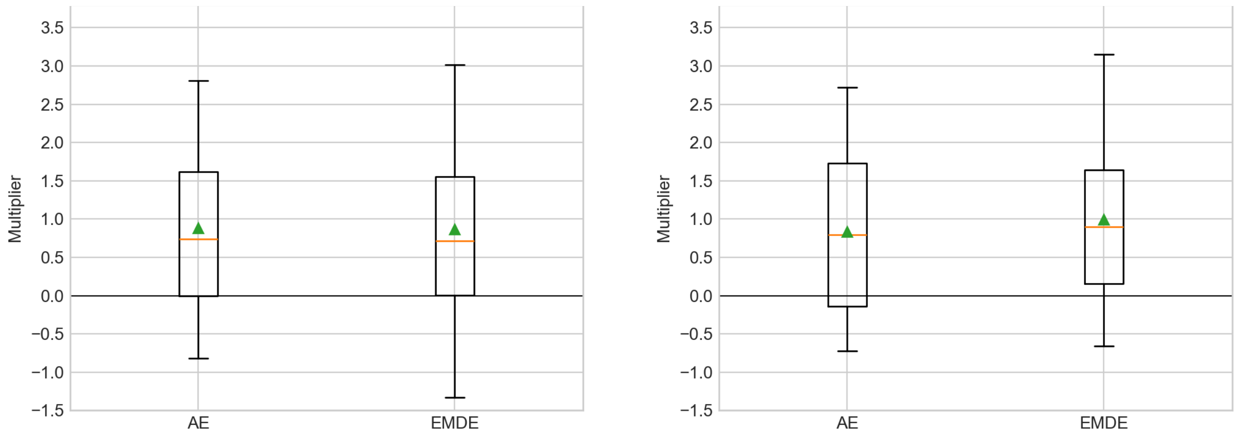
$$\widehat{IRF}_g^X(h) = \frac{\sum_{i\in g} w_i^X(h)\, I\tilde{R}F_i^X(h)}{\sum_{i\in g} w_i^X(h)}, \qquad w_i^X(h) = \frac{1}{\widehat{V}_i^X(h)}.$$

The fiscal multiplier at horizon h is then computed as the cumulative ratio of the aggregated responses scaled by the group-level median government expenditure share:

$$M_g(H) = \frac{1}{\bar{s}_{G,g}} \frac{\sum_{h=0}^{H} \widehat{IRF}_g^Y(h)}{\sum_{h=0}^{H} \widehat{IRF}_g^G(h)},$$

where $\bar{s}_{G,g} = \text{median}_{i\in G}\bar{s_G}$ across all countries i in group g.

Figure 4: Distribution of country-level median cumulative multipliers



(a) 1 year horizon

(b) 2 year horizon

Note: Boxplots summarize the distribution across countries of country-level posterior-median cumulative multipliers at horizons H=4 (left) and H=8 (right). Boxes denote the interquartile range; the horizontal line is the median; whiskers follow the plotting convention used in the code; markers denote mean. Source: Authors' calculations.

We conclude this section by illustrating the advantage of using quarterly data, as opposed to the annual frequency that is typically employed in the literature using data for many countries.

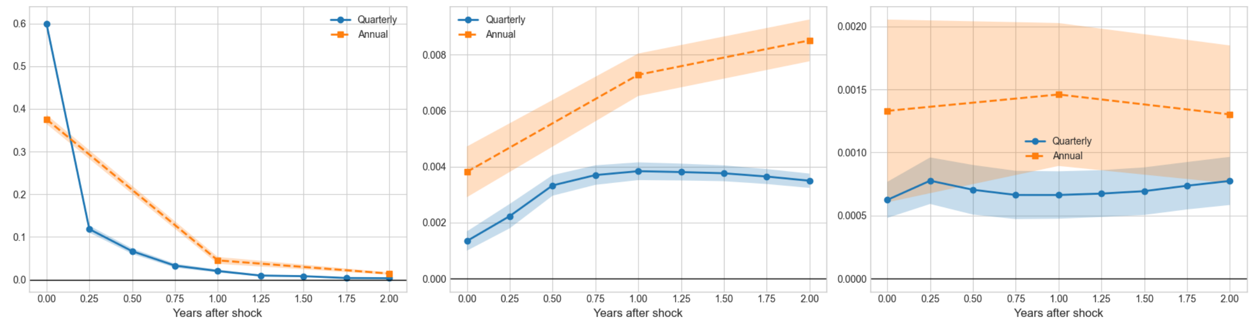
Table 11 compares the inverse variance-weighted multipliers estimated from quarterly and annual data across different country groups. The results reveal substantial differences between estimates computed at the two frequencies, even when aggregated across countries. These discrepancies underscore the potential bias inherent in using annual data. Two important factors can influence this bias. First, as noted in Section 4, annual shocks tend to be more predictable—not only in our study but more broadly in the literature—leading to bias due to omitted macroeconomic variables. Second, even if shocks are unpredictable, aggregation itself introduces bias. This point is emphasized by Stram and Wei (1986), who show that if the true data-generating process at the quarterly frequency is autoregressive (AR), aggregation to annual frequency transforms it into an autoregressive moving average (ARMA) process. Estimating a VAR at the annual level while ignoring the moving average (MA) component leads to model misspecification and biased estimates of impulse responses and fiscal multipliers.

Table 11: Inverse Variance- Weighted Multipliers Across Samples: Quarterly vs. Annual

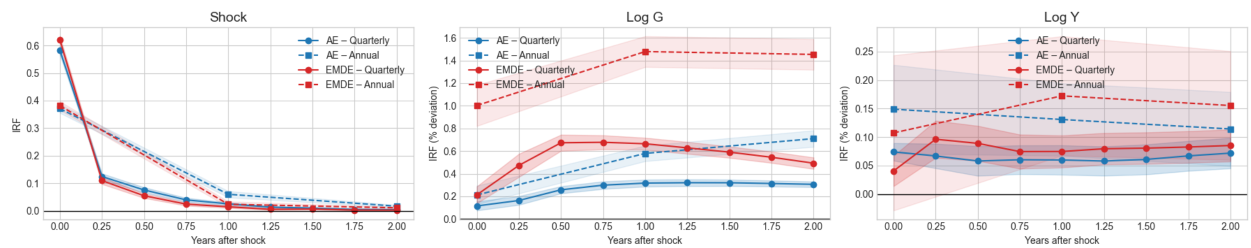
Sample	1-year multiplier		2-year multiplier	
	Quarterly	Annual	Quarterly	Annual
Full Sample	0.74	0.72	0.67	0.61
Advanced Economies	0.70	0.84	0.61	0.63
EMDEs	0.80	0.58	0.81	0.60

Notes: "Quarterly" multipliers are estimated from the baseline quarterly VAR using quarterly data. "Annual" multipliers are estimated from the corresponding annual-frequency specification using annualized data constructed from the same underlying series, with the number of lags changed from 8 to 2. Entries report inverse-variance weighted average fiscal multipliers across countries . For each horizon, the multiplier is computed as the cumulative ratio of the median inverse-variance-weighted impulse responses of output and government spending to an expenditure shock , scaled by the median government expenditure share. Horizons correspond to 1 year and 2 years at the relevant data frequency. Source: Authors' calculations.

Figure 5: Quarterly vs. Annual IRFs



(a) Full sample of countries



(b) Subsamples of AE and EMDE countries

Note: “Quarterly” multipliers are estimated from the baseline quarterly VAR using quarterly data. “Annual” multipliers are estimated from the corresponding annual-frequency specification using annualized data constructed from the same underlying series, with the number of lags changed from 8 to 2. Solid lines report inverse-variance–weighted averages of country-specific impulse responses. For each horizon, country-level impulse responses are summarized by their posterior medians and pooled across countries within each group using weighted least squares with a constant only, where weights are given by the inverse of the estimated variance of the country-specific response. Shaded areas denote ± 1 standard error confidence bands computed from the weighted regression. Horizons are in years. Source: Authors’ calculations.

The country-by-country internal instrument VAR produces country-specific estimates of spending multipliers and it is useful to describe their *distribution* across countries. However, this approach is less suited to examine whether multipliers differ systematically with *slow-moving* structural characteristics (e.g., openness, exchange-rate regimes, labor market institutions) because the limited number of observations and degrees of freedom in the cross-country analysis constrain the precision of estimated factors affecting multipliers. To address this limitation and benchmark our narrative shocks against the canonical cross-country evidence, we complement this approach with a panel exercise that targets *group-average* dynamics using the full country–quarter panel within each group.

5 Cross-country heterogeneity

This section evaluates whether multipliers implied by our quarterly narrative proxy exhibit the cross-country regularities emphasized by Ilzetzki et al. (2013) (IMV). Following IMV, we work with a pooled quarterly panel and estimate the dynamic effects of government spending separately for groups of countries defined by slow-moving structural and institutional characteristics. The resulting estimates are *group-average* multipliers and are intended as benchmark evidence on systematic heterogeneity across durable country characteristics.

5.1 Relation to IMV and to our baseline country-by-country estimates

Ilzetzki et al. (2013) (IMV) identify government spending shocks as structural innovations in an SVAR estimated on fiscal aggregates and macro outcomes in a pooled panel setting. Our identifying variation differs in nature. The narrative series we construct from *EIU* Country Reports is a qualitative indicator of discretionary fiscal actions (expansion, contraction, or no action) and does not measure the intensity of spending. For this reason, it is not naturally treated as a quantitative endogenous fiscal aggregate in the manner of IMV. We therefore retain IMV's pooled-panel SVAR design for comparability, but identify the spending shock using the narrative series as a *proxy* (external instrument) for the innovation to measured government spending.

The pooled benchmark also complements our baseline country-by-country results in Section 4.3. There, we *internalize* the narrative proxy by including it as the first variable in a country-specific recursive VAR and using the first Cholesky innovation as the identified shock. This "internal instrument" VAR targets the LP-IV estimand and—under invertibility—coincides with the usual proxy-SVAR/SVAR-IV estimand for relative impulse responses. In the pooled panel, by contrast, we implement the external-instrument representation directly. This is convenient with two-way fixed effects and aligns naturally with the qualitative nature of the proxy.

5.2 A panel proxy-SVAR framework with a qualitative instrument

This subsection lays out a general panel proxy-SVAR framework used here and in the following sections.

5.2.1 Panel VAR with fixed effects

Let $i \in \{1, \dots, N\}$ index countries and $t \in \{1, \dots, T_i\}$ quarters (the panel may be unbalanced). Let

$$x_{i,t} \equiv \begin{pmatrix} g_{i,t} \\ y_{i,t} \end{pmatrix}$$

collect transformed real government spending and real GDP. The transformation is application-specific (e.g., $\Delta \log$ or log levels) and is always stated with the results.

Fix a set of observations $\mathcal{S}$ (e.g., a subsample defined by a characteristic or state). Within $\mathcal{S}$, we estimate a pooled panel VAR with fixed effects,

$$x_{i,t} = \sum_{\ell=1}^{p} A_\ell \, x_{i,t-\ell} + \mu_i + \tau_t + u_{i,t}, \qquad x_{i,t} = \begin{pmatrix} g_{i,t} \\ y_{i,t} \end{pmatrix}, \tag{10}$$

where μ_i are country fixed effects, τ_t are time fixed effects (included when indicated), and $u_{i,t}$ are reduced-form innovations. Estimation of (10) is by least squares after removing the included fixed effects (within transformation). The estimation sample consists of all $(i, t) \in \mathcal{S}$ for which p lags of $x_{i,t}$ are available.

Remark (state dependence and interactions). Later sections allow the law of motion for $x_{i,t}$ to depend on predetermined state variables through sample restrictions (estimating (10) on state-specific subsamples) or through interactions (e.g., adding regressors of the form $s_{i,t}x_{i,t-\ell}$). These variants remain linear in parameters and are estimated by least squares after removing fixed effects. Conditional on a specified VAR, the proxy identification, impulse-response construction, and inference described below are unchanged.

5.2.2 Structural representation and proxy-SVAR identification

Let $u_{i,t}$ denote the reduced-form innovations from the estimated panel VAR (or its stated variant). We posit a structural representation

$$u_{i,t} = B\,\varepsilon_{i,t}, \qquad \mathbb{E}[\varepsilon_{i,t}\varepsilon'_{i,t}] = I,$$

and define the *spending shock* as the first structural innovation, $\varepsilon^g_{i,t}$, with impact vector $b \equiv Be_1$ (so $u_{i,t} = b\,\varepsilon^g_{i,t} +$ other shocks). Let $z_{i,t}$ denote the narrative proxy.

To encompass both sample splits and smooth weighting schemes used later, we allow proxy moments to be formed with nonnegative weights $w_{i,t}$ (which may encode subgroup membership, state weights, or missingness of the proxy moment). Define an *effective proxy* $v_{i,t}$ as a scalar instrument constructed from $z_{i,t}$ (see below). Throughout, all proxy moments are computed after residualizing $v_{i,t}$ (and any instrument components used to construct it) with respect to the *same* fixed effects included in the VAR. We denote the resulting residual by $\widetilde{v}_{i,t}$.

Assumption PS (proxy relevance and exogeneity). For the environment indexed by $(\mathcal{S}, w)$, the effective proxy satisfies: (i) *relevance* $\mathbb{E}\big[w_{i,t}\,\widetilde{v}_{i,t}\,\varepsilon^g_{i,t}\big] \neq 0$; (ii) *exogeneity* $\mathbb{E}\big[w_{i,t}\,\widetilde{v}_{i,t}\,\varepsilon^j_{i,t}\big] = 0$ for all structural shocks $j \neq g$.

Under Assumption PS, the impact vector is identified up to scale by the weighted proxy moment

$$b \;\propto\; \mathbb{E}[w_{i,t}\,u_{i,t}\,\widetilde{v}_{i,t}]\,. \tag{11}$$

In practice, we compute the sample analogue using VAR residuals $\widehat{u}_{i,t}$ and the residualized proxy $\widetilde{v}_{i,t}$ on $(i,t) \in \mathcal{S}$:

$$\widehat{c} \equiv \widehat{\mathbb{E}}_{\mathcal{S}}[w_{i,t}\,\widehat{u}_{i,t}\,\widetilde{v}_{i,t}]\,, \qquad \widehat{b} \equiv \frac{\widehat{c}}{\widehat{c}_g},$$

where $\widehat{c}_g$ denotes the first element of $\widehat{c}$ (the weighted covariance with the spending residual). This normalization sets the impact response of spending to unity, $b_g = 1$, thereby fixing the scale of the identified shock.

With two endogenous variables, the normalization yields the familiar covariance ratio representation:

$$\widehat{b} = \begin{pmatrix} 1 \\ \dfrac{\widehat{\mathrm{Cov}}_{\mathcal{S},w}(\widehat{u}^y_{i,t},\ \widetilde{v}_{i,t})}{\widehat{\mathrm{Cov}}_{\mathcal{S},w}(\widehat{u}^g_{i,t},\ \widetilde{v}_{i,t})} \end{pmatrix}, \tag{12}$$

where $\widehat{\mathrm{Cov}}_{\mathcal{S},w}(\cdot,\cdot)$ denotes the weighted sample covariance over $(i,t) \in \mathcal{S}$.[8]

5.2.3 Instrument processing for a qualitative and sparse proxy

In our setting, the raw narrative proxy $z_{i,t}$ is qualitative and sparse. We therefore allow the effective proxy $v_{i,t}$ to be a constructed scalar instrument formed from a (possibly multivariate) instrument vector $q_{i,t}$ that is itself a function of $z_{i,t}$ (e.g., sign indicators and lags). Let $q_{i,t}$ be a $K \times 1$ vector of instrument components, and let $\widetilde{q}_{i,t}$ denote the residualized version obtained by removing the same fixed effects as in the VAR. We strengthen relevance by projecting the spending residual onto $\widetilde{q}_{i,t}$ within $\mathcal{S}$:

$$\widehat{u}^g_{i,t} = \pi' \widetilde{q}_{i,t} + e_{i,t}, \tag{13}$$

and define the fitted component

$$v_{i,t} \equiv \widehat{\pi}' \widetilde{q}_{i,t}, \qquad \widetilde{v}_{i,t} \equiv v_{i,t} \ \ \text{(since } v_{i,t} \text{ is already residualized through } \widetilde{q}_{i,t}\text{)}.$$

We then use $\widetilde{v}_{i,t}$ in the proxy moment (11) and in the impact mapping (12). This two-step construction is algebraically equivalent to using the multivariate instrument $\widetilde{q}_{i,t}$ directly in the proxy-SVAR moment conditions, but is convenient in implementation and yields a scalar effective proxy.

Zeros and missing proxy information. Because $z_{i,t}$ may equal zero in quarters with no narrative action, we distinguish two empirically relevant conventions that recur in the paper: (i) *zeros-as-zero*, which treats $z_{i,t} = 0$ as a literal value contributing to the proxy moment; and (ii) *zeros-as-missing*, which treats such quarters as containing no instrument information

[8] Equivalently, (12) is the just-identified IV estimand obtained by instrumenting $\widehat{u}^g_{i,t}$ with $\widetilde{v}_{i,t}$ in a regression of $\widehat{u}^y_{i,t}$ on $\widehat{u}^g_{i,t}$ (with the same fixed-effect residualization).

by setting the proxy-moment weight to zero, $w_{i,t} = 0$ when $z_{i,t} = 0$, while retaining those quarters in the VAR estimation. Each application states which convention is used.

5.2.4 Impulse responses and multiplier construction

Let $\{\widehat{A}_\ell\}_{\ell=1}^{p}$ denote the estimated VAR coefficients (or their stated state-dependent counterparts evaluated at the relevant state). The implied moving-average representation yields matrices $\{\widehat{\Psi}_h\}_{h=0}^{H}$ with $\widehat{\Psi}_0 = I$ and recursion $\widehat{\Psi}_h = \sum_{\ell=1}^{\min\{p,h\}} \widehat{A}_\ell \, \widehat{\Psi}_{h-\ell}$. Impulse responses to the identified spending shock are

$$\widehat{\mathrm{IRF}}(h) = \widehat{\Psi}_h \, \widehat{b}, \qquad h = 0, 1, \ldots, H. \tag{14}$$

We map impulse responses into *dollar multipliers* using the average spending share computed from the underlying levels data. Let $s_{i,t} \equiv G_{i,t}/Y_{i,t}$ and define the environment-specific average spending share as a (weighted) trimmed mean,

$$\bar{s} \equiv \mathrm{TrimMean}_{\mathcal{S},w}(s_{i,t}),$$

where trimming limits sensitivity to outliers.[9]

Multiplier formulas depend on the transformation used in the VAR. In the benchmark $\Delta \log$ representation, impulse responses are for growth rates and are cumulated over horizons. Let $\widehat{\mathrm{IRF}}_g(h)$ and $\widehat{\mathrm{IRF}}_y(h)$ denote the spending and output responses in (14). The cumulative multiplier at horizon H is

$$\widehat{\mathcal{M}}(H) = \frac{1}{\bar{s}} \frac{\sum_{h=0}^{H} \widehat{\mathrm{IRF}}_y(h)}{\sum_{h=0}^{H} \widehat{\mathrm{IRF}}_g(h)}. \tag{15}$$

For a VAR in log levels, the corresponding multiplier uses level-log impulse responses at horizon H (i.e., without cumulation); each application reports the relevant formula when log levels are used.

[9] The trim rate and the construction of quarterly G/Y (including any smoothing/interpolation used to align levels) are recorded in the replication code.

5.2.5 Bootstrap inference

Inference is based on a country-cluster bootstrap designed to preserve within-country serial dependence. For each bootstrap replication, we resample countries with replacement, reconstruct the bootstrap panel by stacking the full time series of sampled countries, re-estimate the panel VAR (or its stated variant), rebuild the effective proxy (including the first-stage (13) when used), recompute the impact vector (12), and recompute impulse responses and multipliers (14)–(15). Reported intervals are 90% bootstrap percentile bands. Two-sided bootstrap p-values for $H_0 : \mathcal{M}(H) = 0$ are computed from the empirical distribution of bootstrap replicates.

Because (15) is a ratio, a small number of bootstrap draws can yield unstable values when the cumulative spending response in the denominator is close to zero. We therefore apply mild acceptance filters that exclude draws with near-zero denominators or implausibly extreme multipliers. The specific bounds are stated with the results and are not consequential for the reported patterns.

5.3 Implementation in this section: pooled benchmarks within groups

We apply the framework above to reproduce IMV-style pooled-panel benchmarks and to study heterogeneity across slow-moving structural characteristics. For a given characteristic, let G denote the corresponding subsample of country–quarter observations.[10]

Within each group G, we set $\mathcal{S} = G$ and estimate the two-way fixed-effects panel VAR (10) on the stacked country–quarter sample. For Table 12, the benchmark transformation is quarterly log changes, $g_{i,t} = \Delta \log G_{i,t}$ and $y_{i,t} = \Delta \log Y_{i,t}$. Identification uses the narrative proxy $z_{i,t} \in \{-1, 0, +1\}$ via the effective proxy construction in Section 5.2.3: we form $q_{i,t}$ from positive/negative narrative indicators and a small number of their lags, estimate the first stage (13) within G after removing the same fixed effects, and use the fitted component

[10] For slow-moving characteristics (trade openness, informality, and labor-market institutions), G is defined at the country level using country-level summaries of each characteristic. For public debt, G is defined at the country–quarter level; we apply a minimum-duration rule to avoid transient reclassification of debt states. Cutoffs are pre-specified and reported in Table 12. For trade openness we use IMV's baseline cutoff (trade/GDP = 60). For informality and the labor reform index we split at the sample median of country-level medians, yielding cutoffs 25.25 and 0.73, respectively; this choice produces balanced groups and avoids ex-post selection over thresholds. The debt cutoff is set at debt/GDP = 70.

$v_{i,t}$ in (12). The benchmark adopts the *zeros-as-zero* convention for $z_{i,t}$.

We report cumulative multipliers at one- and two-year horizons, $\widehat{\mathcal{M}}_G(4)$ and $\widehat{\mathcal{M}}_G(8)$, computed from (15) and scaled by the group-specific trimmed-mean spending share $\bar{s}_G = \text{TrimMean}_G(G/Y)$. Uncertainty bands are 90% country-cluster bootstrap percentile intervals constructed as in Section 5.2.5.

Table 12: Benchmark Panel Heterogeneity by Structural Characteristics

	Countries	$\mathcal{M}_4$	$\mathcal{M}_8$
Panel A. Trade openness			
Closed	26	1.425**	1.364**
		(0.437, 1.993)	(0.418, 1.935)
Open	34	0.558**	0.649**
		(0.075, 1.094)	(0.088, 1.165)
Panel B. Public debt			
Low/No high debt	56	0.738**	0.893**
		(0.118, 1.481)	(0.146, 1.669)
High debt	26	0.821**	0.851**
		(0.185, 1.856)	(0.188, 2.008)
Panel C. Exchange-rate regime split			
Fixed	51	1.104***	1.119***
		(0.372, 1.624)	(0.379, 1.653)
Float	8	0.715	0.703
		(-0.053, 1.881)	(-0.019, 1.745)
Panel D. Informality split			
Low informality	28	1.173**	1.195***
		(0.403, 1.504)	(0.434, 1.494)
High informality	28	0.788**	0.798***
		(0.092, 1.669)	(0.097, 1.687)
Panel E. Labor flexibility			
Low	25	1.439***	1.347***
		(0.555, 1.796)	(0.523, 1.613)
High	24	1.560**	1.662**
		(0.410, 1.993)	(0.427, 2.140)

Notes: Group assignments are based on the following pre-defined split rules. *Trade openness:* each country is classified as *open* if its median trade-to-GDP ratio (over the estimation sample) exceeds 60 (percent), and *closed* otherwise. *Public debt:* the *high-debt* state is defined as country–quarters that belong to an episode in which the debt-to-GDP ratio exceeds 70 (percent); all remaining country–quarters are classified as *low/no high debt*. *Exchange-rate regime:* each country is classified as *floating* if the share of quarters coded as floating (`float_dummy` = 1) exceeds 0.25, and *fixed/non-floating* otherwise. *Informality:* each country is classified as *high informality* if its median informality measure is above the overall median. *Labor-market flexibility:* each country is classified as *high flexibility* if its median labor-reform index exceeds the overall median. Countries without sufficient data to compute the relevant classification are excluded. Entries report cumulative multipliers computed as the ratio of the cumulative response of real GDP to the cumulative response of real government spending, scaled by the group-specific average spending share (Equation (9)). Parentheses report 90% country-cluster bootstrap confidence intervals. *, **, and *** denote rejection of $H_0 : \mathcal{M} = 0$ at the 10%, 5%, and 1% levels. *Source:* Authors' calculations.

Table 12 yields five benchmark results. We begin with the dimensions emphasized by Ilzetzki et al. (2013) (IMV). First, the trade-openness split reproduces IMV's central regularity that multipliers are smaller in more open economies. The one-year multiplier is 1.425 for relatively closed economies and 0.558 for more open economies, with the same ranking at two years ($\mathcal{M}_8 = 1.364$ versus 0.649). This pattern is consistent with import-leakage logic: a larger share of additional demand is absorbed by imports in more open economies, attenuating the domestic output response.

Second, the debt split indicates limited debt-state heterogeneity in this benchmark. The one- and two-year multipliers are similar in low/no-high-debt and high-debt states ($\mathcal{M}_4 = 0.738$ versus 0.821 and $\mathcal{M}_8 = 0.893$ versus 0.851). In our pooled panel design, multipliers vary more sharply with openness and the exchange-rate regime than with the debt state.

Third, splitting by exchange-rate regime delivers the other canonical IMV benchmark: multipliers are larger under fixed exchange-rate arrangements. For fixed regimes, the cumulative multiplier is close to unity and precisely estimated ($\mathcal{M}_4 = 1.104$ and $\mathcal{M}_8 = 1.119$, both statistically different from zero). For floating regimes, point estimates are smaller ($\mathcal{M}_4 = 0.715$ and $\mathcal{M}_8 = 0.703$), but inference is substantially weaker because the floating-regime subsample contains relatively few countries and the 90% intervals include zero. We therefore view the regime split primarily as strong evidence on the *ranking*—fixed above float—which is the key regularity emphasized in IMV.

We next turn to two additional dimensions highlighted in the broader cross-country fiscal-multiplier literature: the degree of labor market flexibility (e.g., Colombo et al. 2024) and the extent of labor market rigidity (e.g., Cacciatore et al. 2021). We measure informality using the index developed by Medina and Schneider (2018) and labor market rigidity using the index of employment protection legislation constructed by Alesina et al. (2024). The results suggest multipliers are smaller in more informal economies. In Panel D, the multiplier declines from 1.173 in the low-informality group to 0.788 in the high-informality group at one year, with a similarly sized gap at two years (1.195 versus 0.798). Thus, informality emerges as a quantitatively important margin for fiscal transmission in the pooled benchmark.

In addition, multipliers are larger in countries with more flexible labor markets. The gap is modest at one year (1.560 versus 1.439) and becomes more pronounced toward the two-year horizon (1.662 versus 1.347).

Taken together, this exercise shows that the quarterly narrative proxy delivers systematic cross-country heterogeneity along economically meaningful dimensions. In particular, the openness and exchange-rate regime splits reproduce the two canonical IMV regularities, providing a stringent external-validity check for the narrative identification in a pooled panel setting.

5.4 Lessons so far

The evidence in this section supports two conclusions. First, the pooled panel benchmarks reproduce the canonical IMV regularities—smaller multipliers in more open economies and larger multipliers under fixed exchange-rate arrangements—, as well as results in the literature relating multpliers to informality and labor market reforms. Taken together, these results provide an external-validity check for our narrative database.

The remainder of the results examine a complementary dimension of heterogeneity: *time-varying* state dependence. Whereas the previous benchmarks condition on slow-moving country characteristics, the next sections ask whether the effectiveness of exogenous spending actions varies with macroeconomic states that evolve at higher frequencies. We consider (i) slack and monetary-policy constraints around the effective lower bound, (ii) alternative measures of policy uncertainty, and (iii) political support and cohesion.

6 State of the business cycle

An extensive theoretical and empirical literature documents that spending multipliers tend to be larger during recessions and when increased spending is followed by monetary policy accommodation. (see, for example Ramey and Zubairy (2018), and the literature cited therein).

This section asks whether the causal effects of our narrative, discretionary spending actions vary systematically with *growth conditions*. We use a continuous, predetermined growth index and estimate a smooth-transition panel proxy-SVAR. This approach is particularly useful in our setting because the narrative proxy is noisy and sample splitting can sharply reduce precision.

6.1 Smooth-transition panel proxy-SVAR under growth conditions

This section applies the panel proxy-SVAR framework developed in Section 5.2. Relative to the pooled benchmark in Section 5.3, there are two modifications only. First, we allow the panel VAR dynamics to vary smoothly with a predetermined growth index. Second, we implement proxy identification using regime weights (high- versus low-growth) in the proxy moment and in the construction of environment-specific spending shares.

Growth index and smooth regime weights. Let $Y_{i,t}$ denote real GDP for country i in quarter t and define quarterly output growth as $\Delta \log Y_{i,t}$. We summarize local growth conditions using the lagged four-quarter moving average

$$s_{i,t} \equiv \frac{1}{4} \sum_{r=1}^{4} \Delta \log Y_{i,t-r}, \tag{16}$$

which is predetermined with respect to quarter-t shocks. To ensure comparability across countries, we standardize $s_{i,t}$ within country,

$$\widetilde{s}_{i,t} \equiv \frac{s_{i,t} - \mathbb{E}(s_{i,\cdot})}{\mathrm{sd}(s_{i,\cdot})}.$$

We then map $\widetilde{s}_{i,t}$ into a smooth *low-growth weight* $F_{i,t} \in (0, 1)$ using the logistic transition function

$$F_{i,t} \equiv \frac{1}{1 + \exp\{\gamma(\widetilde{s}_{i,t} - c)\}}, \qquad \gamma > 0, \tag{17}$$

so that $F_{i,t}$ is larger in weak-growth environments. In the benchmark we set $c = 0$ and $\gamma = 1.5$, so that $F_{i,t} \approx 0.5$ at average standardized growth and transitions smoothly across growth conditions.[11]

[11]For reporting, we evaluate impulse responses and multipliers at two representative values of $F_{i,t}$: the 25th and 75th percentiles of the sample distribution. In the data these correspond to $F^{\mathrm{HighGrow}} = 0.3167$ and $F^{\mathrm{LowGrow}} = 0.6399$, respectively (denoted "HighGrowth_p25" and "LowGrowth_p75").

Smooth-transition panel VAR. Let $z_{i,t} \in \{-1, 0, +1\}$ denote the narrative spending-action proxy. We estimate a pooled smooth-transition panel VAR with country fixed effects. For each transformation we consider the two-dimensional endogenous vector

$$x_{i,t} \equiv \begin{bmatrix} x^g_{i,t} \\ x^y_{i,t} \end{bmatrix}, \qquad \text{where either } (x^g_{i,t}, x^y_{i,t}) = (\Delta \log G_{i,t}, \Delta \log Y_{i,t}) \text{ or } (\log G_{i,t}, \log Y_{i,t}),$$

with $G_{i,t}$ real government spending and $Y_{i,t}$ real GDP. With $p = 8$ lags, the smooth-transition panel VAR is

$$x_{i,t} = \alpha_i + \sum_{\ell=1}^{p} A_\ell x_{i,t-\ell} + \sum_{\ell=1}^{p} B_\ell \big(F_{i,t}\, x_{i,t-\ell}\big) + u_{i,t}, \tag{18}$$

where α_i are country fixed effects and $u_{i,t}$ are reduced-form residuals. When $F_{i,t} = 0$ (high growth) the lag coefficients equal A_ℓ, and when $F_{i,t} = 1$ (low growth) they equal $A_\ell + B_\ell$. Estimation is by least squares after removing α_i.

Proxy identification with regime weights. Identification follows Section 5.2.2, using the narrative proxy as an external instrument and allowing the proxy moment to vary with the growth environment through weights. Define regime weights

$$w^{\text{Low}}_{i,t} \equiv F_{i,t}, \qquad w^{\text{High}}_{i,t} \equiv 1 - F_{i,t}.$$

Consistent with the *zeros-as-missing* convention of Section 5.2.3, quarters with $z_{i,t} = 0$ are retained in the VAR estimation but contribute no instrument information in the proxy moment; operationally, we replace the regime weights by $w^\kappa_{i,t} \cdot \mathbf{1}\{z_{i,t} \neq 0\}$. Let $\widetilde{z}_{i,t}$ denote the proxy residualized with respect to the included country fixed effects. For each regime $\kappa \in \{\text{High}, \text{Low}\}$, the impact vector is identified (up to scale) by the weighted proxy moment

$$b^\kappa \propto \mathbb{E}\big[w^\kappa_{i,t}\, u_{i,t}\, \widetilde{z}_{i,t}\big], \qquad \text{normalized by } b^\kappa_g = 1, \tag{19}$$

so the spending impact response is normalized to one and the output impact response is scaled relative to spending.

Our smooth-transition specification closely follows the approach of Auerbach and Gorodnichenko (2012), who model fiscal multipliers as varying smoothly with the business cycle using a logistic transition function applied to lagged output growth. As in their framework, the transition variable is predetermined and standardized, and the logistic function governs the degree to which the economy is in a "low-growth" versus "high-growth" regime. The key difference in our setting is that we embed this smooth-transition structure within a panel proxy-SVAR framework and allow the external-instrument moment itself to vary with regime weights.

6.2 Multipliers across growth conditions

Table 13 reports cumulative real spending multipliers at one- and two-year horizons for high-growth and low-growth conditions. We focus on two pooled-panel specifications that are scale-invariant across countries: a benchmark VAR in growth rates ($\Delta \log$) and a VAR in log levels. In both cases we treat $z_{i,t} = 0$ as missing in the proxy moment, which yields substantially stronger weighted proxy relevance than treating zeros as literal values.

Two patterns emerge. First, the point estimates suggest stronger multipliers in weak-growth environments. Under the $\Delta \log$ specification, the high-growth multiplier is essentially zero (-0.01 at both $H = 4$ and $H = 8$), while the low-growth multiplier is about 0.77 at both horizons. Under the log-level specification, multipliers are positive in both growth conditions but again larger under low growth: at $H = 4$ the multiplier rises from 0.25 in high growth to 0.74 in low growth, and at $H = 8$ it rises from 0.58 to 1.39.

Second, these differences are suggestive but not precisely estimated. We would only be able to reject the hypothesis of equal values at about 10 percent confidence only for the baseline specification at one year ahead. For the benchmark $\Delta \log$ specification, at $H = 4$ (bootstrap $p = 0.11$) and at $H = 8$ ($p = 0.14$). For the log-level specification, the corresponding intervals are at $H = 4$ ($p = 0.39$) and at $H = 8$ ($p = 0.24$).

Table 13: Panel Multipliers Across Growth Conditions

Specification	1-year ($H = 4$) High growth	1-year ($H = 4$) Low growth	2-year ($H = 8$) High growth	2-year ($H = 8$) Low growth
Real (Δ log VAR)	-0.01	0.77	-0.01	0.77
Real (log-level VAR)	0.25	0.74	0.58	1.39

Notes: Entries report cumulative real spending multipliers from a smooth-transition panel proxy-SVAR with country fixed effects and $p = 8$ lags. Growth conditions are indexed by a predetermined continuous growth measure: $s_{i,t}$ is the lagged four-quarter moving average of $\Delta \log$ real GDP, standardized within country; $F_{i,t}$ is the logistic low-growth weight in (17) with $\gamma = 1.5$ and $c = 0$. "High growth" and "Low growth" evaluate the VAR at $F^{\text{HighGrow}} = 0.3167$ (25th percentile of F) and $F^{\text{LowGrow}} = 0.6399$ (75th percentile).

7 Policy uncertainty and fiscal multipliers

A large macro and public-finance literature emphasizes the role of economic and policy uncertainty for investment, consumption, and asset prices, and documents that fiscal multipliers can be smaller in periods of elevated uncertainty or policy risk.[12] Much of this work relies on uncertainty measures constructed from news, financial markets, or survey data and estimates multipliers that vary with such uncertainty states. By contrast, our narrative identification strategy delivers a quarterly panel of discretionary spending actions anchored in *EIU* Country Reports. This section studies whether the causal effects of these exogenous spending actions vary systematically with alternative measures of *policy uncertainty* constructed from the same *EIU* Country Reports, ranging from broad uncertainty to fiscal-policy-focused uncertainty indices derived from textual sources.

7.1 Uncertainty measures

We use uncertainty measures built from *EIU* Country Reports, ensuring coverage that closely matches the country coverage of our narrative spending-action proxy.

[12]See, for example, Bachmann and Sims (2012), Fernández-Villaverde et al. (2015), Baker et al. (2016), Azzimonti (2018), Fritsche et al. (2021), and Alloza (2022) for VAR- and DSGE-based evidence on how uncertainty and confidence shape the transmission of fiscal shocks.

***EIU*-text uncertainty: broad and fiscal uncertainty.** We draw on the World Uncertainty Index project of Ahir et al. (2022) and download the country-level quarterly series from their public data repository.[13] The broad World Uncertainty Index (WUI) is computed by counting the frequency of the word "uncertain" (and variants) in the *EIU* reports, scaled by report length and rescaled for convenience. We complement this with a fiscal-policy-focused uncertainty index (FUI) that applies the same *EIU* text-mining approach but restricts attention to fiscal-policy and public-finance passages.

High/low uncertainty states. We classify quarters into high/low uncertainty states using transparent threshold rules applied to smoothed versions of the uncertainty indices. Let $U_{i,t}^{m}$ denote the quarterly uncertainty index for country i in quarter t, for concept $m \in \{\mathrm{WUI}, \mathrm{FUI}\}$. For each index we compute a four-quarter moving average,

$$\widetilde{U}_{i,t}^{m} \equiv \frac{1}{4} \sum_{s=0}^{3} U_{i,t-s}^{m}.$$

We then define a binary high-uncertainty indicator $\mathrm{Unc}_{i,t}^{m} \in \{0,1\}$ using 75th-percentile thresholds. In the benchmark WUI specification, the threshold is computed within-country:

$$\mathrm{Unc}_{i,t}^{\mathrm{WUI}} \equiv \mathbf{1}\left\{\widetilde{U}_{i,t}^{\mathrm{WUI}} \geq Q_{0.75}\left(\widetilde{U}_{i,\cdot}^{\mathrm{WUI}}\right)\right\},$$

where $Q_{0.75}(\widetilde{U}_{i,\cdot}^{\mathrm{WUI}})$ is the country-specific 75th percentile. In the benchmark FUI specification we use an analogous 75th-percentile rule applied to the pooled (global) distribution of $\widetilde{U}_{i,t}^{\mathrm{FUI}}$.[14] Quarters not classified as high uncertainty are assigned to the low-uncertainty state. In the data, these benchmark definitions classify roughly one quarter of observations as high uncertainty for each index (among defined observations).

[13] See `https://worlduncertaintyindex.com/data/`.

[14] In practice, the fiscal-policy-focused index features more limited coverage and a more discrete distribution than WUI; a pooled threshold yields a stable and comparable high/low split.

7.2 Panel proxy-SVAR under uncertainty

This subsection applies the panel proxy-SVAR framework developed in Section 5.2. Uncertainty is used solely to index the environment in which an already-identified exogenous spending action occurs. Relative to the pooled benchmark, the only change is that we estimate the panel VAR and the proxy moment separately for high- and low-uncertainty quarters, as defined in Section 7.1.

State-contingent panel VAR in growth rates. Fix an uncertainty concept $m \in \{\mathrm{WUI}, \mathrm{FUI}\}$ and define the state indicator $\mathrm{Unc}^m_{i,t} \in \{0,1\}$ as in Section 7.1. For each state $\kappa \in \{\mathrm{Low}m, \mathrm{High}m\}$, let

$$\mathcal{S}^{\kappa,m} \equiv \{(i,t):\ \mathrm{Unc}^m_{i,t} = \mathbf{1}\{\kappa = \mathrm{High}m\}\}$$

denote the corresponding state subsample. We then estimate a separate pooled panel VAR on $\mathcal{S}^{\kappa,m}$. In the benchmark specification, the endogenous panel vector is

$$x_{i,t} \equiv \begin{bmatrix} \Delta \log g_{i,t} \\ \Delta \log y_{i,t} \end{bmatrix},$$

where $g_{i,t}$ is real government spending and $y_{i,t}$ is real GDP. For each (κ, m) we estimate

$$x_{i,t} = A_1^{\kappa,m} x_{i,t-1} + \cdots + A_p^{\kappa,m} x_{i,t-p} + \alpha_i^{\kappa,m} + u_{i,t}^{\kappa,m}, \qquad \text{using observations in } \mathcal{S}^{\kappa,m}, \tag{20}$$

where $\alpha_i^{\kappa,m}$ are country fixed effects and we set $p = 4$.

Proxy identification within state. Within each state subsample $\mathcal{S}^{\kappa,m}$, identification follows Section 5.2.2 with the effective proxy set to the narrative proxy, $v_{i,t} = z_{i,t}$ (residualized with respect to the included fixed effects). Let $u_{i,t}^{\kappa,m} = (u_{g,i,t}^{\kappa,m}, u_{y,i,t}^{\kappa,m})'$ denote the reduced-form residuals from (20). The impact vector of the spending shock in state (κ, m) is identified (up to scale) from the state-specific proxy moment and normalized so the impact response of spending equals one:

$$b^{\kappa,m} \propto \mathbb{E}\left[u_{i,t}^{\kappa,m} z_{i,t} \;\middle|\; (i,t) \in \mathcal{S}^{\kappa,m}\right], \qquad \text{normalized by } b_g^{\kappa,m} = 1. \tag{21}$$

Impulse responses are obtained from the moving-average representation of the estimated state-specific VAR, as in Section 5.2.4.

Cumulative multipliers in dollars. Because (20) is estimated in $\Delta \log$, multiplier construction follows the cumulative-response mapping in Section 5.2.4. Let $\widehat{\mathrm{IRF}}^{\kappa,m}_{\Delta \log y}(h)$ and $\widehat{\mathrm{IRF}}^{\kappa,m}_{\Delta \log g}(h)$ denote the impulse responses in growth rates at horizon h to the identified spending shock. Define the state-specific average spending share as a trimmed mean of G/Y within $\mathcal{S}^{\kappa,m}$,

$$\bar{s}_G^{\kappa,m} \equiv \mathrm{TrimMean}\left(\frac{G_{i,t}}{Y_{i,t}} \;\middle|\; (i,t) \in \mathcal{S}^{\kappa,m} \right).$$

The cumulative multiplier at horizon H is then

$$\widehat{\mathcal{M}}^{\kappa,m}(H) = \frac{1}{\bar{s}_G^{\kappa,m}} \frac{\sum_{h=0}^{H} \widehat{\mathrm{IRF}}^{\kappa,m}_{\Delta \log y}(h)}{\sum_{h=0}^{H} \widehat{\mathrm{IRF}}^{\kappa,m}_{\Delta \log g}(h)}. \tag{22}$$

7.3 Multipliers across uncertainty states

Table 14 reports cumulative multipliers at one- and two-year horizons ($H = 4, 8$) in low- and high-uncertainty quarters. Two patterns emerge. First, for broad uncertainty (WUI), spending multipliers are lower in high-uncertainty quarters than in low-uncertainty quarters at both horizons. At $H = 4$, the cumulative multiplier declines from about 1.04 in low-WUI quarters to 0.51 in high-WUI quarters, and at $H = 8$ it declines from 1.09 to 0.42.

Second, fiscal-policy-focused uncertainty (FUI) yields a similar qualitative conclusion: in the preferred FUI specification, cumulative multipliers are also smaller in high-uncertainty quarters. At $H = 4$, the multiplier declines from 0.95 in low-FUI quarters to 0.21 in high-FUI quarters, and at $H = 8$ it declines from 1.13 to 0.18. Thus, under both uncertainty concepts, elevated uncertainty is associated with weaker fiscal transmission.

Bootstrap-based uncertainty intervals for the high-minus-low differences are wide and include zero, reflecting limited precision when splitting the sample into uncertainty states.

Table 14: Panel Multipliers Across Uncertainty States

	WUI			FUI		
Horizon	Low	High	High−Low	Low	High	High−Low
1-year ($H = 4$)	1.04	0.51	−0.52	0.95	0.21	−0.74
2-year ($H = 8$)	1.09	0.42	−0.67	1.13	0.18	−0.94

Notes: Entries report cumulative spending multipliers from state-contingent panel proxy-SVARs estimated separately in low- and high-uncertainty quarters. WUI is the broad World Uncertainty Index of Ahir et al. (2022) constructed from *EIU* Country Reports; FUI is the fiscal-policy-focused uncertainty index constructed from the same *EIU* text. The benchmark state construction smooths each index with a four-quarter moving average and defines the high-uncertainty state using 75th-percentile thresholds (see Section 7.1). The panel VAR is estimated in first differences of logs for real spending and real GDP with $p = 4$ lags and country fixed effects, separately by state.

Benchmark specifications: For WUI, the preferred specification uses the within-country threshold definition in Section 7.1 and identifies the spending shock using the signed narrative proxy $z_{i,t}$. For FUI, the preferred specification uses a pooled 75th-percentile threshold and a flexible proxy first stage based on positive/negative components of $z_{i,t}$ with one lag; all other elements of the estimation are unchanged.

8 Political support and fiscal multipliers

A large political-economy literature argues that the macroeconomic effects of fiscal policy, and the success of consolidations in particular, depend on the prevailing political environment. Majority support in parliament, cabinet stability, and the timing of elections shape both the composition and durability of fiscal packages, as well as the likelihood that governments stay the course when measures become unpopular (see, among others, Alesina et al. (1995); Alesina and Ardagna (2010); Alesina et al. (2019)).

More recently, macro and public-finance studies have started to measure fiscal multipliers explicitly as a function of political conditions. At the sub-national level, Carlino et al. (2023) use close U.S. gubernatorial elections as a quasi-random shock to state party control and show that the output response to federal transfers is substantially larger when governors are more inclined to spend the aid. In cross-country work, Coulombe (2021) estimates that a large share of post-war spending variation is driven by electoral politics and government ideology, while Duque Gabriel et al. (2025) document that multipliers around consolidation episodes are markedly higher when parliamentary fragmentation is low. Relatedly, Carmignani (2022) finds that government spending multipliers vary systematically with election timing. These contributions motivate treating political support and cohesion as potential state variables for fiscal transmission.

8.1 Narrative measures of political support and its primitives

We augment our spending-shock database with a quarterly narrative indicator of whether the *political environment is favorable* for adopting and sustaining fiscal measures. For each country-quarter, we use the *Domestic politics* and related sections of the *EIU Country Reports* to code a coarse support indicator $\text{Supp}_{i,t} \in \{0, 1\}$, where $\text{Supp}_{i,t} = 1$ indicates a broadly favorable political backdrop and $\text{Supp}_{i,t} = 0$ an unfavorable one. The coding is performed with a fixed, pre-specified large-language-model prompt in the same secure, non-adaptive environment used for the fiscal narrative series. The model returns short verbatim excerpts supporting each classification, which we retain for auditability.

A key challenge is that "support" is conceptually multi-dimensional: electoral timing, legislative cohesion, and unrest/gridlock may matter differently for fiscal implementation and for private-sector responses. To separate these mechanisms, we additionally code three *political primitives* from the same EIU text, each taking values in $\{\text{Yes}, \text{No}, \text{Unclear}\}$ with accompanying verbatim evidence: (i) *majority/cohesion* (working legislative majority or stable coalition), (ii) *election proximity* (national election scheduled within the current or next quarter), and (iii) *unrest or severe gridlock* (material unrest or legislative paralysis). This decomposition is used solely for better understanding the mechanisms behind the key results.

Missing political text is non-trivial in the EIU archive. Rather than imputing, our baseline specifications treat missing support as its own state. This prevents the estimates from implicitly attributing missingness to either $\text{Supp}_{i,t} = 0$ or $\text{Supp}_{i,t} = 1$.

8.2 Empirical design: pooled panel local projections with IV

Let $z_{i,t}$ denote our narrative spending-action proxy. Let $G_{i,t}$ be real government spending and $Y_{i,t}$ real GDP. We estimate pooled panel local projections with country and global time fixed effects and use $z_{i,t}$ as an instrument for the within-country change in government spending.

Define $\Delta g_{i,t} \equiv \Delta \ln G_{i,t}$ and $\Delta y_{i,t} \equiv \Delta \ln Y_{i,t}$. For each horizon $h \geq 0$, we form cumulative log changes from $t-1$ to $t+h$:

$$\Delta^h y_{i,t} \equiv \ln Y_{i,t+h} - \ln Y_{i,t-1}, \tag{23}$$

$$\Delta^h g_{i,t} \equiv \ln G_{i,t+h} - \ln G_{i,t-1}. \tag{24}$$

To allow for political-state dependence and to address missingness explicitly, we define two indicators: $S^1_{i,t} \equiv \mathbf{1}\{\text{Supp}_{i,t} = 1\}$ and $S^m_{i,t} \equiv \mathbf{1}\{\text{Supp}_{i,t} \text{ missing}\}$. The omitted baseline state is observed $\text{Supp}_{i,t} = 0$.

For each horizon h, we estimate the pooled LP-IV equation

$$\Delta^h y_{i,t} = \alpha_i + \tau_t + \beta_h \, \Delta g_{i,t} + \theta^1_h \, (\Delta g_{i,t} \times S^1_{i,t}) + \theta^m_h \, (\Delta g_{i,t} \times S^m_{i,t}) + \Gamma_h X_{i,t-1} + \varepsilon_{i,t+h}, \tag{25}$$

where α_i are country fixed effects, τ_t are global time effects, and $X_{i,t-1}$ includes four lags of $\Delta y_{i,t}$ and $\Delta g_{i,t}$ and standard macro controls (inflation, unemployment, policy rate, and a below-trend indicator). We instrument the endogenous regressors $\{\Delta g_{i,t}, \Delta g_{i,t} \times S^1_{i,t}, \Delta g_{i,t} \times S^m_{i,t}\}$ with $\{z_{i,t}, z_{i,t} \times S^1_{i,t}, z_{i,t} \times S^m_{i,t}\}$. Standard errors are clustered at the country level.

We summarize fiscal transmission by a cumulative multiplier at horizon h, computed as the ratio of reduced-form cumulative responses scaled by the sample median of G/Y:

$$\widehat{m}_h(s) \equiv \frac{1}{s_G} \frac{\widehat{\text{RF}}^y_h(s)}{\widehat{\text{RF}}^g_h(s)}, \qquad s \in \{0, 1, m\}, \tag{26}$$

where $\widehat{\text{RF}}^y_h(s)$ and $\widehat{\text{RF}}^g_h(s)$ denote the reduced-form effects of $z_{i,t}$ on $\Delta^h y_{i,t}$ and $\Delta^h g_{i,t}$ in state s. At horizons where the cumulative spending reduced form is extremely small in absolute value, the ratio is not informative; in figures and tables we therefore suppress $\widehat{m}_h(s)$ when $|\widehat{\text{RF}}^g_h(s)| < 0.005$. s_G is the sample median of G/Y.

8.3 Results: political support amplifies spending pass-through and multipliers

Table 15 reports cumulative multipliers by political support. In favorable political environments ($\text{Supp}_{i,t} = 1$), multipliers are consistently positive and economically meaningful: at horizons of one to three years, the scaled dollar multiplier is approximately 1.0–1.1. In contrast, under unfavorable support ($\text{Supp}_{i,t} = 0$) multipliers are close to zero (and occasionally slightly negative).

The main driver of this state dependence is the *realized spending response.* In low-support quarters the cumulative spending reduced form is small and unstable at short horizons, so that multipliers are not well-defined there. Under high support, by contrast, the same narrative shock translates into a substantially larger cumulative spending response, and output responds positively. Table 16 quantifies this implementation channel: the difference in the cumulative spending reduced form between support and no-support quarters is large and statistically significant at medium horizons (e.g., $\Delta\text{RF}^g_h(1 - 0) = 0.0182$ at $h = 4$, $t = 2.94$). These patterns are consistent with political support increasing the likelihood that announced or legislated spending actions are executed and sustained.

Table 15: Cumulative multipliers by political support (pooled LP-IV)

Horizon h	Support=0	Support=1	Support missing
$h = 4$	–	1.0590	1.1078
$h = 8$	0.1880	0.9627	-0.6868

Notes: Entries report pooled LP-IV multipliers at horizons $h \in \{4, 8\}$ (quarters). Support is the narrative political-environment indicator from EIU reports. Multipliers are suppressed ("–") when the cumulative spending reduced form satisfies $|\widehat{\text{RF}}^g_h(s)| < 0.005$; this occurs for Support=0 at $h = 4$. All specifications include country and time fixed effects, four lags of $\Delta \ln Y$ and $\Delta \ln G$, and baseline macro controls; standard errors are clustered by country. Source: Authors' calculations.

Table 16: Reduced-form implementation channel by political support

Horizon h	$\text{RF}^y_h(0)$	$\text{RF}^y_h(1)$	$\Delta\text{RF}^y_h(1-0)$	$\text{RF}^g_h(0)$	$\text{RF}^g_h(1)$	$\Delta\text{RF}^g_h(1-0)$
$h=4$	0.00048	0.00524	0.00476 (1.71)	-0.00330	0.01490	0.01819 (2.94)
$h=8$	0.00061	0.00621	0.00560 (1.45)	0.00979	0.01943	0.00964 (1.65)

Notes: Reduced-form coefficients from pooled local projections with country and time fixed effects and baseline controls. $\text{RF}^y_h(s)$ and $\text{RF}^g_h(s)$ denote the cumulative reduced-form effects of $z_{i,t}$ on $\Delta^h y_{i,t}$ and $\Delta^h g_{i,t}$ in support state $s \in \{0,1\}$. $\Delta\text{RF}(1-0)$ is the coefficient on the interaction $z_{i,t} \times S^1_{i,t}$. t-statistics in parentheses use country-clustered standard errors. Source: Authors' calculations.

8.4 Which political dimension matters? Elections attenuate spending pass-through

We next ask whether the support effect is mediated by specific political dimensions. We estimate reduced-form local projections that include both support interactions and primitive interactions simultaneously (majority/cohesion, election proximity, and unrest/gridlock), along with controls for missingness and "unclear" classifications. Table 17 reports key coefficients.

Two findings stand out. First, even controlling for primitives, favorable support remains associated with a significantly stronger cumulative spending response at medium horizons: the coefficient on $z_{i,t} \times S^1_{i,t}$ in the spending reduced form is about 0.02 at $h = 4$ and $h = 8$ (both $t \approx 2.2$). Second, election proximity is independently associated with substantially weaker spending pass-through: the coefficient on $z_{i,t} \times \text{ElectYes}_{i,t}$ in the spending reduced form is around -0.018 to -0.020 at $h = 4$ and $h = 8$ (both statistically significant). The magnitudes are comparable, implying that election proximity can nearly offset the implementation advantage associated with political support.

A potential concern is that shocks may differ across political states. We therefore verify that the sign composition of narrative shocks is essentially identical across observed support states: the share of $z = -1$ shocks is 0.551 when Support=0 and 0.546 when Support=1, suggesting that the support-state results are not driven by simple sign selection.

Table 17: Decomposing political support into primitives: reduced-form spending pass-through

Horizon h	$z \times S^1$ in RF^g	$z \times$ ElectYes in RF^g	Net: $(z \times S^1) + (z \times \text{ElectYes})$
$h = 4$	0.02044 (2.17)	-0.01956 (-2.20)	0.00088
$h = 8$	0.01955 (2.26)	-0.01834 (-2.75)	0.00121

Notes: Entries come from reduced-form pooled local projections for cumulative spending $\Delta^h g_{i,t}$ on $z_{i,t}$ interacted with (i) observed support and (ii) primitive indicators, including missing/unclear controls, country and time fixed effects, and baseline macro controls. The sample is restricted to observed support (Support $\in \{0, 1\}$). t-statistics in parentheses are clustered by country. The final column reports the point-estimate net effect of support and election proximity on spending pass-through (sum of coefficients). Source: Authors' calculations.

9 Conclusion

This paper develops the first AI-assisted narrative approach to identifying government spending shocks at quarterly frequency and uses it to build a new global database of fiscal actions. Building on Romer and Romer (2010, 2019), we show that an off-the-shelf large language model (GPT–4.1), operated under a fixed prompt in a secure, non-adaptive environment, can reliably extract fiscal stance and motivations from textual sources. The procedure is conservative: prompts are pre-specified and uniform across countries and time, the model is used without task-specific fine-tuning, and only actions with motives clearly orthogonal to contemporaneous macroeconomic conditions are retained as exogenous. The resulting database provides quarterly sign shocks—expansions and contractions—for an unbalanced panel of 64 countries beginning in 1952:Q1.

To estimate the macroeconomic effects, we embed the qualitative shock series in country-specific Bayesian VARs. The baseline model orders the narrative proxy first, followed by real government spending and real GDP. The first orthogonalized innovation serves as the proxy-identified fiscal shock; cumulative spending multipliers are calculated as ratios of cumulated GDP to spending responses. Since the proxy is qualitative, the shock's magnitude is pinned down by spending responses, making multipliers invariant to rescaling of the narrative series. Baseline estimates yield moderate multipliers within familiar ranges: about 0.74 at one year and 0.67 at two years for the full sample. Advanced economies show multipliers below one at both horizons (0.70 at one year and 0.61 at two years), consistent with the range of existing time-series evidence. Emerging market and developing economies exhibit slightly larger median multipliers (0.80 and 0.81), but with wider dispersion, highlighting that output effects vary significantly by country context and policies.

The cross-country and quarterly scope of the database also enables systematic evidence on heterogeneity in fiscal transmission. Along standard dimensions emphasized in the open-economy literature, multipliers are smaller in more open economies, and under fixed exchange-rate regimes compared to floating exchange-rate regimes. Consistent with state-dependent fiscal transmission, multipliers are higher during downturns. Finally, while multipliers are smaller when broad-economic-policy uncertainty or fiscal policy-specific uncertainty is high. Favorable political support, in turn, is associated with larger *conditional* multipliers. The evidence indicates that this state dependence operates primarily through implementation: narrative spending shocks translate into a substantially stronger realized government-spending response when political support is high, with correspondingly larger output responses. Moreover, election proximity is associated with a markedly weaker spending pass-through, suggesting that electoral timing can attenuate the effectiveness of discretionary spending actions.

Methodologically, the paper demonstrates how large language models can be integrated into empirical macroeconomics without compromising identification. The AI's role is limited to a transparent, pre-specified classification task; all prompts and coding rubrics are documented; and the resulting machine-readable database, scripts, and country listings are provided for replication. The approach relies on the qualitative sign of exogenous actions for identification in the internal-instrument VAR, with the scale of structural shocks set by spending responses. Thus, the narrative layer is already rich, and the main constraint on expanding coverage lies not in AI capabilities but in the availability of consistent quarterly macro-fiscal data.

Several directions for future work follow naturally. First, the methodology can be adapted to other narrative sources and policy domains, including tax changes, credit and regulatory interventions, and monetary or macroprudential actions, where relevant textual material is abundant but costly to code manually. Second, the cross-country database of exogenous spending actions could be integrated with richer micro-level data—on firm investment, household consumption, or distributional outcomes—to study the incidence and composition of fiscal shocks. Third, the political and uncertainty dimensions explored here could be expanded to include additional institutional features (such as fiscal rules or central bank independence) and to investigate interactions between fiscal and monetary policy.

References

Adler, G., Allen, C., Ganelli, G., and Leigh, D. (2024). An Updated Action-based Dataset of Fiscal Consolidation. *IMF Working Papers*, 2024(210):1.

Ahir, H., Bloom, N., and Furceri, D. (2022). The world uncertainty index. *NBER Working Paper Series*, (29763).

Alesina, A. and Ardagna, S. (2010). Large changes in fiscal policy: Taxes versus spending. *Tax Policy and the Economy*, 24(1):35–68.

Alesina, A., Favero, C., and Giavazzi, F. (2019). Effects of austerity: Expenditure- and tax-based approaches. *Journal of Economic Perspectives*, 33(2):141–62.

Alesina, A., Furceri, D., Ostry, J. D., Papageorgiou, C., and Quinn, D. P. (2024). Structural reforms and elections: Evidence from a world-wide new dataset. *Journal of the European Economic Association*, 22(4):1936–1980.

Alesina, A., Perotti, R., Giavazzi, F., and Kollintzas, T. (1995). Fiscal expansions and adjustments in oecd countries. *Economic Policy*, 10(21):207–248.

Alloza, M. (2022). Is fiscal policy more effective during recessions? *International Economic Review*, 63(3):1271–1292.

Aruoba, S. B. and Drechsel, T. (2026). Identifying monetary policy shocks: A natural language approach. *American Economic Journal: Macroeconomics*. Forthcoming.

Auerbach, A. J. and Gorodnichenko, Y. (2011). Fiscal multipliers in recession and expansion. Working Paper 17447, National Bureau of Economic Research.

Auerbach, A. J. and Gorodnichenko, Y. (2012). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy*, 4(2):1–27.

Azzimonti, M. (2018). Partisan conflict and private investment. *Journal of Monetary Economics*, 93:114–131. Carnegie-Rochester-NYU Conference on Public Policy held at the Stern School of Business at New York University.

Bachmann, R. and Sims, E. R. (2012). Confidence and the transmission of government spending shocks. *Journal of Monetary Economics*, 59(3):235–249.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Cacciatore, M., Duval, R., Furceri, D., and Zdzienicka, A. (2021). Fiscal multipliers and job-protection regulation. *European Economic Review*, 132:103616.

Carlino, G. A., Drautzburg, T., Inman, R. P., and Zarra, N. (2023). Partisanship and fiscal policy in economic unions: Evidence from us states. *American Economic Review*, 113(3):701–737.

Carmignani, F. (2022). The electoral fiscal multiplier. *Economic Analysis and Policy*, 76:938–945.

Cloyne, J., Dimsdale, N., and Postel-Vinay, N. (2024). Taxes and growth: New narrative evidence from interwar britain. *The Review of Economic Studies*, 91(4):2168–2200.

Colombo, E., Furceri, D., Pizzuto, P., and Tirelli, P. (2024). Public expenditure multipliers and informality. *European Economic Review*, 164:104703.

Coulombe, R. G. (2021). The electoral origin of government spending shocks. *Journal of Economic Dynamics and Control*, 129:104167.

Duque Gabriel, R., Klein, M., and Nöller, M. (2025). Fiscal multipliers and political fragmentation. Ruhr Economic Papers 1175, RWI – Leibniz Institute for Economic Research.

Fernández-Villaverde, J., Guerrón-Quintana, P., Kuester, K., and Rubio-Ramírez, J. F. (2015). Fiscal volatility shocks and economic activity. *American Economic Review*, 105(11):3352–3384.

Fritsche, J. P., Klein, M., and Rieth, M. (2021). Government spending multipliers in (un)certain times. *Journal of Public Economics*, 203:104513.

Guajardo, J., Leigh, D., and Pescatori, A. (2014). Expansionary austerity? international evidence. *Journal of the European Economic Association*, 12(4):949–968.

Ilzetzki, E., Mendoza, E. G., and Vegh, C. A. (2013). How big (small?) are fiscal multipliers? *Journal of Monetary Economics*, 60(2):239–254.

Jordà, and Taylor, A. M. (2015). The time for austerity: Estimating the average treatment effect of fiscal policy. *The Economic Journal*, 126(590):219–255.

Li, D., Plagborg-Møller, M., and Wolf, C. K. (2024). Local projections vs. vars: Lessons from thousands of dgps. *Journal of Econometrics*, 244(2):105722.

Mertens, K. and Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the united states. *American Economic Review*, 103(4):1212–47.

Plagborg-Møller, M. and Wolf, C. K. (2021). Local projections and vars estimate the same impulse responses. *Econometrica*, 89(2):955–980.

Ramey, V. A. (2019). Ten years after the financial crisis: What have we learned from the renaissance in fiscal research? *Journal of Economic Perspectives*, 33(2):89–114.

Ramey, V. A. and Zubairy, S. (2018). Government spending multipliers in good times and in bad: Evidence from us historical data. *Journal of Political Economy*, 126(2):850–901.

Romer, C. D. and Romer, D. H. (2010). The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks. *American Economic Review*, 100(3):763–801.

Romer, C. D. and Romer, D. H. (2019). Fiscal space and the aftermath of financial crises: How it matters and why. Technical Report 25768, National Bureau of Economic Research.

Romer, C. D. and Romer, D. H. (2023). Presidential address: Does monetary policy matter? the narrative approach after 35 years. *American Economic Review*, 113(6):1395–1423.

Stock, J. H. and Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *Economic Journal*, 128(610):917–948.

Stram, D. O. and Wei, W. W. (1986). Temporal aggregation in the arima process. *Journal of Time Series Analysis*, 7(4):279–292.

Appendices

A Posterior computation and the set of admissible restrictions

This appendix describes posterior computation for the reduced-form VAR parameters and the accept–reject implementation of the sign and multiplier restrictions. We use a standard Bayesian VAR toolkit for computing draws of (B, Σ_u) and then map each draw into impulse responses and multipliers.

A.1 Working likelihood and conjugate prior for reduced-form draws

To generate posterior draws of the reduced-form parameters, we adopt a standard Gaussian likelihood as a *working likelihood* for the linear-projection parameters of (6). Concretely, conditional on the regressors X in (7), the working model assumes the stacked innovation matrix U has conditionally independent rows with common covariance Σ_u, i.e.

$$\mathrm{vec}(U) \;\Big|\; (B, \Sigma_u, X) \;\sim\; \mathcal{N}(0,\; \Sigma_u \otimes I_{T_{\mathrm{eff}}})\,.$$

Equivalently,

$$Y \;\Big|\; (B, \Sigma_u, X) \;\sim\; \mathcal{MN}(XB,\; \Sigma_u,\; I_{T_{\mathrm{eff}}})\,,$$

where $\mathcal{MN}$ denotes the matrix-normal distribution.

We pair this working likelihood with a conjugate matrix-normal/inverse-Wishart prior (Minnesota-style variants are special cases):

$$B \mid \Sigma_u \sim \mathcal{MN}(B_0,\; \Sigma_u,\; \Omega_0)\,, \qquad \Sigma_u \sim \mathcal{IW}(S_0,\; \nu_0)\,,$$

where B_0 is a $k \times n$ prior mean, Ω_0 is a $k \times k$ prior covariance over coefficients, and (S_0, ν_0) are the inverse-Wishart scale matrix and degrees of freedom.

Under these conjugate assumptions, the conditional posteriors are available in closed form. Define

$$\Omega_1 = \left(\Omega_0^{-1} + X'X\right)^{-1}, \qquad B_1 = \Omega_1\left(\Omega_0^{-1}B_0 + X'Y\right).$$

Then

$$B \mid \Sigma_u, Y, X \sim \mathcal{MN}(B_1,\ \Sigma_u,\ \Omega_1)\,.$$

For Σ_u, let

$$\widehat{U}(B) \equiv Y - XB.$$

Then the inverse-Wishart update takes the form

$$\Sigma_u \mid B, Y, X \sim \mathcal{IW}(S_1,\ \nu_1)\,, \qquad \nu_1 = \nu_0 + T_{\text{eff}}, \qquad S_1 = S_0 + \widehat{U}(B)'\widehat{U}(B).$$

(Equivalent expressions that integrate out B can also be used; the implementation in our code follows standard Bayesian VAR practice.)

A.2 Gibbs sampling for reduced-form parameters

We generate posterior draws using a Gibbs sampler that alternates:

1. Draw $\Sigma_u^{(s)} \sim \mathcal{IW}(S_1^{(s-1)}, \nu_1)$ conditional on $B^{(s-1)}$.

2. Draw $B^{(s)} \sim \mathcal{MN}(B_1^{(s)}, \Sigma_u^{(s)}, \Omega_1)$ conditional on $\Sigma_u^{(s)}$.

For each draw $(B^{(s)}, \Sigma_u^{(s)})$, we compute the Cholesky factor $P^{(s)}$ from $\Sigma_u^{(s)} = P^{(s)}P^{(s)\prime}$ and then compute orthogonalized impulse responses and multipliers.

A.3 Impulse responses, multipliers, and accept–reject restrictions

Given a draw $(B^{(s)}, \Sigma_u^{(s)})$, we compute the orthogonalized impulse responses to the first Cholesky shock using the companion-form representation implied by (6) and the impact matrix $P^{(s)}$. Denote by $\theta_{j,h}^{(s)}$ the response of variable j at horizon h to shock ε_{1t}.

We then apply the restrictions described in Section 4.2: We supplement the proxy-based ordering with weak admissibility conditions that rule out draws in which the identified innovation produces negligible spending responses and, as a consequence, economically implausible (small-denominator) multipliers.

Formally, letting g denote the index of government spending and y the index of real GDP, we impose a weak lower bound on the average short-run spending response,

$$\frac{1}{H_g+1}\sum_{h=0}^{H_g}\theta_{g,h}(B,\Sigma_u) \;\geq\; \underline{g}, \tag{27}$$

and we screen multipliers by requiring a positive cumulative spending response at the screening horizons and by restricting the implied cumulative multipliers to lie in broad pre-specified ranges at a short horizon H_S and a long horizon H_L:

$$\mathcal{M}(H;B,\Sigma_u)=\frac{1}{\bar{s}_G}\frac{\sum_{h=0}^{H}\theta_{y,h}(B,\Sigma_u)}{\sum_{h=0}^{H}\theta_{g,h}(B,\Sigma_u)}.$$

In the baseline specification, the average response of government spending is constrained to be above 0 for the first 8 quarters ($H_g = 8, \underline{g} = 0$), while cumulative multipliers are constrained to lie within the range $[-2, 5]$, with short run horizon being 8 and long run horizon being 20 respectively ($H_s = 8, H_L = 20$).

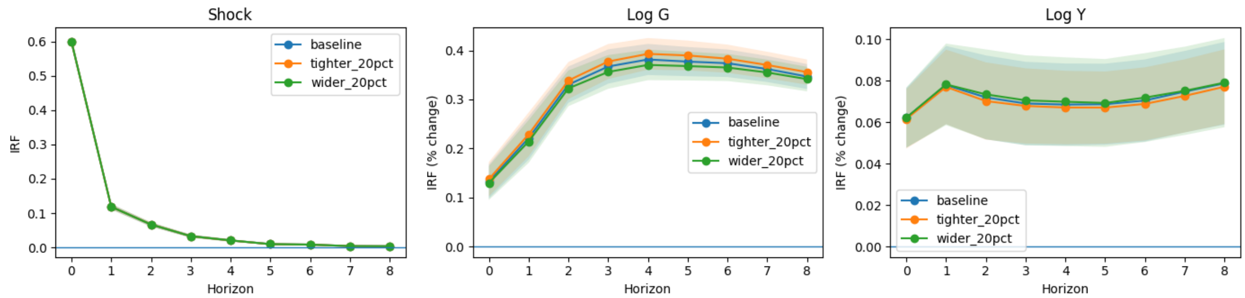
A.4 Robustness to alternate multiplier ranges

To assess sensitivity to the admissibility bounds on cumulative multipliers, we widen and tighten the baseline interval $[-2, 5]$ by 20%, yielding $[-2.4, 6]$ and $[-1.6, 4]$,respectively, while holding the spending-response condition ($H_g = 8$, $\underline{g} = 0$) fixed.

The resulting pooled IVW impulse responses (figure 6) are nearly indistinguishable from the baseline, with substantial overlap in posterior bands across all horizons. The acceptance rate changes only modestly—from 47% in the baseline specification to 50% under wider bounds and 43% under tighter bounds—indicating that inference is not driven by the exact choice of multiplier endpoints.

Overall, the results are robust to economically reasonable perturbations of the admissible multiplier range.

Figure 6: IRFs of a shock to the fiscal variable: different multiplier ranges



Note: Solid lines report inverse-variance–weighted averages of country-specific impulse responses. For each horizon, country-level impulse responses are summarized by their posterior medians and pooled across countries within each group using weighted least squares with a constant only, where weights are given by the inverse of the estimated variance of the country-specific response. Shaded areas denote ± 1 standard error bands computed from the weighted regression. Horizons are in quarters. Source: Authors' calculations.

B Comparing Bayesian and frequentist estimates

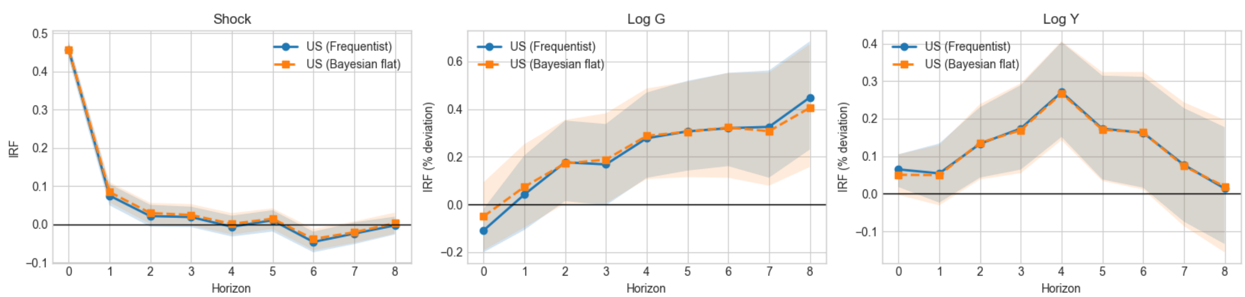
Here, we compare the results from estimating Bayesian VARs and their frequentist counterparts. The motivation behind this exercise is to look at whether the assumption of Gaussianity implicit in the Bayesian estimates holds or not in practice. The frequentist estimation does not impose a parametric distribution on innovations, and instead relies on bootstrapping using the empirical distribution of the residuals.

We first estimate our baseline three-variable Bayesian VAR, keeping all restrictions on government spending responses and the implied fiscal multipliers unchanged, but setting the prior hyperparameters to approximate a flat (diffuse) prior. We then estimate the frequentist counterpart by estimating the VAR via OLS, resampling reduced-form residuals to generate artificial data using the estimated coefficients, and re-estimating the VAR on these simulated samples. Bootstrap draws are retained only if they satisfy the same stability and admissibility constraints imposed in the Bayesian analysis.

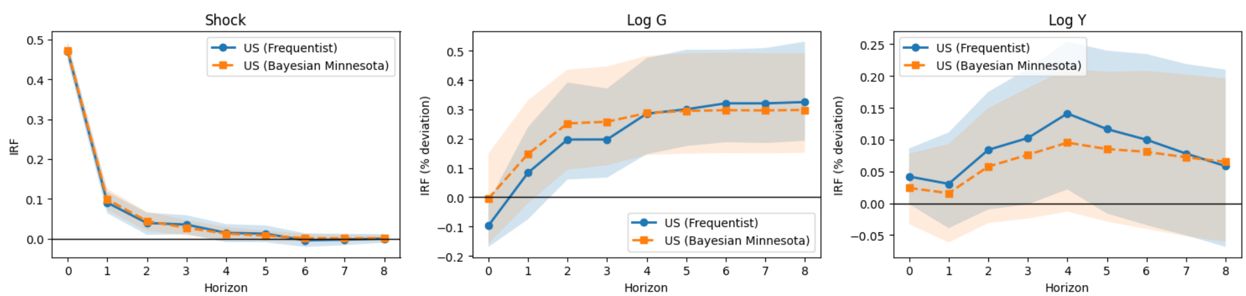
For comparison, we focus on a single country—the United States—and examine whether impulse response distributions differ across the two approaches. As shown in Figure 7a, the resulting IRF distributions under the Bayesian and frequentist procedures are nearly identical.

We then estimate a frequentist Minnesota counterpart by estimating the VAR via penalized least squares using a Minnesota-style ridge penalty, resampling reduced-form residuals to generate artificial data, and re-estimating the penalized VAR on these simulated samples. As in the Bayesian case, bootstrap draws are retained only if they satisfy the same stability and admissibility constraints. Figure 7b shows that the Bayesian and frequentist Minnesota IRFs are again very similar, suggesting that the Gaussianity assumption underlying Bayesian inference provides a good approximation in this application.

Figure 7: IRFs from Bayesian and Frequentist counterparts



(a) Flat Bayesian Priors



(b) Minnesota Bayesian Frequentist

Note: Medians and 68th percentile credible intervals. Impulse responses are weighted using inverse variance weights.

C Benchmark pooled panel heterogeneity: specification, splits, and inference

This appendix provides the implementation details for the pooled-split benchmark reported in Table 12. The goal is to mirror the pooled quarterly perspective of Ilzetzki et al. (2013) as closely as possible while remaining consistent with our identification strategy, which relies on a qualitative quarterly narrative proxy used as an external instrument.

C.1 Panel VAR specification and estimation

For each split, we estimate a separate pooled panel VAR on the stacked country–quarter observations in each group G:

$$x_{i,t} = \sum_{\ell=1}^{p} A_\ell \, x_{i,t-\ell} + \mu_i \; (+\, \tau_t) + u_{i,t}, \qquad x_{i,t} = \begin{pmatrix} g_{i,t} \\ y_{i,t} \end{pmatrix}. \tag{28}$$

Country fixed effects μ_i are always included; time fixed effects τ_t are included in the benchmark specification for Table 12 and varied in robustness checks.

We estimate (28) by least squares after removing the included fixed effects (i.e., within transformation). With two-way fixed effects, demeaning is carried out over country and time. The estimation sample is unbalanced; within each group G we use all available country–quarter observations that permit p lags of $x_{i,t}$.

Transformations. For Table 12, the benchmark transformation is quarterly log changes ($\Delta \log$) for both spending and output.[15] Formally, in the benchmark $g_{i,t} = \Delta \log G_{i,t}$ and $y_{i,t} = \Delta \log Y_{i,t}$. Because constant rescalings drop out of log differences, any country-specific normalization of levels has no effect on the benchmark $\Delta \log$ specification.

[15] Alternative transforms used in robustness include detrended log levels. In all cases, the reported multiplier is constructed from the estimated impulse responses of the transformed variables; see below.

C.2 Proxy-SVAR identification in the pooled panel

Let $\widehat{u}_{i,t} = (\widehat{u}^g_{i,t}, \widehat{u}^y_{i,t})'$ denote the residuals from (28) in a given group G. Let $z_{i,t}$ denote the quarterly narrative proxy, coded as an ordered sign indicator. Identification proceeds group-by-group using $z_{i,t}$ as an external instrument for the spending shock.

Baseline covariance mapping. Under the standard proxy-SVAR conditions (instrument relevance and exogeneity with respect to other structural shocks), the impact vector of the spending shock is proportional to the covariance between the proxy and the reduced-form innovations within group G:

$$b_G \propto \mathbb{E}_G[u_{i,t} z_{i,t}] \,. \tag{29}$$

In practice we compute the sample analogue $\widehat{c}_G = \widehat{\mathbb{E}}_G[\widehat{u}_{i,t} z_{i,t}]$ and normalize so that the impact response of spending equals one:

$$\widehat{b}_G = \frac{\widehat{c}_G}{\widehat{c}_{G,g}} = \begin{pmatrix} 1 \\ \dfrac{\widehat{\mathrm{Cov}}_G(\widehat{u}^y_{i,t},\, z_{i,t})}{\widehat{\mathrm{Cov}}_G(\widehat{u}^g_{i,t},\, z_{i,t})} \end{pmatrix}, \tag{30}$$

where $\widehat{c}_{G,g}$ denotes the first element of $\widehat{c}_G$ (the covariance with $\widehat{u}^g_{i,t}$). This normalization fixes the scale of the identified shock.

Benchmark instrument construction. For Table 12, the benchmark implementation uses a positive/negative instrument construction with lags. Specifically, we form indicators for positive and negative narrative innovations and include a small number of lags of these indicators in a first-stage projection for the spending residual. Denote by $q_{i,t}$ the vector collecting these instrument components. We estimate, within group G and after removing the same fixed effects as in (28),

$$\widehat{u}^g_{i,t} = \pi' q_{i,t} + e_{i,t}, \tag{31}$$

and define the fitted component $\widehat{v}_{i,t} = \widehat{\pi}' q_{i,t}$ as the effective proxy. We then replace $z_{i,t}$ by $\widehat{v}_{i,t}$ in (29)–(30). Treating zeros in $z_{i,t}$ as literal zeros (rather than missing) is the benchmark choice for the table.

C.3 Impulse responses and multiplier construction

Given $\{\widehat{A}_\ell\}_{\ell=1}^{p}$ and $\widehat{b}_G$, we compute the moving-average coefficients $\{\widehat{\Psi}_h\}_{h=0}^{H}$ implied by the estimated VAR and obtain impulse responses

$$\widehat{\mathrm{IRF}}_G(h) = \widehat{\Psi}_h \, \widehat{b}_G, \qquad h = 0, 1, \ldots, H. \tag{32}$$

Let $\widehat{\mathrm{IRF}}_{g,G}(h)$ and $\widehat{\mathrm{IRF}}_{y,G}(h)$ denote the spending and output responses, respectively. We report cumulative multipliers at horizon H as

$$\widehat{\mathcal{M}}_G(H) = \frac{1}{\bar{s}_G} \frac{\sum_{h=0}^{H} \widehat{\mathrm{IRF}}_{y,G}(h)}{\sum_{h=0}^{H} \widehat{\mathrm{IRF}}_{g,G}(h)}. \tag{33}$$

Table 12 reports $\widehat{\mathcal{M}}_G(4)$ and $\widehat{\mathcal{M}}_G(8)$, and also the medium-run summary $\overline{\mathcal{M}}_{G,4:8} = \frac{1}{5}\sum_{H=4}^{8} \widehat{\mathcal{M}}_G(H)$.

Spending share. The scaling term $\bar{s}_G$ is a group-specific average spending share constructed from the underlying levels data as $\bar{s}_G = \mathrm{TrimMean}_G(G/Y)$, where the trimmed mean drops a small fraction of the largest and smallest observations to limit sensitivity to outliers.[16]

C.4 Split definitions and cutoffs.

We implement heterogeneity by estimating (28) separately on groups defined by structural characteristics. The table reports the number of countries contributing observations to each group.

Trade openness. Trade openness is measured as trade-to-GDP (imports plus exports divided by GDP). For the openness split, countries are assigned to groups based on their country-level summary openness (computed over the sample period). Countries with openness below the cutoff are classified as *Closed* and those at or above the cutoff as *Open*.

[16] The trim rate and the exact levels definitions (including any smoothing or interpolation steps used to construct quarterly shares) are recorded in the replication code.

Public debt. Public debt is measured as debt-to-GDP. Because debt is state-like and varies over time, the debt split is defined at the country–quarter level. We classify observations into a *High debt* state when debt-to-GDP exceeds the cutoff, subject to a minimum-duration rule to avoid transient reclassifications; all remaining observations are assigned to the *Low/No high debt* group.

Informality. Informality is measured using a country-level informality index, coded so that higher values correspond to more informality. Countries are split into low- and high-informality groups using the cutoff reported in Table 12.

Labor-market liberalization. Labor-market institutions are measured using a labor reform (liberalization) index, coded so that higher values correspond to more liberal labor markets. Countries are split into less- and more-liberal groups using the cutoff reported in Table 12.

C.5 Bootstrap inference and acceptance

Inference for each group-specific multiplier is based on a country-cluster bootstrap designed to preserve within-country serial dependence. For each bootstrap replication b:

1. Sample countries with replacement from the set of countries in group G.
2. Construct the bootstrap panel by stacking the full time series for each sampled country (retaining the original within-country time ordering).
3. Re-estimate (28) and recompute the proxy-SVAR mapping (29)–(30) (including the first stage (31) when applicable).
4. Recompute impulse responses and multipliers (32)–(33).

For each reported statistic, the 90% confidence interval is the bootstrap percentile band. Bootstrap p-values for testing $H_0 : \mathcal{M} = 0$ are computed from the empirical distribution of the bootstrap replicates (two-sided).

Acceptance and plausibility filters In a small number of replications, the cumulative spending response in the denominator of (33) can be close to zero, yielding unstable ratios. We therefore apply mild plausibility filters to bootstrap draws (e.g., excluding draws with near-zero denominators or extreme ratios). The reported acceptance rate is the fraction of bootstrap replications retained after these filters. Table 12 reports results for specifications with high acceptance rates, and our baseline results are not sensitive to reasonable variations in these filters.

D Building the political support narrative data

This appendix documents the fixed prompts, coding criteria, and mapping rules used to construct (i) the quarterly indicator of whether the political environment is favorable for passing and sustaining fiscal measures and (ii) its three primitive components (majority/cohesion, election proximity, unrest/gridlock). The prompts are applied uniformly across countries and time in a non-adaptive pipeline; all model outputs are stored together with short verbatim evidence excerpts to enable ex-post auditing.

D.1 Overview of the pipeline

The political-context data are built in two stages using fixed prompts.

1. **Stage 1 (support indicator + verbatim evidence).** From each country–quarter *EIU Country Report* excerpt, an off-the-shelf model produces: (i) a binary indicator `political_environment_favorable_indicator` $\in \{1, 0, \texttt{null}\}$, (ii) a short narrative field `political_environment_favorable_explanation` (for readability), and (iii) a compact *verbatim evidence bundle* consisting only of short quotations copied exactly from the underlying EIU excerpt and referring to majority/cohesion, election timing, and unrest/gridlock.

2. **Stage 2 (mechanism primitives).** To separate mechanisms, we run a second fixed prompt that reads the Stage-1 verbatim evidence bundle (not macro outcomes and not

external sources) and codes three political primitives: `maj_cohesion`, `election_proximity`, and `unrest_gridlock`, each in {YES, NO, UNCLEAR} with a supporting quote.

Stage 2 does *not* affect the baseline political-environment indicator or the fiscal narrative classification. It is used only for mechanism and "horse race" diagnostics in the pooled panel analysis.

All non-UNCLEAR primitive labels are anchored to the underlying report text: before running Stage 2, we mechanically verify that each quotation in the Stage-1 evidence bundle is an exact substring of the original EIU excerpt and otherwise drop it (see Section D.4 and post-processing rules below).

D.2 Stage 1: Fixed prompt template for the political environment indicator

Task. Read the *Economist Intelligence Unit* country report excerpt for country `C` and quarter `Q`, focusing on the *Domestic politics* and related sections. Based solely on the information in this text, assess (i) the overall strength of public support for the national government and (ii) whether the political environment is favourable for passing and sustaining fiscal measures, using the definitions below.

Public support classification. Classify the overall strength of public support for the national government in `Q` as:

- `HIGH`: the government is clearly popular or enjoys strong public backing.
- `LOW`: the government is clearly unpopular or faces widespread public discontent.
- `MIXED`: the report contains explicit evidence pointing in both directions.
- `UNCLEAR`: the report does not provide enough information to form a view.

Base this classification on explicit language in the report regarding election outcomes, approval or disapproval, protest activity, opinion polling, and other stated indicators of popularity or discontent.

Favourable political environment indicator. Define an indicator of whether the political environment is favourable for designing, passing, and sustaining fiscal measures. Set:

- `political_environment_favorable_indicator` = 1 (*favourable*) if the report clearly indicates that all three of the following conditions hold jointly in `Q`:
  1. the executive is backed by a working majority in the legislature (single party or stable coalition);
  2. no national election is scheduled in the current or next quarter; and
  3. the report does not highlight material political unrest, sustained gridlock, or legislative paralysis that threatens the government's capacity to act.
- `political_environment_favorable_indicator` = 0 (*unfavourable*) if the report clearly indicates that at least one of these conditions fails (for example, the government lacks a working majority or faces an unstable coalition, a national election occurs or is scheduled in the current or next quarter, or there is serious unrest or persistent legislative gridlock).
- `political_environment_favorable_indicator` = `null` if the report does not provide sufficient information to determine whether the three conditions hold.

Outputs. Return:

- `public_support` $\in$ {`HIGH`, `LOW`, `MIXED`, `UNCLEAR`}.
- `public_support_evidence`: at most one short phrase (≤ 30 words) copied *verbatim* from the report excerpt; if `public_support` is `UNCLEAR`, return the empty string.
- `political_environment_favorable_indicator` $\in$ {1, 0, `null`}, defined as above.
- `political_environment_favorable_evidence`: up to three short phrases (≤ 30 words each) copied *verbatim* from the report excerpt, providing the key textual basis for the classification and explicitly referring to (i) majority/cohesion, (ii) election timing, and/or (iii) unrest/gridlock; if the indicator is `null`, return the empty list.
- `political_environment_favorable_explanation`: 2–4 sentences summarising why the indicator takes the value 1, 0, or `null`. This field may paraphrase, but it is used for documentation only.

Notes. Base all classifications strictly on the provided report excerpt; do not infer unstated events or motives, and do not use external knowledge. Evidence fields must be copied *verbatim* from the excerpt. When the information is incomplete or ambiguous, prefer `MIXED`, `UNCLEAR`, or `null` over strong statements.

D.3 Stage 2: Fixed prompt template for political primitives

Stage 2 codes political primitives from the *verbatim evidence bundle* produced in Stage 1, `political_environment_favorable_evidence`. This design (i) standardizes the input into a short, auditable set of quotations and (ii) reduces token usage relative to passing the full report excerpt.

> **Task.** You are coding political context for a macroeconomic panel dataset.
>
> **Input.** A JSON object with key `items`, where each item contains: an `id`, the country identifier, the current-quarter date window, and an `evidence` string formed by concatenating the elements of `political_environment_favorable_evidence` (verbatim quotations from the EIU excerpt).
>
> **Definitions.** Using *only* the `evidence` string:
>
> 1. `maj_cohesion`: Is there a working legislative majority or stable coalition during the quarter?
> 2. `election_proximity`: Does a national election occur or is it scheduled within the window `[q_start, next_q_end]` (i.e. current quarter through next quarter)?
> 3. `unrest_gridlock`: Is there major unrest *or* severe political/legislative gridlock that materially hinders policy?
>
> **Outputs.** Return *JSON only* with schema:
>
> $$\{ \texttt{"results"} : [\cdots] \}.$$
>
> Each element must contain: `maj_cohesion`, `election_proximity`, `unrest_gridlock`

$\in \{\text{YES}, \text{NO}, \text{UNCLEAR}\}$ and a corresponding quote field for each primitive.

Exact-quote rule. If a label is YES or NO, the quote must be copied *exactly* as a substring from `evidence`. If a label is UNCLEAR, the quote must be the empty string.

All prompts are run with fixed parameters and without any tuning on outcomes.

D.4 Mapping to the econometric variables

Political environment indicator (support).

Let $\texttt{political\_environment\_favorable\_indicator}_{i,t} \in \{1, 0, \texttt{null}\}$ denote the Stage 1 output for country i and quarter t. The econometric support indicator in the main text is defined as:

$$\text{Supp}_{i,t} = \begin{cases} 1, & \text{if } \texttt{political\_environment\_favorable\_indicator}_{i,t} = 1, \\ 0, & \text{if } \texttt{political\_environment\_favorable\_indicator}_{i,t} = 0, \\ \text{missing}, & \text{if } \texttt{political\_environment\_favorable\_indicator}_{i,t} = \texttt{null}. \end{cases}$$

The baseline panel estimates treat missing support as a distinct state, rather than imputing or collapsing it into the unfavorable category. Robustness checks may drop missing-support quarters.

Political primitives. Stage 2 produces three quarter-level primitive indicators:

$$\texttt{maj\_cohesion}_{i,t}, \quad \texttt{election\_proximity}_{i,t}, \quad \texttt{unrest\_gridlock}_{i,t} \in \{\text{YES}, \text{NO}, \text{UNCLEAR}\},$$

together with the corresponding quote fields (`maj_cohesion_quote`, `election_quote`, `unrest_quote`). In the pooled regressions, these primitives are used to form indicator variables (e.g. $\mathbf{1}\{\text{YES}\}$), with additional controls for missing/unclear classifications as described in the main text.

D.5 Separation from fiscal narrative coding

Political-context variables are never used to classify fiscal actions as exogenous or endogenous. The narrative exogeneity screen for fiscal measures is based solely on the stated motivations in the fiscal text, and political variables are used only to index the environment in which already-identified exogenous spending actions occur and to explore mechanism through the primitives.

PUBLICATIONS

AI Meets Fiscal Policy: Mapping Government Spending Actions Across 64 Countries

Working Paper No. WP/2026/043